ROCKWOOL FOUNDATION
BERLIN
INSTITUTE FOR THE ECONOMY AND THE FUTURE OF WORK

CReAM | Centre for Research & Analysis of Migration

DISCUSSION PAPER SERIES 23/25

# Do Judges Exhibit Gender Bias? Evidence from the Universe of Divorce Cases in China

Xiqian Cai, Pei Li, Qinyue Luo, Hong Song, Huihua Xie

JUNE 2025

**ROCKWOOL Foundation Berlin**
www.rfberlin.com

**Centre for Research & Analysis of Migration**
www.cream-migration.org

# Do Judges Exhibit Gender Bias? Evidence from the Universe of Divorce Cases in China

Xiqian Cai[*], Pei Li[†], Qinyue Luo[‡], Hong Song[§], Huihua Xie[¶]

**Abstract**

Does gender identity affect judicial decisions? This paper provides novel evidence of in-group gender bias in the judicial decisions for almost all divorce cases in China. Exploiting the effectively random assignment of cases to judges, the analysis finds that female judges are 1.2 percentage points more likely to grant divorce petitions filed by female plaintiffs compared to male plaintiffs, relative to male judges. This bias primarily reflects female judges' harsher treatment of male plaintiffs. The bias is significantly weaker in regions with stronger traditional gender norms, indicating that conservative cultural attitudes may constrain overt displays of in-group gender favoritism. Institutional legal development has little moderating effect, underscoring the primary role of culture. These findings highlight the importance of complementing efforts to promote judicial diversity with safeguards to detect and mitigate implicit bias.

**Keywords: gender, in-group bias, gender discrimination, judicial decisions**

**JEL Codes: J16, J14, J10**

We thank Sascha O. Becker, Catharina Behrens, Thomas Cornelissen, Hyejin Ku, Yi Lu, Valentina Melentyeva, Andrea Weber and Josef Zweimüller for their helpful comments. Any errors are our own.

[*]School of Economics and WISE, Xiamen University, China. caixiqian@xmu.edu.cn.

[†]School of Economics, Zhejiang University, China. bengcome@gmail.com.

[‡]ROCKWOOL Foundation Berlin, Germany. ql@rfberlin.com.

[§]School of Economics, Fudan University, China. songhong@fudan.edu.cn.

[¶]School of Economics, Zhejiang University, China. huihuaxie@zju.edu.cn.

# 1 Introduction

Does gender identity affect judicial decisions? From the perspective of the public, the justice system is a context in which society particularly values fair, impartial, and rational decision-making. Hence, gender identity should have little or no influence on judicial decisions. However, economic studies have found gender bias in multiple professions and settings, such as teaching evaluations, career promotion, and job recruitment. Can judges, who are traditionally viewed as objective and neutral, be an exception? If not, to what extent? How is this bias integrated with existing social culture, attitudes, and institutions? What measures can be adopted in the judicial process to bolster fairness and accuracy in the decision-making process? The answers to these questions are particularly important for those in developing countries or regions in which corruption, traditional customs, and poor institutions are still widespread. However, there is limited evidence on these questions.

In this paper, we investigate gender bias in judicial decisions. We use comprehensive online transcripts of the judicial decisions for almost all divorce cases in China. The institutional judicial setup in China provides a unique opportunity to explore gender bias. During recent decades, the Chinese judiciary has increasingly adopted measures to boost fairness and accuracy in the decision-making process. To increase the transparency and public supervision of judicial decisions, in late 2013, the Supreme People's Court of China released regulations on publishing the verdicts of the People's Courts online. According to official regulations, the verdicts of the People's Courts at various levels should be published online except for those that (1) involve state secrets or individual privacy, (2) involve juveniles, or (3) involve other "improper" situations. Furthermore, to avoid potential manipulation of the case assignment, which may result in unfairness and public controversy, the Supreme People's Court implemented an official announcement in early 2014 that requires that within the same lower court, each case, in principle, should be randomly assigned to a judge who will preside over the trial, plea bargaining, and sentencing process.[1]

These features offer an ideal setup to investigate gender bias in judicial decisions. First, the online transcripts of judicial decisions allow us to obtain a naturally occurring dataset that contains rich information on judges, plaintiffs, and cases. Second, the random assignment strategy required by the Supreme People's Court allows us to exploit the effectively exogenous gender shock of judges to cases. Third, substantial regional variation in social norms and institutional environments across China allows us to examine how cultural context shapes the expression of in-group gender bias and to assess the broader generalizability of our findings.

[1] See Section 2 for a detailed discussion of the institutional background.

Our empirical analysis proceeds in three stages. First, we document the significant existence of in-group gender bias in judicial decisions. Specifically, the in-group gender bias comes primarily from the negative bias of female judges toward male plaintiffs, which is aligned with physiological and sociological research on gender differences in in-group bias. In terms of economic magnitude, with an average of acceptance ratio of 0.265 points for divorce claims, we document an in-group gender bias of 0.012 points. The estimates are robust across different specifications, including a placebo test with fake assignment of judges to plaintiffs, randomly dropping a court each time to address the concern of outlier influence, using an alternative inference method, keeping courts with at least one female judge, and examining the sample selection issue due to missing gender information of the plaintiff.

Second, we detect several mechanisms that may account for this pattern. The first is in-group gender bias: people care not only about themselves, but also about the groups to which they belong. In our focal setting, female judges show preferential treatment toward members of their group (female plaintiffs) and offer them greater lenience in judicial decisions. We also consider a number of alternative explanations for the empirical patterns. One possibility is that although judicial decisions vary with the plaintiff's gender, this variation is not due to preferential treatment toward members of one's own group, but rather to differences in the information available to the judge. In other words, a female judge may pay equal attention to the arguments made by both sides, yet may better understand the arguments made by members of her own group. The second possibility is that the plaintiff's—not the judge's—behavior is driving the results. For example, a female plaintiff may perform better in the courtroom when she faces a female judge. The third possibility is that it is not the judges' gender, but other characteristics that are correlated with gender that drive the bias. The fourth possibility is that judges may not be biased toward the plaintiff's gender but are influenced by other plaintiff characteristics. Although definitively ruling out these alternative explanations is difficult, we provide multiple pieces of evidence that suggest that these possibilities are unlikely to fully account for the observed patterns.

Third, we explore heterogeneity patterns in in-group gender bias across regions to deepen our understanding of the in-group gender bias and gain insight into external validity. Specifically, estimated in-group gender bias in judicial decisions is less pronounced in regions with stronger traditional gender norms (that is, initially higher gender discrimination against women). These results show that traditional culture and social norms shape the expression of in-group gender bias. Meanwhile, we find that there are no substantial differences in the magnitude of in-group gender bias across regions with different levels of legal development. These results suggest that our documented in-group gender bias by judges is universal across levels of legal development.

These findings have important implications for the design of equitable judicial systems in the context of increasing gender diversity on judicial benches. According to data from China's National Bureau of Statistics and the Chinese Judicial Statistics Yearbook, the share of female judges in China rose from 23.5 percent in 2009 to 28.8 percent in 2013, reaching 32 percent by 2022.[2] These shifts represent meaningful progress toward gender parity in judicial institutions and broader inclusion in state governance. However, increased gender diversity alone does not guarantee judicial impartiality. Implicit identity-based biases, such as in-group gender favoritism (identified in our analysis), may inadvertently emerge and affect judicial outcomes. Effective judicial diversity should be accompanied by institutional safeguards designed to identify and mitigate implicit bias. For example, our study finds that such biases are significantly moderated by prevailing cultural norms, indicating that they reflect socially conditioned beliefs rather than intrinsic gender-based traits. Meanwhile, recent empirical evidence from the education sector demonstrates that bias-awareness programs can attenuate discriminatory behavior.[3] Analogous bias-awareness programs in judicial settings could similarly mitigate implicit gender biases, helping to reinforce judicial neutrality and fairness.

*Literature Review.* This study echoes the theory of social identity—namely, that people care not only about themselves, but also about the group to which they belong, and show preferential treatment toward members of their own group. Within the context of gender, this theory suggests that decision-makers may systematically favor individuals of their own gender. In-group bias has been extensively studied in the literature using the minimal group paradigm in experimental setting. Some recent research relies on naturally occurring data to explore in-group bias in real-world contexts. These studies span diverse decision-making domains, including judicial decisions (Shayo and Zussman 2011, 2017; Anwar et al. 2012), policing (Antonovics and Knight 2009; West 2018), academic journals refereeing (Chari and Goldsmith-Pinkham 2017; Card et al. 2020; Van Der Lee and Ellemers 2015), student evaluations of teachers (Boring 2017; Mengel et al. 2019), and hiring decisions (Bagues and Esteve-Volart 2010; Bagues et al. 2017).[4]

Our study is among the first to use naturally occurring data and exploit a random assign-

[2]International trends reflect similar patterns: the UN Women reports that the global share of female judges or magistrates increased from under 32 percent in 2010 to 43 percent in 2021.

[3]For instance, Alesina et al. (2024) show that revealing stereotypes among teachers leads to measurable reductions in discriminatory behavior toward immigrant students.

[4]A large strand of the literature has examined in-group social identity and bias in lab or field experiments; see Charness and Chen (2020) for a review. Whereas experiments conducted in lab settings suffer from weaknesses—for example, group identities are artificially generated and the stakes are low—a growing amount of research in recent years has used naturally occurring data to overcome these limitations and improve identification.

ment strategy to investigate in-group gender bias in judicial decisions. While several studies have explored gender bias in judicial decisions, many are case studies or descriptive analyses (Bauer and Dawuni 2015; Kenney 2013), or they fail to account for potential nonrandom assignment between judges and cases, which limits their ability to causally identify gender bias (Bindler and Hjalmarsson 2020; Mustard 2001; Kulik et al. 2003; Fix and Johnson 2017). Recent advances in the literature leverage quasi-random assignment mechanisms to address these challenges. For example, Hoekstra and Street (2021) exploits the random ordering of jury pools in U.S. misdemeanor and felony cases to demonstrate that jurors of the same gender as the defendant reduce the likelihood of conviction. Similarly, Didwania (2022) examine U.S. federal prosecutors and shows that gender-matched prosecutors and defendants lead to more lenient sentencing outcomes, providing robust evidence of implicit gender preferences. In another context, Ash et al. (2024) use Natural Language Processing to quantify gender attitudes expressed in judicial opinions from U.S. appellate courts, finding that judges with traditional gender attitudes are more likely to issue rulings unfavorable to women and less likely to engage with female colleagues (e.g., fewer citations of female authors). While in the developing country context, Ash et al. (2025) analyze over 5 million Indian court cases and find little evidence of gender, religion, and caste in-group bias, even when identity is salient.

While these studies have significantly advanced our understanding of in-group bias, the literature remains predominantly focused on developed economies. Research in developing economies, where cultural norms and institutional constraints may lead to distinct patterns of bias, is notably limited. Furthermore, most existing studies concentrate on criminal cases, leaving civil cases—especially those involving family law—underexplored. Family law cases are deeply intertwined with societal values and gender norms, making them particularly important for studying how culture and identity shape judicial decisions.

Our study addresses these gaps by providing the first large-scale evidence explicitly focused on in-group gender bias in judicial decision-making within a civil law context in a developing economy. By taking advantage of the institutional setting in China and leveraging a national-level, comprehensive dataset covering 47,603 Basic Civil Court judges in 2,696 locations spread across the whole of China, we reveal in-group gender bias in judicial decision-making as well as its integration with existing cultural and institutional characteristics. Importantly, we find that the magnitude of the bias is smaller in regions with stronger traditional gender norms against women, suggesting that conservative social expectations may constrain the expression of gender favoritism among female judges. Additionally, our results indicate that institutional factors, such as levels of regional legal development, show little impact on mitigating this bias, emphasizing the critical role of cultural attitudes over formal institutions. These findings extend the literature by shifting the focus to civil law in

a developing economy and uncovering the critical role of cultural norms in shaping judicial outcomes. By offering robust evidence and ruling out alternative explanations, our study complements existing research and provides new insights into the interplay between culture, identity, and judicial behavior, solidifying its contribution to a broader understanding of institutional fairness and bias.

This paper also contributes to the broader literature on gender bias. Gender bias has been well-established as one of the determinants of women's underperformance across multiple societal domains. It has been examined extensively in the literature, including in teacher evaluations (Boring 2017; Mengel et al. 2019), hiring decisions (Bagues and Esteve-Volart 2010; Bagues et al. 2017; Reuben et al. 2014), the peer review process (Chari and Goldsmith-Pinkham 2017; Card et al. 2020; Van Der Lee and Ellemers 2015), and evaluation in driving tests (Bar and Zussman 2020). The empirical conclusions, however, are mixed. While certain studies detect no gender bias in evaluations (Blank 1991), many others identify substantial biases adversely affecting women (Bagues et al. 2017; Mengel et al. 2019; Card et al. 2020; Van Der Lee and Ellemers 2015). For example, Bagues et al. (2017) find that the gender composition of the committees does not affect hiring decisions, whereas evidence shows that candidates are less likely to be hired if the committee contains a higher share of evaluators of the same gender as the candidate. Mengel et al. (2019) investigate gender bias in university teaching evaluations. They find that women receive systematically lower teaching evaluations than their male colleagues, and this bias is driven by male students' evaluations. Although this strand of the literature has documented that gender bias has a significantly negative impact on women's career development—including hiring, tenure, promotion, and so forth—the evidence on women's broader social welfare remains limited. Our paper extends this literature by providing novel evidence of gender bias in judicial decisions, a domain closely tied to women's societal welfare, legal protection, and subjective well-being, and traditionally regarded as impartial and procedurally fair. Importantly, we identify significant asymmetric in-group favoritism among female judges. Our analysis uncovers an indirect but critical channel through which institutional gender composition can affect distributive outcomes, even in settings governed by formal rules and professional norms of impartiality. These results carry important policy implications that promoting judicial impartiality requires targeted policies and programs designed to mitigate implicit gender biases.

# 2 Institutional Background, Data, and Summary Statistics

## 2.1 Institutional Background

To provide a comprehensive picture of the institutional background of the divorce cases, we first describe the institutional background of the legal system in China, then present the judicial procedure for divorce cases and finally, discuss the features and advantages of the institutional setting for our study.

*The Court System in China.* There are four levels of courts in China: Supreme, Superior, Intermediate, and Basic Courts. The Supreme People's Court is the highest judicial organ of the People's Republic of China. Superior Courts are established in province-level regions (including provinces, autonomous regions, and municipalities) and are the highest judicial organization at the provincial level. They exercise judicial powers in accordance with the law and supervise the judicial work of lower-level courts. Intermediate Courts are established in prefectures, with their superior units being the Superior Courts. Basic Courts are established at the county level and are supervised by Intermediate Courts. Basic Courts are responsible, in the first instance, for most civil cases, including divorce cases. The main data (first-instance judicial decisions) used in our study are from the Basic Courts.

*Judicial Procedure of Divorce Cases in China.* According to Chinese law, divorce cases proceed in three stages. In the first stage (the prosecution stage), a husband or wife can make an appeal to the Basic Court to dissolve their marriage. In principle, the plaintiff in the divorce proceedings—that is, the party who filed for divorce—should file a lawsuit with the Basic Court in the defendant's registered permanent residence, which may differ from that of the plaintiff. When filing a lawsuit, the plaintiff should submit the complaint and relevant documents, including personal information for both parties, the claim and causes for the divorce, and relevant evidence and witness information.[5]

In the second stage (the investigation stage), the corresponding Basic Court takes charge of the case and proceeds in two phases: preparation for trial and mediation. In the preparation phase, the court reviews the relevant documents, conducts an investigation, and collects

[5]According to the Civil Procedure Law of the People's Republic of China, the plaintiffs are allowed to withdraw the case after learning the judge. Specifically, the plaintiff can withdraw the action after the court has accepted the case until the judgment has been pronounced. The court decides whether to grant permission or not. In most cases, the court grants a dismissal. It is worth noting that once the case is withdrawn, the court will not accept the case again within six months until there are new reasons or circumstances.

evidence. After that, courts are required to attempt mediation in divorce cases before proceeding to trial, according to Article 9 of the Civil Procedure Law (amended in 2012). This reflects a systematic effort to prioritize alternative dispute-resolution mechanisms. Typically, mediation is not conducted by judges, and each court has a dedicated mediation committee to handle pre-trial mediation. These are staffed by a diverse group of professionals, including retired judges, trained mediators, legal advisors, social workers, community volunteers, and psychological counselors. Their primary role is to facilitate communication between parties and encourage reconciliation. In practice, mediation teams are randomly formed and not assigned based on gender or other characteristics of judges and plaintiffs.

If mediation leads both parties to reach a mutual agreement, the case is closed. If mediation fails, the court proceeds to the third stage, the trial stage. The court assigns presiding judge(s), who examines the evidence, ascertains the facts, distinguishes right from wrong, and affirms the rights and obligations of both parties. Details regarding types and assignment of judges are provided in the following paragraphs. The plaintiff and defendant are also allowed to make arguments and ask questions.

After the trial, the judge pronounces the result in public, which can occur in the court immediately after the trial or at another time. The predominant legal standard for granting divorce in China, as stipulated by the Article 32 of the Marriage Law of the People's Republic of China, is whether "the emotional relationship is truly ruptured." This criterion is intentionally broad, providing judges with significant discretions in evaluating individual cases. However, the law also specifies certain circumstances that may be considered evidence of a ruptured relationship, including: (1) bigamy or cohabitation with another person; (2) domestic violence or maltreatment; (3) gambling, drug use, or other criminal activities; (4) long-term (more than 2 years) separation without reconciliation; and (5) other serious circumstances that damage the marriage relationship.

In practice, judges interpret "truly ruptured emotional relationship" based on evidence provided during the trial. For instance, a plaintiff may present written evidence, police reports, or witness testimony to support claims of abuse or infidelity. Communication records, such as messages or recordings, may also be used to demonstrate irreparable conflict. Testimonies from family members or community representatives may supplement the evaluation of the relationship's status.[6]

If either party disagrees with the result, they can appeal to a higher level court and enter

[6]For example, in a case where the plaintiff presented hospital records documenting injuries from domestic violence and police intervention reports, the judge ruled that the emotional relationship was irreparably damaged. Another example involves a long-term separation with documented attempts at reconciliation by one party, which were consistently rejected by the other. This was deemed sufficient to grant a divorce.

the second trial stage within 15 days of the date of receiving the written judgment.[7] The judicial process for the second trial is the same as for the first trial. For divorce cases, the second trial stage is the final instance. The proportion of appeals is generally low for divorce cases; for example, it is 4.1% in our dataset.

For divorce cases in the first instance, the time limit is six months. If an extension is needed under special circumstances, the deadline may be extended for six or nine months upon approval of the president of the court or the next higher-level court. The time limit for hearing appeals against civil judgments is three months. If an extension is needed under special circumstances, the deadline may be extended by three months upon the approval of the president of the court.

*Judges in Divorce Cases.* In China, there are two types of civil case procedure: summary procedures and general procedures. Most divorce cases apply summary procedures. According to the Civil Procedural Law, for the first instance in the summary procedure, one judge appears at the trial and is solely in charge of the trial (assisted by a clerk who records the proceedings). In some complicated divorce cases, general procedures are applied in which there are three judges: a principal judge and two juror judges, who form a collegial panel. In our dataset, 69.3% of the cases apply summary procedures, and the remaining 30.7% apply general procedures.[8]

In addition, while most judges in the court system are formal judges, some are acting judges. Specifically, according to the Organization Law of the People's Courts in China, once an individual passes the judicial examination and starts to work in the court system, they become an acting judge. After engaging in judicial work for at least two years, they can be appointed by the People's Congress to be a formal judge. Acting judges can play the same role in the divorce case courts as formal judges. In our dataset, for the first instance, 16.5% of the cases were heard by acting judges.

*Advantage of the Institutional Setting for the Study.* The judicial procedural details of divorce cases in China provide a unique opportunity to examine the possible in-group gender bias in judicial decisions. First, unlike criminal cases, the applicable legal standards in divorce cases are rather loose, and the legal provisions that specify when a judge grants a divorce are extremely vague. For example, the most predominant legal criterion for divorce is "whether the emotional relationship is truly ruptured". This criterion is elusive, as it is

[7] In general, the appeal letter will be transferred to the People's Court at the next higher level by the Basic Court that originally tried the case. Alternatively, the appellant may directly file the appeal with the People's Court at the next higher level.

[8] For cases that apply general procedures, our divorce case dataset only records the principal judge's information.

difficult to determine whether it is truly ruptured or not. Instead of bright-line rules, judges are given vast discretion in interpreting the seriousness of a ruptured emotional relationship. Hence, unlike the decision system in the United States, judges in China have a high degree of personal discretion and independence in how they evaluate divorce cases. Trial court decisions are generally final, as appeals are only available on a limited basis, occurring infrequently and seldom leading to reversal. The degree of discretion and the compelling force of the decisions empowered by the local court judges allow us to examine whether the judges exhibit gender bias in their rulings.

Second, the Chinese judicial system has increasingly adopted measures to boost fairness and precision in the decision-making process. Specifically, the law implemented by the Supreme People's Court in 2014 requires that within the same lower court, each case is in principle, randomly assigned to a judge who presides over the trial, plea bargaining, and sentencing process.[9] Meanwhile, judges are unaware of the details of the upcoming cases. In most provinces, the cases are randomly assigned by a computer system. Once it is assigned, the judge cannot reject a case unless the judge has a particular reason (for example, a conflict of interest). To enforce the implementation of the random assignment, the whole process is supervised by the Discipline Inspection and Supervision Departments of China.[10] The random assignment generates an exogenous shock of judges to cases, which provides the estimation power for unbiasedly identifying the effect of the judges' gender on court decisions.

## 2.2 Data and Variables

The primary source of data is online transcripts of judicial decisions. In late 2013, the Supreme People's Court promulgated the *Regulations on Publishing Verdicts of Peoples' Courts Online*. According to this document, the verdicts of People's Courts at various levels should be published online, allowing us to obtain comprehensive records of judicial cases.[11] These documents first became available online in late 2013 for a handful of courts, and the coverage widened over time.

[9]For the official document, see the announcement issued by the Supreme People's Court in China: http://www.court.gov.cn/zixun-xiangqing-13520.html.

[10]See, for example, the requirement of random assignment in the prefecture of Guangdong: http://www.gzhzcourt.gov.cn/news/50007629.cshtml.

[11]There are limited exceptions to this policy, including cases involving state secrets, individual privacy, or juvenile. The first two categories are often not publicly disclosed due to the sensitive nature of the information. According to the China Judicial Statistical Yearbook, juvenile criminal cases accounted for only 0.5% of all adjudicated cases adjudicated in 2012. This percentage has remained consistently low over the years, although more recent data is unavailable. These cases are excluded from our dataset as they fall under a separate judicial system.

In our dataset, 5.65% of cases involved plaintiffs withdrawing their claims. We exclude these cases from the analysis as they do not contain judicial ruling information.[12] Meanwhile, we exclude collegial panel cases (31.67%) from the main analysis and focus on single-judge trials. Collegial panels, typically composed of one principal judge and two jurors, present two estimation complications; that is, only the principal judge's name is recorded, and collective decision-making may differ from that of individual judges. To avoid ambiguity in assigning responsibility and ensure clean identification, we restrict our main analysis to cases adjudicated by a single judge.[13] Our final analysis sample consists of 540,525 divorce cases decided between 2014 and 2016 by 42,766 judges in Basic Civil Courts across 2,639 locations nationwide.[14]

From the judicial transcripts, we extract detailed case-level variables, including the verdict date, adjudicating court, judge's name, plaintiff's gender, and the cause of action and the outcome of the claim. The cause of action is captured through a set of binary indicators, such as whether the defendant engaged in gambling, was incarcerated, committed domestic violence, used drugs, or had an extramarital affair. To capture the context of marital conflict, we extract additional indicators such as whether the defendant left the home, whether the parties frequently quarreled over trivial matters, whether the defendant's whereabouts were unknown, and whether the couple had been living apart for an extended period. We further document whether the couple met through a blind date, whether the case involved disputes over child support, and the number of sons and daughter mentioned in the judgment.

The primary outcome variable, *Case Acceptance*, is a binary indicator which equals to one if the ruling explicitly grants the plaintiff's request for divorce (e.g., "divorce granted", "divorce application approved", or "plaintiff prevails") and zero if the request is denied (e.g., "divorce not granted", "application denied", or "plaintiff's request not supported"). This classification ensures consistency and and accuracy across judicial rulings.

[12]This raises a potential concern of sample selection bias if withdrawal decisions are systematically related to judge or plaintiff gender. To assess this, we examine whether withdrawal rates vary by judge-plaintiff gender combinations in Column 1 of Appendix Table C2. The estimated coefficients are statistically and economically insignificant, suggesting that selection bias from case withdrawal is unlikely.

[13]In a robustness check, we experiment with full sample including both individual judges and collegial panels. While our main findings remain robust, the estimated coefficients change, implying the substantial difference between cases adjudicated by single judge and those by panels.

[14]Cases resolved through mediation are excluded, as they do not proceed to trial and do not involve judicial rulings. According to the China Judicial Statistical Yearbook, approximately 46% of first-instance civil cases involving marriage and family disputes were resolved through mediation between 2012 and 2014. Because mediation bypasses judicial adjudication, it does not compromise the random assignment assumption underlying our identification. To further assess potential selection concerns, we examine whether in-group gender bias affects case duration (from filing to closure or trial). Columns 2 and 3 of Appendix Table C2 show that the interaction between male judge and male plaintiff has no significant effect on either measure, alleviating concerns about sample selection due to excluded mediated cases.

The key information for our analysis is the gender of the plaintiff and the gender of the judge. Plaintiff gender is typically reported in the case documents, and cases missing this information(3.56%) are excluded[15]. However, the transcripts include only the judges' names, but not their gender. We then manually identify the gender of the judges in the following steps. We first search and explore each court's website. In China, some courts have websites that list the demographic information of judges (including gender), their position, and the work for which they are responsible, and we identify the gender of these judges directly. For courts that do not have a website or do not provide gender information, we search the judge's name together with the court's name on search engines such as *Baidu* or *Google*. Some news reports on judges' achievements or experiences state their gender directly. Lastly, we supplement our identification of the gender of the judges by using the China Court Trial Online website. On 11 December 2013, the China Court Live Trial website was officially launched to stream court trials.[16] As the website contains videos of parts of the trials, we visually identify the gender of the judges by watching the videos. With these steps, we are able to identify the gender of most of the judges. For the few remaining judges, we identify their gender by exploring their names; that is, those with recognizable female names. Online Appendix A provides detailed descriptions and examples.

One limitation of the publicly available judicial decision transcripts is that they typically provide only the names of judges and their acting status (i.e., whether the judge is a junior judge who has recently obtained adjudicative authority). To complement our analysis, we employ a systematic approach using multiple publicly accessible sources to collect more judge-level characteristics. Data Appendix B provides the details of data collection. Specifically, using publicly accessible sources such as court websites, news reports and public profiles, judicial selection announcements, university alumni networks, and search engines, we systematically collect and cross-check data over several iterations. These efforts allow us to obtain additional characteristics for a subset of judges, including their year of birth (8.13% of judges), judicial ranks (13% of judges), educational attainment (7.6% of judges) and whether the judge is local (2.98% of judges).

From this enriched dataset, we create four new judge-level variables: (1) judge age, defined as the year in which the case was adjudicated minus the judge's year of birth; (2) senior judge status, a binary indicator equal to one if the judge holds a rank of Level 4 Senior Judge or higher, which corresponds to the median rank in our sample;[17] and (3)

[15]In Section 3.3 and Appendix Table C4, we conduct several checks that sample attrition due to missing plaintiff gender information does not pose a threat to our identification strategy, and our findings remain robust to concerns about sample selection bias.

[16]China Court's Live Trial Website: http://tingshen.court.gov.cn.

[17]China's judicial ranking system comprises twelve levels: Chief Justice, Level 1 Supreme Court Justice,

graduate degree status, indicating whether the judge holds a postgraduate (master's or doctoral) degree; and (4) local judge, indicating whether the judge is a local resident of the jurisdiction in which the case is tried.

In addition, using the judge's name and court affiliation, we construct two further variables: (1) number of cases handled, defined as the cumulative number of cases adjudicated by the judge from their first appearance in the dataset up to and including the current case; and (2) experienced judge, a binary indicator equal to one if the number of cases handled exceeds the sample median.

From the judicial decision transcripts, we further extract several plaintiff-level characteristics. These include whether the plaintiff holds an urban hukou (i.e., urban residence registration), whether the plaintiff hired an attorney, whether the plaintiff appeared in court, as well as the plaintiff's year of birth and year of marriage. Using this information, we construct the following variables: (1) Urban hukou, a binary indicator equal to one if the plaintiff holds urban residence registration; (2) Attorney, indicating whether the plaintiff hired legal representation; (3) Present in court, a binary indicator for whether the plaintiff appeared in court; (4) Age, calculated as the year the case was handled minus the plaintiff's year of birth; (5) Age at marriage, calculated as the year of marriage minus the year of birth; and (6) Marriage duration, defined as the year of case handling minus the year of marriage.

Table 1 presents summary statistics for cases, judges, and plaintiffs separately by plaintiff gender, along with tests of differences between these groups. Panel A reports characteristics of divorce cases. Cases filed by female plaintiffs are significantly more likely to involve allegations of domestic violence, drug addiction, incarceration, gambling, frequent quarrels, long-term separation, and blind-date marriages. They also more often involve disputes over child custody and tend to include a greater number of sons. In contrast, cases filed by male plaintiffs are more likely to cite the spouse's extramarital affairs, disappearance, or leaving home as the cause of action, and they involve a greater number of daughters on average.

[Insert Table 1 here]

Figure 1 compares divorce grant rates by judge and plaintiff gender. Panel A presents raw means. Among male judges, the divorce grant rate is approximately 27.2% for male plaintiffs and 28.1% for female plaintiffs. Among female judges, the difference is more pronounced:

Level 2 Supreme Court Justice, Level 1 Senior Judge, Level 2 Senior Judge, Level 3 Senior Judge, Level 4 Senior Judge, and Level 1 through Level 5 Judge. Higher ranks typically reflect greater experience and assignment to more complex cases. Our dataset does not include the top three ranks—Chief Justice and Supreme Court Justices—who primarily serve in the Supreme People's Court and do not preside over first-instance trials.

23.2% versus 25.8%. Panel B shows conditional means after controlling for court-by-time fixed effects. The pattern remains: female plaintiffs are more likely to receive favorable rulings, especially from female judges. These results indicate a remarkable difference in the divorce grant rates across gender identities, especially among female judges. However, one concern with this raw comparison is that the differences in acceptance ratio across judge gender and across plaintiff gender may reflect the differences in cases filed by different plaintiffs (as documented in panel A of Table 1). To address this identification issue, we present our estimation strategy using the random assignment of cases to judges in the Chinese divorce cases in the next section.

[Insert Figure 1 here]

# 3 Empirical Analyses

## 3.1 Estimation Strategy

To estimate the extent of judicial in-group gender bias, we leverage the random case assignment system implemented in Chinese courts, as described in Section 2.1. According to the official mandate of China's Supreme People's Court, within the same lower court, each case is randomly assigned to a judge who presides over the trial, plea bargaining, and sentencing process. This implies that the assignment of judges to cases within a given court is in principle orthogonal to characteristics of the case, i.e., cases assigned to male and female judges are comparable. The differences in judicial decisions on similar cases but different plaintiff genders can then be attributed to the gender bias of the judges.

Specifically, we estimate the following specification:

$$\begin{aligned} Y_{ijct} = \beta_0 &+ \beta_1 FemalePlaintiff_i \times FemaleJudge_j \\ &+ \beta_2 FemalePlaintiff_i + \beta_3 FemaleJudge_j + \lambda_{ct} + \varepsilon_{ijct}, \end{aligned} \tag{1}$$

where $Y_{ijct}$ is the outcome, indicating whether the divorce request in case $i$ assigned to judge $j$ in court $c$ at time $t$ was granted; The variable $FemalePlaintiff_i$ equals one if case $i$ involves a female plaintiff; $FemaleJudge_j$ equals one if the presiding judge $j$ of case $i$ is female; $\lambda_{ct}$ denotes court fixed effect interacted with time (year-month of court hearing) fixed effects; and $\varepsilon_{ijct}$ is the error term. To account for potential heteroscedasticity and serial autocorrelation, we cluster the standard errors at the judge level (as recommended in Bertrand et al. 2004).

In spirit, our identification takes the form of a difference-in-differences (DD) design, accounting for two possible differences across gender groups that may not indicate in-group gender bias. First, $\beta_2$ captures differences in case composition across gender plaintiffs. That is, $\beta_2$ may be nonzero even in the absence of in-group bias if cases submitted by female plaintiffs systematically differ from those submitted by male plaintiffs. Second, the judge strictness may differ across gender, captured by $\beta_3$. In other words, $\beta_3$ may be nonzero if female judges are overall more or less strict than male judges, independent of the plaintiff's gender. Our DD-type estimation isolates whether, conditional on these factors, female judges exhibit systematically more favorable (or unfavorable) treatment toward female plaintiffs relative to male plaintiffs—a direct test of in-group bias (see Appendix Table C1 for a simple illustration of the DD design). Specifically, the coefficient of interest $\beta_1$ captures the in-group gender bias, i.e., whether female plaintiffs are granted divorce at a higher (or lower) rate than male plaintiffs when randomly assigned to female judges.

To further address potential confounding from unobserved judge or case characteristics, we estimate an augmented model:

$$\begin{aligned} Y_{ijct} = \beta_0 + \beta_1 FemalePlaintiff_i \times FemaleJudge_j + \beta_2 FemalePlaintiff_i \\ + \lambda_j + ActingJudge_{jt} + X_i' \times FemaleJudge_j + \lambda_{ct} + \varepsilon_{ijct}, \end{aligned} \tag{2}$$

where $\lambda_j$ denotes judge fixed effect, enabling the identification from case decisions across plaintiff gender within the same judge.[18] Meanwhile, to contain the possibility that junior judges may behave differently in the early stages of their careers, we include an indicator $ActingJudge_{jt}$ that takes the value one if the case is assessed by an acting judge $i$ in time $t$. The vector $X_i'$ is a list of case-specific characteristics listed in panel A of Table 1, interacted with with *FemaleJudge* indicator to further account for the possibility that female judges respond differently to the types of cases brought by female plaintiffs, rather than it being due to in-group bias. Specifically, $X_i'$ include whether the defendant engaged in gambling, was incarcerated, committed domestic violence, used drugs, or had an extramarital affair, whether the defendant left the home, whether the parties frequently quarreled over trivial matters, whether the defendant's whereabouts were unknown, and whether the couple had been living apart for an extended period, whether the couple met through a blind date, whether the case involved disputes over child support, and the number of sons and daughters.

[18]The judge fixed effect picks up any time invariant judge characteristics that may affect her rulings. Note that including judge fixed effects absorbs the main effect of judge gender *FemaleJudge* from the equation(1).

**Balancing test.** Unbiased estimation of $\beta_1$ requires that, conditional on the controls, the interaction term $FemalePlaintiff_i * FemaleJudge_j$ is uncorrelated with the error term $\varepsilon_{ijct}$.To assess this identifying assumption, we conduct a balancing test following following Bagues et al. (2017), testing whether predetermined case characteristics are randomly distributed across judge-plaintiff gender pairings.

Table 2 reports these results. In Panel A, we regress case characteristics on judge and plaintiff gender interactions with court-by-time fixed effects; Panel B adds judge fixed effects and controls for acting judge status. Across both panels, most coefficients are economically and statistically insignificant, indicating balance in observable case characteristics. The only exception is the indicator for extramarital affairs (*HaveAffair*), which is statistically significant but small in magnitude.

[Insert Table 2 here]

In Column 14, we test whether the case characteristics are jointly significant. The estimation result shows that the coefficient is statistically significant but very small in magnitude (in the absolute term and relative to the sample mean). Together, these sets of results support the validity of our identification strategy. As an additional precaution, our main specifications include interactions of case characteristics with the female judge indicator to further isolate the in-group gender effect.

## 3.2 Baseline Findings

Table 3 presents the baseline regression results. Column (1) reports estimates from equation (1), which includes only court-by-time fixed effects as controls.[19] Interpretation of coefficients in this difference-in-differences framework is provided in Appendix Table C1. Specifically, to our central interest, the coefficient ($\beta_1$) on the interaction term between *FemalePlaintiff* and *FemaleJudge* is estimated at 1.8 percentage points and is highly statistically significant. This result provides causal evidence of in-group gender bias in judicial decisions, consistent with the raw patterns shown in Figure 1. That is, female judges are 1.8 percentage points more likely than male judges to grant divorce claims when the plaintiff is female rather than male.

[Insert Table 3 here]

[19]The inclusion of court-time fixed effects leads to the exclusion of 13,246 singleton observations, resulting in a final estimation sample of 527,279 cases.

In the meantime, the individual coefficient on *FemaleJudge* ($\beta_3$) is negative and statistically significant, indicating that female judges are, on average, less likely than male judges to grant divorce claims submitted by male plaintiffs. In contrast, the coefficient on *Female-Plaintiff* ($\beta_2$) is close to zero and statistically insignificant, suggesting that among male judges, there is no systematic difference in the acceptance rate between male and female plaintiffs.[20] Taken together, these estimates imply that female judges are more restrictive overall than male judges, and this restrictiveness is especially pronounced for male plaintiffs.

Columns (2) through (4) of Table 3 report results from augmented specifications; that is, stepwisely adding judge fixed effects, case characteristics (interacted with the female judge indicator), and an indicator for acting judge status. Across all specifications, the estimated interaction effect remains positive, statistically significant, and quantitatively stable, reinforcing the robustness of the in-group gender bias finding.

In terms of economic significance, the estimated magnitude of the interaction term from the fully specified model (in the last column) suggests that, relative to male judges, female judges are 1.2 percentage points more likely to grant divorce claims submitted by female (as opposed to male) plaintiffs.

These findings are consistent with insights from psychology and sociology, which document stronger in-group bias among women than men. For example, Pratto et al. (1997) find that women tend to identify more strongly with their gender group. Similarly, Rudman and Goodwin (2004) show that while both men and women express favorable views toward women, women exhibit in-group bias that is 4.5 times stronger than that of men. They also find that only women exhibit cognitive alignment among in-group bias, identity, and self-esteem, suggesting that men lack a mechanism that fosters automatic in-group preference. Evidence from other domains, such as policy-making, supports this pattern. For instance, Chattopadhyay and Duflo (2004) show that female leaders in India were more likely to prioritize policies benefiting women.

*Comparison with the literature.* This study's estimated in-group gender bias of 1.2 percentage points—corresponding to a 4.5% increase relative to the mean probability of granting divorce (26.5%)— is modest compared to recent studies employing quasi-random assignment methods with naturally occurring data to investigate judicial in-group bias. For instance, Hoekstra and Street (2021), using data from jury trials in two large Florida counties, document a substantial effect, finding that having at least one own-gender juror reduces convic-

[20] It is important to note that although the baseline acceptance rate for female plaintiffs is higher when cases are assigned to male judges (as $\beta_1 + \beta_3 < 0$), this reflects the overall more restrictive adjudication style of female judges (reflected by $\beta_3 < 0$), consistent with prior evidence on gender differences in judicial strictness (Songer and Crews-Meyer 2000; Boyd et al. 2010).

tion rates in drug trials by about 30 percentage points (a 43% relative reduction). Similarly, Didwania (2022) examines sentencing outcomes in U.S. federal courts and finds that gender-matching between prosecutors and defendants reduces sentence length by approximately 5 months, or about 8% relative to the mean sentence length. These studies highlight how gender-matched dynamics between jurors or prosecutors and defendants significantly influence outcomes, particularly in high-stakes criminal trials and sentencing decisions.[21]

Several contextual and institutional factors likely account for the modest magnitude observed in our study, aligning more closely with Ash et al. (2025) who find negligible in-group biases in the Indian judiciary despite examining over five million criminal cases. First, cultural and societal attitudes prevalent in developing countries like China may temper overt demonstrations of gender bias. Specifically, our heterogeneity analysis below shows that in-group gender bias is indeed less pronounced in regions with stronger traditional gender norms—that is, higher initial gender discrimination against women—suggesting conservative cultural attitudes may restrain explicit favoritism. Second, both studies examine judicial systems within large, diverse developing countries (China and India), where aggregation across heterogeneous regions likely dilutes localized bias and then reduces observed magnitudes at the national level. Third, both studies use binary judicial outcome variables (i.e., divorce granted versus rejected in our study, and conviction outcomes in Ash et al. (2025)), potentially understating the extent of bias compared to continuous measures such as sentence length (Didwania (2022)) or citation patterns (Ash et al. (2024)), which capture subtler decision-making nuances. Finally, publication bias may amplify the visibility of larger, statistically significant effects documented in existing literature, as noted by Ash et al. (2025), underscoring the importance of contextualizing modest yet meaningful findings like ours.

## 3.3 Robustness checks.

We conduct several tests to examine the robustness of our results. These include a placebo test with simulated assignment of judges to plaintiffs, a leave-one-court-out analysis to address outlier influence, using an alternative inference method, keeping courts with at least one female judge, and examining the potential sample selection issue due to missing plaintiff gender information.

*Placebo test with simulate case assignment.* We conduct a placebo test in which we

[21] Ash et al. (2024) examines judicial decisions in U.S. Circuit Courts and focuses on the role of judges' implicit gender attitudes in rulings and professional interactions. They find that judges with traditional gender attitudes are 4.1 percentage points (7% of mean) less likely to support expanding women's rights in gender-related cases and 1.7 percentage points (4.5% of mean) less likely to assign majority opinions to female colleagues.

fabricate the assignment of judges to plaintiffs while preserving the number of cases assigned to each judge (following approaches in Chetty et al. 2009; La Ferrara et al. 2012). Under this placebo design, the constructed regressor of interest should not exhibit any significant effect on case outcomes; otherwise, it would suggest confounding factors. To enhance statistical power, we repeat the simulation 1,000 times. Appendix Figure C1 plots the distribution of these placebo estimates alongside our benchmark estimate (0.012 from column (4) in Table 3). The distribution of the placebo estimates is tightly centered around zero, while our actual estimate lies far outside this distribution, lending further support to our research design and baseline estimates.

*Outlier influence.* To assess whether our results are driven by specific courts, we conduct a leave-one-court-out analysis, re-estimating equation (1) while iteratively dropping each court from the sample. Appendix Figure C2 presents the results. In all iterations, the estimated coefficient on the interaction term $MalePlaintiff_i * MaleJudge_j$) remains statistically significant and sharply clustered around the baseline estimate, indicating that our results are not driven by outliers.

*Alternative inference approach.* In our baseline, we estimate the standard error clustered at the judge level. To assess whether our results are sensitive to the way we cluster the standard error, we experiment with an alternative inference approach; that is, cluster the standard error at the court-time level. Estimation results are reported in column 4 of Appendix Table C2. We continue to find statistical significance, strengthening the confidence of our estimates.

*Courts without female judges.* Our identification relies on random assignment of cases to both male and female judges within a court. In courts with no female judges, this strategy cannot be implemented. To address this, we re-estimate the model excluding courts without any female judges. Results, shown in column in column 5 of Appendix Table C2, remain statistically significant and comparable in magnitude to our baseline estimates, further supporting the robustness of our conclusions.

*Collegial Panel Cases.* Our baseline analysis focuses on the cases presided by single judges to avoid the complications introduced by collegial panel cases (i.e., due to the missing information of other panel judges and the collective decision). To assess the external validity of our findings, we replicate the analysis with full sample including collegial panel cases, and report the results in Appendix Table C3. Across all specifications, the estimated interaction term remains positive and statistically significant, reinforcing the robustness of the in-group gender bias result.

In terms of magnitude, the estimated in-group gender bias is slightly larger in the full sample including collegial panels than in single-judge cases. This pattern is likely explained by the structure of collegial panels: while the data record only the gender of the principal judge, actual decisions reflect the deliberation of the full panel. If the gender composition of the panel is systematically correlated with that of the principal judge (e.g., female principal judges may more often serve on panels with a higher proportion of female members), then the interaction between the principal judge's gender and the plaintiff's gender may partly capture the collective in-group bias of the entire panel, rather than a pure principal-judge effect. This compositional effect would naturally lead to an upward bias in the estimated in-group gender bias for collegial panel cases.

*Missing gender information of the plaintiff.* Approximately 3.28% of the cases in our dataset lack information on the plaintiff's gender, primarily due to incomplete entries in the judicial decision transcripts. In Appendix Table C4, we examine whether cases with missing plaintiff gender information systematically differ from those with complete information Specifically, Columns 1 to 7 examine differences across judge characteristics. Cases missing plaintiff gender information are somewhat less likely to be presided over by male judges or by judges who have handled a larger number of cases, and more likely to be presided over by older and more senior judges. While these differences are statistically significant in some cases, their magnitudes are small relative to the sample means. Columns 8 to 20 further analyzes case characteristics, finding that cases lacking plaintiff gender information differ significantly on most of these dimensions, though again the magnitudes of these differences are relatively small.

These results may raise concerns about potential sample attrition, since our analysis sample is effectively selected on the presence of plaintiff gender information. Note that the proportion of cases missing plaintiff gender information is relatively small and any significant differences are small in magnitude, implying that the sample attrition issue may not cause significant estimation biases on our baseline estimates.

Nonetheless, we further conduct two exercises to quantitatively examine the relevance of the sample attrition issue in our setting. First, we implement the inverse-probability-weighted (IPW) method developed by Wooldridge (2007). That is, we first estimate the probability that a case does not contain plaintiff gender information, and then use the inverse of this estimated probability as a weight in our baseline specification. The results, reported in column (6) of Appendix Table C2, yield a coefficient nearly identical to our baseline estimate.

Second, we consider two extreme cases following the spirit by Lee (2009) to obtain the

bounds of our potential effects. Specifically, we construct extreme bounds by alternatively assuming that all cases with missing plaintiff gender information correspond either to female or male plaintiffs (reported in columns (7) and (8) of Appendix Table C2). The resulting estimates of 0.011 and 0.012, respectively, bracket our baseline estimate and suggest limited sensitivity to this source of attrition.

Taken together, these analyses suggest that the sample attrition from missing plaintiff gender information does not cause any significant problems with our estimation.

## 3.4 Competing Interpretations

The preceding results document a clear pattern of in-group gender bias in judicial decisions: female judges treat female plaintiffs more favorably relative to male plaintiffs in comparable cases, while male judges exhibit no significant differential treatment across plaintiff gender. Our preferred explanation is that this pattern reflects in-group gender bias—that is, judges' differential treatment is influenced by shared gender identity with the plaintiff. In this section, we consider three alternative explanations for the observed empirical pattern.

*Information and communication.* One possible alternative mechanism is that variation in judicial decisions across judge and plaintiff gender reflects differences in the quality of communication or information transmission. In this view, while judges may apply neutral standards, they may more readily understand or interpret arguments presented by members of their own gender due to subtle linguistic or communicative dynamics, particularly in cases requiring nuanced factual assessment or credibility judgments.[22] If this mechanism were important, we would expect the in-group gender bias to be more pronounced in cases that are more complex—those that require greater judicial reliance on the plaintiff's narrative and the quality of communication—than in cases where decisions are based on clearly defined statutory criteria.

We test this potential mechanism empirically by examining whether the in-group gender bias varies across cases of differing complexity that require different efforts of arguments. Specifically, according to Article 32 Chapter IV of the Marriage Law of the People's Republic of China, divorce shall be granted if mediation fails and the case involves certain circumstances, including bigamy, domestic violence, drug abuse, imprisonment, or gambling. These circumstances are typically more straightforward for judges to assess, and the law allows them to grant a divorce directly based on these grounds. In contrast, cases lacking

[22]In practice, we believe that the differences in the quality of in-group-in-group versus in-group-out-group communication are unlikely to drive our results, because divorce cases tend to be simple and judicial decisions essentially concern whose version of the events to accept instead of sophisticated lines of argumentation.

these grounds—typically involving disputes over emotional relationships or financial matters—require the judge to rely on more information from both parties and often require additional communication and supporting evidence.

We classify cases into "straightforward" and "complicated" categories based on whether the judgment includes keywords corresponding to statutory grounds for divorce (domestic violence, drug abuse, imprisonment, gambling, or extramarital affairs). We then estimate a specification that interacts the straightforward case indicator with our main regressor of interest to test for differential in-group bias across case types.

Results, reported in column (1) of Table 4, show that the coefficient on the interaction between female judge and female plaintiff remains statistically significant and of similar magnitude to our baseline estimate. Importantly, the triple interaction term is both statistically and economically insignificant. These findings indicate that in-group gender bias does not vary meaningfully between straightforward and complicated cases, suggesting that differences in communication or information transmission are unlikely to account for our results.

[Insert Table 4 here]

Meanwhile, if differences in the quality of communication or information transmission were driving the observed pattern, we would expect heavier caseloads and associated cognitive fatigue to impair judges' ability to fully process case information, making them more prone to rely on stereotypes and thereby amplifying in-group gender bias. To test this possibility, we use a judge's monthly workload (measured as whether the number of cases closed in the corresponding month is above the sample median), interacted with the main regressor to capture the heterogeneity in-group gender bias. The results, reported in column 2 of Table 4, do not support this explanation. The estimated coefficient on the triple interaction term (between female judge, female plaintiff, and high workload) is close to zero and statistically insignificant. This finding indicates that the magnitude of in-group gender bias does not systematically vary with judicial workload, suggesting that cognitive fatigue or information processing differences are unlikely to be the primary driver of the bias.

*Plantiff behavior.* The second alternative interpretation is that it is plaintiffs' behavior, rather than judges' behavior, that drives the results. Price and Wolfers (2010) provides an intuitive example to illustrate this alternative interpretation. They examine National Basketball Association (NBA) refereeing decisions and find that more personal fouls are called against players when the games are officiated by an opposite-race refereeing crew than a same-race crew. An important difficulty that arises in the NBA context is that player

behavior may depend on the racial composition of the refereeing crew. For example, white players may play more aggressively when the refereeing crew consists of more black referees. In the NBA setting, when a player makes a foul, he already knows the racial composition of the refereeing crew, which may conceivably affect his behavior. Analogously, in our setting, plaintiffs may perform better in the courts when they face a same-gender judge.

However, we do not observe the plaintiff's behavior in the courtroom (we only have access to the decision document written by the judge), we cannot directly rule out this possibility. Instead, we conduct some indirect tests to shed light on this alternative. Specifically, we implement a heterogeneity analysis by judges' work experience. The premise of the test is that if plaintiffs' behavior is the main channel, we may expect a less experienced judge to be more likely to be affected by the plaintiff's court performance.[23] To this end, we interact our main regressor with an indicator for judge experience, defined as whether the number of cases handled by the judge between their first appearance in our dataset and the case date exceeds the sample median. Estimation results are reported in column 3 of Table 4. The estimated coefficient of the interaction between female judge and female plaintiff remains positive and statistically significant, confirming our main findings. Meanwhile, the triple interaction among female judge, female plaintiff, and the high-experience indicator is small in magnitude and statistically insignificant, indicating that the in-group gender bias is similar for judges with both greater and lesser levels of experience.

One concern with this analysis is that judges may have prior unobserved experience before entering our data. Given that the full career information of judges is unavailable to us, we conduct another heterogeneity analysis by using the seniority information of judges in a subsample to shed light on the check on the plaintiff's behavior story. Estimation results for the heterogeneity analysis across judicial seniority are reported in column 4 of Table 4. We continue to find a positive coefficient of the interaction between female judge and female plaintiff with a similar magnitude but less precise due to the small sample size. Meanwhile, the triple interaction among female judge, female plaintiff, and the indicator for senior judge is highly insignificant with a magnitude close to zero, indicating the estimated in-group gender bias is similar for judges with different seniority.

Alternatively, we examine the heterogeneity effects across judge age, as older judges tend to be more experienced. Indeed, in the subsample with judge age information (around 8.13% of the full regression sample), we find a positive correlation between judge experience and age. Estimation results are reported in column 5 of Table 4. We again find a positive

[23] One caveat with this test is that judge experience may affect other things, such as how judges process information. We examine the relevance of the information channel in the previous analysis, and also examine whether experience correlates with age and how age affects the heterogeneity effects.

estimated coefficient for the interaction term between male judge and male plaintiff despite being imprecisely estimated due to the small sample. Meanwhile, the triple interaction term is small in magnitude and statistically insignificant, indicating similar effects for old and young judges.

Furthermore, we examine heterogeneity across acting and formal judges. Acting judges are those who have passed the judicial examination and recently started working in the court system, while formal judges are more experienced. In column 6 of Table 4, we interact our main regressor with an indicator for acting judge status. The triple interaction term is statistically insignificant, and the main interaction effect remains positive and significant, indicating that the in-group gender bias does not vary between acting and formal judges.

Taken together, these results suggest that it is unlikely that plaintiffs' behavior explains the main pattern.

*Characteristics correlated with judge gender.* A third alternative interpretation is that the estimates are likely to be driven by some other judge characteristics associated with gender. As shown in Panel B of Table 1, male judges tend to be older and more experienced; female judges are more likely to be acting judges, more senior, and have higher educational attainment. To address this, we follow Shayo and Zussman (2011) and control for these judge characteristics by introducing interaction terms between judge characteristics and plaintiff's gender.

Estimation results are reported in Table 5. With the inclusion of these additional controls for judge characteristics that may confound the effect of judge gender, our main regressor of interest (the interaction term between female judge and female plaintiff) remains positive, although some estimates are less precise due to smaller sample sizes.[24] These results suggest that our main findings on in-group gender bias are not significantly confounded by other judge characteristics correlated with judge gender.[25]

[Insert Table 5 here]

[24]We acknowledge substantial missing data for several judge characteristics—age, seniority, educational attainment, and nonlocal judge status—in columns (4)-(7). To assess whether this missing data introduces sample selection concerns, we examine whether the probability of missingness systematically varies across female judge and female plaintiff. Results, reported in columns (1)-(4) of Appendix Table C5, show statistically and economically insignificant coefficients with magnitudes close to zero, suggesting no substantial sample attrition problem.

[25]To further address the possibility that the documented in-group gender bias is driven by other judge characteristics, such as acting judge status or judge age, we examine heterogeneity in treatment effects across young and old judges and across acting and formal judges (columns (4)-(5) of Table 5). In both cases, the triple interaction terms are statistically and economically insignificant, indicating that the in-group gender bias is similar across judge age groups and judge types.

*Characteristics correlated with plaintiff gender.* It is also possible that the documented in-group gender bias reflects not plaintiff gender per se, but other plaintiff characteristics correlated with gender. As shown in Panel C of Table 1, male plaintiffs are more likely to hold an urban hukou, are older, marry later, have longer marriage durations, are less likely to appear in court, and are more likely to hire legal representation than female plaintiffs.

Using an approach analogous to that in Table 5, we control for the potential confounding influence of these characteristics by including interaction terms between plaintiff characteristics and judge gender. The estimation results are reported in Table 6.[26]

Across all specifications, the coefficient on our main interaction term remains positive and statistically significant, with only small variations in magnitude. These findings indicate that our documented in-group gender bias is not significantly confounded by other plaintiff characteristics correlated with gender.

[Insert Table 6 here]

To summarize, while we cannot definitively rule out all alternative explanations, the evidence presented in this subsection suggests that they are unlikely to fully account for the observed pattern of in-group gender bias in judicial decisions.

## 3.5 Heterogeneity Analysis

Our analyses establish the presence of in-group gender bias in divorce cases. To further understand this bias and assess its external validity, we examine patterns of heterogeneity across regions and individuals. Specifically, we first explore whether the magnitude of in-group gender bias varies across regions with differing prevailing gender attitude, to help assess whether cultural context shapes the extent to which such bias is expressed in judicial decisions. We then investigate whether in-group gender bias varies with regional differences in legal conditions, providing insight into the generality and institutional robustness of our findings.

*Initial attitudes toward gender.* In China, provinces differ substantially in prevailing gender attitudes. In regions with stronger gender discrimination, men occupy higher social and familial status, and son preference remains strong; in other provinces, gender roles

[26]Note that there are substantial reductions in sample size due to missing data on certain plaintiff characteristics in columns (2) and (5)-(8). To address potential sample attrition concerns, we examine whether the probability of missing data systematically varies across male judge and male plaintiff. Results, reported in columns (5)-(9) of Appendix Table C5, show statistically and economically insignificant coefficients with magnitudes close to zero, suggesting no substantial sample attrition problem.

are more egalitarian. These patterns reflect deep-rooted Confucian values that historically emphasized male dominance and female submission within families and society. Despite significant economic and social development, traditional Confucian values that prioritize male dominance persist, particularly in rural and more conservative provinces. This enduring cultural context shapes societal expectations around gender roles, potentially conditioning the expression of gender biases, including in judicial decisions.

We use two complementary measures to capture regional variation in gender discrimination. The first measure is the province-level sex ratio from the 2010 census. Given the biological consistency of the natural sex ratio at birth, observed variations largely reflect gender-selective abortions driven by strong son preference (Almond et al. 2019). The second measure derives from respondents' answers to gender-related questions from the 2014 wave of the China Family Panel Studies (CFPS 2014). Respondents rated their agreement (on a five-point Likert scale) with four statements reflecting traditional gender roles. Specifically: (1) Men should prioritize careers, women the family; (2) Marrying well is more important for women than professional success; (3) Women should have at least one child; and (4) Men should share half the housework (reverse-coded). Higher scores indicate stronger endorsement of traditional gender attitudes. We aggregate these responses by province to create a regional index of gender attitudes.

Both measures reflect societal gender biases, emphasizing different mechanisms. A highly skewed sex ratio signals entrenched son preference, creating pronounced societal expectations that women serve primarily as family caregivers. Meanwhile, regions with strong traditional gender attitudes explicitly reinforce normative beliefs about appropriate gender roles, often emphasizing women's dependence and vulnerability. Judges, despite professionally trained, are also socialized within these regional cultural contexts, which may influence their empathetic or evaluative responses toward plaintiffs of different genders.

Our data show that a substantial majority of judges (approximately 75.5%) serve in their home regions, making local gender norms highly salient to judicial behavior. Moreover, even non-local judges typically adapt to local social and cultural expectations to perform their duties effectively. Primary court proceedings frequently incorporate local dialects and customs, requiring judges—regardless of origin—to internalize local norms that likely influence their adjudicative perspectives.

Estimation results on the heterogeneity in the in-group gender bias across regional gender attitudes are reported in Columns 1 and 2 of Table 7. We find that the in-group gender bias among female judges is significantly weaker in regions with more traditional gender attitudes; specifically, the triple interaction terms are negative, with statistical significance for the survey-based measure. One possible interpretation is that conservative regional gender norms

impose social pressures on female judges to demonstrate formal impartiality and professional neutrality in patriarchal environments. As a result, female judges may consciously moderate overt displays of in-group favoritism toward female plaintiffs. Additionally, female judges in more traditional settings may adapt more strongly to patriarchal expectations—either by internalizing prevailing norms or by taking extra care to avoid perceptions of bias—thus limiting the expression of in-group gender bias. In contrast, male judges, who are generally less influenced by gender identification (as documented by Rudman and Goodwin 2004; Pratto et al. 1997), show no corresponding pattern, consistent with our baseline findings.

Overall, these results highlight a nuanced interaction between cultural context and the expression of in-group gender bias. Rather than uniformly amplifying bias, traditional gender norms appear to constrain its overt manifestation among female judges. These findings underscore the complex relationship between regional culture and judicial behavior, and the broader importance of addressing cultural norms and attitudes as part of efforts to promote gender equity within judicial institutions.

[Insert Table 7 here]

*Initial level in legal development.* The level of legal system development varies substantially across regions in China. In some provinces—particularly coastal areas—authorities exhibit greater adherence to legal procedures and stronger institutional capacity, resulting in more standardized trials. In other provinces—such as parts of the northeast—trial practices remain more discretionary and less formalized. We examine whether the in-group gender bias documented above varies systematically with regional differences in legal system development.

To this end, we construct two measures of regional legal development. The first is based on responses to legal-related questions in the 2010 wave of the China General Social Survey (CGSS), including: (1) trust in the courts and judicial system, (2) legal knowledge, and (3) frequency of legal compliance.[27] Responses are recorded on a five-point Likert scale (1 = *completely unknown* to 5 = *completely known*), with higher scores indicating stronger legal development. We compute province-level averages to form an index of regional legal development. Our second measure draws on a subindex of the widely used National Economic Research Institute (NERI) Marketization Index, which captures provincial variation in legal environment quality based on a combination of original survey data and official statistics.

Estimation results are reported in columns 3 and 4 of Table 7. The estimated interaction

[27]Legal knowledge covers basic legal understanding, familiarity with court functions, and knowledge of how to access legal services such as hiring a lawyer, seeking legal assistance, or filing a lawsuit.

between female judge and female plaintiff remains similar in magnitude and significance to our baseline results. The triple interaction terms with both legal development measures are statistically and economically insignificant, indicating that the degree of in-group gender bias does not vary meaningfully across regions with different levels of legal system development. These findings suggest that the observed in-group gender bias is pervasive and largely invariant to institutional legal quality.[28]

# 4 Conclusion

In this paper, we provide novel evidence of in-group gender bias in judicial decision-making using a large-scale dataset comprising nearly all divorce cases handled by basic civil courts across China. Exploiting the effectively random assignment of cases to judges, we identify a clear pattern of gender-based favoritism. Specifically, we find that female judges are 1.2 percentage points more likely than their male counterparts to grant divorce petitions filed by female plaintiffs relative to those filed by male plaintiffs. This effect is robust to a variety of checks, including alternative specifications, placebo tests, and controls for extensive judge and plaintiff characteristics. Our findings rule out alternative explanations such as access to information available to the judge or plaintiffs' courtroom behaviors, pointing clearly toward female judges' in-group gender bias as the most plausible explanation.

To further contextualize and examine the external validity of this bias, we examine heterogeneity across different regions and judicial contexts. We document that the magnitude of the bias varies systematically with regional gender attitudes but not with regional levels of legal system development. Interestingly, the observed gender bias among female judges is significantly smaller in regions with more traditional gender norms, suggesting that conservative cultural contexts may suppress overt expressions of gender-based favoritism. In contrast, institutional factors such as legal development appear to have minimal moderating effects, underscoring the stronger influence of cultural attitudes over formal institutions in shaping judicial behavior.

Our study contributes to the literature by providing one of the first comprehensive analyses of in-group gender bias in judicial decisions using naturally occurring large-scale judicial

[28]The contrast between the heterogeneity patterns for legal development and those for cultural gender attitudes highlights the distinct geographic distributions of these two dimensions (correlations ranging from -0.11 to -0.35). This divergence reflects the fact that legal development and gender norms are shaped by different historical and institutional processes. Specifically, the legal origins literature (Glaeser and Shleifer 2002; Lu and Tao 2009) shows that regional variation in legal development in China is historically linked to foreign influence during the late Qing Dynasty (1840–1911), whereas gender norms are rooted in local cultural practices, family structures, and agricultural traditions (Alesina et al. 2013).

data and a rigorous identification strategy based on random case assignment. While existing literature has extensively documented gender biases in various professional and societal contexts, evidence on such biases within high-stakes judicial decisions—especially within civil law contexts such as divorce in developing economies—has remained limited. Our findings underscore that even judges, typically viewed as impartial arbiters, may exhibit subtle biases shaped by shared gender identities.

These results carry important policy implications. As gender diversity on judicial benches continues to rise globally, our findings underscore that increasing gender representation, while desirable for equity and legitimacy, it may also introduce new challenges to impartial adjudication if identity-based biases persist. Without complementary safeguards, identity-based biases may persist or emerge. Effective responses include targeted judicial training on implicit bias, enhanced transparency and accountability in court procedures, and broader efforts to reshape cultural norms through education and institutional reform. Such interventions are essential to ensuring equal treatment under the law and to advancing substantive gender equity within judicial systems and beyond.

# References


**Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti**, "Revealing stereotypes: Evidence from immigrants in schools," *American Economic Review*, 2024, *114* (7), 1916–1948.

**_ , Paola Giuliano, and Nathan Nunn**, "On the origins of gender roles: Women and the plough," *The quarterly journal of economics*, 2013, *128* (2), 469–530.

**Almond, Douglas, Hongbin Li, and Shuang Zhang**, "Land reform and sex selection in China," *Journal of Political Economy*, 2019, *127* (2), 560–585.

**Antonovics, Kate and Brian G Knight**, "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics*, 2009, *91* (1), 163–177.

**Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson**, "The Impact of Jury Race in Criminal Trials," *The Quarterly Journal of Economics*, 2012, *127* (2), 1017–1055.

**Ash, Elliott, Daniel L Chen, and Arianna Ornaghi**, "Gender Attitudes in the Judiciary: Evidence from US Circuit Courts," *American Economic Journal: Applied Economics*, 2024, *16* (1), 314–350.

**_ , Sam Asher, Aditi Bhowmick, Sandeep Bhupatiraju, Daniel L Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi**, "In-Group Bias in the Indian Judiciary: Evidence from 5 Million Criminal Cases," *Review of Economics and Statistics*, 2025, p. Forthcoming.

**Bagues, Manuel F and Berta Esteve-Volart**, "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *The Review of Economic Studies*, 2010, *77* (4), 1301–1328.

**Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva**, "Does the Gender Composition of Scientific Committees Matter?," *American Economic Review*, 2017, *107* (4), 1207–38.

**Bar, Revital and Asaf Zussman**, "Identity and Bias: Insights from Driving Tests," *The Economic Journal*, 2020, *130* (625), 1–23.

**Bauer, Gretchen and Josephine Dawuni**, *Gender and the Judiciary in Africa: from Obscurity to Parity?*, Routledge, 2015.

**Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan**, "How Much Should We Trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

**Bindler, Anna and Randi Hjalmarsson**, "The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions," *The Journal of Law and Economics*, 2020, *63* (2), 297–339.

**Blank, Rebecca M**, "The Effects of Double-Blind Versus Single-Blind Reviewing: Experi-

mental Evidence from the American Economic Review," *The American Economic Review*, 1991, pp. 1041–1067.

**Boring, Anne**, "Gender Biases in Student Evaluations of Teaching," *Journal of Public Economics*, 2017, *145*, 27–41.

**Boyd, Christina L., Lee Epstein, and Andrew D. Martin**, "Untangling the Causal Effects of Sex on Judging," *American Journal of Political Science*, 2010, *54* (2), 389–411.

**Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri**, "Gender Neutrality in Economics: The Role of Editors and Referees," 2020.

**Chari, Anusha and Paul Goldsmith-Pinkham**, "Gender Representation in Economics across Topics and Time: Evidence from the NBER Summer Institute," Technical Report, National Bureau of Economic Research 2017.

**Charness, Gary and Yan Chen**, "Social Identity, Group Behavior, and Teams," *Annual Review of Economics*, 2020, *12.*

**Chattopadhyay, Raghabendra and Esther Duflo**, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

**Chetty, Raj, Adam Looney, and Kory Kroft**, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–77.

**Didwania, Stephanie Holmes**, "Gender Favoritism among Criminal Prosecutors," *The Journal of Law and Economics*, 2022, *65* (1), 77–104.

**Ferrara, Eliana La, Alberto Chong, and Suzanne Duryea**, "Soap Operas and Fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, 2012, *4* (4), 1–31.

**Fix, Michael P and Gbemende E Johnson**, "Public Perceptions of Gender Bias in the Decisions of Female State Court Judges," *Vand. L. Rev.*, 2017, *70*, 1845.

**Glaeser, Edward L and Andrei Shleifer**, "Legal origins," *The Quarterly Journal of Economics*, 2002, *117* (4), 1193–1229.

**Hoekstra, Mark and Brittany Street**, "The Effect of Own-Gender Jurors on Conviction Rates," *The Journal of Law and Economics*, 2021, *64* (3), 513–537.

**Kenney, Sally Jane**, *Gender and Justice: Why Women in the Judiciary Really Matter*, Routledge, 2013.

**Kulik, Carol T, Elissa L Perry, and Molly B Pepper**, "Here Comes the Judge: The Influence of Judge Personal Characteristics on Federal Sexual Harassment Case Outcomes," *Law and Human Behavior*, 2003, *27* (1), 69–86.

**Lee, David S.**, "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

**Lee, Romy Van Der and Naomi Ellemers**, "Gender Contributes to Personal Research Funding Success in The Netherlands," *Proceedings of the National Academy of Sciences*,

2015, *112* (40), 12349–12353.

**Lu, Yi and Zhigang Tao**, "Contract enforcement and family control of business: Evidence from China," *Journal of Comparative Economics*, 2009, *37* (4), 597–609.

**Mengel, Friederike, Jan Sauermann, and Ulf Zölitz**, "Gender Bias in Teaching Evaluations," *Journal of the European Economic Association*, 2019, *17* (2), 535–566.

**Mustard, David B**, "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the US Federal Courts," *The Journal of Law and Economics*, 2001, *44* (1), 285–314.

**Pratto, Felicia, Lisa M Stallworth, and Jim Sidanius**, "The Gender Gap: Differences in Political Attitudes and Social Dominance Orientation," *British Journal of Social Psychology*, 1997, *36* (1), 49–68.

**Price, Joseph and Justin Wolfers**, "Racial Discrimination among NBA Referees," *The Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

**Reuben, Ernesto, Paola Sapienza, and Luigi Zingales**, "How Stereotypes Impair Women's Careers in Science," *Proceedings of the National Academy of Sciences*, 2014, *111* (12), 4403–4408.

**Rudman, Laurie A and Stephanie A Goodwin**, "Gender Differences in Automatic in-group Bias: Why do Women Kike Women More Than Men like Men?," *Journal of Personality and Social Psychology*, 2004, *87* (4), 494.

**Shayo, Moses and Asaf Zussman**, "Judicial Ingroup Bias in the Shadow of Terrorism," *The Quarterly Journal of Economics*, 2011, *126* (3), 1447–1484.

**_ and _**, "Conflict and the Persistence of Ethnic Bias," *American Economic Journal: Applied Economics*, 2017, *9* (4), 137–65.

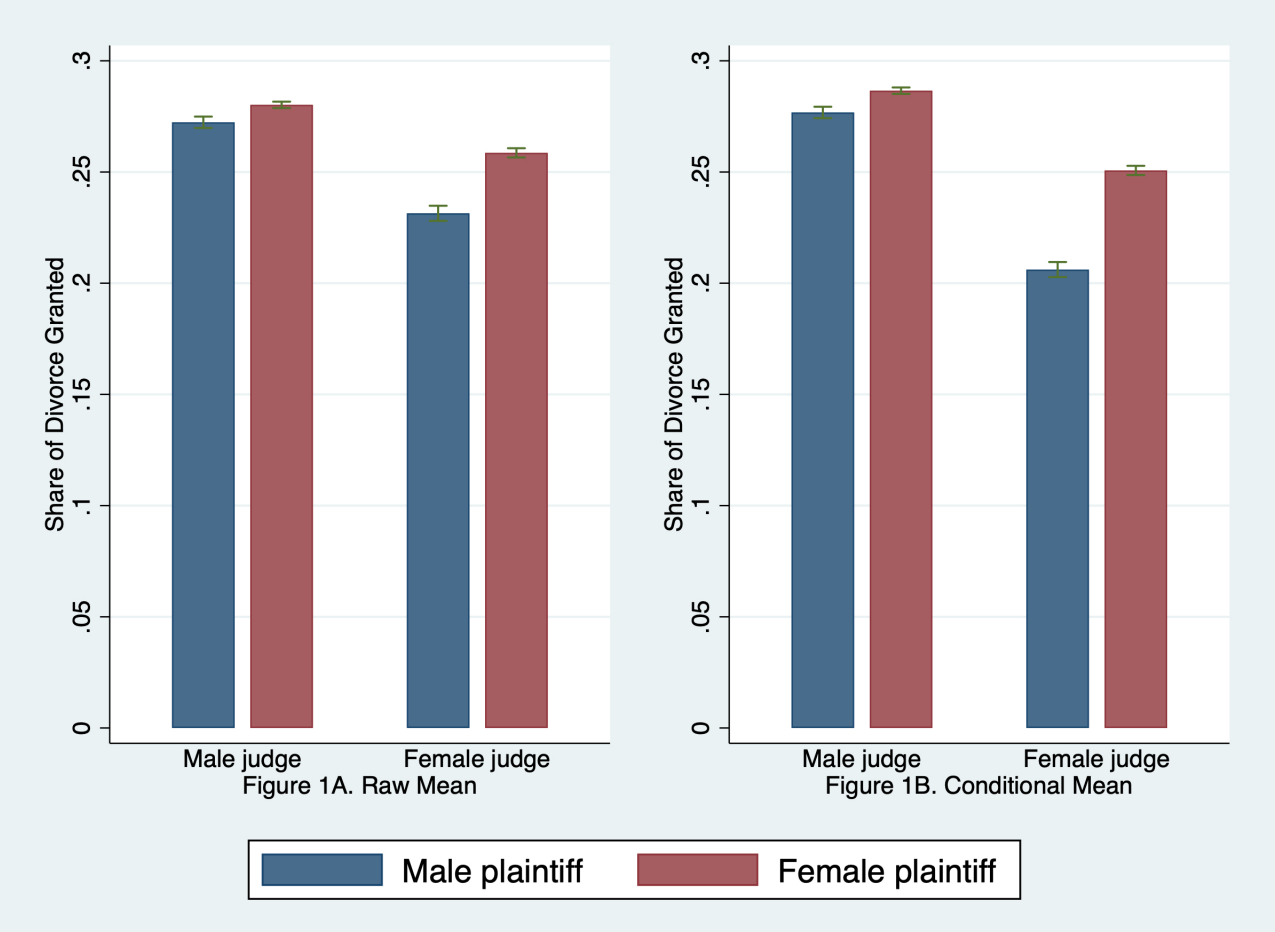
**Songer, Donald R. and Kelley Crews-Meyer**, "Does Judge Gender Matter? Decision Making in State Supreme Courts," *Social Science Quarterly*, 2000, *81* (3), 750–762.

**West, Jeremy**, "Racial Bias in Police Investigations," 2018.

**Wooldridge, Jeffrey M.**, "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 2007, *141* (2), 1281–1301.

# Figures and Tables

Figure 1: Divorce Grant Rates by Judge and Plaintiff Gender




*Notes:* This figure compares the divorce grant rates across judge and plaintiff gender combinations. Panel A displays the raw mean grant rates for each group. Panel B presents the conditional mean grant rates after accounting for court-time fixed effects.

Table 1: Summary Statistics

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
|---|---|---|---|---|---|---|---|
| **Panel A: Cases** | Male Plaintiffs | | | Female Plaintiffs | | | [2]-[5] |
| | Obs | Mean | SD | Obs | Mean | SD | Diff |
| Divorce (Accept) | 134,072 | 0.2584 | 0.4378 | 406,453 | 0.2734 | 0.4457 | -0.015*** |
| Female judge | 134,072 | 0.3414 | 0.4742 | 406,453 | 0.3141 | 0.4642 | 0.027*** |
| Custody | 134,072 | 0.3991 | 0.4897 | 406,453 | 0.4564 | 0.4981 | -0.057*** |
| Domestic Violence | 134,072 | 0.0252 | 0.1568 | 406,453 | 0.0686 | 0.2527 | -0.043*** |
| Drug Addiction | 134,072 | 0.0139 | 0.1172 | 406,453 | 0.0233 | 0.1508 | -0.009*** |
| Have an Affair | 134,072 | 0.0162 | 0.1264 | 406,453 | 0.0128 | 0.1124 | 0.003*** |
| Incarceration | 134,072 | 0.0036 | 0.0602 | 406,453 | 0.0100 | 0.0996 | -0.006*** |
| Gambling | 134,072 | 0.0250 | 0.1562 | 406,453 | 0.0485 | 0.2148 | -0.023*** |
| Blind Date | 134,072 | 0.5194 | 0.4996 | 406,453 | 0.5402 | 0.4984 | -0.021*** |
| Missing | 134,072 | 0.0028 | 0.0527 | 406,453 | 0.0011 | 0.0330 | 0.002*** |
| Lived Apart | 134,072 | 0.3445 | 0.4752 | 406,453 | 0.3710 | 0.4831 | -0.026*** |
| Quarrel | 134,072 | 0.5965 | 0.4906 | 406,453 | 0.6099 | 0.4878 | -0.013*** |
| Runaway | 134,072 | 0.0742 | 0.2621 | 406,453 | 0.0665 | 0.2491 | 0.008*** |
| Sons | 128,961 | 0.3604 | 0.5310 | 393,383 | 0.5124 | 0.5571 | -0.152*** |
| Daughters | 128,815 | 0.6885 | 0.6204 | 393,205 | 0.6564 | 0.6340 | 0.032*** |
| **Panel B: Judges** | Male Judges | | | Female Judges | | | [2]-[5] |
| | Obs | Mean | SD | Obs | Mean | SD | Diff |
| Acting judge | 367,093 | 0.1370 | 0.3439 | 173,432 | 0.2393 | 0.4266 | -0.102*** |
| Num of cases handled | 361,786 | 23.9695 | 28.0634 | 170,710 | 20.0585 | 24.3289 | 3.911*** |
| Age | 29,063 | 39.2787 | 7.2863 | 15,764 | 35.8698 | 7.2977 | 3.409*** |
| Senior judge | 46,892 | 0.1739 | 0.3791 | 25,449 | 0.2141 | 0.4102 | -0.040*** |
| Graduate degree | 26,031 | 0.1768 | 0.3815 | 16,254 | 0.3250 | 0.4684 | -0.148*** |
| Nonlocal | 10,799 | 0.2084 | 0.4062 | 5,300 | 0.3119 | 0.4633 | -0.103*** |
| **Panel C: Plaintiffs** | Male Plaintiffs | | | Female Plaintiffs | | | [2]-[5] |
| | Obs | Mean | SD | Obs | Mean | SD | Diff |
| Urban | 56,500 | 0.6958 | 0.4601 | 185,335 | 0.6024 | 0.4894 | 0.040*** |
| Attoney | 134,072 | 0.5976 | 0.4904 | 406,453 | 0.5691 | 0.4952 | 0.014*** |
| Present in court | 131,705 | 0.9974 | 0.0512 | 401,517 | 0.9981 | 0.0437 | -0.000*** |
| Age | 60,238 | 38.4136 | 10.0732 | 179,661 | 34.8018 | 8.5455 | 3.179*** |
| Age at marry | 50,050 | 28.3026 | 7.1548 | 149,067 | 24.9937 | 5.5482 | 2.944*** |
| Marriage duration | 92,042 | 10.3004 | 8.5375 | 293,902 | 10.0265 | 8.0480 | 0.231*** |

*Notes:* Panel A summarizes case characteristics, Panel B summarizes judge-level characteristics, and Panel C reports plaintiff-level characteristics. Standard errors are reported for differences in means, with significance levels denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 2: Balancing Test for Male Judge and Male Plaintiff Interaction

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] | [10] | [11] | [12] | [13] | [14] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Case Characteristics | | | | | | | | | | | | | |
| VARIABLES | Custody | Domestic Violence | Drug Addiction | Have an Affair | Incarceration | Gambling | Blind Date | Missing | Substation | Quarrel | Runaway | Sons | Daughters | Predicted Divorce |
| **Panel A** | | | | | | | | | | | | | | |
| Female judge * Female plaintiff | 0.003 | -0.001 | 0.000 | 0.002** | 0.000 | 0.001 | -0.006 | 0.000 | -0.004 | -0.004 | 0.002 | -0.002 | -0.008 | 0.000 |
| | [0.004] | [0.002] | [0.001] | [0.001] | [0.001] | [0.001] | [0.004] | [0.000] | [0.004] | [0.004] | [0.002] | [0.004] | [0.005] | [0.001] |
| Court ×Month FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 | 508,956 | 508,624 | 508,621 |
| R-squared | 0.207 | 0.250 | 0.352 | 0.134 | 0.143 | 0.264 | 0.243 | 0.159 | 0.255 | 0.263 | 0.189 | 0.196 | 0.156 | 0.243 |
| Sample Mean of Y | 0.442 | 0.0574 | 0.0208 | 0.0137 | 0.00833 | 0.0425 | 0.533 | 0.00146 | 0.364 | 0.607 | 0.0678 | 0.475 | 0.665 | 0.266 |
| **Panel B** | | | | | | | | | | | | | | |
| Female judge * Female plaintiff | 0.003 | -0.001 | -0.001 | 0.002** | 0.000 | 0.000 | -0.006 | 0.000 | -0.001 | -0.005 | 0.003 | -0.004 | -0.004 | 0.001 |
| | [0.004] | [0.002] | [0.001] | [0.001] | [0.001] | [0.001] | [0.004] | [0.000] | [0.003] | [0.003] | [0.002] | [0.004] | [0.005] | [0.001] |
| Court ×Month FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Judge FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Acting Judge | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 522,295 | 503,749 | 503,418 | 503,415 |
| R-squared | 0.308 | 0.385 | 0.531 | 0.219 | 0.245 | 0.399 | 0.374 | 0.245 | 0.412 | 0.478 | 0.306 | 0.270 | 0.234 | 0.392 |
| Sample Mean of Y | 0.442 | 0.0572 | 0.0208 | 0.0136 | 0.00827 | 0.0424 | 0.532 | 0.00146 | 0.363 | 0.608 | 0.0675 | 0.475 | 0.665 | 0.266 |

*Notes:* In Panel A, the regressions include controls for court and time fixed effects. In Panel B, the regressions additionally include controls for judge fixed effects and an indicator for whether the judge is acting. The dependent variable in Column 9 represents the predicted probability of divorce, calculated as the fitted value from an ordinary least squares (OLS) regression of divorce on all case characteristics listed in Columns 1–8. In Column 10, the dependent variable is the predicted probability of divorce based on an OLS regression of divorce on all case characteristics, excluding the presence of a child. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3: Baseline Results

| VARIABLES | [1] Divorce | [2] Divorce | [3] Divorce | [5] Divorce |
|---|---|---|---|---|
| Female judge * Female plaintiff | 0.018*** | 0.017*** | 0.012*** | 0.012*** |
| | [0.003] | [0.003] | [0.003] | [0.003] |
| Female plaintiff | 0.003 | 0.002 | 0.003 | 0.003 |
| | [0.002] | [0.002] | [0.002] | [0.002] |
| Female judge | -0.040*** | | | |
| | [0.003] | | | |
| Court × Month FE | Y | Y | Y | Y |
| Judge FE | | Y | Y | Y |
| Female judge × Case characteristics | | | Y | Y |
| Acting judge | | | | Y |
| Observations | 527,279 | 522,295 | 503,415 | 503,415 |
| R-squared | 0.223 | 0.321 | 0.391 | 0.391 |
| Sample Mean of Y | 0.267 | 0.266 | 0.265 | 0.265 |

*Notes:* Column 1 controls for court-by-month fixed effects. Column 2 adds judge fixed effects. Column 3 includes eight case characteristics interacted with the male judge indicator, and Column 4 further incorporates an indicator for whether the judge is acting. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Competing Interpretations

| | [1] | [2] | [3] | [4] | [5] | [6] |
|---|---|---|---|---|---|---|
| Female judge * Female plaintiff | 0.013*** | 0.012** | 0.012** | 0.015 | 0.020 | 0.012*** |
| | [0.003] | [0.005] | [0.005] | [0.012] | [0.016] | [0.004] |
| Female judge * Female plaintiff * Straightforward case | 0.000 | | | | | |
| | [0.023] | | | | | |
| Male judge * Male plaintiff * High workload | | -0.000 | | | | |
| | | [0.007] | | | | |
| Female judge * Female plaintiff * Experienced judge | | | -0.000 | | | |
| | | | [0.007] | | | |
| Female judge * Female plaintiff * Senior judge | | | | 0.003 | | |
| | | | | [0.025] | | |
| Female judge * Female plaintiff * Older judge | | | | | -0.022 | |
| | | | | | [0.038] | |
| Female judge * Female plaintiff * Acting judge | | | | | | -0.000 |
| | | | | | | [0.009] |
| Court × Month FE | Y | Y | Y | Y | Y | Y |
| Judge FE | Y | Y | Y | Y | Y | Y |
| Male judge x Case characteristics | Y | Y | Y | Y | Y | Y |
| Acting judge | Y | Y | Y | Y | Y | Y |
| Observations | 477,873 | 480,140 | 488,392 | 59,429 | 33,889 | 489,846 |
| R-squared | 0.417 | 0.452 | 0.447 | 0.446 | 0.516 | 0.415 |
| Sample Mean of Y | 0.262 | 0.264 | 0.264 | 0.246 | 0.262 | 0.264 |

*Notes:* Column 1 explores the heterogeneous effects across simple and straightforward cases. Columns 2-4 examine the heterogeneous effects based on the judge's experience, with column 2 measuring tenure, column 3 using the senior judge indicator, and column 4 measuring judge's age. Column 5 explores the heterogeneous effects across acting and formal judges. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Judge Characteristics Interacted with Plaintiff Gender

| VARIABLES | [1] Divorce | [2] Divorce | [3] Divorce | [4] Divorce | [5] Divorce | [6] Divorce | [7] Divorce | [8] Divorce |
|---|---|---|---|---|---|---|---|---|
| Female judge * Female plaintiff | 0.012*** | 0.013*** | 0.011*** | 0.020 | 0.011 | 0.016 | 0.034 | 0.011 |
| | [0.003] | [0.003] | [0.003] | [0.014] | [0.010] | [0.014] | [0.027] | [0.043] |
| Female plaintiff | 0.003 | 0.008*** | 0.001 | -0.004 | -0.001 | -0.017 | -0.002 | -0.128 |
| | [0.002] | [0.003] | [0.002] | [0.037] | [0.006] | [0.010] | [0.018] | [0.134] |
| Female plaintiff * Experienced judge | | -0.010*** | | | | | | 0.060 |
| | | [0.003] | | | | | | [0.041] |
| Female plaintiff * Acting judge | | | 0.014*** | | | | | 0.048 |
| | | | [0.004] | | | | | [0.053] |
| Female plaintiff * Older judge | | | | -0.000 | | | | 0.003 |
| | | | | [0.001] | | | | [0.003] |
| Female plaintiff * Senior judge | | | | | -0.004 | | | -0.040 |
| | | | | | [0.012] | | | [0.044] |
| Female plaintiff * Graduate degree | | | | | | 0.023 | | 0.007 |
| | | | | | | [0.016] | | [0.052] |
| Female plaintiff * Nonlocal | | | | | | | -0.016 | -0.029 |
| | | | | | | | [0.029] | [0.046] |
| Court × Month FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Judge FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Male judge x Case characteristics | Y | Y | Y | Y | Y | Y | Y | Y |
| Acting judge | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 503,415 | 503,415 | 503,415 | 35,583 | 61,014 | 33,677 | 12,050 | 3,817 |
| R-squared | 0.387 | 0.390 | 0.387 | 0.500 | 0.428 | 0.460 | 0.496 | 0.466 |
| Sample Mean of Y | 0.265 | 0.265 | 0.265 | 0.265 | 0.246 | 0.265 | 0.274 | 0.246 |

*Notes:* The regression in column 1 reports the baseline results. Regressions in columns 2-6 further control for various judge characteristics interacted with the male plaintiff dummy. Specifically, column 2 includes the judge's tenure, column 3 includes an indicator for acting judges, column 4 considers the judge's age, column 5 includes an indicator for senior judges, column 6 includes an indicator for those with a graduate degree and column 7 includes an indicator for nonlocal judges. The regression in column 7 incorporates all additional judge characteristics interacted with the male plaintiff dummy. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Plaintiff Characteristics Interacted with Judge Gender

| VARIABLES | [1] Divorce | [2] Divorce | [3] Divorce | [4] Divorce | [5] Divorce | [6] Divorce | [7] Divorce | [8] Divorce |
|---|---|---|---|---|---|---|---|---|
| Female judge * Female plaintiff | 0.012*** | 0.010 | 0.012*** | 0.012*** | 0.010* | 0.016*** | 0.015*** | 0.018* |
| | [0.003] | [0.006] | [0.003] | [0.003] | [0.006] | [0.006] | [0.004] | [0.009] |
| Female plaintiff | 0.003 | -0.005 | 0.003 | 0.003 | 0.006* | 0.001 | -0.004 | 0.004 |
| | [0.002] | [0.003] | [0.002] | [0.002] | [0.003] | [0.004] | [0.003] | [0.006] |
| Female judge * Urban | | -0.019*** | | | | | | -0.020* |
| | | [0.007] | | | | | | [0.011] |
| Female judge * Attorney | | | -0.000 | | | | | 0.005 |
| | | | [0.003] | | | | | [0.008] |
| Female judge * Present | | | | 0.066** | | | | 0.196** |
| | | | | [0.031] | | | | [0.089] |
| Female judge * Age | | | | | -0.001** | | | 0.012** |
| | | | | | [0.000] | | | [0.005] |
| Female judge * Age at Marry | | | | | | 0.001 | | -0.011** |
| | | | | | | [0.000] | | [0.005] |
| Female judge * Marriage Duration | | | | | | | -0.001*** | -0.013*** |
| | | | | | | | [0.000] | [0.005] |
| Observations | 503,415 | 209,268 | 503,415 | 498,436 | 213,374 | 175,182 | 350,754 | 89,910 |
| R-squared | 0.391 | 0.490 | 0.391 | 0.392 | 0.425 | 0.436 | 0.422 | 0.505 |
| Sample Mean of Y | 0.265 | 0.299 | 0.265 | 0.264 | 0.272 | 0.272 | 0.273 | 0.292 |

*Notes:* The regression in column 1 reports the baseline results. Regressions in columns 2-7 further control for various plaintiff characteristics interacted with the male judge dummy. Specifically, column 2 includes an indicator for urban plaintiffs, column 3 indicates whether the plaintiff used an attorney, column 4 considers the judge's presence in court, column 5 includes the plaintiff's age, column 6 includes the plaintiff's age at marriage, and column 7 considers the duration of marriage. The regression in column 8 incorporates all additional plaintiff characteristics interacted with the male judge dummy. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 7: Heterogeneous Effects

| | [1] | [2] | [3] | [4] |
|---|---|---|---|---|
| Male judge * Male plaintiff | 0.015*** | 0.018*** | 0.013** | 0.015** |
| | [0.004] | [0.005] | [0.005] | [0.007] |
| Male judge * Male plaintiff * Imbalanced sex ratio | -0.007 | | | |
| | [0.007] | | | |
| Male judge * Male plaintiff * Biased gender attitudes | | -0.011* | | |
| | | [0.007] | | |
| Male judge * Male plaintiff * High legal index I | | | -0.002 | |
| | | | [0.007] | |
| Male judge * Male plaintiff * High legal index II | | | | -0.006 |
| | | | | [0.008] |
| Court × Month FE | Y | Y | Y | Y |
| Judge FE | Y | Y | Y | Y |
| Male judge x Case characteristics | Y | Y | Y | Y |
| Acting judge | Y | Y | Y | Y |
| Observations | 471,951 | 501,038 | 471,951 | 471,951 |
| R-squared | 0.364 | 0.390 | 0.364 | 0.364 |
| Sample Mean of Y | 0.250 | 0.265 | 0.250 | 0.250 |

*Notes:* Columns 1 and 2 analyze the heterogeneous effects related to local levels of gender discrimination, as measured by the sex ratio in column 1 and reported gender attitudes in column 2. Columns 3 and 4 assess the heterogeneous effects based on the level of local legal development. Finally, column 5 reports the heterogeneous effects related to the judge's workload, measured by the number of closed cases each judge handled in the corresponding month. Standard errors are clustered at the judge level. Significance levels are denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

# Do Judges Exhibit Gender Bias? Evidence from the Universe of Divorce Cases in China

# ONLINE APPENDIX

Xiqian Cai, Pei Li, Qinyue Luo, HOng Song, Huihua Xie

## Appendix A Strategies to Determine Judge's Gender

To ensure the accuracy of the gender information for judges in our dataset, we employed a multi-step manual verification process using a variety of publicly available sources. Below, we detail the four-step methods and examples used in this process.

### 1. Search on Baidu/google using the judge's name + "judge" + county name

#### 1.1 Lists of Judges:

For some courts, searching the judge's name alongside "judge" and the county name on Baidu or Google yields direct lists of judges, which often include their gender.

**Example:** The official website of Wuxi City, Jiangsu Province, provides a comprehensive list of judges, including their gender.

| | | | | |
|---|---|---|---|---|
| 王一川 | 男 | 江苏 | 立案一庭 | 法官 |
| 施美华 | 女 | 江苏 | 审监庭 | 法官 |
| 李海林 | 女 | 江苏 | 审监庭 | 法官 |
| 张天浪 | 男 | 江苏 | 审监庭 | 法官 |
| 过坚列 | 男 | 江苏 | 审监庭 | 法官 |
| 姜丽丽 | 女 | 江苏 | 审监庭 | 法官 |
| 羊羚 | 女 | 江苏 | 审监庭 | 副庭长 |
| 徐海宏 | 女 | 江苏 | 刑一庭 | 法官 |
| 蒋璟 | 男 | 江苏 | 刑一庭 | 法官 |
| 周华 | 女 | 江苏 | 刑一庭 | 法官 |
| 蔡连德 | 男 | 江苏 | 刑一庭 | 法官 |
| 孙炜 | 男 | 江苏 | 刑一庭 | 庭长 |
| 韩锋 | 男 | 江苏 | 刑一庭 | 副庭长 |
| 马小卫 | 男 | 江苏 | 刑一庭 | 法官 |
| 周群 | 男 | 江苏 | 刑二庭 | 法官 |
| 杨温蕊 | 女 | 江苏 | 刑二庭 | 法官 |

*Note:* Wuxi Judge Directory. Source: https://splcgk.court.gov.cn/gzfwww/fgml.

### 1.2 News Reports:

In some cases, news articles either explicitly reveal the judge's gender or do so indirectly through the use of pronouns like "he" or "she."

**Example:** Searching for "Jiang Qingchun judge" led to a report on his achievements that uses the pronoun "he," confirming the judge is male.

> 蒋青春1981年毕业后，分到开鲁县人民法院工作。“做一名好法官”是蒋青春刚参加工作时立下的誓言。32年来，蒋青春不负重托，钻研业务，心存高远，先后任开鲁县人民法院书记员、助审员、审判员、副庭长、庭长、审委会委员、评查室主任和审委会专职委员。一路走来，蒋青春的职务多次转换，可是他立志“做一名好法官”的誓言始终没有变。32年，他承办各类案件1600余件，没有出现一件错案，没有一件被投诉，也没有一件引发上访。用法院内部人的话说：蒋青春心中有根红线，那就是法律和良知；用当地老百姓的话说，蒋青春就是一杆秤，他办的案子，咱都信服。蒋青春先后获得“全国优秀法官”、“内蒙古自治区十大法治人物”、“优秀政法干警，二等功获得者”“内蒙古自治区创先争优优秀共产党员”、“通辽市道德模范和劳动模范”等多项荣誉，年近花甲、重病缠身的蒋青春仍然兢兢业业工作在评查岗位上，为他心中的誓言，默默的奉献着……。

*Note:* News Report on Jiang Qingchun. Source: http://www.ewkfy.gov.cn/show.asp?id=2535.

### 1.3 Photos:

For some judges, online photos can be used to visually identify their gender.

**Example:** A report on Judge Zhou Bin, including his photo, confirms his gender as male.


*Note:* Sohu Article on Zhou Bin. Source: https://www.sohu.com/a/304644474_120059495.

## 2. Explore Courts' Official Websites

### 2.1 Judges' Personal Information on Court Websites:

Some courts provide detailed personal information about judges on their official websites, including gender.

**Example:** The Taizhou City Court website lists judges' personal information, including gender.


*Note:* Taizhou City Court. Source: http://www.tzfyw.gov.cn/InfoPub/CategoryView.aspx?CategoryID=71.

# 3. Browse the China Court Trial Online

### 3.1 Trial Videos:

In cases where judges preside over trials broadcast on the China Court Trial Online platform, their gender can often be identified by watching the trial video.

**Example:** A trial video available on the platform confirmed the gender of a female judge presiding over the case.


*Note:* China Court Trial Online. Source: http://tingshen.court.gov.cn.


*Note:* China Court Trial Online. Source: http://tingshen.court.gov.cn.

# 4. Manual Identification Using Name Characteristics

For the remaining judges (approximately 8% of the total), whose gender could not be determined through the above methods, we inferred their gender based on name characteristics,

particularly for names that are typically associated with one gender in Chinese culture.

**Example:** A list of judges' names from Hanyang District, Wuhan City, Hubei Province, includes names like "Wang Qin," "Zhu Yini," and "Sun Jiao," which are commonly recognized as female names in Chinese culture.

| | | | | | |
|---|---|---|---|---|---|
| 18 | 刘先兵 | | 湖北 | 刑一庭 | 法官 |
| 19 | 朱燕飞 | | 湖北 | 未综庭 | 庭长 |
| 20 | 祝旖旎 | | 湖北 | 未综庭 | 法官 |
| 21 | 叶凌芳 | | 湖北 | 民一庭 | 庭长 |
| 22 | 张朝管 | | 湖北 | 民一庭 | 法官 |
| 23 | 王琴 | | 湖北 | 民一庭 | 法官 |
| 24 | 崔玉华 | | 湖北 | 民一庭 | 法官 |
| 25 | 董泽锋 | | 湖北 | 民二庭 | 法官 |
| 26 | 黄丽明 | | 湖北 | 民二庭 | 法官 |
| 27 | 孙娇 | | 湖北 | 民二庭 | 法官 |
| 28 | 何俊杰 | | 湖北 | 民二庭 | 庭长 |
| 29 | 刘跃昌 | | 湖北 | 民二庭 | 书记 |
| 30 | 熊敏 | | 湖北 | 民二庭 | 庭长 |
| 31 | 王侃 | | 湖北 | 民二庭 | 副庭长 |
| 32 | 柯亚兰 | | 湖北 | 民二庭 | 无 |
| 33 | 李琳 | | 湖北 | 民三庭 | 法官 |

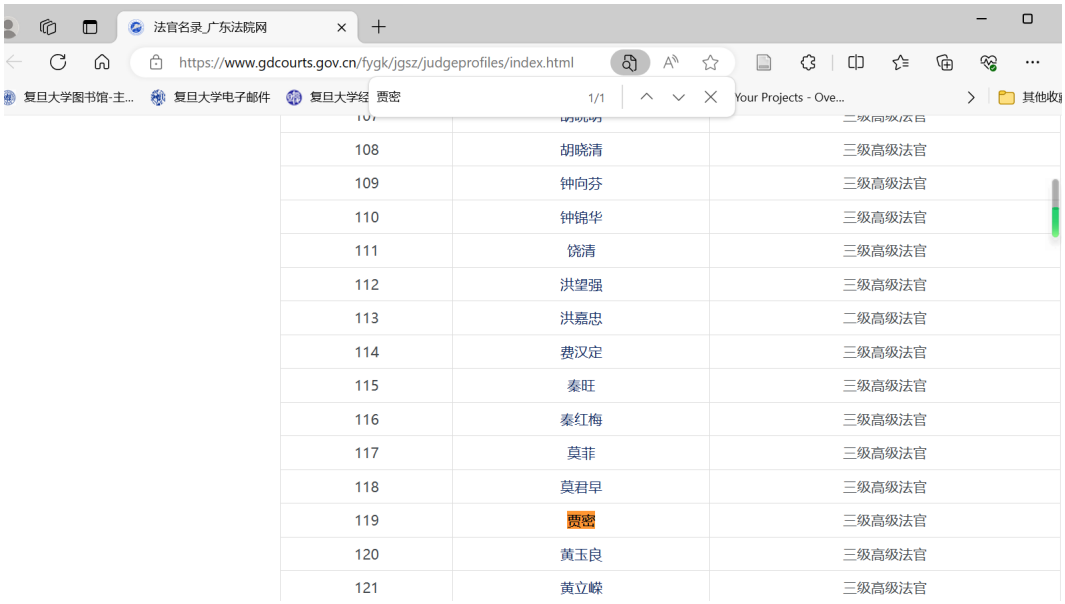
*Note:* Hanyang Judge Directory. Source: https://splcgk.court.gov.cn/gzfwww/fgml.

# Appendix B Judge Information Collection Methods

To collect additional judge-level information for our study, we employed a systematic approach using multiple publicly accessible sources. Below, we outline the key methods and channels used by our research team:

1. Court Websites: The primary source of information was the official websites of the courts where the judges were employed. Many court websites provide public disclosures or personnel directories. For example, on the Guangdong High People's Court website, we identified Judge Jia Mi through the "Judge Directory" section using the browser's search tools. From these listings, we gathered personal information disclosed by the court, such as educational background and professional roles.

Example:




*Note:* This image shows the Guangdong High People's Court website, where we used the "Judge Directory" section to identify Judge Jia Mi


*Note:* This image displays publicly disclosed information about Judge Jia Mi obtained from the same website.

2. Search Engines (e.g., Baidu): We used Baidu to search for "Judge Name + Court Name," which sometimes led to Baidu Baike (the Chinese equivalent of Wikipedia) entries. If a judge had a Baidu Baike profile, it often contained supplementary information, such as education and career details. However, since the accuracy of Baidu Baike entries depends on cited sources, we cross-verified the information by reviewing the referenced websites to ensure reliability and check for updates.


*Note:* This image illustrates a search result for a judge with a Baidu Baike entry. The profile contains supplementary information, such as educational background and career details.

3. Judicial Selection Announcements: Another valuable resource was public announcements of judicial selection or promotion, which often disclose information such as academic qualifications and alma maters. For example, we referenced the 2023 "Judicial Selection Announcement for the Sichuan High People's Court," which included detailed profiles of candidates for judicial positions.

4. News Reports and Public Profiles: We searched for news articles and other public

**四川省高级人民法院 2023 年度面向中级、基层法院遴选法官拟任职人员名单**

| 遴选单位 | 遴选职位 | 姓名 | 性别 | 学历学位 | 毕业院校和专业 | 工作单位 |
|---|---|---|---|---|---|---|
| 省法院 | 员额法官 | 夏艳 | 女 | 研究生<br>法律硕士 | 四川大学法律硕士专业 | 成都市中级人民法院 |
| | | 许筱然 | 女 | 研究生<br>法律硕士 | 四川轻化工大学法律（法学）专业 | 南部县人民法院 |
| | | 李维 | 女 | 研究生<br>法学硕士 | 西南政法大学国际法专业 | 绵阳市中级人民法院 |
| | | 梁楷 | 男 | 大学<br>法律硕士 | 大学：中国政法大学政治学与行政学专业<br>法律硕士：西南财经大学法律专业 | 大邑县人民法院 |
| | | 曹余曦 | 男 | 研究生<br>法学硕士 | 华东政法大学刑法专业 | 成都市中级人民法院 |
| | | 何美琪 | 女 | 研究生<br>法律硕士 | 西南财经大学法律硕士专业 | 成都市中级人民法院 |

*Note:* This is an image of the "2023 Judicial Selection Announcement for the Sichuan High People's Court," detailing candidates' profiles and disclosing key personal information.

platforms that showcased judges' achievements or profiles. For instance, we identified judges featured in alumni spotlights, law school newsletters, or professional achievement reports.


*Note:*This image is from Shandong University's 2023 announcement featuring award-winning alumni, providing information about their education, specialization, and achievements.

5. University Alumni Networks: We reviewed alumni association directories from law schools and legal universities. Although these lists were often incomplete, they occasionally included details about alumni who are now judges. For example, the East China University of Political Science and Law's alumni council member list provided information on judges' education and alma maters.

Despite our comprehensive efforts, some judges' information remains incomplete due to factors such as personnel changes or limited disclosures by certain courts. While we aimed

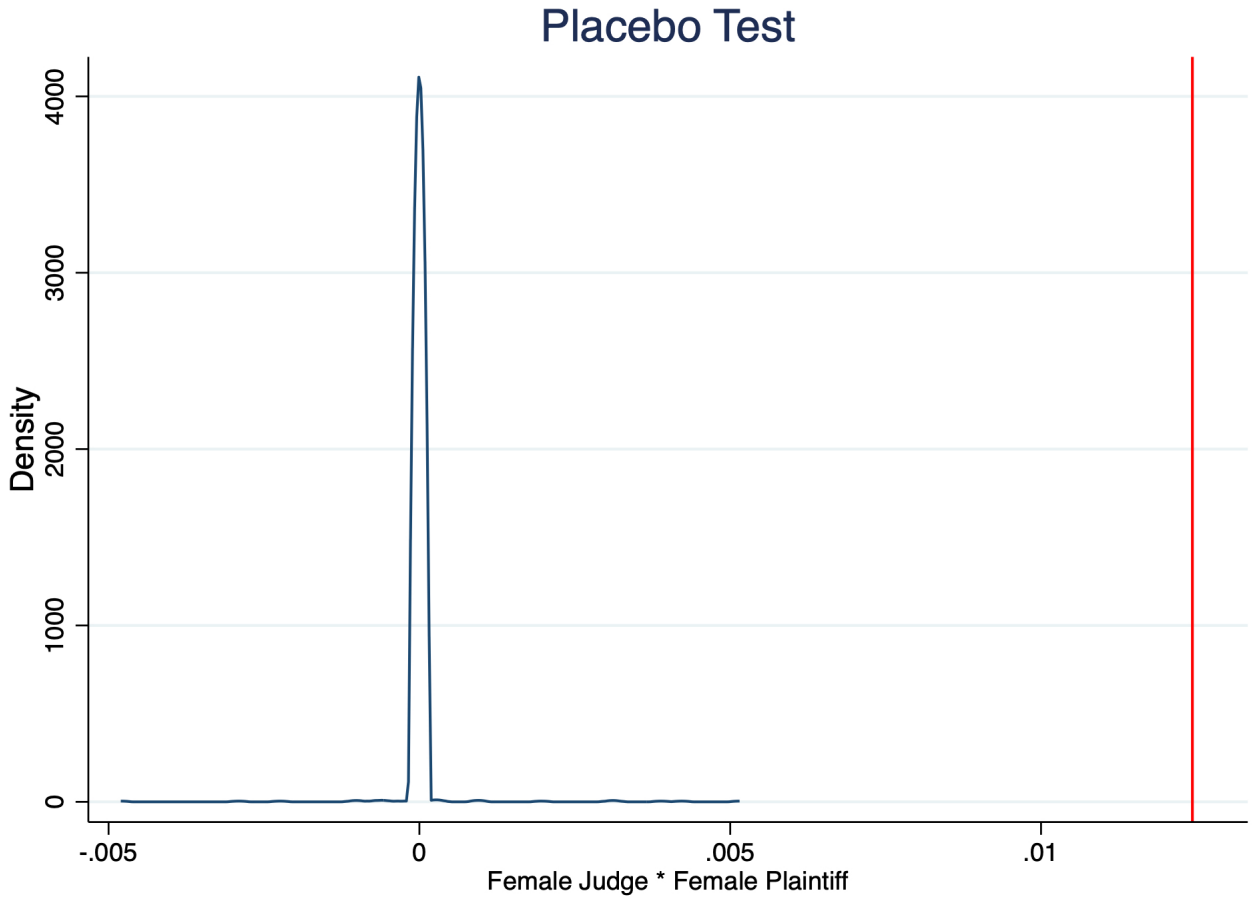

*Note:* This image shows Henan Finance and Law University's report on prominent alumni appointments, which includes details on judges' graduation years and qualifications.


*Note:* This is an image of the East China University of Political Science and Law's alumni council member list, which offered insights into judges' education and alma maters.

to gather as much information as possible, we supplemented the dataset only with verified and reliable information to maintain data accuracy.

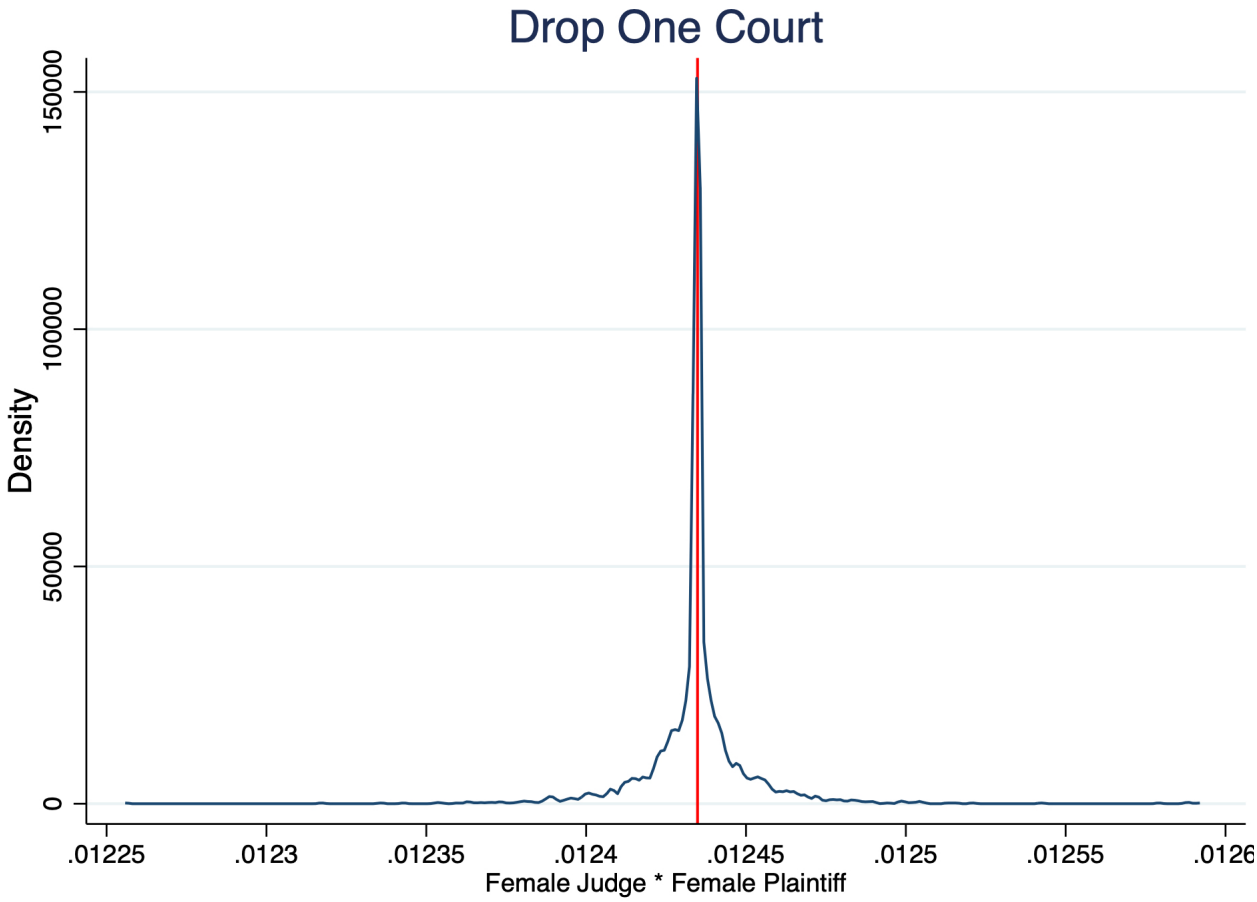
# Appendix C Additional Figures and Tables

Figure C1: Placebo Test, Randomly Assignment of Gender




*Notes:* This shows the distribution of the estimates from the 1,000 runs of placebo test regression, along with the benchmark estimate. The placebo test is to randomly assign judges to plaintiffs, keeping the ratio of each judge's cases unchanged (for similar practices, see, e.g., Chetty etal.,2009; La Ferrara et al., 2012). To increase the identification power of this placebo test, it is repeated 1,000 times.

Figure C2: Robustness Test, Drop One Court




*Notes:* This illustrates the distribution of estimates from the robustness test regressions, alongside the benchmark estimate. The robustness test involves sequentially dropping one court at a time.

Table C1: Predicted Differences in Divorce Acceptance Rate for Judge-Plaintiff Gender Pairs

| | [1] Female Plaintiff | [2] Male Plaintiff | [3] Difference |
|---|---|---|---|
| Female Judge | $\beta_0 + \beta_1 + \beta_2 + \beta_3$ | $\beta_0 + \beta_3$ | $\beta_1 + \beta_2$ |
| Male Judge | $\beta_0 + \beta_2$ | $\beta_0$ | $\beta_2$ |
| Difference | $\beta_1 + \beta_3$ | $\beta_3$ | $\beta_1$ |

Table C2: Robustness Checks

| VARIABLES | [1] Withdraw | [2] Case Duration I | [3] Case Duration II | [4] Clustering S.E. at Court-Time Level | [5] Courts with at least one female judge | [6] Inverse Probability Weights | [7] Missing Plaintiff Gender=Females | [8] Missing Plaintiff Gender=Males |
|---|---|---|---|---|---|---|---|---|
| Female judge * Female plaintiff | -0.001 | -0.100 | 0.118 | 0.012*** | 0.010*** | 0.011*** | 0.012*** | 0.011*** |
| | [0.002] | [0.415] | [0.393] | [0.003] | [0.003] | [0.003] | [0.003] | [0.003] |
| Court × Month FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Judge FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Female judge × Case characteristics | Y | Y | Y | Y | Y | Y | Y | Y |
| Acting judge | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 532,378 | 404,885 | 396,214 | 503,415 | 420,946 | 503,415 | 521,959 | 521,959 |
| R-squared | 0.291 | 0.398 | 0.389 | 0.391 | 0.376 | 0.357 | 0.386 | 0.387 |
| Sample Mean of Y | 0.0507 | 51.93 | 36.96 | 0.265 | 0.259 | 0.265 | 0.263 | 0.263 |

*Notes:* The dependent variable in Column 1 indicates whether the plaintiff withdrew the lawsuit between the filing date and the determination of the trial outcome. The samples in Columns 2 and 3 exclude withdrawn cases. In Column 2, the dependent variable is the case duration, defined as the number of days from the filing date to the closure date. In Column 3, the dependent variable is the case duration, defined as the number of days from the filing date to the hearing date. Standard errors are clustered at the judge level. Standard errors in column 4 are clustered at the court-time level, while those in other columns are clustered at the judge level. The regression in column 5 includes only courts that had at least one female judge present when cases were submitted. Column 6 presents results estimated using inverse probability weights to address missing plaintiff gender data. Columns 7 and 8 report estimation results under the assumptions that all missing plaintiff genders are male and female, respectively. Significance levels are indicated as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table C3: Including Collegial Panel Cases

| VARIABLES | [1] Divorce | [2] Divorce | [3] Divorce | [4] Divorce |
|---|---|---|---|---|
| Female judge * Female plaintiff | 0.044*** | 0.042*** | 0.018*** | 0.018*** |
| | [0.003] | [0.003] | [0.003] | [0.003] |
| Female plaintiff | -0.072*** | -0.064*** | -0.015*** | -0.014*** |
| | [0.002] | [0.002] | [0.002] | [0.002] |
| Female judge | -0.073*** | | | |
| | [0.003] | | | |
| Court × Month FE | Y | Y | Y | Y |
| Judge FE | | Y | Y | Y |
| Female judge × Case characteristics | | | Y | Y |
| Acting judge | | | | Y |
| Observations | 791,858 | 788,953 | 762,634 | 762,634 |
| R-squared | 0.202 | 0.299 | 0.415 | 0.415 |
| Sample Mean of Y | 0.364 | 0.363 | 0.362 | 0.362 |

*Notes:* Column 1 controls for court-by-month fixed effects. Column 2 adds judge fixed effects. Column 3 includes eight case characteristics interacted with the female judge indicator, and Column 4 further incorporates an indicator for whether the judge is acting. Significance levels are indicated as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table C4: Missing Plaintiff Gender

| | Judge and Case Characteristics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] | [10] |
| VARIABLES | Male | Num of Cases Handled | Acting Judge | Senior Judge | Age | Graduate Degree | Nonlocal | Custody | Domestic Violence | Drug Addiction |
| Missing plaintiff gender | -0.069*** | -0.011** | -0.350 | 0.022*** | 0.027*** | -0.100 | 0.000 | -0.019** | -0.012*** | 0.000 |
| | [0.004] | [0.005] | [0.246] | [0.004] | [0.008] | [0.185] | [0.011] | [0.010] | [0.002] | [0.001] |
| Court ×Month FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 547,309 | 547,309 | 538,927 | 547,309 | 67,568 | 38,946 | 37,592 | 13,109 | 547,309 | 547,309 |
| R-squared | 0.220 | 0.290 | 0.572 | 0.318 | 0.566 | 0.659 | 0.679 | 0.786 | 0.243 | 0.350 |
| Sample Mean of Y | 0.264 | 0.678 | 23.16 | 0.171 | 0.191 | 37.97 | 0.237 | 0.232 | 0.0567 | 0.0207 |
| | [11] | [12] | [13] | [14] | [15] | [16] | [17] | [18] | [19] | [20] |
| VARIABLES | Have an Affair | Incarceration | Gambling | Blind Date | Missing | Substation | Quarrel | Runaway | Sons | Daughters |
| Missing plaintiff gender | -0.003*** | -0.003*** | -0.004** | 0.017*** | 0.000 | -0.036*** | -0.025*** | -0.007*** | -0.169*** | -0.192*** |
| | [0.001] | [0.001] | [0.002] | [0.004] | [0.000] | [0.005] | [0.004] | [0.002] | [0.006] | [0.008] |
| Court ×Month FE | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 547,309 | 547,309 | 547,309 | 547,309 | 547,309 | 547,309 | 547,309 | 547,309 | 527,315 | 526,981 |
| R-squared | 0.131 | 0.140 | 0.258 | 0.240 | 0.155 | 0.252 | 0.261 | 0.185 | 0.188 | 0.156 |
| Sample Mean of Y | 0.0135 | 0.00823 | 0.0424 | 0.535 | 0.00146 | 0.363 | 0.606 | 0.0676 | 0.468 | 0.657 |

*Notes:* Significance levels are indicated as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table C5: Data Missingness

| | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] |
|---|---|---|---|---|---|---|---|---|---|
| | Missing Indicator for Judge's | | | | Missing Indicator for Plaintiff's | | | | |
| VARIABLES | Age | Rank | Education | Nonlocal | Urban | Present | Age | Age at Marry | Marriage Duration |
| Female judge * Female plaintiff | 0.002 | -0.001 | 0.001 | 0.002 | 0.000 | 0.000 | 0.002 | 0.002 | 0.002 |
| | [0.002] | [0.002] | [0.002] | [0.001] | [0.004] | [0.001] | [0.002] | [0.002] | [0.002] |
| Court × Month FE | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 567,081 | 567,081 | 567,081 | 567,081 | 527,279 | 527,279 | 527,279 | 527,279 | 527,279 |
| R-squared | 0.227 | 0.387 | 0.306 | 0.209 | 0.246 | 0.156 | 0.712 | 0.640 | 0.613 |
| Sample Mean of Y | 0.443 | 0.866 | 0.922 | 0.970 | 0.442 | 0.0135 | 0.557 | 0.632 | 0.288 |

*Notes:* Significance levels are indicated as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.