Women in the Courtroom*

HENG CHEN
University of Hong Kong
YUYU CHEN
Peking University
QINGXU YANG
Peking University

September 5, 2022

Abstract. Examining roughly 6 million civil judgments in China during the period 2014–2018, we document that gender disparities in litigation outcomes are present and prevalent and that male judges discriminate against female plaintiffs more than female judges do. Exploiting an open justice reform where an increasing fraction of trials were broadcast online, we find that the disadvantage of female plaintiffs (relative to male plaintiffs) becomes smaller when the broadcast intensity increases and that male judges reduce the gender gap to a larger extent than female judges do. We further show that litigants' plausible behavioral changes during broadcasted trials cannot explain our findings and that the reform is effective because judges' decision-making processes become more visible. Evidence taken together suggests that when the group identity, based on which the victim group is discriminated against, is observable, then external monitoring devices that scrutinize the decision-making process can be effective.

Keywords: Behavioral Bias; Gender Bias; Judicial Reform; Judge; Open Justice.

JEL Classification. K10 K4 O12 P16 P48

*Heng CHEN: The Faculty of Business and Economics, The University of Hong Kong, Email: hengchen@hku.hk; Yuyu CHEN: Guanghu School of Management, Peking University, Email: chenyuyu@gsm.pku.edu.cn; and Qingxu Yang: Guanghu School of Management, Peking University, Email: qxyang1996@pku.edu.cn. This research project is partly funded by the General Research Fund of the Research Grants Council of Hong Kong (Project No. HKU 17500421) and by the Seed Fund for Basic Research of Hong Kong University (Project No. HKU 201811159002).

1. Introduction

Discrimination against women is ubiquitous in societies. While they are well documented in the labor market, education, and the political sphere, gender disparities in judicial systems are much less researched, especially in the context of developing countries. Considering prior studies that have documented substantial economic gains from reducing gender-based disparities and barriers in various areas (Duflo 2012; Falk and Hermle 2018; Doepke and Tertilt 2019; Hsieh, Hurst, Jones, and Klenow 2019; and Tang and Zhang 2021), gender bias in judicial systems may not only be a social justice issue but may also have significant economic consequences.

The goal of this paper is to document gender-based discrimination in the judicial system in China and unveil some of its sources. We study a unique policy experiment in which information technology (IT) is used to improve judicial transparency and examine its effects on reducing gender disparities. The analysis of the underlying mechanisms yields insights into the role of IT in public affairs.

For our purpose, we obtained civil court judgments with individual litigants from *China Judgements Online*, a unique platform that publishes legal documents across all courts in China. We analyze approximately 6 million such judgments published during the period 2014–2018, which provide detailed information on civil litigation (e.g., litigation outcomes and litigant characteristics) and, thus, allow us to study gender bias in litigations. Furthermore, we examine the impacts of a nationwide *open justice* reform, in which surveillance technologies are used to improve the visibility of litigation proceedings in the courtroom. Specifically, in September of 2016, the Supreme Court began requiring that all courts in China broadcast trials live on a centralized online platform and that all the recorded proceedings could be viewed by the public. The reform was implemented in a staggered fashion, with large variations across courts and over time, which offers us a setting in which to examine the impact of judicial transparency on litigation outcomes.

Our analysis yielded salient findings on gender disparities. Throughout the paper, we focus on the extent to which plaintiffs' claims are supported and how this is affected by plaintiff gender because plaintiffs initiate lawsuits and make claims in civil litigations, while defendants are in a more passive position. We find that female plaintiffs are 5 percentage points less likely to win than male plaintiffs, after controlling for a number of case-level control variables ranging from defendant gender to legal costs and region-level control variables as well as court, area, and year-quarter fixed effects. As the average chance of winning for plaintiffs in civil cases is approximately 70 percent in China, this effect is large.

In general, it is difficult to interpret the association between litigant gender and

litigation outcomes as a causal effect. To address this concern, we take advantage of the open justice reform and exploit the variations in policy intensity over time and across courts. We test the effect of this reform on gender differences in litigation outcomes and interpret our findings under the assumption that the new surveillance system improves judicial fairness.

Specifically, using a generalized difference-in-differences (DID) empirical design, we find that female plaintiffs are at less of a disadvantage after the trial broadcasting reform is introduced and when the broadcasting intensity increases. When the intensity of live broadcasting increases from zero to 10 percent (i.e., the mean value of the intensity at the court level since the last quarter of 2016 when the reform started), the gap between female and male plaintiffs' chances of winning declines by 1 percentage point. This effect is pervasive: it can be found in 5 out of the 9 issue areas, even in those unrelated to gender issues, such as contract disputes.

There are justifiable concerns about two potential endogeneity issues. First, whether a case is broadcast is subject to the selection of the court where the case is adjudicated. Our DID design, which exploits variations in broadcast intensity at the court-area level (i.e., each issue area within a court), avoids this issue. That is, if the broadcasting mechanism had no impact on litigation outcomes, we would not find any effects with our approach, despite that courts select certain cases into the broadcast category.

Second, one may worry that the rollout of the reform may not be purely exogenous. At the intensive margin, the strength of the reform, i.e., the broadcasting intensity, could be endogenous. To address this concern, we resort to both the randomization inference (RI) and Bartik instrumental variable (IV) approaches, which provide reassuring evidence that the potential threats to exogeneity are not too worrisome. At the extensive margin, if the timing of the reform in each region is strategically chosen, the parallel trends assumption may not hold. We estimate an event study model to show that no pretrend in the gender gap exists before the reform was introduced.

We are interested not only in whether there are gender-based disparities in litigation outcomes and, if so, in which areas they exist but also the sources of the disparities. Therefore, we turn to the effects of judge gender in section 6, with the conjecture that female litigants would be, on average, in an unfavorable position when male judges make decisions. Although the assignment of cases to judges is legally required to be random in China, we worry that the implementation may not be perfect in practice. Therefore, we construct a propensity score matched sample, and using that sample, we find that female plaintiffs' disadvantage relative to males is indeed larger when male judges rule. Furthermore, in response to the open justice reform, male judges reduce this gender-based differential more than female judges, suggesting that

the reform has a larger disciplinary effect on male judges.

However, it is not unlikely that broadcasting in fact alters the behaviors of litigants during trials, instead of disciplining judges. In section 6.2, we leverage the records of litigants' appearances in court to investigate this alternative explanation. We study a subsample where litigants on both sides do not appear in court, i.e., the concerning confounding factor is absent. We still find that the gender-based differential is reduced when the intensity of live broadcasting increases.

Does this reform indeed improve judicial fairness? We provide further evidence that female litigants are more willing to accept the litigation outcomes when their claims are completely rejected by the court, on condition that judges are monitored by the new surveillance system (see section 6.3).

Why is this monitoring mechanism effective at reducing gender disparities? In general, judges behave more professionally in courtroom by adhering procedures when they know they are surveilled, which leads to more impartial litigation outcomes. In particular, the reform raises the visibility of the judges' decision-making process and exposes them to potential penalties (e.g., reputation loss and informal evaluation within the legal profession). Therefore, when the cost of discrimination is higher, judges tend to act out their biased preferences to a lesser extent. The effectiveness of the open justice reform constitutes a useful lesson for judicial reforms and the use of information technology on public affairs in general: if the group identity on the basis of which a victim group is discriminated against is observable, then a monitoring system that enhances the transparency of the decision-making process can be effective. Furthermore, it is worthwhile to point out that the cost of implementing this reform is rather low and not prohibitive for developing countries to adopt.[1]

Our results are consistent with an interpretation of social prejudice and in-group bias among male judges. Therefore, they lend support to affirmative action policies for diversifying judge gender in the courtroom, which resembles the gender-based quota system in governments and legislation.[2]

Our work contributes to three bodies of literature in economics. First, there is a bourgeoning literature that examines legal bias based on attitudes and stereotypes that one may have about social categories, such as race and minorities (Abrams, Bertrand, and Mullainathan 2012; Alesina and La Ferrara 2014; Arnold, Dobbie, and Yang 2018; Bielen, Marneffe, and Mocan 2021; Kastellec 2013; and Hou and Truex 2019). However,

[1]The open justice reform was a recent innovation which differs from other low-cost judicial reforms studied in previous studies and alternative settings (Chemin 2009).

[2]Our policy recommendations for diversifying gender on the bench is inspired by the previous literature on the impact of gender quota systems, such as Chattopadhyay and Duflo (2004), Beaman, Chattopadhyay, Duflo, Pande, and Topalova (2009) and Beath, Christia, and Enikolopov (2013).

less attention is paid to gender-based discrimination. Most studies in this domain focus on the effects of judge gender (Songer and Crews-Meyer 2000; Ash, Chen, and Ornaghi 2021) and jury gender (Anwar, Bayer, and Hjalmarsson 2019a; Hoekstra and Street 2021). Among others, Bindler and Hjalmarsson (2020) document that female defendants are treated more leniently than males in criminal cases. However, little is known from the existing literature about the gender disparities of litigants in civil cases, which may have direct economic consequences. We explicitly focus on the gender effects of litigants in civil cases.[3]

In addition, the analysis of judge gender in this paper complements and expands the existing literature with new findings on the context of China. Rarely has the impact of judge gender been found on litigation outcomes, particularly in the context of the U.S. common law system, when the sample is properly rebalanced (Boyd, Epstein, and Martin 2010 and Lim, Silveira, and Snyder 2016). Ash, Asher, Bhowmick, Bhupatiraju, Chen, Devi, Goessmann, Novosad, and Siddiqi (2021) also found that there is no gender in-group bias in the Indian judiciary, based on 5 million criminal cases. However, in the Chinese civil court system, it appears to be prevalent not only in gender-sensitive areas but also economically salient areas.

Second, our work joins a small but growing literature on the impact of IT on public affairs. Pierce, Snow, and McAfee (2015) show that technology-based employee monitoring systems increase perceived general oversight and reduce misconduct. There are studies focusing on the impact of IT on law enforcement such as Mastrobuoni (2020) and Zamoff, Greenwood, and Burtch (2021). Our work differs in that we study the impact of IT on the judicial system and judges. Adams, Adams-Prassl, and Adams-Prassl (2022) address the issue that publishing judicial decisions online can lead to abuses of privacy, while our work mainly focuses on benefits of enhanced judicial visibility.

Finally, given that we identify gender effects in litigation outcomes in China, our findings contribute to the literature on gender inequality in China (Coale and Banister 1994; Qian 2008; and Ebenstein 2010), and extend our knowledge on the scope of gender bias.[4]

[3]Broadly speaking, our work contributes to the sizable literature on the determinants of judicial bias, such as nonrational factors in judge decisions (Dobbie, Goldin, and Yang 2018; Eren and Mocan 2018), peer effects (Eren and Mocan 2020), the impacts of media exposure (Lim, Snyder Jr, and Stromberg 2015 and Philippe and Ouss 2018), political beliefs (Anwar, Bayer, and Hjalmarsson 2019b and Beim, Clark, and Lauderdale 2021), anti-foreign sentiment (Mai and Stoyanov 2019; Bhattacharya, Galpin, and Haslem 2007), in-group bias (Shayo and Zussman 2011) and political or even personal connections (Firth, Rui, and Wu 2011; Ang and Jia 2014; Liu and Zhang 2020; Cai, Chen, Chen, and Zhang 2022).

[4]Recent evidence has emerged primarily from studies on the labor market in China, such as Kuhn and Shen (2013) and Zhang, Jin, Li, and Wang (2021), and in the education sector, such as Brown and Park (2002). Researchers have also accumulated quality evidence that females are not on a level playing field with males in labor market (e.g., Black and Strahan 2001; Bagues and Esteve-Volart 2010; Card, Cardoso, and Kline 2016; Blau and Kahn 2017; Bertrand, Cortes, Olivetti, and Pan 2021; Charles,

2. Background

2.1. Courts and Judges in China

The hierarchy of the court system in China has four layers: the local court, the intermediate court, the high court and the Supreme Court.By January 2021, there were 3,087 local courts, 416 intermediate courts, 33 higher courts and 1 Supreme Court in China. Typically, there is one local court in each county, one intermediate court in each prefecture, and one high court in each province.[5] The superior courts are obligated to supervise and monitor subordinate courts in their jurisdiction.

Generally, within each court, there are a number of subcourts equipped with judges specializing in certain issue areas, such as subcourts of criminal cases, civil cases, business cases and juvenile cases.[6] Civil cases, the focus of our paper, are divided into 9 major issue areas, i.e., (1) personal rights, (2) marriage, family and inheritance, (3) property, (4) contracts, (5) intellectual property, (6) industrial disputes, (7) finance, security and insurance, (8) tort liability and (9) special procedures.[7] Subsets of areas are grouped and handled by subcourts.

In China, judges are predominantly powerful in adjudication and are subject to very limited supervision during trials. Typically, there are two types of procedures: summary procedures and general procedures. The former involve a single judge and are applied to simple cases with "clear facts and few disputes". The latter involve 3, 5 or 7 judges and are applied to "complex cases" (e.g., cases with a larger number of litigants). The attitudes and preferences of judges matter greatly for decisions in the court and, thus, litigation outcomes, particularly for cases with one judge in charge (i.e., accounting for more than 70 percent of all civil cases from 2014 to 2018). Similarly, the chief judge responsible for general procedure cases has dominant power over litigation outcomes (Chen 2014; Fang 2015). Juries are sometimes configured in

Guryan, and Pan 2018; and Adams-Prassl, Huttunen, Nix, and Zhang 2022), education (e.g., Mengel, Sauermann, and Zolitz 2019 and Carlana 2019), in the marriage market (e.g., Bertrand, Kamenica, and Pan 2015), in politics (e.g., Chattopadhyay and Duflo 2004; Hughes 2011), in medicine and health (e.g., Hamberg 2008; Lin, Liu, and Qian 2014) or in access to government services or justice (e.g., Jassal 2020).

[5] According to Articles 17 and 23 of the Organization Law of People's Court in China, the local courts are mainly responsible for hearing and ruling general cases. The intermediate courts handle cases of larger size or widespread social influence and cases transferred or appealed from their subordinate local courts. The high courts are the highest judicial organ in each province and each province-level municipality, responsible for the cases transferred or appealed from the intermediate courts and a small proportion of cases that attract considerable attention. High courts also take responsibility for reviewing cases of subordinate courts and making retrial decisions on those with ambiguous or erroneous judgments. The Supreme Court is the highest judicial institution in China and the main driver of the legal system reforms studied in this paper.

[6] The general staffing of a local court includes a head, several deputy heads, several judges, clerks, executors, forensic experts and judicial police, etc.

[7] Cases that cannot be categorized are pooled together as "others", accounting for approximately 9 percent of the total cases. See details in the Regulation of Civil Case Areas issued by the Supreme Court.

the trial process but play at best a marginal role.[8] Relevant to our analysis, it is noteworthy that the gender composition of judges is not balanced. According to our data, female judges are the minority, accounting for only 27 percent, whereas 34 percent of judges in the U.S. are female.[9]

2.2. Judicial Reforms

In view of longstanding issues such as low transparency, bureaucratism and local protectionism, the Supreme Court has been implementing a series of reforms since 2014, with the objective of promoting judicial independence, trial openness and judgment fairness.[10] One of the most prominent changes is the open justice reform, during which 4 major information disclosure platforms were created, specializing in publishing legal documents, the live broadcasting of trials, providing access to information about litigation procedures, and displaying blacklists of defaulters.[11]

The content and documents of court files (with some exceptions) must be published on the website "China Judgements Online,"to make past decisions available for review in an easy-to-access manner. Obtaining free access to legal documents online is consistent with one of the core objectives of judicial reform.[12] This website was officially launched in July 2013, and as of December 2021, more than 120 million documents had been posted on this website.

In addition, "China Court Trial Online"was officially launched in September 2016,

[8]The differences in the jury system between China and common law countries such as the U.S. are fundamental. First, in the U.S., the 7th Amendment permits litigants in civil litigation to decide whether to invite a jury. In contrast, in China, whether to arrange a jury in court is decided by the court itself, while litigants have no choice in the matter. Second, in the United States, judges and juries have a clear division of functions in the trial process, where the jury takes responsibility for recognizing facts and the judge is responsible for applying the law. In China, the jury system lost its constitutional basis after the revision of China's Constitution in 1982: all related provisions regarding a jury's rights, role and position in court were removed (Chen 2008). In practice, the right of the jury is not guaranteed, and jury members play a minor role, nearly always following the opinion of judges (Guo 2016).

[9]See detailed descriptions in *The American Bench 2021*. https://www.ojp.gov/ncjrs/virtual-library/abstracts/american-bench-judges-nation-2nd-annual-edition

[10]The Central Task Force for Judicial Reforms, which was responsible for this implementation, advanced a batch of reforms, such as restricting the personnel quota of judges and prosecuting attorneys, transferring authority over administrative issues of local courts (e.g., personnel and funding) from local governments to the high courts, and establishing circuit courts.

[11]According to Wikipedia, the term open justice refers to "a legal principle that requires that judicial proceedings be conducted in a transparent manner and with the oversight of the people, so as to safeguard the rights of those subject to the power of the court and to allow for the scrutiny of the public in general."

[12]The motives for such a reform are twofold, as analyzed and summarized by Liebman, Roberts, Stern, and Wang (2020). First, legal documents of previous cases can be useful guidelines for judges and prosecutors to make decisions, and the documents provide easy access to decisions on similar cases, thereby enhancing the quality of court decisions. Second, the availability of court judgment documents improves the transparency of the litigation process to the public and potentially provides an additional mechanism to curb the misbehavior of judges in courts. That is, identifying outlier court decisions is easier when the legal outcomes of similar cases can be readily obtained.

which allows the public to observe trials as they happen in real time on the internet or to view videotaped proceedings later. By the end of 2021, more than 16 million cases had been broadcast.[13] The platform attracted considerable attention from the public, including citizens, journalists, and practitioners in the legal profession.[14]

2.3. Reform Implementation of Trial Broadcasting

The Supreme Court has enhanced its efforts to promote trial broadcasting since 2016. Courts at all levels have been requested to live broadcast trials on the website of "China Court Trial Online", with the long-term goal that every trial (excluding exceptional cases) shall be broadcast.[15] However, due to technical and financial constraints, the timing of connecting to the website has varied greatly across courts. By September 2016, 383 courts were connected (accounting for 10.89 percent of all courts); by January 2017, another 762 courts were connected to the website; and by February 2018, all 3,517 courts were connected.[16] In addition to the variation in timing, large differences exist in the proportion of cases broadcast across courts too.

The Supreme Court established lofty goals for the reform, demonstrating the authorities' determination to modernize the legal system. However, lower level courts did not necessarily have the same ambition, despite their cooperation being essential for the reform's implementation. The solution is characteristically Chinese: higher level courts quantify the desired progress and include it in the rubric used to assess the performance of subordinate courts.

The reform of live broadcasting of trials is no exception. Although practices vary across localities, it is common for intermediate courts to specify a quota of broadcasted trials for local courts and include it as part of the evaluation system. For local courts, the strategy is to divide and assign the quota to subcourts within courts, and the number or proportion of cases broadcast becomes what is graded in performance evaluations of the subcourts.

Usually, the court head first selects the quota of cases to be broadcast based on the

[13]The other two platforms, i.e., "China Judicial Process Information Online,"and "China Executive Information Online," were launched in November 2014 and October 2013, respectively. The former allows litigants to access information about the procedures of their cases, and the latter discloses the particulars of defaulters who failed to fulfill contracts or implement court orders.

[14]According to the Supreme Court, there were more than 10 cases broadcast with over 10 criminal million views. Some civil cases also attracted extensive attention. For example, on January 11, 2018, the Shenzhen Intellectual Property Court broadcast two cases of Huawei suing Samsung for patent rights infringement. More than 1 million users watched the trial broadcast in real time.

[15]See The Revision of the Supreme Court's Regulations on Court Broadcasting and Video Recording, which was issued in February 2017. Before the fall of 2016, courts in several prefectures were instructed to experiment with broadcasting trials live.

[16]See the evaluation report on "Trial Openness" issued by the Institute of Law of the Chinese Academy of Social Sciences.

requirements of the superior court and then allocates assignments to each subcourt. The head judge of each subcourt then decides which cases to broadcast and submits the plan to the court head. In principle, the head reviews the plan and makes a final decision; in practice, however, approval is typically automatic (He and Wang 2015; Liu and Zhu 2020). Our identification strategy is motivated by the institutional features outlined above: we expect the number and fraction of cases to broadcast at the subcourt level to be exogenous to judges' decisions.

3. Data Construction

3.1. Sample Construction

To construct our sample, we obtain the legal documents published on the website of *China Judgements Online*, with assistance from a commercial data company. For our purposes, we restrict our sample to civil litigations that involve only individual litigants. We focus on cases with trial dates ranging from January 2014 to December 2018. The choice of this time period is intentional. On the one hand, it is important that we focus on the period over which live broadcasting was gradually introduced with large variations across courts and over time. On the other hand, we choose to end the data period in the final quarter of 2018 to ensure that the obtained cases are as comprehensive as possible. In practice, legal documents are uploaded by local courts, which may not be always timely, e.g., judgements for cases are sometimes published a few months late. Given that we acquired this set of legal documents in 2020, the quantity of documents should be reasonably close to the universe: A total number of 6,424,324 civil judgements are in our possession.

The full sample consists of cases with a single litigant on both sides and cases with multiple litigants on at least one side. In our empirical explorations, we always analyze both the main sample with all litigants and the subsample with only two litigants (hereafter, two-litigant subsample). Results arising from the latter should be useful, especially when we focus on the size of the estimated effects, because the gender of litigants is clearly defined and easily measured in this subsample.

The data on whether a case is broadcast live were acquired from the website of *China Court Trial Online*, with the assistance of a commercial data company. The key is that we obtain case codes of cases that were broadcast, which have unique identifiers, e.g., 2018 Hebei-0722 Civil-1st No.1170. The total number of cases broadcast on China Court Trial Online as of April 2021 was 11,016,416, and we have obtained all of them. We match those cases with the full sample, using the case code.

We also collect region-level information from the China City Statistical Yearbook, such as the annual GDP per capita, population and internet penetration rate of each

prefecture in each year. The internet penetration rate is calculated as the ratio of the cumulative number of households with access to the internet to the total number of households.

3.2. Constructing Variables

We glean case-level information from the judgements, which are semistructural with a fixed layout. The standard judgement consists of five sections: the basic information on the litigants, claims of litigants, facts recognized by the court, principles applied and outcomes of the litigation. We provide an example in Figure 9 in Appendix D.

From the basic information section, we extract a host of variables about each case. For example, we acquire the case number for each case from this section, which is useful when we merge the data with other datasets. The instance indicates whether a case is in the first instance heard in the original jurisdiction or in the second instance of rehearing an appeal. The litigant information reveals the names of litigants, based on which we identify whether the case involves individuals, organizations or enterprises. We can also obtain the number of plaintiffs and defendants, litigant gender (if applied), and whether plaintiffs and defendants appear in court. The information about lawyers includes the names of lawyers on both sides and the number of lawyers for the plaintiffs and for the defendants. From the outcome section, we extract information regarding the amount of legal costs and how they are divided among the litigants. Additionally, from the signature located at the very end of the legal document, we obtain the names of the judge(s).

One key variable to our analysis is the litigation outcome, i.e., the chances of winning for the relevant legal parties. According to Chapter 11 of China's Civil Procedure Law, the extent to which the court supports one side is in inverse proportion to the share of the legal costs that he or she undertakes. For instance, if the judge adjudicates that the plaintiff prevails entirely, then 100 percent of the legal costs should be borne by the defendant. If the judge supports the plaintiff's claim only partially, such as 80 percent, then 20 percent of the legal costs should be borne by the plaintiff. Therefore, we define the chances of winning for the plaintiffs as the share of the legal costs borne by the defendants. Based on this measure, we have a coarser definition of the litigation outcome, i.e., a dummy *Win* that takes value 1 when the plaintiffs' chances of winning are greater than 50 percent, 0 otherwise.

Regarding the gender of litigants, approximately 72 percent of judgments in our full sample include the gender information of all litigants in the text, while 26 percent of them omit the plaintiff's gender, and 27 percent omit the defendant's gender. In our analysis, we exclude the cases with missing gender information, which leads to our main sample of 4,601,718 cases. When multiple plaintiffs or defendants are involved,

we use the gender of the first plaintiff or defendant to code the gender for each side of the litigation. This is a justified approximation, because it is a norm for civil litigation in China that the litigants' names appear in the legal documents according to their importance and relevance in the case concerned. The dummy variable Female is constructed to represent the gender of the plaintiff, taking value 1 if the plaintiff is female, 0 otherwise. We similarly construct a dummy for defendant gender.

Although judge gender is not disclosed in the legal documents, the names of judges are available. We deduce genders of judges from their names, using an algorithm *Ngender*, which is widely used to predict profiles of Chinese individuals in industry. We externally validate this algorithm with a judge database with names and genders, and the accuracy rate is 92%.[17] Therefore, we define Female Chief as the chief judge's gender and assign value 1 if the chief judge is deduced to be female.

Next, we turn to measurements of policy implementation. The information about which cases were broadcast allows us to construct the ratio of the number of cases broadcast to the total number of cases at any aggregate level. We refer to it as intensity. The default measure of policy intensity is to construct such a ratio at the court-area $\times$ year-quarter level. Based on the intensity at the court-area $\times$ year-quarter level, we construct another dummy variable, i.e., *Top-25*, and assign it value 1 if the intensity of a certain court in a certain area ranks in the top 25 percent among courts (of the same level) in the same issue area during the same year-quarter. That is, we utilize the relative ranking of courts, instead of the absolute value of intensity, to avoid the influence of outliers. In addition, we also construct a variable–*Overall intensity*–to capture the average broadcasting intensity across issue areas at the court $\times$ year-quarter level.

3.3. Summary Statistics

Table 1 presents the summary statistics of the case-level variables employed in our analysis. Columns (1) and (2) of Table 1 report the means and standard deviations of those variables in cases that were broadcast and in cases that were not, respectively. Column (3) of Table 1 presents the means and standard deviations of those variables in the main sample. Regarding the chances of winning for plaintiffs, the mean for broadcast cases is 82 percent, substantially higher than that for cases not broadcast, which is 74 percent. The fact that plaintiffs are 8 percentage points more likely to win in broadcast cases may not necessarily be interpreted as the impact of live broadcasting. It is likely that subcourt heads select straightforward cases to broadcast, a topic to which we return in the section on 4.

[17] We collect the annual list of "Merit Judges" issued by the Supreme Court from 2000 to 2020, from which we extract names and gender information for nearly 2,000 judges.

Table 1. *Summary Statistics: Case Level Variables*

	Main sample: All Litigants			Sub-sample: Two Litigants		
	Live	Non-live	Total	Live	Non-live	Total
	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd
	(1)	(2)	(3)	(4)	(5)	(6)
Chances of winning for plaintiff	0.82	0.74	0.74	0.84	0.70	0.71
	(0.32)	(0.40)	(0.40)	(0.32)	(0.44)	(0.44)
Plaintiff's gender (1=female)	0.27	0.32	0.31	0.25	0.33	0.32
	(0.44)	(0.47)	(0.46)	(0.43)	(0.47)	(0.47)
Defendant's gender (1=female)	0.17	0.19	0.19	0.17	0.20	0.20
	(0.38)	(0.40)	(0.39)	(0.37)	(0.40)	(0.40)
Number of plaintiffs	1.11	1.11	1.11	1.00	1.00	1.00
	(0.68)	(0.73)	(0.73)	(0.00)	(0.00)	(0.00)
Number of defendants	1.65	1.72	1.71	1.00	1.00	1.00
	(1.13)	(1.26)	(1.25)	(0.00)	(0.00)	(0.00)
Plaintiff's appearance (1=Yes)	0.49	0.48	0.48	0.59	0.60	0.60
	(0.50)	(0.50)	(0.50)	(0.49)	(0.49)	(0.49)
Defendant's appearance (1=Yes)	0.07	0.08	0.08	0.07	0.10	0.09
	(0.25)	(0.27)	(0.27)	(0.26)	(0.29)	(0.29)
Chief judge female (1=Yes)	0.29	0.27	0.27	0.30	0.27	0.27
	(0.46)	(0.45)	(0.45)	(0.46)	(0.44)	(0.44)
Number of lawyers for plaintiff	0.47	0.46	0.46	0.40	0.38	0.38
	(0.66)	(0.67)	(0.67)	(0.62)	(0.60)	(0.60)
Number of lawyers for defendant	0.25	0.27	0.27	0.12	0.14	0.14
	(0.59)	(0.62)	(0.62)	(0.37)	(0.40)	(0.40)
Instance (1=first instance)	0.93	0.88	0.88	0.95	0.90	0.90
	(0.25)	(0.33)	(0.32)	(0.22)	(0.30)	(0.29)
Legal cost in log	6.59	6.35	6.36	6.23	5.88	5.90
	(2.01)	(2.12)	(2.11)	(1.94)	(1.97)	(1.97)
Appeal (1=Yes)	0.09	0.08	0.08	0.07	0.06	0.06
	(0.29)	(0.27)	(0.28)	(0.25)	(0.24)	(0.24)
Observations	331,132	4,359,678	4,690,810	175,452	2,244,335	2,419,787

Notes: Columns (1) and (2) report the means and standard deviations of those variables in cases that were broadcast and in cases that were not, respectively. Column (3) presents the means and standard deviations of those variables in the main sample. Columns (4), (5) and (6) summarize the means and standard deviations of relevant variables for cases broadcast, not broadcast, and all cases in the two-litigant subsample, respectively.

Females are observed to be less present in the courtroom. Regarding plaintiff gender, on average, 31 percent of cases have female plaintiffs, including 27 percent of live-broadcast cases and 32 percent of cases that were not broadcast. For defendant gender, only a minority of civil cases are initiated against female defendants; cases with female defendants account for only 19 percent, including 17 percent of live-broadcast cases and 19 percent of cases that were not broadcast. In other words, the vast majority (more than 80 percent) of defendants are male.

The average number of plaintiffs in cases, whether broadcast or not, is 1.11, indicating that most civil cases involving individuals have only one plaintiff per case. The number of defendants is greater, on average, in either case. Nearly half of plaintiffs appear in the courtroom during trials, i.e., 48 percent on average, while only 8 percent of defendants appear.

Table 2. *Summary Statistics: Distribution of Cases Across Areas By Plaintiff Gender*

		Main sample: All Litigants						Sub-sample: Two Litigants					
		Female and Male Plaintiffs		Female Plaintiff		Male Plaintiff		Female and Male Plaintiffs		Female Plaintiff		Male Plaintiff	
N0.	Area	Num.	Prop. (%)	Num.	Prop. (%)	Num.	Prop. (%)	Num.	Prop. (%)	Num.	Prop. (%)	Num.	Prop. (%)
1	Personality rights	153,368	3.27	62,623	4.26	90,745	2.82	94,055	3.89	40,262	5.12	53,793	3.29
2	Marriage and inheritance	374,246	7.98	250,519	17.05	123,727	3.84	334,295	13.82	232,959	29.63	101,336	6.2
3	Property	167,039	3.56	50,344	3.43	116,695	3.62	83,627	3.46	22,818	2.9	60,809	3.72
4	Contract	2,742,162	58.46	668,944	45.52	2,073,218	64.36	1,540,371	63.66	355,786	45.25	1,184,585	72.52
5	Intellectual property	1,775	0.04	382	0.03	1,393	0.04	991	0.04	192	0.02	799	0.05
6	Employment	45,462	0.97	7,769	0.53	37,693	1.17	26,041	1.08	3,682	0.47	22,359	1.37
7	Finance, security and insurance	26,609	0.57	6,401	0.44	20,208	0.63	8,819	0.36	1,857	0.24	6,962	0.43
8	Tort liability	718,980	15.33	256,344	17.44	462,636	14.36	90,407	3.74	29,902	3.8	60,505	3.7
9	Special procedure	45,039	0.96	15,598	1.06	29,441	0.91	12,342	0.51	4,578	0.58	7,764	0.48
10	Other	416,130	8.87	150,573	10.25	265,557	8.24	228,839	9.46	94,191	11.98	134,648	8.24
	Total	4,690,810	100	1,469,497	100	3,221,313	100	2,419,787	100	786,227	100	1,633,560	100

Notes: This table summarizes the distribution of legal areas by plaintiff's gender in both main sample and two-litigant subsample.

Cases adjudicated by female judges only account for, on average, less than one-third, i.e., 27 percent. As the number of female judges accounts for 27 percent of all judges, it appears that female and male judges' workloads are similar.

The number of lawyers for plaintiffs is, on average, 0.46, with little difference between cases that are broadcast and those that are not. The number of lawyers for defendants is even lower, i.e., 0.27 on average. That implies that the majority of litigants do not employ legal representation in civil cases in China.

The vast majority of cases in our sample are those in the first instance, i.e., 88 percent. The proportion of first-instance cases among those broadcast live is even higher, i.e., 93 percent. The average amount of legal costs is similar for cases that are and are not broadcast. The legal costs are in proportion to the size of the case (or the amount of money involved).[18] Approximately 8% of cases in the first instance are appealed to higher level courts. Of those broadcast live, the ratio is slightly higher, i.e., 9 percent.

We also present summary statistics for those aforementioned variables in the subsample with two litigants (i.e., one plaintiff and one defendant). Columns (3), (4) and (5) of Table 1 summarize the means and standard deviations of relevant variables for cases broadcast, not broadcast, and all cases in the subsample, respectively. The two-litigant subsample has similar characteristics to the main sample.[19]

[18] According to Chapter 3 of The Regulations on Legal Costs in China, legal costs are determined by the amount of money in dispute. Consider property dispute cases as an example. If the amount in dispute is less than 10 thousand yuan, 50 yuan shall be charged for legal costs; for amounts exceeding 10 thousand yuan to 100 thousand yuan, 2.5% shall be charged; and for amounts exceeding 100 thousand yuan to 200 thousand yuan, 2% shall be charged and so on.

[19] Unsurprisingly, we observe that the number of lawyers and the average legal costs are both lower

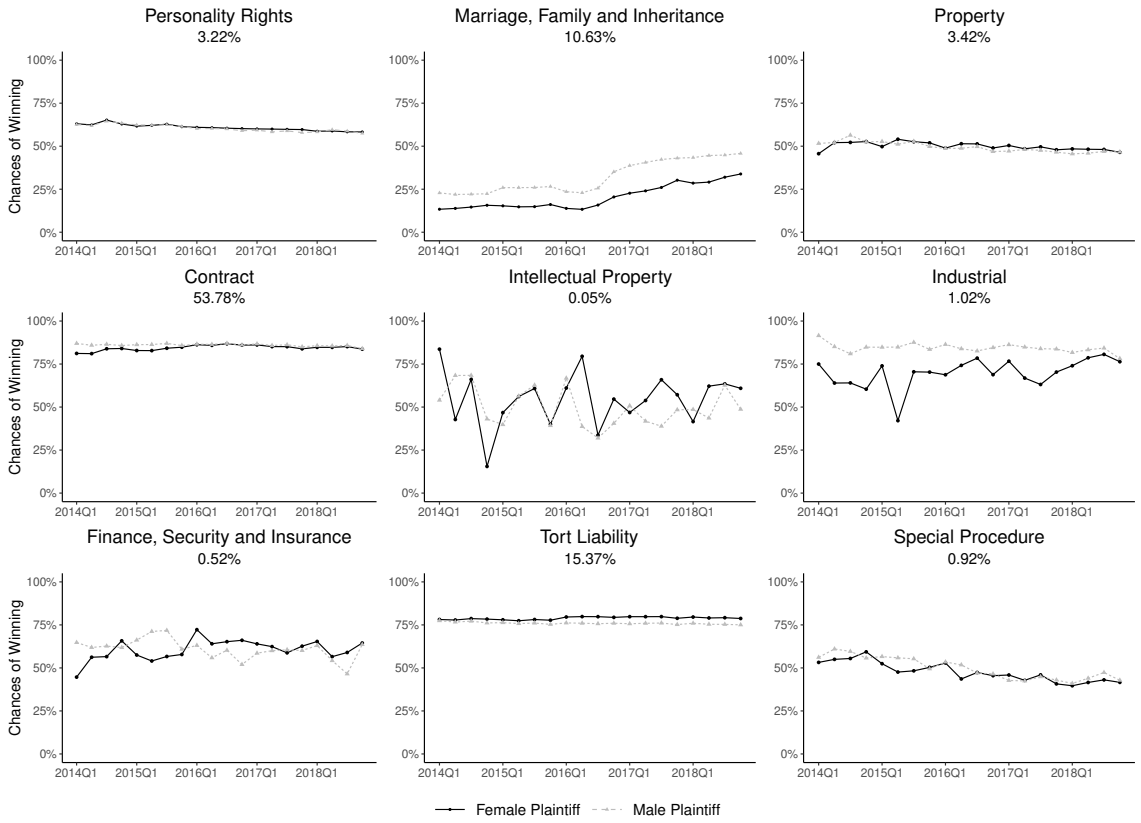
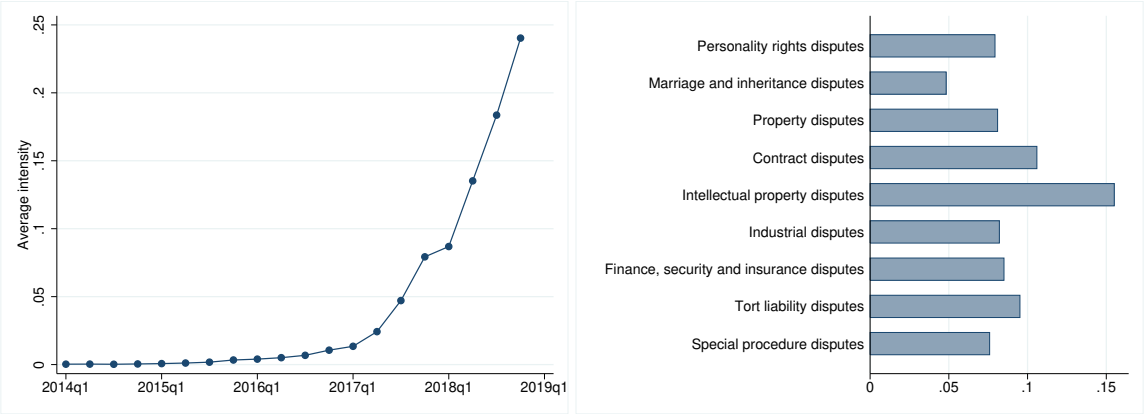


Figure 1. *Plaintiff's Chances of Winning over Time and by Area. These figures show the average chances of winning for plaintiff by gender, in each issue area and over the period 2014-2018. There is no obvious trend over time, except in the areas of marriage, family and inheritance (more than 10 percent of all cases) and special procedures (less than 1 percent of all cases). Considering that cases in area of marriage, family and inheritance may be gender sensitive, we estimate the effects of litigant and judge gender with and without cases in this area.*

Table 2 presents the distribution of cases across areas. In our empirical analysis, we investigate the role of gender by issue area. Column (1) presents the distribution in the main sample. We also present the quantity of cases and their proportions in each area, depending on plaintiff gender, shown in columns (2) and (3) of Table 2, respectively. Columns (4), (5) and (6) of Table 2 present the counterparts for the two-litigant subsample. There are large variations across areas. Cases in the area of contract disputes are the most common civil cases. The area of tort disputes also accounts for a substantial fraction of all cases, despite plaintiff gender. Marriage, family and inheritance cases are another major area in civil litigation, accounting for 10 percent of all cases in the main sample.

Next, we turn to the pattern of litigation outcomes. Figure 1 illustrates the plaintiff's chances of winning over time, by area and by plaintiff gender. There is no obvious trend over time, except in the areas of marriage, family and inheritance (more than 10

in the two-litigant subsample, given that the size of cases might correlate with the number of litigants.



(a) Over Time (b) By Areas

Figure 2. *Average Intensity of Live Broadcasting over Time and by Area. The left plot illustrates the intensity of broadcast civil cases over time. The right plot shows the intensity across across issue areas.*

percent of all cases) and special procedures (less than 1 percent of all cases). In our empirical analysis, we will investigate the impacts of litigant and judge gender with and without the area of marriage, family and inheritance and contrast them based on the conjecture that cases in this area may be gender sensitive.

In addition to the chances of winning, we also present summary information about the policy intensity across courts, over time and by area. Figure 2(a) illustrates the average intensity of courts across China over time: the average intensity of live broadcasting is growing exponentially and reaches roughly 0.25 in the final quarter of 2018.

The histogram 2(b) on the right shows the average intensity of live broadcasting for each issue area from September 2016 to December 2018. The areas of marriage, family and inheritance disputes record the lowest intensity, i.e., 5 percent. That is reasonable, given civil cases involving personal privacy are generally not broadcast. Cases in the area of intellectual property disputes are broadcast the most often, i.e., 15 percent. For the rest of the areas, the variation in broadcast intensity is modest.

4. Empirical Strategy

We begin our investigation by examining the association of a plaintiff's gender with her or his chances of winning. We estimate the following equation:

$$y_{ijkt} = \beta_0 + \beta \text{Female}_i + \omega_j^{area} + \omega_k^{court} + \lambda_t + \varepsilon_{ijkt}, \tag{1}$$

where y_{ijkt} is the plaintiff's chances of winning in case i, issue area j, court k and year-quarter t and Female_i is a dummy variable that takes value 1 if the plaintiff is female in case i, 0 otherwise. We further include area, court, and year-quarter fixed effects—

ω_j^{area}, ω_k^{court} and λ_t, respectively—to control for area-, court- and time-specific factors that affect the plaintiff's chances of winning. All standard errors are clustered at the court level. The coefficient β is our main interest, capturing the gender difference in the chance of winning.

However, it is difficult to interpret the gender difference as causal: there might be unobservable fixed effects that correlate with gender, affecting litigation outcomes. Therefore, we exploit the variation in the policy implementation of broadcast trials and infer the gender bias, under the assumption that the open justice reform improves the fairness of litigation outcomes.

Specifically, we estimate a generalized DID model by controlling for court-area and year-quarter fixed effects as follows:

$$y_{ijkt} = \beta_0 + \beta \text{Female}_i + \gamma \text{intensity}_{jkt} + \theta \text{Female}_i * \text{intensity}_{jkt} + \omega_{j\times k} + \lambda_t + \varepsilon_{ijkt} \quad (2)$$

where intensity_{jkt} is the ratio of live-broadcast cases to the total cases in issue area j, court k and year-quarter t and $\omega_{j\times k}$ is the court-area fixed effect. The coefficient θ for the interaction term is our main interest, capturing the impact of live broadcasting on the gender difference in chances of winning, conditional on fixed differences across court-areas and fixed differences across year-quarters.

The baseline model allows us to identify the causal effect of the reform on gender difference. Fixed differences across court-areas (or subcourts within a court) cannot drive our estimated effects, because we control for court-area fixed effects and exploit variation over time within court areas. Similarly, we difference out cross-year-quarter changes by controlling for year-quarter fixed effects.

We prefer this generalized DID model (i.e., Equation (2)) for a number of reasons. First, the DID specification differences out potential fixed effects that are correlated with gender. Second, it is the head judge of each subcourt who decides which cases to broadcast; therefore it is a legitimate concern that the head judge selects straightforward cases to broadcast, i.e., those that can easily be adjudicated fairly. However, such a case-level selection issue is less relevant in this specification. If the reform has no impact at all on litigation outcomes, then reshuffling cases across the categories of being broadcast or not would not affect the estimated DID coefficient, because the treatment—intensity_{jkt}—is defined at the court-area $\times$ year-quarter level, rather than at the case level. Finally, the policy impact of enhanced transparency in the courtroom on litigation outcomes is in itself interesting.

To be cautious, in addition to the baseline, we will also estimate a specification with a set of control variables X_i at both the case and region levels. We include the gender

of the defendant, which may affect the plaintiff's chances of winning, the number of plaintiffs and the number of defendants, which may correlate with the size and sophistication of the case and whether they appear in court at all, which in turn might affect the winning odds. The number of lawyers for the plaintiff and defendant is considered, which is used to capture the legal resources of both sides. Judge gender is controlled for in all estimations in section 5, whose effects will be analyzed in detail in section 6. In addition, a legal fees variable (in logs) is included to proxy for the size of the case and instance is also included since the characteristics of cases may differ across the first and second instances. For prefecture-level variables, we include GDP per capita and population to proxy for the development and size of a region and the internet penetration rate to proxy for technological status.

We also include court $\times$ year-quarter fixed effects to alleviate the concern that unobserved factors may vary at the court $\times$ year-quarter level (e.g., changes in court heads and other reforms implemented during the same period). We also estimate a specification that includes court-specific time trends, which is able to capture smooth, court-specific changes.

Our identification strategy implicitly depends on an assumption that the rollout of the reform is exogenous. However, this assumption can be violated at both the intensive and extensive margins. First, it is possible that the choice of intensity by the court head is affected not only by the technical capacity or pressure from the superior court but also by factors that influence plaintiffs' chances of winning. To mitigate this concern, we resort to the randomization inference (RI) approach. We randomize the intensity of live broadcasting at the court-area $\times$ year-quarter level. Specifically, we first randomize the treatment of live broadcasting among cases in each cell of issue area $\times$ year-quarter $\times$ level of courts (i.e., local court, intermediate court and high court). With such a randomized sample, we can reconstruct the broadcasting intensity at the court-area $\times$ year-quarter and re-estimate the DID specification (Equation 2). This exercise is repeated 2,000 times, which allows us to compare the actual treatment effect to the distribution of placebo treatment effects and compute the RI based p-value. The comparison would leave us more confident that the concern is not too worrisome, if the pseudo p-value is indeed low.

There might still be concerns that changes in intensity at the court-area level can be driven to some extent by changes in gender differences in the litigation outcomes. To relieve this concern, we resort to the Bartik instrumental variable (IV) approach. The Bartik instrument we construct exploits two sources of variation: the distribution of cases across issue areas in each court before the reform (specifically, 2014); and the time-varying average broadcasting intensity for each issue area and each prefecture

(excluding the focal court). The predicted intensity is given by:

$$Z_{kt} = \sum_{j \in \text{issue area}} \alpha_{jk}^{2014} \, \text{intensity}_{jp_{-k}t} \tag{3}$$

where Z_{kt} is the instrument that predicts the proportion of cases broadcast in court k and year-quarter t. The term α_{jk}^{2014} is the share of issue area j for court k in year of 2014, i.e., 2 years before the implementation of the reform. The term $\text{intensity}_{jp_{-k}t}$ is the average intensity in issue area j and year-quarter t of prefecture p (where court k is located), excluding court k.

As discussed in Goldsmith-Pinkham, Sorkin, and Swift (2020), the validity of this Bartik instrument depends on the exogeneity of the early shares of issue areas in 2014. That is, the exclusive restriction condition requires that conditional on court and year-quarter fixed effects, third factors affecting the litigation outcome after the reform should not be simultaneously correlated with both the early issue area share in 2014 and the post-reform prefecture-level average broadcasting intensity by areas.

To illustrate the exogeneity of the early share, we show that the composition of civil cases in each court is very stable over time in our data period and not affected by the reform.[20] To do so, we design an event-study model to test the pre-trend of issue area share and show that the changes in shares of areas are very small and not significant after the reform:

$$Share_{kt} = \beta_0 + \sum_{\tau=-8}^{8} \gamma_\tau \text{Reform}_{k,t}^{\tau} + \omega_k^{court} + \lambda_t + \varepsilon_{kt}, \tag{4}$$

where $Share_{kt}$ is the share of cases in one particular issue area in court k and year-quarter t and $\text{Reform}_{k,t}^{\tau}$ (for τ = -8,...,8) is a sequence of dummy variables, indicating that in court k, quarter t is τ quarters away from the introduction of the reform. To absorb the effects outside of the estimation window, we additionally set $\tau = -8$ for $\tau \leq -8$ and $\tau = 8$ for $\tau \geq 8$, respectively. The quarter prior to the reform (i.e., $\tau = -1$) is dropped as the reference period. We control court and year-quarter fixed effects and standard errors are clustered at the court level.

The constant legal shares lay a favorable basis for the validity of our instrument: even if there exist confounders correlated with the early share, they are unlikely to be

[20] First, the distribution of cases across issue areas is mainly determined by fixed regional characteristics such as economic structure and local culture, etc, which is unlikely to change much over a 5-year period. Second, plaintiffs are not likely to choose which court to initiate their litigation, based on whether their cases would be broadcasted. The courts that plaintiffs can initiate their litigation are strictly regulated and plaintiffs do not have room to choose. According to Chapter 2 of China Civil Procedure Law, a lawsuit shall be brought in the place where the plaintiff is located, the defendant is located or where the dispute occurs. The default is that the plaintiff initiates the lawsuit where the most important defendant is located, unless it is otherwise specified in the contracts.

time-variant and therefore can be absorbed by fixed effects in our specification.

To further strengthen our confidence that restriction condition of our instrument holds, we conduct two sets of placebo tests. If there were long-run factors that contaminated the predicted intensity we construct, we would expect that litigation outcomes *before* the reform should be correlated with the predicted intensity *after* the reform. Specifically, we test whether litigation outcomes in 2014, 2015 and the first 3 quarters of 2016 (prior to the reform) respond to the predicted intensity in 2018. Similarly, if there were short-run factors that contaminated the instrument (e.g., earlier litigation outcomes motivate courts to speed up or slow down implementation of the reform), we would expect litigation outcomes to be correlated with lagged predicted intensity. This possibility is investigated by estimating our baseline model with one- or two-quarter lagged predicted intensity. We can be more confident about the validity of this Bartik instrument, if there exists no such correlation.

Second, at the extensive margin, the timing of the introducing the reform in each prefecture is determined by the corresponding intermediate court and may not be purely random either. For instance, imagine a certain unobserved social or cultural change in some regions had reduced the gender gap in litigation outcomes prior to the reform and simultaneously caused early adoption of the reform. To examine whether the parallel trends assumption holds, we estimate an event study model with the following Equation (5):

$$\begin{aligned} y_{ijkt} =& \beta_0 + \beta \text{Female}_i + \sum_{\tau=-8}^{8} \gamma_\tau \text{Reform}_{p,t}^{\tau} + \sum_{\tau=-8}^{8} \theta_\tau \text{Female}_i * \text{Reform}_{p,t}^{\tau} + X'\zeta \\ &+ \omega_j^{area} + \omega_k^{court} + \lambda_t + \varepsilon_{ijkt}, \end{aligned} \tag{5}$$

where y_{ijkt} is the outcome in case i, issue area j, court k and year-quarter t and Female_i is a dummy variable that takes value 1 if the plaintiff is female in case i, 0 otherwise. $\text{Reform}_{p,t}^{\tau}$ for τ = -8,...,8 is a sequence of dummy variables, indicating that in prefecture p (where court k is located), quarter t is τ quarters away from the introduction of the reform. To absorb the effects outside of the estimation window, we additionally set $\tau = -8$ for $\tau \leq -8$ and $\tau = 8$ for $\tau \geq 8$, respectively. To study the impact of the reform on the gender differential, we include the interaction terms between Female_i and $\text{Reform}_{p,t}^{\tau}$. We further include area, court, and year-quarter fixed effects, i.e., ω_j^{area}, ω_k^{court} and λ_t, respectively. The same set of case- and prefecture-level control variables X as in our baseline specification is also included. The quarter prior to the reform (i.e., $\tau = -1$) is dropped as the reference period. All standard errors are clustered at the court level.

A remaining concern relates to the change in the composition of litigants in re-

sponse to the implementation of this reform: it might encourage women with less winnable cases to initiate litigation, who would not have initiated those cases otherwise. This mechanism is likely, as women are found to be more risk averse (Niederle and Vesterlund 2007; Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner 2011; Noussair, Trautmann, and Van de Kuilen 2014) and tend to file litigation when having a higher winning probability than men do. If it indeed exists, our estimate would be biased toward zero. Therefore, we investigate how the proportion of female plaintiffs changes in response to the open justice reform using a two-way fixed effects model.

What would be the sources of discrimination against women in the courtroom? We are interested in the role of judge gender, despite other likely contributing factors. Do male judges, on average, act less impartially and treat female litigants unfairly? The major obstacle to our causal inference is that the assignment of cases to female or male judges may not be entirely random in practice. Given that female judges are responsible for less than one-third of cases in our sample, it could be case that judge gender is inconsequential but rather that it is the assignment of cases that drives the difference across judge genders, if any.

To mitigate such a concern, we resort to the propensity score matching approach. The aim of the matching procedure is to find a set of cases adjudicated by male judges that display the same characteristics as the set of cases adjudicated by female judges. Toward this end, we employ the nearest-neighbor method without replacement. The balancing property of the propensity score will be tested before we construct the matched sample. With the matched sample, we estimate the following specification:

$$\begin{aligned} y_{ijkt} = {} & \beta_0 + \beta \text{Female}_i + \delta \text{Female Chief}_i + \theta \text{Female}_i * \text{Female Chief}_i \\ & + \omega_j^{area} + \omega_k^{court} + \lambda_t + \varepsilon_{ijkt} \end{aligned} \tag{6}$$

where Female Chief_i is the gender of the chief judge for case i that takes value 1 if the chief judge is female, 0 otherwise. If the coefficient on the interaction term θ is significantly positive, this implies that the disadvantage of female plaintiffs relative to male plaintiffs is smaller when the chief judge is female.

A natural follow-up question arises: do male judges respond more to the open justice reform than female judges? To provide an answer, we employ a difference-in-difference-in-differences model,

$$\begin{aligned} y_{ijkt} = {} & \beta_0 + \beta_1^1 \text{Female}_i + \beta_2^1 \text{intensity}_{jkt} + \beta_3^1 \text{Female Chief}_i \\ & + \beta_{12}^2 \text{Female}_i * \text{intensity}_{jkt} + \beta_{13}^2 \text{Female}_i * \text{Female Chief}_i + \beta_{23}^2 \text{intensity}_{jkt} * \text{Female Chief}_i \\ & + \theta \text{Female}_i * \text{intensity}_{jkt} * \text{Female Chief}_i + \omega_{j \times k} + \lambda_t + \varepsilon_{ijkt} \end{aligned} \tag{7}$$

Table 3. *Gender and Litigation Outcomes*

	Outcome Variable: Chances of Winning					
	All Litigants			Sub-sample: Two Litigants		
	All Areas		Excl Marriage	All Areas		Excl Marriage
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0293***	-0.0267***	-0.0166***	-0.0492***	-0.0500***	-0.0391***
	(0.00134)	(0.00139)	(0.00143)	(0.00197)	(0.00214)	(0.00227)
Controls	N	Y	Y	N	Y	Y
Area FE	Y	Y	Y	Y	Y	Y
Court FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Mean of outcome	0.744	0.744	0.790	0.707	0.707	0.792
N	4,755,860	4,174,135	3,867,789	2,419,745	2,141,814	1,864,599
R^2	0.334	0.351	0.261	0.434	0.453	0.328

Notes: This table presents the results for gender differentials in the plaintiff's chances of winning. Columns (1), (2) and (3) report the results with main sample and columns (4), (5) and (6) present the results with two-litigant subsample. In columns (3) and (6), we further exclude cases in the area of marriage and family disputes. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance, legal cost (log), and whether the case is broadcast. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

The coefficient on the triple interaction term captures how live broadcasting intensity affects the gender difference in the chances of winning when the chief judge is female, relative to that when the chief judge is male. If it is negative and significant, this would imply that the impact of the open justice reform on the gender difference in chances of winning is greater when the chief judge is male.

5. Gender, Reform and Litigation Outcomes

5.1. Baseline Results

Are female plaintiffs less likely to win civil litigation than male plaintiffs? We first estimate Equation (1) with our main litigation sample (i.e., all litigants), in which the number of litigants on each side can be more than one. Table 3 shows the results.

Table 4. *Difference-in-Differences Estimation: Impacts of Trials Broadcasting*

	Outcome Variable: Chances of Winning					
	All Litigants			Sub-sample: Two Litigants		
	All Areas		Excl Marriage	All Areas		Excl Marriage
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Intensity	0.0632***	0.0639***	0.0472***	0.0929***	0.105***	0.0875***
	(0.00345)	(0.00373)	(0.00376)	(0.00486)	(0.00562)	(0.00574)
Female	-0.0257***	-0.0237***	-0.0159***	-0.0427***	-0.0452***	-0.0381***
	(0.00143)	(0.00149)	(0.00153)	(0.00205)	(0.00229)	(0.00244)
Intensity	-0.0216***	-0.0241***	-0.0183***	-0.0195***	-0.0265***	-0.0193***
	(0.00418)	(0.00427)	(0.00434)	(0.00606)	(0.00592)	(0.00602)
Controls	N	Y	Y	N	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Mean of outcome	0.744	0.744	0.790	0.707	0.707	0.792
N	4,664,389	4,114,282	3,816,171	2,365,478	2,105,831	1,836,302
R^2	0.374	0.388	0.291	0.483	0.498	0.366

Notes: This table presents the main results for the DID specification. Columns (1),(2) and (3) report the results with main sample and columns (4),(5) and (6) report the results with the two-litigant subsample. In columns (3) and (6), we further exclude cases in the area of marriage and family disputes. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

Column (1) reports the estimated coefficient on the dummy Female, which is negative and statistically significant. Column (2) of Table 3 reports the estimates of Equation (1) with the set of control variables X_i discussed in section 4 and a dummy variable Live_i, which takes value 1 if case i is broadcast, 0 otherwise. The magnitude and significance of the coefficients do not change substantially.

For comparison, we repeat our estimations with the two-litigant subsample in which only one litigant is on either side. We report the results in columns (4) and (5) of Table 3, which are similar to those in columns (1) and (2) of the same table. A notable difference is that the coefficients of interest are consistently larger in size in this subsample. The gender difference in the chances of winning is nearly 5 percent in this subsample (i.e., column (5) of Table 3), compared with 2.67 percent in the main sample (i.e., column (2) of Table 3). The difference is not surprising: In the full sample

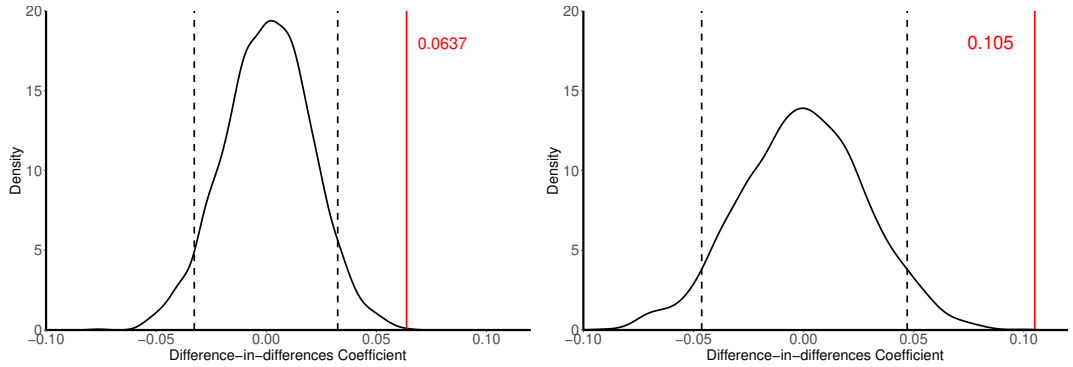
with all litigants, the gender of other litigants (besides the first litigant on each side) is not controlled for in our estimations. As a result, the mixture of gender would bias our estimate toward zero.

To ensure that the association is not driven only by gender-salient issues, such as cases in the area of marriage, family and inheritance, we drop those cases and repeat our exercises with the two samples. The results are reported in columns (3) and (6) of Table 3. Both estimated coefficients are smaller in size but still very significant, suggesting such an association exists in areas unrelated to gender.

Next, we estimate the generalized DID model specified by Equation (2). The estimation results using the main sample without control variables and when using control variables are reported in column (1) and column (2) of Table 4, respectively. For comparison, we also report their counterparts using the two-litigant subsample in columns (4) and (5) of Table 4, respectively. Furthermore, we drop cases in the area of marriage, family and inheritance and repeat our exercises with the two samples. The results are reported in columns (3) and (6) of Table 4, respectively. All specifications produce consistent results: the disadvantage of female plaintiffs relative to males decreases when a larger fraction of cases are broadcast live. The magnitude of this effect is fairly large: using the estimate in column (5) of Table 4 as a basis, one-fifth of the female plaintiffs' disadvantage (i.e., $0.105 \times 10\% \approx 1$ percentage point) would be removed if the broadcast intensity rose to 10 percent from zero percent. The fact that the disadvantage of female plaintiffs relative to males shrinks when litigations become fairer is evidence of the presence of gender discrimination.

Prevalence Does this pattern exist in areas that are not gender salient but economically relevant? To investigate, we re-estimate our generalized DID model (i.e., Equation (2)) for cases in each area in our main sample and the two-litigant subsample. We find that the estimated coefficients on the interaction term, i.e., Female $\times$ Intensity, are positive and significant for 5 areas—marriage, family and inheritance disputes, contract disputes, intellectual property disputes, industrial disputes and tort liability disputes—and that they are insignificant in the remaining 4 areas. This suggests that the bias against female plaintiffs is rather prevalent, not only in gender sensitive areas. The results are relegated to Table 19 of Appendix B.1.

Defendant Gender As discussed in the introduction, plaintiffs and defendants are in asymmetric positions in civil litigations, i.e., judges make decisions on the extent to which plaintiffs' claims are supported. How does defendant gender play a role? Not surprisingly, the gender difference of plaintiffs is even larger when the defendant is male than when it is female. Furthermore, when the defendant is male, the impact of the reform is stronger and the gender difference can be reduced by a larger degree.



(a) RI with the main sample (b) RI with the two-litigant sample

Figure 3. *Randomization Inferences. We randomize the court-area level intensity of live broadcasting 2,000 times and re-estimate Equation (2). Kernel density distributions (i.e., the solid lines) are close to symmetric around zero in both samples. The estimated actual treatment effect, marked by the vertical solid line, is far outside of the 95% confidence interval (vertical dashed lines).*

The results are relegated to Table 20 of Appendix B.2.

5.2. Identification Issues

Time-varying Shocks In this section, we address a host of potential concerns about our identification strategy raised in section 4. To rule out confounding factors such as time-varying shocks at the court level, we estimate two additional specifications. That is, we add to Equation (2) court $\times$ year-quarter fixed effects and a court-specific time trend. The results, relegated to Table 10 (see Appendix A.1), are rather similar to those of the baseline specifications.

Omitted Variables Although anecdotal evidence (documented in section 2) suggests that the intensity of live broadcasting (determined at the court level) is largely exogenous to litigation outcomes (determined by judges), it is still legitimate to question whether there are unobserved factors that drive both variables. To address this issue, we adopt the RI approach (elaborated in section 4). Figure 3 illustrates the kernel densities of those estimated coefficients. First, the RI-based p-value is 0.003 in the main sample and even lower in the subsample (less than 0.001). Second, if the issue of omitted variables were worrisome, we would expect the distribution to shift toward the right, with a significantly positive mean. It is not the case: the 95% confidence intervals in both samples contain zero.

Bartik Instrument Another challenge to the exogeneity of the live broadcasting intensity is the possibility of reverse causality. To mitigate this concern, we resort to the Bartik IV approach (elaborated in section 4). Table 5 reports our IV estimates, using the main sample and two-litigant subsample. For the main sample, columns (1) and

(2) of Table 5 present results from the first stage and column (3) shows our key estimate of interest from the second stage. For the two-litigant sample, the corresponding results from the first and second stages are shown in columns (4), (5) and (6) of Table 5. The coefficients in the first stage is highly significant, implying that the instrument has high predictive power. As shown in columns (1) and (4), every percentage in predicted broadcasting intensity is associated with 0.702 and 0.714 percentage in actual intensity. The LM and Wald tests imply that underidentification and weak identification are unlikely to be concerns for our IV estimations. The IV estimates for the interaction term shown in columns (3) and (6) remain positive and highly significant. This is reassuring, because our IV approach exploits variation in the intensity of live broadcasting that is, arguably, independent of litigation outcomes.

That the size of the IV estimates is only slightly larger than the corresponding estimates in the baseline generalized DID model (reported in columns (2) and (5) of Table 4) also warrants attention. The fact that the magnitudes of the estimates are comparable across specifications with and without the instruments suggests that the conjectured endogeneity issue is less worrisome.

We use an event study model specified by Equation (4) to test the pre-trend of issue area shares. As shown in Figure 5 and 6 in Appendix A.3, after controlling the court and year-quarter fixed effects, shares of all areas except the marriage disputes, are very stable over time and changes in shares over time are not statistically different from zero. Considering that the share of marriage disputes shows a moderate decrease before the reform, we re-do our exercises by excluding cases in this area from our sample. The IV estimation results are still robust (see details in Appendix A.3).

To mitigate the concern that there are long-run factors that contaminated our predicted intensity, we estimate Equation (2) using litigation outcomes from the 2014, 2015 and 2016 (first three quarters) subsamples and the predicted intensity constructed for the subsample of 2018. The results are shown in Table 13 of Appendix A.3. All the coefficients on the interaction term are statistically insignificant. To mitigate the concern that there are short-run factors that contaminated our predicted intensity, we estimate Equation (2) using one- or two-quarter lagged predicted intensity. The results are shown in Table 14 of Appendix A.3: the coefficients of our interests are statistically indistinguishable from zero. The results combined indicate that it is less likely that the early shares we use to construct the instrument are endogenous.

Pre-existing Trend We estimate the event study model specified by Equation (5) in section 4 with both the main sample with all litigants and the two-litigant subsample. The sequence of coefficients θ_τ is our main interest. Figure 4 shows the estimated coefficients θ_τ over τ. In both samples, there is no pre-trend in the gender differential

Table 5. *Instrumental Variable Estimations*

	Bartik Instrument Estimation					
	All Litigants			Sub-sample: Two Litigants		
	1st Stage		2nd Stage	1st Stage		2nd Stage
Outcome Variable	Intensity	Female × Intensity	Chances of Winning	Intensity	Female × Intensity	Chances of Winning
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Intensity			0.111***			0.168***
			(0.00823)			(0.0112)
Female	0.000406	0.0216***	-0.0268***	0.000512	0.0208***	-0.0493***
	(0.000298)	(0.00159)	(0.00160)	(0.000335)	(0.00160)	(0.00245)
Intensity			-0.0212***			-0.0260***
			(0.00722)			(0.00749)
Predicted Intensity	0.702***	-0.0677***		0.714***	-0.0629***	
	(0.0410)	(0.00638)		(0.0413)	(0.00593)	
Female × Predicted Intensity	-0.0164***	0.933***		-0.0160**	0.956***	
	(0.00618)	(0.0333)		(0.00661)	(0.0319)	
Controls	Y	Y	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
K-P LM Statistics			149.737			137.550
K-P Wald Statistics			146.155			150.913
N	4,114,282	4,114,282	4,114,282	2,105,831	2,105,831	2,105,831
R^2	0.692	0.541	0.027	0.713	0.560	0.040

Notes: This table presents the first stage and second stage results for the IV regressions. We construct a Bartik instrument by using the early shares of legal areas in each court in 2014 and the time-varying prefecture-level average broadcasting intensity for each legal area (excluding the focal court). For the main sample, columns (1) and (2) present results from the first stage and column (3) shows our the estimate of interest from the second stage. For the two-litigant sample, the corresponding results from the first and second stages are shown in columns (4), (5) and (6). Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

before the reform was introduced, leaving us more confident about the parallel trends assumption. In the post-reform period, the estimated coefficients increase gradually and become significantly positive when τ is large. This pattern is not surprising, as the intensity of the reform increases over time.

5.3. Robustness

Definitions of Independent Variables To examine whether our results are robust to the effects of live broadcasting, we re-estimate Equation (2) with alternative measures discussed in section 3.2. Table 11 (in Appendix A.2) reports the results using Top-25 and the overall intensity respectively. Those estimates are reported in columns (1) and (2) for the main sample and in columns (5) and (6) for the two-litigant subsample. They are consistent with those estimated using the default court-area level intensity

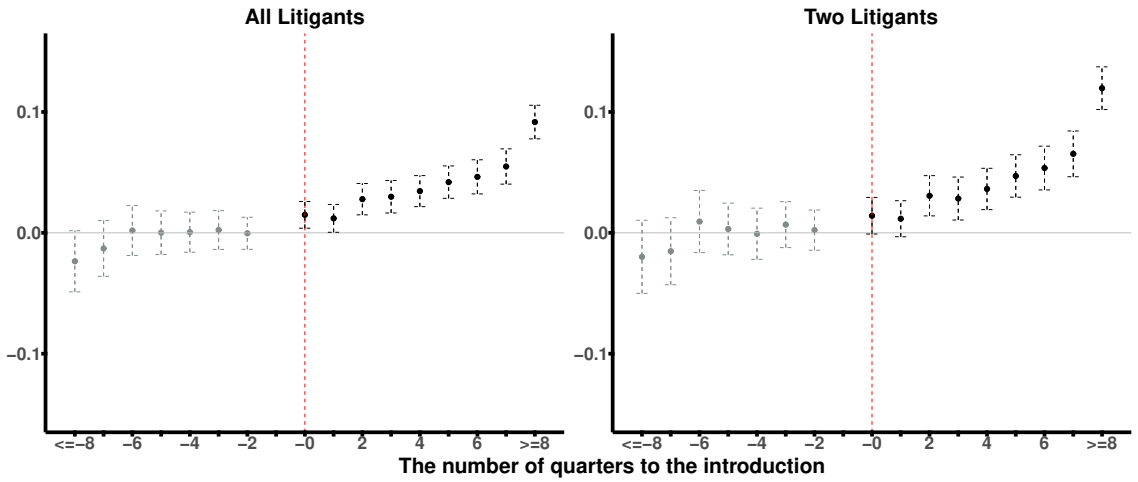


Figure 4. *Event study. This figure depicts the estimated effects of the reform on the gender differential in the chances of winning over τ, i.e., from 8 quarters before the introduction of the reform ($\tau = -8, ..., -1, 0$ with $\tau = -1$ dropped as the reference point) to 8 quarters after the reform ($\tau = 1, ..., 8$). The grey and black dots represent the estimated coefficients for each period, before and after the reform, respectively; the dashed lines represent 95 percent confidence intervals. No pre-trend is found in either the main sample or the two-litigant subsample.*

measure (columns (2) and (5) of Table 4), although, unsurprisingly, the magnitude and significance vary.

Definitions of Dependent Variables Next, we test whether our results are sensitive to the measure of winning litigation. We re-estimate Equation (2) by using the dummy Win as the dependent variable. The estimates, shown in columns (3) and (7) of Table 11 (in Appendix A.2), are also close to those estimated using the default chances of winning (columns (2) and (5) of Table 4), suggesting that our results are robust.

Samples In the main sample and the subsample with two litigants, we drop the cases with missing gender information. Using the name-based gender deduction method, we deduce the gender of litigants in those cases and supplement the two samples with that particular set of cases (which were dropped otherwise). The results are reported in columns (4) and (8) of Table 11 (in Appendix A.2), which are barely different from the baseline results.

Extensive Margin Our findings show that the open justice reform reduces female plaintiffs' disadvantage in court, i.e., a change on the intensive margin. A natural question arises about likely changes on the extensive margin. Do more women initiate litigation in response to the reform because they perceive the judicial system to be less biased against them? To provide an answer, we investigate how the proportion of female plaintiffs changes in response to the open justice reform using a two-way fixed effects model. We find no evidence that the reform encourages women to initiate liti-

gation and that the reform changed the gender composition of plaintiffs in each area. All the analysis and results are relegated to Appendix A.4.

Alternative Specification In our main analysis, we have utilized variations in reform treatment at the court-area and year-quarter level to identify the effects of the reform. Can we find consistent evidence by using variations in treatment at the case level that the reform reduces the gender-based difference in litigation outcomes? In Appendix A.5, we show that the effect of the reform is still robust when we conduct our analysis using the treatment at the case level.

6. Judge Matters

Having documented gender-based disparities in litigation outcomes, we turn to the role of judge gender in our further pursuit of the sources of such disparities. The conjecture is warranted that the gender difference in plaintiffs' winning odds is lower when female judges decide the cases. In general, numerous studies have shown that female legislators and politicians understand and represent women's interests better than their male counterparts (e.g., Chattopadhyay and Duflo 2004; Beaman, Chattopadhyay, Duflo, Pande, and Topalova 2009; Clots-Figueras 2012). Similar mechanisms could also imply that female plaintiffs are better supported when a female judge is on the bench. In particular, societal gender-based preferences in the form of men's prejudice likely influence how male judges interpret evidence, parse legal principles, and make judgments, preventing them from judging impartially (Patton and Smith 2017; Knepper 2018).

6.1. Does Judge Gender Matter?

Two specific questions emerge. First, we are interested in the question of whether male chief judges discriminate against female plaintiffs to a greater extent than female chief judges do. Second, do male judges react to the open justice reform more than female judges? In China, it had long been a principle that cases should be randomly assigned to judges. In 2015, the Supreme Court implemented this rule more strictly and explicitly requested that courts at all levels adopt the "random allocation rule" and allocate cases to judges with computerized algorithm.[21] If more than one judge is involved, as in the general procedure, the chief judge shall be determined at random.

However, one may worry that the implementation of the random allocation rule is not perfect across courts and over time. We take three measures to alleviate this

[21]See Article 2 of The Several Acts of Enhancing People's Court Judges' Accountability issued by the Supreme Court in September 2015 and Article 3 of The Acts of Enhancing Judges' Accountability and Trial Supervision System issued by the Supreme Court in April 2017.

Table 6. *Judge Gender and Litigation Outcomes: Matched Sample*

	Outcome Variable: Chances of Winning					
	All Litigants			Sub-sample: Two Litigants		
	All Areas	Excl Marriage	All Areas	All Areas	Excl Marriage	All Areas
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Female Chief	0.0161***	0.00885***	0.0143***	0.0207***	0.0135***	0.0188***
	(0.00156)	(0.00163)	(0.00139)	(0.00210)	(0.00243)	(0.00196)
Female	-0.0352***	-0.0225***	-0.0283***	-0.0587***	-0.0461***	-0.0475***
	(0.00174)	(0.00184)	(0.00154)	(0.00250)	(0.00276)	(0.00224)
Female Chief	-0.00173*	-0.000860	0	-0.00199*	-0.00144	0
	(0.000998)	(0.000996)	(.)	(0.00115)	(0.00113)	(.)
Controls	Y	Y	Y	Y	Y	Y
Judge FE	N	N	Y	N	N	Y
Area FE	Y	Y	Y	Y	Y	Y
Court FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Mean of outcome	0.744	0.744	0.744	0.707	0.707	0.707
N	3,741,092	3,442,232	3,700,449	1,994,585	1,721,601	1,958,253
R^2	0.367	0.267	0.424	0.466	0.333	0.532

Notes: This table presents the results for the in-group bias analysis. We use the matched sample and regress the plaintiff's chances of winning on the dummy Female, i.e. whether the plaintiff is female, the dummy Female Chief, i.e. whether the chief judge is female, and their interaction term, after adding control variables and fixed effects. We report the regression results for the main sample in columns (1), (2) and (3) and results for two-litigant subsample in columns (4), (5) and (6). In columns (2) and (4), we exclude cases in the area of marriage and family disputes. In columns (3) and (6), we include judge fixed effects. Case-level control variables include the defendant gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance, legal cost (log), and whether the case is broadcast. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

concern. First, we resort to the propensity score matching approach. The matching procedure and the resulting balancing property are relegated to Appendix C. Second, we conduct a randomization test, in which we randomize and reassign the gender of chief judges of the cases in our sample and implement our propensity score matching procedure to the placebo data. The idea is to examine the distributions of propensity scores of female and male judges using the placebo data and contrast them with their counterparts in our actual data. If they are rather similar, we should be more confident that the selection issue is not a crucial concern. Appendix C shows that it is indeed the case with details of the test and results. Finally, we additionally control for judge fixed effects so as to further mitigate the concern that the assignment of cases to judge

gender is largely driven by unobservables.

With the matched sample, we turn to our research questions in sequence. To provide an answer to the first question, we add the dummy Female Chief and its interaction with the dummy Female to Equation (1). The results estimated with the main sample and two-litigant subsample are reported in columns (1) and (4) of Table 6, respectively. A full set of control variables and whether the case is broadcast are included. To ensure that our results are not driven solely by gender-sensitive areas, we exclude cases in the area of marriage, family and inheritance disputes and re-estimate the aforementioned exercise with cases in the remaining areas. The results are reported in columns (2) and (5) of Table 6 for the main sample and two-litigant subsample, respectively. We re-do the exercise with cases of all areas but add judge fixed effects. The respective results can be found in columns (3) and (6) of Table 6.

If the plaintiff is male, it at best marginally matters whether a female or male judge decides the case; the estimate for the Female chief dummy is negative and only marginally significant in columns (1) and (4). This small effect disappears once we drop cases in the area of marriage, family and inheritance (i.e., columns (2) and (5)).

However, even in areas unrelated to marriage and family (i.e., column (5) of Table 6), female plaintiffs are still at a disadvantage relative to males by more than 4 percentage points if the chief judge is male (i.e., the estimate for the Female dummy is –4.6 percentage points and significant in column (5)). The coefficient on the interaction term is 1.35 percentage points and significant, suggesting that the gender difference would be reduced by more than one quarter if the chief judge were female. Our results suggest that female and male judges decide cases with female plaintiffs distinctly in both gender-related and gender-unrelated areas. Once we include the judge fixed effects, the magnitude of coefficients on the interaction term changes only slightly.

To investigate the second question, we first subdivide the matched main sample into separate samples for female and male judges and re-estimate Equation (2). The results are reported in columns (1) and (2) of Table 7, respectively. The estimated coefficients for the interaction term, i.e., Female $\times$ Intensity, are both positive, consistent with those estimated with our baseline sample. However, the size of this estimate in the subsample with male chief judges is substantially larger, suggesting that the impact of the reform may be larger for male judges.

To confirm this, we estimate the difference-in-difference-in-differences model specified with Equation (7). The result is shown in column (3) of Table 7: The estimated coefficient on the triple interaction term is indeed significantly negative. To see whether such a pattern remains if we rely on the two-litigant subsample, we re-estimate the

Table 7. *Differential Responses across Judge Genders*

	Outcome Variable: Chances of Winning			
	All Areas			Excl Marriage
	Sample with Chief Judge Gender:			
	Female	Male	Both	
All Litigants	(1)	(2)	(3)	(4)
Female × Intensity	0.0492***	0.0763***	0.0787***	0.0584***
	(0.00475)	(0.00472)	(0.00471)	(0.00480)
Female × Intensity × Female Chief			-0.0309***	-0.0223***
			(0.00545)	(0.00559)
Mean of outcome	0.755	0.744	0.747	0.749
N	1,122,547	2,561,442	3,686,913	3,396,155
R^2	0.384	0.424	0.406	0.298
Sub-Sample: Two Litigants	(5)	(6)	(7)	(8)
Female × Intensity	0.0847***	0.116***	0.119***	0.100***
	(0.00675)	(0.00678)	(0.00671)	(0.00705)
Female × Intensity × Female Chief			-0.0342***	-0.0301***
			(0.00695)	(0.00729)
Mean of outcome	0.726	0.705	0.711	0.796
N	565,548	1,392,019	1,960,607	1,695,236
R^2	0.491	0.530	0.512	0.372
Controls	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Notes: This table shows the differential responses of female judges and male judges to the reform. We subdivide the matched sample into separate samples for female and male judges and re-estimate the in-group bias regression, as shown in columns (1) and (2) with the main sample and columns (5) and (6) with the two-litigant subsample. We further use a triple difference specification to examine the coefficients on Female × Intensity across judge genders and the results are shown in columns (3) and (4) with the main sample and columns (7) and (8) with the two-litigant subsample. Case-level control variables include intensity, whether plaintiff is female or not, the defendant gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms and other interaction terms in both specifications are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

aforementioned exercises and report the corresponding estimated coefficients in columns (5), (6) and (7). The finding is rather similar. We also examine whether such a pattern exists in areas other than marriage, family and inheritance disputes by dropping cases in this area. We re-estimate the difference-in-difference-in-differences model specified with Equation (7) with cases in the remaining areas of the main sample and the two-litigant subsample. The results are reported in columns (4) and (8), respectively, and the coefficients for the triple interaction term are both negative and significant.

One interpretation of those results is that male judges respond to the open justice reform to a greater extent than female judges, given that they are more discriminative against female plaintiffs. In other words, the disciplinary effect of live broadcasting is greater for male judges.

Our findings on judge behavior contrasts with the existing literature. The question of whether female and male judges decide cases distinctly has been well researched in the context of the U.S. judicial system, where judge characteristics and decision data are well archived and easily available. However, only in very restricted areas can researchers find an effect of judge gender, such as in the area of sex discrimination in the employment context (Boyd, Epstein, and Martin 2010). Against the backdrop of this literature, the reincarnation of this research question in the context of the Chinese legal system is even more intriguing. We found that women and men indeed judge differently and the difference is not confined to gender-sensitive areas but rather is consistently present in other areas. Furthermore, the extent to which male and female judges respond to the policy intervention also varies. The two findings complement each other, suggesting that judge gender matters in general for litigation outcomes in the Chinese context.

6.2. Judges or Litigants?

Thus far, the interpretation of our findings is based on the assumption that broadcasting trials would enhance the transparency of the courtroom and effectively discipline judges' behavior by allowing the public, including other legal professionals, to monitor the adjudication process. Nevertheless, a plausible alternative is that litigants may change their behavior during trials after they learn that their cases are being broadcast; accordingly, judges may change their decisions in response to litigants' behavioral changes, instead of minding the broadcasting cameras in front of them.

However, it is unlikely that such a hypothesis explains away our findings. If this idea were to hold, the latent assumption would have to be that female plaintiffs do not choose to behave to their advantage when their trials are not broadcast. Incidences that are consistent with such an assumption of irrationality could arise idiosyncratically but are less likely to occur systematically.

Nevertheless, we implement an empirical strategy to re-examine our findings in a subsample where we can eliminate the potential confounding effect of litigants' behavioral changes. Our dataset is so rich that it includes appearance records of plaintiffs and defendants, which allows us to build a subsample of cases in which none of the plaintiffs or defendants appear in court. To address the aforementioned concern, this subsample is useful, as changes in the litigants' behaviors cannot play a role in the trial, given that none of them appear.

It is noteworthy that, in China, civil-case litigants have the right to not appear in court and do not have to inform the court in advance.[22] Therefore, judges do not

[22] According to Articles 143 and 144 of China Civil Procedure Law, the only exception is in cases

Table 8. *DID Estimations with Subsample of No Litigant Appearing in Court.*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	(1)	(2)	(3)	(4)
Female × Intensity	0.0272***	0.0316***	0.0576***	0.0661***
	(0.00369)	(0.00441)	(0.00537)	(0.00673)
Female × Intensity × Female Chief		-0.0125*		-0.0249**
		(0.00667)		(0.00969)
Controls	Y	Y	Y	Y
Area FE	Y	Y	Y	Y
Court FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.750	0.750	0.735	0.735
N	1,742,126	1,742,126	718,340	718,340
R^2	0.347	0.347	0.465	0.465

Notes: This table presents DID estimation results with a subsample that none of litigants appears in court. Both estimated coefficients are still positive and highly significant, although the size of these estimates is only half of those in the baseline case (see Table 4). Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms and other interaction terms in both specifications are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

know in advance whether litigants will appear for the trial. Furthermore, the courts do not inform the litigants whether the upcoming cases will be broadcast live, leaving litigants unaware of the court choice in advance.

We construct such a subsample in which no litigant appears from the matched sample (constructed in section 6) and estimate Equation (2) with it. As usual, we re-estimate it with cases only involving two litigants in this subsample. The results are reported in columns (1) and (3) of Table 8, respectively. Both estimated coefficients are still positive and highly significant, although the size of these estimates is only half of those in the baseline case (see Table 4).

We proceed to estimate Equation (7) with the two samples and report the respective results in columns (2) and (4) of Table 8. The sign of the estimated coefficient on the triple interaction is negative, still in line with those estimated with the full sample (with and without litigants' appearance, reported in Table 7). The coefficient on the triple interaction is less significant, suggesting that the difference between female and

involving custody in the area of marriage, family and inheritance disputes, where litigants are required to appear in court.

male judges in response to live broadcasting is smaller when litigants are absent.

Those results, taken together, suggest that the live broadcasting of trials does affect litigation outcomes through changes in judges' behavior.

6.3. Does the Open Justice Reform Promote Judicial Fairness?

Our empirical analysis shows that female plaintiffs' disadvantage has decreased since the open justice reform was introduced. We interpret this as evidence of gender bias in the judicial system, under the assumption that the reform improves the fairness of litigation outcomes. In this section, we provide additional evidence in support of this assumption by using information on litigants' appeal decisions following the first instance trial. According to Article 164 of the China Civil Procedural Law, if one party of the litigation is not satisfied with the outcome in first instance and believes it is unfair, she or he shall have the right to appeal to a higher level court within 15 days after the first instance. We investigate the impacts of the trial broadcasting reform on female plaintiffs' appeal decisions.

We do not have information on whether it is the plaintiff or the defendant who appeals. Therefore, we focus on a subsample where plaintiffs lose the case overwhelmingly in the first instance (i.e., the plaintiffs' share of legal costs is 100%) and plaintiffs' claims are completely rejected by judges. Therefore, if such a case is appealed, the appellant must be the plaintiff. Cases in this category account for 28 percent of first instance cases.

To investigate, we estimate the generalized DID model specified by Equation (2) with the dependent variable *Appeal*, i.e., a dummy variable, to which we assign value 1 if the plaintiffs appeal. To alleviate the concern that litigants' appeal decisions might be affected by other reforms implemented during the same period, we add court$\times$year-quarter fixed effect to capture unobserved factors varying with court and time.[23] The results are reported in column (1) of Table 9. We estimate the same equation with cases with only two litigants and report the results in column (4). Both estimated coefficients for the interaction term are negative and highly significant, indicating that female plaintiffs, relative to male plaintiffs, are less likely to appeal when a larger fraction of cases are broadcast. In other words, as the live broadcasting reform is in-

[23]Since August of 2014, the Supreme Court in China started promoting a fast arbitration procedure. If a civil case is ruled based on the procedure of fast arbitration, litigants are not allowed to appeal once the judge's decision during the first instance takes effect. The fast arbitration procedure is applicable to civil cases with clear facts, little dispute and involving a small amount of money (see the Opinions on Further Promoting the Diversion of Complicated and Simple Cases and Optimizing the Allocation of Judicial Resources issued by the Supreme People's Court on September 12, 2016 and the Operating Procedures for the Diversion of Complicated and Simple Civil Cases and Fast Arbitration Procedure issued by the Supreme People's Court on May 10, 2017).

Table 9. *Impacts of Trials Broadcasting on Appealing*

	Outcome Variable: Appeal (1=Yes)		
	Matched Sample with Chief Judge Gender:		
	Both	Female	Male
All Litigants	(1)	(2)	(3)
Female × Intensity	-0.0150***	-0.0180**	-0.0129**
	(0.00495)	(0.00898)	(0.00574)
Female	-0.000377	-0.000341	-0.000418
	(0.000795)	(0.00140)	(0.000938)
Intensity	0.00708	0.0265	0.00203
	(0.00930)	(0.0192)	(0.0110)
Mean of outcome	0.119	0.128	0.115
N	1,207,632	366,974	832,655
R^2	0.157	0.213	0.176
Sub-sample: Two Litigants	(4)	(5)	(6)
Female × Intensity	-0.0202***	-0.0163	-0.0230**
	(0.00756)	(0.0148)	(0.00929)
Female	0.00309***	0.00614***	0.00182
	(0.00114)	(0.00220)	(0.00132)
Intensity	0.0125	0.0275	0.0213
	(0.0146)	(0.0308)	(0.0180)
Mean of outcome	0.089	0.101	0.844
N	631,102	169,746	451,615
R^2	0.211	0.302	0.236
Controls	Y	Y	Y
Court-Area FE	Y	Y	Y
Court×Year-Quarter FE	Y	Y	Y

Notes: This table presents results for appealing decisions. We focus on a subsample where plaintiffs lose the case overwhelmingly in the first instance and estimate the generalized DID model specified by Equation (2) with the dependent being variable *Appeal*, i.e., a dummy variable, to which we assign value 1 if the plaintiffs appeal. The results are shown in column (1) with the main sample and column (4) with the two-litigant subsample. We also separate the sample into cases with female judge and with male judge, as shown in columns (2) and (3) and columns (5) and (6). Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

troduced, female plaintiffs (relative to males) are more willing to accept the litigation outcomes, even if they lose the lawsuit.

Does judge gender matter for the appeal decisions? We divide the constructed sample into separate decisions for female and male chief judges and then re-estimate Equation (2) with court×year-quarter fixed effects. The results for all litigants and two litigants are reported in columns (2) and (3) and (5) and (6) of Table 9, respectively.

For cases adjudicated by male judges, the estimated coefficients for the interac-

tion term in column (3) (all litigants sample) and column (6) (two-litigant sample), are consistent, i.e., both are negative and significant. The results imply that fewer female (relative to male) plaintiffs appeal after the reform is introduced and the broadcast intensity is raised, provided that male judges make decisions. However, for cases adjudicated by female judges, as shown in columns (2) and (4), the results are less clear-cut: while the coefficient on the interaction term is significantly negative in the all litigants sample, no significant impact is found in the subsample with two litigants. That is, in the two-litigant cases with female judges, the difference between male and female plaintiffs' appeal decisions do not react to the reform. In summary, the evidence is stronger that male judges became less unfair in the face of the reform and female plaintiffs are more willing to accept their decisions.

6.4. The Nature of the Discrimination and Effectiveness of the Reform

If discrimination exists in courts, is it purely statistical? Taken together, the evidence presented in this section suggests that statistical discrimination likely exists but that it is unlikely to explain away all the documented litigation disadvantages that women endure. Female and male judges do not treat male plaintiffs differently, suggesting that it is reasonable to assume that they possess similar sets of information. Then, the difference between male and female judges regarding cases involving female plaintiffs contradicts the interpretation that statistical discrimination is the only driving factor underlying our findings. Furthermore, if minimizing errors in adjudication is the only objective of judges based on available statistics, male and female judges should not react differently to the policy intervention. Therefore, our analysis suggests that prejudice against women may underlie our findings, at least to some extent.[24]

Is this a form of in-group bias? Shayo and Zussman (2011) find evidence that one offers favorable treatment to members of his or her own group in the context of small claims courts in Israeli. Bar and Zussman (2020) study the driver's license tests in Israeli and find that there exist both in-group and out-group bias. In our study, judges of both genders treat male plaintiffs equally, but male judges treat female plaintiffs worse than female judges. The evidence suggests that gender-based group identity could be one force underlying the male judges' decisions and less so for female judges. Such a contrast is consistent with recent economics and psychology findings that group-based bias is less pronounced among females (Lavy, Sand, and Shayo 2022 and Balliet, Wu, and De Dreu 2014).

Furthermore, such prejudice could be amplified by the salience effect (Bordalo, Gennaioli, and Shleifer 2015): A prejudiced male judge could discriminate against

[24]Giuliano (2017) and Giuliano (2020) provide excellent reviews of studies that explore the roots of bias against women.

female litigants to a larger extent when he overweighs the effects of litigant gender in his decision-making. The litigants' appearance in court could potentially make gender more salient. This is consistent with our findings that gender-based disparities can still be found when litigants do not appear in court and that the magnitude of the effect becomes larger when litigants do appear.

The effectiveness of the monitoring mechanism at reducing the gender disparities is not unexpected. The judicial reform was intended to improve judicial visibility: It makes visible judges' decision-making processes. First, based on our interviews with legal professionals, such as judges, lawyers and prosecutors, we learned that if judges favored litigants on one side, either intentionally or not, they would bend the procedural rules against the other side. For example, judges can choose to offer the other side less time to respond to questions or present evidence, which would be integral and important inputs for adjudication. In our setting, the enhanced transparency may incentivize judges to act more professionally by adhering court rules, which leads to more impartial litigation outcomes.

Second, the higher level judicial visibility also likely raises the price of discrimination, when judges act out their biased preferences. Parsons, Sulaeman, Yates, and Hamermesh (2011) found a similar mechanism in the setting of baseball games. Baseball umpires do express their racial preference when they evaluate pitchers, but such bias tends to be smaller when ballparks are outfitted with cameras to monitor umpires' decisions. Similarly, Zamoff, Greenwood, and Burtch (2021) found that body-worn cameras (BWCs) may cause police officers to regulate their own behavior and make less questionable arrests, because officers know that the encounter is being filmed.

Our findings about the effectiveness of the open justice reform, together with prior studies, suggest that when the group identity, based on which the victim group is discriminated against, is observable, then external monitoring devices that scrutinize the decision-making process (e.g., judges, umpires or police officers) can be effective.

7. Concluding Remarks: Affirmative Action

In this paper, we show that the gender-based disadvantage is pervasive and can be found in areas of business and economic issues, such as contract disputes, industrial disputes and tort liabilities disputes. In China, contract disputes alone account for more than 50 percent of all civil litigation. If entrepreneurs are discriminated against in the courtroom because they are women, this gender bias would constitute an additional friction, impeding women's economic decisions (e.g., career choices). In light of this possibility, gender bias in the courtroom may not only be a social equality issue but may also have substantial economic consequences.

We also show that the open justice reform reduced this bias by disciplining judges, especially male judges. The effectiveness of the judicial reform offers new insights into how enhanced transparency can help mitigate judicial bias in particular and behavioral biases in the decision-making process in general.

Furthermore, diversifying the gender of judges through affirmative action is warranted based on our evidence that discrimination may be a reflection of social prejudice and the in-group bias of male judges. Implementing affirmative action policies in this regard is not only relevant but feasible. In 2018, more than 70 percent of law school students and over 60 percent of the judicial examination candidates in China were female. Nevertheless, women account for only 27 percent of judges and an even smaller fraction of lawyers. Balancing judge gender composition and reserving quotas for women on the bench could be beneficial to both social equality and economic efficiency.

References

Abrams, D. S., M. Bertrand, and S. Mullainathan (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies 41*(2), 347–383.

Adams, Z., A. Adams-Prassl, and J. Adams-Prassl (2022). Online tribunal judgments and the limits of open justice. *Legal Studies 42*(1), 42–60.

Adams-Prassl, A., K. Huttunen, E. Nix, and N. Zhang (2022). Violence against women at work. *Oxford Discussion Paper*.

Alesina, A. and E. La Ferrara (2014). A test of racial bias in capital sentencing. *American Economic Review 104*(11), 3397–3433.

Ang, Y. Y. and N. Jia (2014). Perverse complementarity: Political connections and the use of courts among private firms in china. *The Journal of Politics 76*(2), 318–332.

Anwar, S., P. Bayer, and R. Hjalmarsson (2019a). A jury of her peers: The impact of the first female jurors on criminal convictions. *The Economic Journal 129*(618), 603–650.

Anwar, S., P. Bayer, and R. Hjalmarsson (2019b). Politics in the courtroom: Political ideology and jury decision making. *Journal of the European Economic Association 17*(3), 834–875.

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics 133*(4), 1885–1932.

Ash, E., S. Asher, A. Bhowmick, S. Bhupatiraju, D. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi (2021). In-group bias in the indian judiciary: Evidence from 5 million criminal cases. *Working Paper*.

Ash, E., D. L. Chen, and A. Ornaghi (2021). Gender attitudes in the judiciary: Evidence from us circuit courts. *Center for Law & Economics Working Paper Series 2019*(02).

Bagues, M. F. and B. Esteve-Volart (2010). Can gender parity break the glass ceiling? evidence from a repeated randomized experiment. *The Review of Economic Studies 77*(4), 1301–1328.

Balliet, D., J. Wu, and C. K. De Dreu (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological Bulletin 140*(6), 1556.

Bar, R. and A. Zussman (2020). Identity and bias: insights from driving tests. *The Economic Journal 130*(625), 1–23.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: does exposure reduce bias? *The Quarterly Journal of Economics 124*(4), 1497–1540.

Beath, A., F. Christia, and R. Enikolopov (2013). Empowering women through development aid: Evidence from a field experiment in afghanistan. *American Political Science Review 107*(3), 540–557.

Beim, D., T. S. Clark, and B. E. Lauderdale (2021). Republican-majority appellate panels increase execution rates for capital defendants. *The Journal of Politics 83*(3), 000–000.

Bertrand, M., P. Cortes, C. Olivetti, and J. Pan (2021). Social norms, labour market opportunities, and the marriage gap between skilled and unskilled women. *The Review of Economic Studies 88*(4), 1936–1978.

Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics 130*(2), 571–614.

Bhattacharya, U., N. Galpin, and B. Haslem (2007). The home court advantage in international corporate litigation. *The Journal of Law and Economics 50*(4), 625–660.

Bielen, S., W. Marneffe, and N. Mocan (2021). Racial bias and in-group bias in virtual reality courtrooms. *The Journal of Law and Economics 64*(2), 269–300.

Bindler, A. and R. Hjalmarsson (2020). The persistence of the criminal justice gender gap: Evidence from 200 years of judicial decisions. *The Journal of Law and Economics 63*(2), 297–339.

Black, S. E. and P. E. Strahan (2001). The division of spoils: rent-sharing and discrimination in a regulated industry. *American Economic Review 91*(4), 814–831.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bordalo, P., N. Gennaioli, and A. Shleifer (2015). Salience theory of judicial decisions. *The Journal of Legal Studies 44*(S1), S7–S33.

Boyd, C. L., L. Epstein, and A. D. Martin (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science 54*(2), 389–411.

Brown, P. H. and A. Park (2002). Education and poverty in rural china. *Economics of Education Review 21*(6), 523–541.

Cai, H., H. Chen, Y. Chen, and S. Zhang (2022). Monitoring Courts: A Top-down or Grassroots Approach? *SSRN: https://ssrn.com/abstract=4183534* .

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Charles, K. K., J. Guryan, and J. Pan (2018). The effects of sexism on american women: The role of norms vs. discrimination. Technical report, National Bureau of Economic Research.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Chemin, M. (2009). The impact of the judiciary on entrepreneurship: Evaluation of pakistan's "access to justice programme". *Journal of Public Economics 93*(1-2), 114–125.

Chen, J. (2008). *Chinese law: Context and transformation*, Volume 654. Martinus Nijhoff Publishers Boston/Leiden.

Chen, W.-d. (2014). Study on the independence of judicial office (in Chinese). *China Legal Science 2*(11), 20–49.

Clots-Figueras, I. (2012). Are female leaders good for education? evidence from india. *American Economic Journal: Applied Economics 4*(1), 212–44.

Coale, A. J. and J. Banister (1994). Five decades of missing females in china. *Demography 31*(3), 459–479.

Dehejia, R. H. and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics 84*(1), 151–161.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–40.

Doepke, M. and M. Tertilt (2019). Does female empowerment promote economic development? *Journal of Economic Growth 24*(4), 309–343.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature 50*(4), 1051–79.

Ebenstein, A. (2010). The "missing girls" of china and the unintended consequences of the one child policy. *Journal of Human resources 45*(1), 87–115.

Eren, O. and N. Mocan (2018). Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics 10*(3), 171–205.

Eren, O. and N. H. Mocan (2020). Judge peer effects in the courthouse. Technical report, National Bureau of Economic Research.

Falk, A. and J. Hermle (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science 362*(6412).

Fang, L. (2015). Resource and pattern of the judicial mechanism reform (in Chinese). *Legal Science 3*(6), 26–40.

Firth, M., O. M. Rui, and W. Wu (2011). The effects of political connections and state ownership on corporate litigation in china. *The Journal of Law and Economics 54*(3), 573–607.

Giuliano, P. (2017). Gender: An historical perspective. *NBER Working Paper*.

Giuliano, P. (2020). Gender and culture. *Oxford Review of Economic Policy 36*(4), 944–961.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Guo, P. (2016). The core problems and reform path of the people's jury system. *Law Science* (8), 92–99.

Hamberg, K. (2008). Gender bias in medicine. *Women's health 4*(3), 237–243.

He, J. and R. Wang (2015). An empirical study on courtroom live broadcasting (in Chinese). *Science of Law 33*(3), 53–63.

Hoekstra, M. and B. Street (2021). The effect of own-gender jurors on conviction rates. *The Journal of Law and Economics 64*(3), 513–537.

Hou, Y. and R. Truex (2019). Ethnic discrimination in criminal sentencing in china. Technical report, working paper, Department of Political Science, University of Pennsylvania

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and us economic growth. *Econometrica 87*(5), 1439–1474.

Hughes, M. M. (2011). Intersectionality, quotas, and minority women's political representation worldwide. *American Political Science Review 105*(3), 604–620.

Jassal, N. (2020). Gender, law enforcement, and access to justice: Evidence from all-women police stations in india. *American Political Science Review 114*(4), 1035–1054.

Kastellec, J. P. (2013). Racial diversity and judicial influence on appellate courts. *American Journal of Political Science 57*(1), 167–183.

Knepper, M. (2018). When the shadow is the substance: Judge gender and the outcomes of workplace sex discrimination cases. *Journal of Labor Economics 36*(3), 623–664.

Kuhn, P. and K. Shen (2013). Gender discrimination in job ads: Evidence from china. *The Quarterly Journal of Economics 128*(1), 287–336.

Lavy, V., E. Sand, and M. Shayo (2022). Discrimination between religious and non-religious groups: Evidence from marking high-stakes exams. *The Economic Journal*.

Liebman, B. L., M. E. Roberts, R. E. Stern, and A. Z. Wang (2020). Mass digitization of Chinese court decisions: How to use text as data in the field of Chinese law. *Journal of Law and Courts 8*(2), 177–201.

Lim, C. S., B. S. Silveira, and J. M. Snyder (2016). Do judges' characteristics matter? ethnicity, gender, and partisanship in texas state trial courts. *American Law and Economics Review 18*(2), 302–357.

Lim, C. S., J. M. Snyder Jr, and D. Stromberg (2015). The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems. *American Economic Journal: Applied Economics 7*(4), 103–35.

Lin, M.-J., J.-T. Liu, and N. Qian (2014). More missing women, fewer dying girls: The impact of sex-selective abortion on sex at birth and relative female mortality in taiwan. *Journal of the European Economic Association 12*(4), 899–926.

Liu, T. and D. H. Zhang (2020). Do judge-lawyer relationships influence case outcomes? *Available at SSRN 3711873*.

Liu, Y. and L. Zhu (2020). The risk and prevention of court live broadcasting in the era of big data (in Chinese). *Law Science Magazine 41*(12), 18–31.

Liu, Z. (2014). Hierachy and competition: the deep structure of court system (in Chinese). *Tsinghua University Law Journal* (4), 146–163.

Mai, J. and A. Stoyanov (2019). Anti-foreign bias in the court: Welfare explanation and evidence from canadian intellectual property litigations. *Journal of International Economics 117*, 21–36.

Mastrobuoni, G. (2020). Crime is terribly revealing: Information technology and police productivity. *The Review of Economic Studies 87*(6), 2727–2753.

Mengel, F., J. Sauermann, and U. Zolitz (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association 17*(2), 535–566.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Noussair, C. N., S. T. Trautmann, and G. Van de Kuilen (2014). Higher order risk attitudes, demographics, and financial decisions. *Review of Economic Studies 81*(1), 325–355.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review 101*(4), 1410–35.

Patton, D. and J. L. Smith (2017). Lawyer, interrupted: Gender bias in oral arguments at the us supreme court. *Journal of Law and Courts 5*(2), 337–361.

Philippe, A. and A. Ouss (2018). "no hatred or malice, fear or affection": Media and sentencing. *Journal of Political Economy 126*(5), 2134–2178.

Pierce, L., D. C. Snow, and A. McAfee (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science 61*(10), 2299–2319.

Qian, N. (2008). Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics 123*(3), 1251–1285.

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

Songer, D. R. and K. A. Crews-Meyer (2000). Does judge gender matter? decision making in state supreme courts. *Social Science Quarterly*, 750–762.

Tang, H. and Y. Zhang (2021). Do multinationals transfer culture? evidence on female employment in china. *Journal of International Economics 133*, 103518.

Yu, L.-g. (2020). Executive ecology and enforcement difficulty of local courts (in Chinese). *Chinese Journal of Law* (1), 102–122.

Zamoff, M. E., B. N. Greenwood, and G. Burtch (2021). Who watches the watchmen: Evidence of the effect of body-worn cameras on new york city policing. *The Journal of Law, Economics, and Organization 38*(1), 161–195.

Zhang, J., S. Jin, T. Li, and H. Wang (2021). Gender discrimination in china: Experimental evidence from the job market for college graduates. *Journal of Comparative Economics*.

Online Appendix

(Not intended for publication)

A.1 Specifications with Various Fixed Effects

We estimate specifications by adding to Equation (2) court × year-quarter fixed effects and court-specific time trends. We report the respective results using the main sample in columns (1) and (2) of Table 10 and their counterparts using the two-litigant sub-sample in columns (3) and (4) of the same table, respectively. They are similar to those reported in Table 4.

Table 10. *Specifications with Various Fixed Effects*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	Court×Time FE	Court Time Trend	Court×Time FE	Court Time Trend
	(1)	(2)	(3)	(4)
Female × Intensity	0.0626***	0.0636***	0.104***	0.105***
	(0.00370)	(0.00373)	(0.00563)	(0.00560)
Controls	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	N	Y	N	Y
Court×Year-Quarter FE	Y	N	Y	N
Court×Time Trend	N	Y	N	Y
Mean of outcome	0.744	0.744	0.707	0.707
N	4,112,704	4,114,282	2,104,044	2,105,831
R^2	0.403	0.391	0.517	0.502

Notes: To rule out confounding factors such as time-varying shocks at the court level, we estimate two additional specifications and add court × year-quarter fixed effects and a court-specific time trend to Equation (2). The results, as shown in columns (1) and (2), for the main sample and columns (3) and (4) for the two-litigant subsample, are rather similar to those of the baseline specifications. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms in this specification are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

A.2 Robustness

Table 11 reports the results of our robustness tests. We use a coarser measure of the live broadcasting intensity: Top-25, which takes 1 when the intensity of court k in quarter t belongs to the top 25% of all courts in the same area, 0 otherwise. We re-estimate equation (2), and column (1) reports the result, which is consistent with our baseline findings. Furthermore, we use court-level (instead of court-area level) broadcasting intensity (i.e., overall intensity) to identify the impact on litigation outcomes. Column (2) indicates that the result is very close to that of the baseline estimation in terms of magnitude and significance.

We also investigate whether our results are robust to alternative measures of litigation outcomes. We replace the dependent variable with Win, which takes value 1 only when the plaintiff's claim is fully supported, 0 otherwise. The regression result in column (3) shows a significantly positive coefficient. It is expected that the coefficient is slightly smaller than that of the baseline result, given the average odds that plaintiffs win entirely are also lower.

We use the gender deduction algorithm (i.e., Ngender) to interpolate the missing cases that do not report the gender of litigants. Our sample size increases by 1,168,856 observations in main sample. The baseline specification is re-estimated with this larger sample. The estimated coefficient is showed in column (4), which is close to that of the baseline regression, suggesting that the missing data have a limited impact on our results.

We re-estimate the aforementioned exercises with our two-litigant subsample. The counterparts of those estimation results are reported in the lower panel of Table 11, in columns (5) to (8). All of them are robust.

Table 11. *Robustness*

Outcome Variable	Chances of Winning		Win	Chances of Winning
All Litigants	(1)	(2)	(3)	(4)
Female × Top25	0.00307**			
	(0.00100)			
Female × Overall Intensity		0.0794***		
		(0.00468)		
Female × Intensity			0.0515***	0.0564***
			(0.00429)	(0.00338)
Mean of outcome	0.744	0.744	0.585	0.750
N	4,114,282	4,116,122	4,114,282	5,224,547
R^2	0.388	0.351	0.304	0.370
Two Litigants	(5)	(6)	(7)	(8)
Female × Top25	0.00646**			
	(0.00291)			
Female × Overall Intensity		0.133***		
		(0.00661)		
Female × Intensity			0.0954***	0.0977***
			(0.00620)	(0.00516)
Mean of outcome	0.707	0.707	0.601	0.720
N	2,105,831	2,108,325	2,105,831	2,598,049
R^2	0.498	0.454	0.377	0.476
Controls	Y	Y	Y	Y
Court FE	N	Y	N	N
Area FE	N	Y	N	N
Court-Area FE	Y	N	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Notes:This table presents the robustness check using alternative variables and sample. In Columns (1)(5) and Columns(2)(6) we use Top-25 and the overall intensity instead of Intensity. In Columns(3)(7) we use dummy Win as the dependent variable. In Columns(4)(8) we supplement our sample by a name-based gender deduction method. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms in all specifications are also controlled. Standard errors are in the parentheses beneath coefficients, clustered at the court level; * p<0.1, ** p<0.05, *** p<0.01.

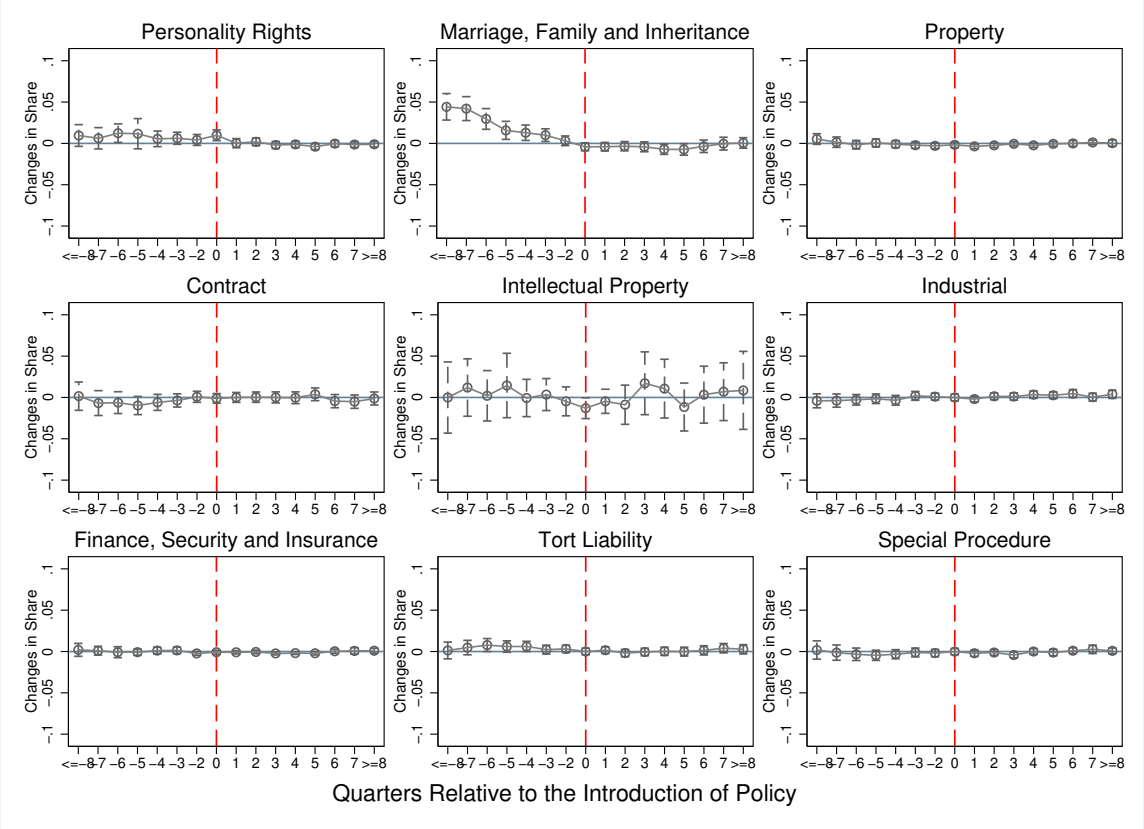


Figure 5. *Event Study for the shares of issue areas over time: all-litigant sample. These figures depict the estimated effects of the reform on the case shares of issue areas, using the main sample. From 8 quarters before the introduction of the reform ($\tau = -8, ..., -1, 0$ with $\tau = -1$ dropped as the reference period) to 8 quarters after the reform ($\tau = 1, ..., 8$). The grey and black dots represent the estimated coefficients for each period, before and after the reform, respectively; the dashed lines represent 95 percent confidence intervals.*

A.3 Addressing Threats to Instrument Validity

In this section, we discuss the validity conditions for our Bartik instrument and provide further robustness checks to deal with the potential threats that might undermine our IV estimation. Our instrument is based on the exogeneity of issue area share. To examine whether the early share is correlated with the later broadcasting reform, we use an event study model and regress shares of issue areas in each court and each quarter on a sequence of dummy variables capturing the number of quarters relative to the introduction of policy (see Equation (4)). The results show that almost all areas have stable shares and changes in share are insignificant over time.

We notice that the share of marriage and family disputes shows a moderate decreasing trend before the introduction of reform. To have cleaner results, we exclude all cases in this area, re-construct the instrument and repeat our IV estimations. Table 12 reports the results and the coefficients of our interest are similar with those of

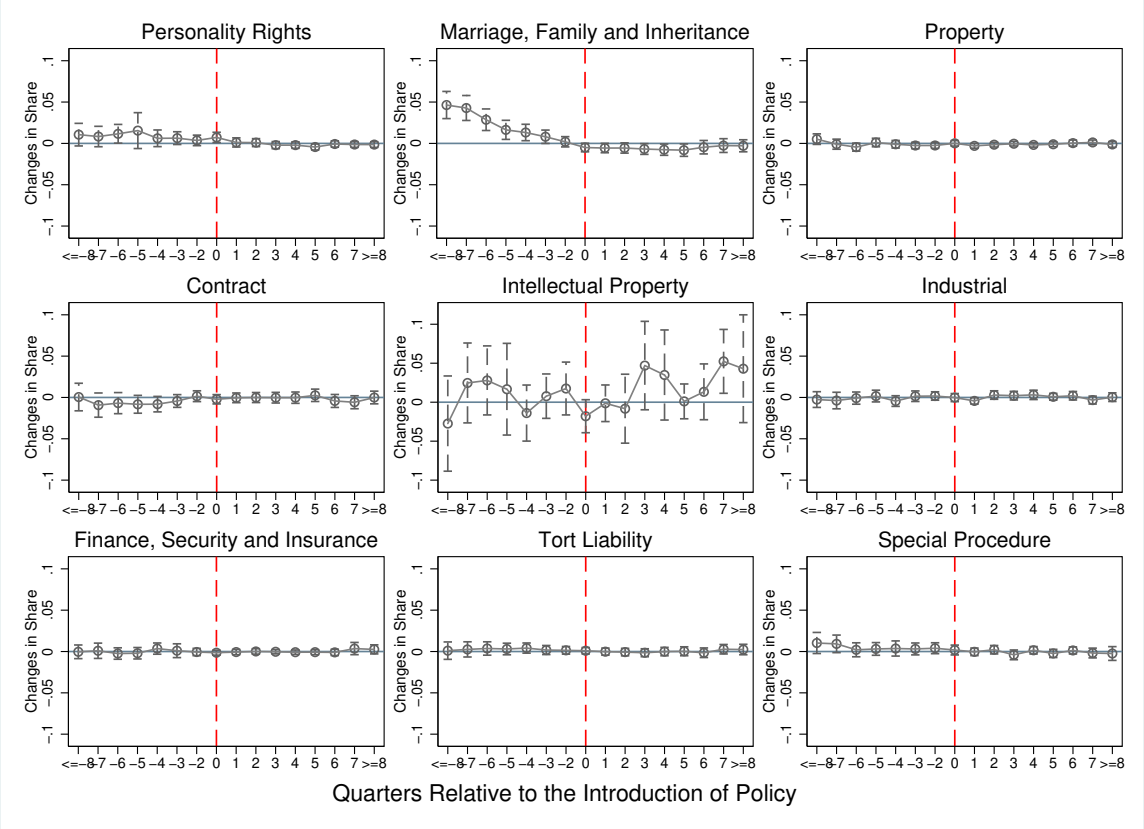


Figure 6. *Event Study for the shares of issue areas over time: two-litigant subsample. These figures depict the estimated effects of the reform on the case shares of issue areas, using the two-litigant subsample. From 8 quarters before the introduction of the reform ($\tau = -8, ..., -1, 0$ with $\tau = -1$ dropped as the reference period) to 8 quarters after the reform ($\tau = 1, ..., 8$). The grey and black dots represent the estimated coefficients for each period, before and after the reform, respectively; the dashed lines represent 95 percent confidence intervals.*

full-sample IV estimations.

Table 12. *Instrumental Variable Estimations: Excluding Marriage*

	Bartik Instrument Estimation					
	All Litigants			Sub-sample: Two Litigants		
	1st Stage		2nd Stage	1st Stage		2nd Stage
Outcome Variable	Intensity	Female × Intensity	Chances of Winning	Intensity	Female × Intensity	Chances of Winning
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Intensity			0.0833***			0.139***
			(0.00766)			(0.0108)
Female	0.000298	0.0209***	-0.0185***	0.000278	0.0210***	-0.0420***
	(0.000314)	(0.00169)	(0.00165)	(0.000364)	(0.00180)	(0.00263)
Intensity			-0.0156**			-0.0205***
			(0.00711)			(0.00741)
Predicted Intensity	0.683***	-0.0580***		0.695***	-0.0527***	
	(0.0381)	(0.00586)		(0.0387)	(0.00546)	
Female × Predicted Intensity	-0.0121**	0.886***		-0.00851	0.910***	
	(0.00552)	(0.0302)		(0.00596)	(0.0295)	
Controls	Y	Y	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
K-P LM Statistics			155.892			141.300
K-P Wald Statistics			160.721			162.263
N	3,816,171	3,816,171	3,816,171	1,836,302	1,836,302	1,836,302
R^2	0.697	0.555	0.031	0.718	0.577	0.051

Notes: This table presents the IV estimation results excluding cases of marriage and family disputes. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

Table 13. *Placebo Test: Early Outcome Against Later Predicted Intensity*

	Outcome Variable: Chances of Winning					
	All Litigants			Sub-sample: Two Litigants		
	2014	2015	2016Q1∼3	2014	2015	2016Q1∼3
	(1)	(2)	(3)	(4)	(5)	(6)
Female × 2018 Predicted Intensity	0.00780	-0.000166	-0.000965	0.00860	0.0152	0.00273
	(0.00905)	(0.00758)	(0.0104)	(0.0138)	(0.0105)	(0.0157)
Female	-0.0985***	-0.0900***	-0.0183***	-0.160***	-0.150***	-0.0307***
	(0.00493)	(0.00448)	(0.00152)	(0.00778)	(0.00675)	(0.00233)
2018 Predicted Intensity	-0.0216**	-0.0229***	0.0211**	-0.0350**	-0.0373***	0.0122
	(0.0104)	(0.00797)	(0.0105)	(0.0142)	(0.0111)	(0.0129)
Controls	Y	Y	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
N	341,752	449,592	429,989	203,625	257,139	227,318
R^2	0.457	0.437	0.534	0.510	0.496	0.630

Note:This table presents results from the first placebo test. Using subsamples in 2014, 2015 and the first 3 quarters of 2016, we regress chances of winning on female dummy and predicted broadcasting intensity in 2018 (according to court and quarter) and their interaction term. If there exist other factors across court driving both the early share and later intensity, we should observe a significant correlation between the gender difference before the implementation of policy and the predicted intensity after the introduction of policy. However, all the coefficients for the interaction terms are insignificant. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; *p<0.1, ** p<0.05, *** p<0.01.

Table 14. *Placebo Test: Using Lagged Predicted Intensity*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	1 Quarter Lagged	2 Quarters Lagged	1 Quarter Lagged	2 Quarters Lagged
	(1)	(2)	(3)	(4)
Female × Lagged Predicted Intensity	0.00294	0.00255	0.00302	0.00215
	(0.00691)	(0.00994)	(0.00701)	(0.00988)
Female	-0.0170***	-0.0153***	-0.0338***	-0.0316***
	(0.00129)	(0.00123)	(0.00195)	(0.00184)
Lagged Predicted Intensity	0.0108	0.0133	0.0158	0.0189
	(0.00750)	(0.0104)	(0.0102)	(0.0148)
Controls	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
N	3,565,200	3,364,649	1,825,307	1,717,050
R^2	0.385	0.381	0.496	0.492

Notes: This table presents results from the placebo test. We estimate the baseline specification by using 1 and 2 quarter lagged predicted intensity. If the predicted intensity is endogenous and partially driven by earlier litigation outcomes, we should find significant correlation between the outcome and lagged predicted intensity. Nevertheless, the results show that the coefficients of our interests are statistically indistinguishable from zero in all estimations. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms in this specification are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

A.4 Extensive Margin

In this section, we investigate whether the proportion of female plaintiffs reacts to the open justice reform, using the two-way fixed effects model with following Equation (8):

$$y_{jkt} = \beta_0 + \beta \times \text{intensity}_{jkt} + \omega_{j\times k} + \lambda_{t\times p} + \varepsilon_{jkt} \tag{8}$$

where y_{jkt} represents the proportion of civil cases with female plaintiffs in issue area j, court k, and year-quarter t. As in the benchmark case, we include court-area $\omega_{j\times k}$ as in the benchmark model. To alleviate the concern that other reforms may take place simultaneously, we also include prefecture $\times$ year-quarter fixed effects. The coefficient β is our main interest, capturing the impact of the reform on the outcome variable.

Table 15 reports the regression results when using the proportion of female plaintiffs as the dependent variable. We consider different specifications with or without control variables, in both the main and two-litigant samples. We find an insignificant impact in all columns. Further, we re-run this regression for each area and report the results in Table 16. There is no evidence that females have stronger incentives to initiate litigation in response to the reform and that the composition of plaintiff gender in each area is affected by the reform.

Table 15. *Impacts of Trials Broadcasting on the Proportion of Civil Cases with Female Plaintiffs*

	Outcome Variable: Proportion of Female Plaintiffs			
	All Litigants		Sub-sample: Two Litigants	
	(1)	(2)	(3)	(4)
Intensity	-0.00599	-0.00307	-0.0124	-0.0115
	(0.00641)	(0.00697)	(0.00822)	(0.00899)
Controls	N	Y	N	Y
Court-Area FE	Y	Y	Y	Y
Prefecture×Year-Quarter FE	Y	Y	Y	Y
N	220,219	197,473	172,830	155,133
R^2	0.335	0.345	0.377	0.382

Notes: This table presents results for the extensive margin. We examine whether more women initiate litigation in response to the reform by regressing the proportion of female plaintiffs on the overall broadcasting intensity. Court-level control variables include the proportion of cases with female defendant, the proportion of cases with female chief judge, the average number of judges, the average number of plaintiffs and defendants respectively, the average number of lawyers for plaintiffs and defendants respectively, the proportion of plaintiffs and defendants appear in court respectively, the proportion of first instance, the mean of legal cost. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

Table 16. *Extensive Margin by Areas*

	Outcome Variable: Proportion of Female Plaintiffs								
	Personality	Marriage	Property	Contract	IP	Industrial	Finance	Tort	Other
All Litigants	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intensity	0.0214	-0.0162	-0.0167	-0.00564	0.625	-0.0150	-0.0430	-0.000857	0.0156
	(0.0261)	(0.0272)	(0.0263)	(0.00920)	(0.712)	(0.0421)	(0.0593)	(0.0151)	(0.0491)
N	28,600	31,819	27,256	44,641	224	9,353	7,821	37,624	12,136
R^2	0.351	0.562	0.316	0.312	0.715	0.528	0.471	0.288	0.427
Two Litigants	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Intensity	0.0632*	-0.0279	-0.0142	-0.00847	-1.939	-0.0238	0.0710	0.0245	-0.167
	(0.0322)	(0.0235)	(0.0341)	(0.0104)	(2.249)	(0.0565)	(0.0978)	(0.0317)	(0.130)
N	24,292	29,180	21,076	41,425	75	6,163	2,548	24,424	3,655
R^2	0.384	0.706	0.352	0.312	0.839	0.559	0.544	0.322	0.572
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Court FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Prefecture×Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents results for the extensive margin by area. We find no evidence that the reform encourages women to initiate litigation and further the reform did not change the gender composition of plaintiffs in each area. Court-level control variables include the proportion of cases with female defendant, the proportion of cases with female chief judge, the average number of judges, the average number of plaintiffs and defendants respectively, the average number of lawyers for plaintiffs and defendants respectively, the proportion of plaintiffs and defendants appear in court respectively, the proportion of first instance, the mean of legal cost. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

A.5 Case-level Analysis

As discussed in section 2, whether a case is broadcast is decided by the court and therefore nonrandom. In the main text, to circumvent the potential selection issue, we utilize the intensity variations in the court-area $\times$ year-quarter level to identify impacts of the reform in our baseline specification. Can we find evidence that the reform reduces gender differentials in litigation outcomes using the case-level treatment variation? To answer this question, we estimate the following specification:

$$y_{ijkt} = \beta_0 + \beta \text{Female}_i + \delta \text{Live}_i + \theta \text{Female}_i * \text{Live}_i + \omega_{j\times k} + \lambda_t + \varepsilon_{ijkt} \tag{9}$$

where Live_i represents whether case i is broadcast, i.e., taking value 1 if broadcast, 0 otherwise. The coefficient on the interaction term θ captures the impact of the reform on the gender disparities in winning litigation.

The regression results are presented in Table 17. Columns (1) and (2) report the estimation results using the main sample, and columns (3) and (4) report those using the two-litigant subsample. In columns (2) and (4), all control variables specified in the benchmark model are included. All estimated coefficients for the interaction term are positive and significant, which does not contradict the interpretation that the disadvantage of female plaintiffs, relative to male plaintiffs, appears to be smaller when cases are broadcast.

However, as discussed earlier, the selection issue is likely to be present and may bias the estimated coefficient on the interaction term either upward and downward. Suppose that courts tend to select cases with male (female) plaintiffs who have strong evidence (e.g., straightforward cases), to broadcast; then the estimate is downward (upward) biased. To address this issue, we resort to the IV approach: we instrument the variable Live_i using the Bartik instrument Z_{kt}. This instrument variable is constructed and discussed in section 4. Table 18 reports the IV regression results at the case level for both the main sample and the two-litigant subsample. The IV estimates for the interaction term shown in columns (3) and (6) remain positive and significant. Those coefficients are larger than those reported in Table 17 without using the IV, which indicates that the selection mechanism likely biases the estimate downward. However, the size of those coefficients is very similar to that of the coefficients when we estimate Equation (2) with the same IV (see Table 18).

Table 17. *Case-level Estimation: Impacts of Trials Broadcasting on Gender Disparities*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	(1)	(2)	(3)	(4)
Female × Live	0.0304***	0.0305***	0.0446***	0.0508***
	(0.00172)	(0.00182)	(0.00238)	(0.00266)
Female	-0.0238***	-0.0218***	-0.0399***	-0.0421***
	(0.00132)	(0.00138)	(0.00189)	(0.00211)
Live	-0.00593***	-0.00801***	-0.00766***	-0.0111***
	(0.00139)	(0.00142)	(0.00191)	(0.00185)
Controls	N	Y	N	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.744	0.744	0.707	0.707
N	4,753,644	4,171,936	2,416,850	2,139,039
R^2	0.374	0.389	0.482	0.498

Notes: This table presents results for case-level estimation. We use the dummy Live indicating whether the case is broadcast or not instead of Intensity in the baseline specification. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

Table 18. *Instrumental Variable Estimations: Case-level analysis*

	Instrumental Variable Estimations					
	All Litigants			Sub-sample: Two Litigants		
	1st Stage		2nd Stage	1st Stage		2nd Stage
Outcome Variable	Live	Female × Live	Chances of Winning	Live	Female × Live	Chances of Winning
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Live			0.113***			0.168***
			(0.00823)			(0.0111)
Female	0.000236	0.0219***	-0.0272***	0.000232	0.0217***	-0.0500***
	(0.000381)	(0.00157)	(0.00159)	(0.000464)	(0.00161)	(0.00242)
Live			-0.0217***			-0.0267***
			(0.00715)			(0.00739)
Predicted Intensity	0.705***	-0.0681***		0.725***	-0.0629***	
	(0.0409)	(0.00630)		(0.0413)	(0.00584)	
Female × Predicted Intensity	-0.0272***	0.927***		-0.0257***	0.962***	
	(0.00804)	(0.0328)		(0.00879)	(0.0322)	
Controls	Y	Y	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
K-P LM Statistics			151.703			139.499
K-P Wald Statistics			148.992			155.580
N	4,171,936	4,171,936	4,171,936	2,139,039	2,139,039	2,139,039
R^2	0.314	0.256	0.026	0.331	0.273	0.038

Notes: This table presents the IV results in case level estimation. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

B. Heterogeneity Analysis

B.1 Area

It is legitimate to conjecture that the gender bias in litigation outcomes may not even exist in other areas but only in gender sensitive areas, such as marriage, family and inheritance disputes. To investigate, we re-estimate our generalized DID model (i.e., Equation (2)) for each issue area in our main sample and the two-litigant subsample. The results are shown in the upper and lower panels of Table 19, respectively. All the estimated coefficients on the interaction term, i.e., Female × Intensity, are either positive and significant or not significant.

Table 19. *Heterogeneity Analysis: Legal Areas*

	Outcome Variable: Chances of Winning								
	Personality	Marriage	Property	Contract	IP	Industrial	Finance	Tort	Other
All Litigants	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female × Intensity	0.00179	0.0548**	-0.0123	0.00275	0.246*	0.0596**	-0.0116	0.00537	0.0120
	(0.0113)	(0.0274)	(0.0161)	(0.00300)	(0.135)	(0.0237)	(0.0364)	(0.00406)	(0.0306)
Mean of outcome	0.60	0.22	0.49	0.85	0.58	0.80	0.59	0.77	0.46
N	132,134	298,111	140,331	2,437,601	1,163	37,826	23,302	646,865	37,755
R^2	0.161	0.220	0.120	0.271	0.440	0.435	0.332	0.135	0.226
Two Litigants	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Female × Intensity	0.00484	0.0576*	0.0207	0.0102***	0.463**	0.119***	0.00505	0.0507***	0.0340
	(0.0141)	(0.0312)	(0.0217)	(0.00288)	(0.197)	(0.0290)	(0.0747)	(0.0116)	(0.0555)
Mean of outcome	0.60	0.22	0.48	0.85	0.50	0.82	0.59	0.77	0.45
N	82,182	269,529	70,562	1,368,313	587	20,950	7,172	80,033	9,737
R^2	0.188	0.202	0.156	0.320	0.502	0.459	0.493	0.193	0.394
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Court FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the heterogenous analysis by legal area. We separate our sample into nine subsample defined by legal areas and for each area we repeat the baseline estimations. Case-level control variables include the defendant gender dummy, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance and legal cost (log). Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms in all specifications are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

For most areas, the results across the two samples agree, although the magnitude and significance differ. The exceptions are the areas of contract disputes and tort liability disputes, where the estimated coefficients are highly significant in the subsample with two litigants but insignificant in the main sample. This is also expected because the number of litigants involved in the cases of these two areas are indeed higher and the greater degree of gender mixture on the plaintiff side biases our estimate toward zero in the main sample. Although the size of coefficient for the area of contract disputes is smaller than those of others, the fraction of cases in this area is the largest, indicating that its economic significance should not be ignored.

For the area of marriage, family and inheritance disputes in particular, the estimated coefficients are close to one another across the two samples but only marginally

significant. Removing this area from our samples does not qualitatively change our estimated coefficients. The size of the estimated coefficients for the area of industrial disputes is the largest: It is 6.42 percent in the main sample and 11.8 percent in the two-litigant subsample. For the area of intellectual property disputes, the estimated coefficients are significant, but the total numbers of cases in both samples are rather small, and hence they are unlikely to affect our results when pooling cases from all areas.

B.2 Defendant Gender

To explore the heterogeneity of defendant gender, we include both a male defendant dummy and its interaction term with the female plaintiff dummy in specification (1) and estimate it with our samples (i.e., all-litigant and two-litigant). The results are reported in Table 20. Columns (1) and (3) show that the coefficients for the interaction term are negative and significant, indicating that a female plaintiff would be even more disadvantaged if she initiated litigation against a male litigant. Based on column (3), the gender gap in winning odds would be 14 percentage points larger when the defendant is male than when the defendant is female.

Additionally, we include the male defendant dummy and its interaction term with all relevant linear and second-order terms in Equation (2) and estimate it with our samples (i.e., all-litigant and two-litigant samples). We are interested in the coefficient on the triple interaction term. Columns (2) and (4) of Table 20 show that the coefficients on the triple interaction term are significant and positive. The estimate in column (4) indicates that the gender gap in winning odds would be only 11 ($\approx$0.14-0.29$\times$0.1) percentage points greater when the defendant is male than when the defendant is female if that the intensity increases from zero percent to 10 percent.

Table 20. *Heterogeneity Analysis: Gender of Defendants*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	OLS	DID	OLS	DID
	(1)	(2)	(3)	(4)
Female × Male Defendant	-0.0917***	-0.0878***	-0.141***	-0.142***
	(0.00378)	(0.00413)	(0.00583)	(0.00648)
Female × Male Defendant × Intensity		0.171***		0.287***
		(0.0103)		(0.0153)
Controls	Y	Y	Y	Y
Area FE	Y	N	Y	N
Court FE	Y	N	Y	N
Court-Area FE	N	Y	N	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.744	0.744	0.707	0.707
N	4,174,135	4,114,282	2,141,814	2,105,831
R^2	0.353	0.390	0.456	0.501

Notes: This table presents the heterogenous analysis by defendant's gender. We add the Male Defendant dummy, indicating whether the defendant is male or not, and its interaction with plaintiff's gender to our OLS and DID specifications. The results show that a female plaintiff would be even more disadvantaged if she initiated litigation against a male litigant. Case-level control variables include whether defendant is male or not, the chief judge gender dummy, the number of judges, the number of plaintiffs and defendants, the number of lawyers for plaintiffs and defendants, whether plaintiffs and defendants appear in court or not, instance, legal cost (log) and whether the case is broadcast. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Linear terms and other interaction terms in both specifications are also controlled. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

C. The Procedure for Propensity Score Matching

When interpreting the differences across judge genders as causal, we are concerned that the assignment of cases is not completely random, despite that it is required to be random. To address this, we use the propensity score matching (PSM) method to select a balanced sample. In this appendix, we present the details of matching procedure that is used in section 6.

Our matching procedure is based on the nearest-neighbor matching approach, the most popular and straightforward matching method. We use the algorithms provided by Dehejia and Wahba (2002). In our context, we treat cases adjudicated by female judges as the treatment group and those adjudicated by male judges as the control group. The assumption underlying the matching is called conditional independence (CIA), requiring that all important differences between the groups of treated and non-treated observations are captured by their observable characteristics. The aim of this matching procedure is to construct a male judge sample, which is comparable to a female judge sample, to approximate the unobserved counterfactual. Toward this end, we need to estimate the propensity score of being adjudicated by a female judge for all cases. For each case with a female judge, we find the nearest case with a male judge that minimizes the distance between the two cases in propensity scores. The matching algorithm results in a matched sample in which cases with female judges have matched cases with male judges. The matched sample allows us to investigate the impacts of judge gender.

The specific matching procedure is as follows. First, we calculate the probability or the propensity score of a case that is judged by female chief judge, using logistic regression on a number of observable characteristics. There are various determinants to include, so we choose the combination satisfying the balancing property with the highest goodness of fit as represented by R^2. The balancing property of the propensity score is tested and satisfied in all estimations, as reported in Table 21. The estimated differences are exceedingly small in magnitude.

Second, all observations in the region of common support are stratified using the estimated propensity scores. We then use the nearest-neighbor without replacement algorithm to match each case that is adjudicated by a female chief judge with an "identical" case that is adjudicated by a male chief judge within the same stratum. Since our data span the years from 2014 to 2018, the matching algorithm is implemented year by year. Third, we drop the unmatched observations from both treatment and control groups, obtaining a balanced sample for further analysis. The density of the propensity score before and after matching is presented in Figure 7. Before matching, the distributions of propensity scores of cases adjudicated by female and male judges are

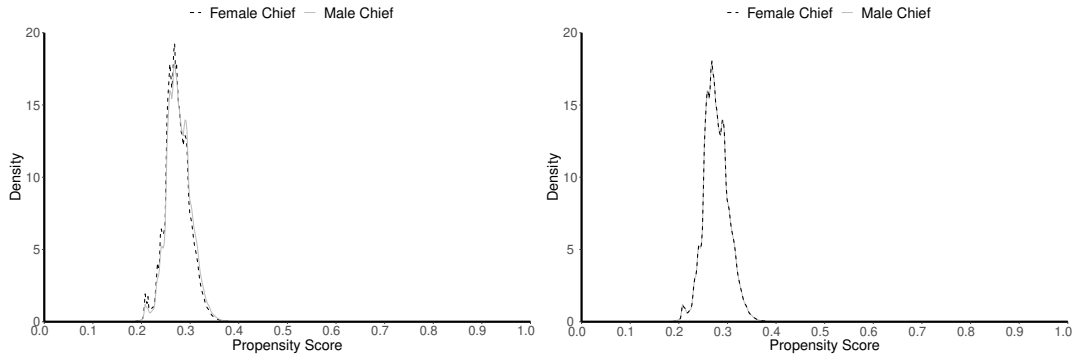
Table 21. *Balancing Tests*

Variable	Mean		Difference in Means		
	Female Judge vs. Male Judge		Fixed Effects Included		
	Female Judge	Male Judge	None	Year-Quarter	Year-Quarter & Court
Live broadcasting (1=Yes)	0.0771	0.0785	-0.00109	-0.00313	1.24E-17
Plaintiff's gender (1=female)	0.3148	0.3125	0.00237	0.00241	-0.000174
Defendant's gender (1=female)	0.2026	0.2028	0.000471	0.00022	1.45E-16
Number of plaintiffs	1.1060	1.1024	0.00352	0.00346	-4.95E-16
Number of defendants	1.7181	1.6993	0.0176***	0.0164**	9.54E-17
Plaintiff's appearance (1=Yes)	0.4747	0.4784	-0.00433	-0.00344	-2.22e-16***
Defendant's appearance (1=Yes)	0.0770	0.0752	0.00127	0.00123	-1.48E-16
Number of lawyers for plaintiff	0.5265	0.5240	0.00423	0.00262	-1.30e-15***
Number of lawyers for defendant	0.2926	0.2810	0.0108**	0.00965*	-2.06E-16
Instance (1=first instance)	0.8879	0.8898	-0.00132	-0.0012	-5.63E-17
Legal cost in log	6.5329	6.5065	0.0274	0.0228	3.15e-15***
Number of Judges	1.3908	1.3850	0.00494	0.00517	-6.60E-16
Personality rights disputes	0.0296	0.0293	0.000218	0.000211	0.00065
Marriage, family and inheritance disputes	0.0659	0.0651	0.000955	0.00162	0.00486***
Property disputes	0.0330	0.0326	0.0002	0.000112	-0.000506
Contracts disputes	0.5992	0.6055	-0.00649	-0.00738	-0.0121***
Intellectual property disputes	0.0004	0.0004	0.0000703	0.0000655	0.0000471
Employment disputes	0.0083	0.0079	0.000769**	0.000776**	0.000729*
Companies, bill-related disputes	0.0062	0.0057	0.000434*	0.000376	0.000204
Tort liability disputes	0.1573	0.1546	0.0025	0.00215	0.00697**
Special procedures cases	0.0093	0.0087	0.000673*	0.000600*	0.000244

Notes: This table summarizes the balancing test between female judge sample and male judge sample after the matching procedure. We find that the difference between the two groups is exceedingly small or not significant from zero.

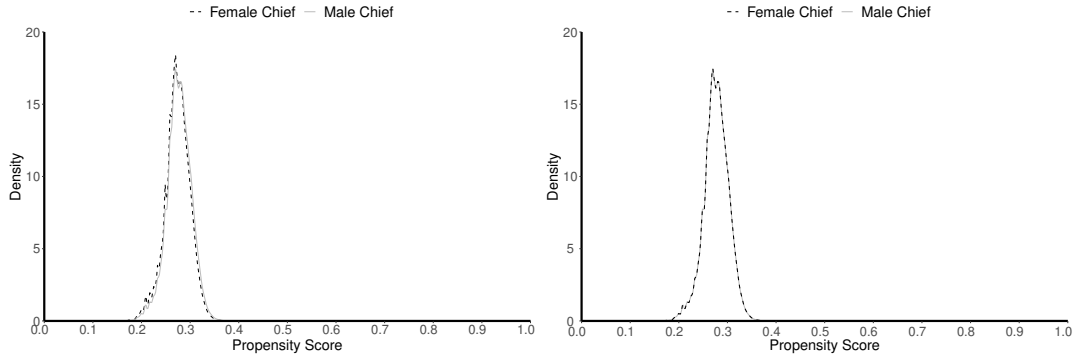
rather similar, which indicates that the random assignment rule is being implemented reasonably well. After matching, the two distributions are even closer.

Furthermore, we perform a placebo test by randomizing the gender of chief judges of the cases in our sample and implement our propensity score matching procedure. The basic idea is to examine the distributions of propensity scores of female and male judges and contrast them with their counterparts in our actual data so we can understand to what extent the actual assignment of judges in the data is subject to selection. The placebo distributions are presented in Figure 8, which are rather similar to those using our actual data in the sample. In other words, the nonrandom assignment of judges should not be a crucial concern.



(a) Before Matching (b) After Matching

Figure 7. *Balancing Property. Before matching, the distributions of propensity scores of cases adjudicated by female and male judges are rather similar, given that judge assignment is officially required to be random. After matching, the two distributions are even closer.*



(a) Before Matching (b) After Matching

Figure 8. *Randomization Test. We randomize chief judge gender to cases within each cell of court-area and year-quarter and then implement propensity score matching. The placebo distributions of propensity scores before and after the matching are very similar to those using actual data (Figure 7).*

D. Description of legal document

The document structure of civil judgments is strictly regulated, which allows us to efficiently extract case information. A standard judgment mainly consists of five sections: the basic information, claims of litigants, facts recognized by the court, legal principles applied and outcomes of the litigation. Take case 2018 Hebei-0722 Civil-1st No.1170 as an example, the document of which is depicted in Figure 9.

The basic information section includes the title of the document, the case number, the corresponding issue area, the dates of trial and publication, the court where the case is adjudicated, the characteristics of each litigant (e.g., name, gender, birth date, ethnicity, address and appearance), and the characteristics of each lawyer (e.g., name and affiliation).

The claims section is about the causes of the litigation, including the claims and reasons of plaintiffs as well as arguments and defenses of defendants. The facts section contains the facts and evidence recognized by the court and clarifies the controversial parts of the case. The next section elucidates the legal principles applied and the judges' justifications.

The last section presents the court decisions and litigation outcomes, e.g., to what extent the plaintiff's claims are supported and how the legal costs are shared. The signatures of the judge and jury members (if any) are located at the very end of the legal document.

Figure 9. *An example of case information extracted from a legal document. This is a sample of civil legal document that includes five main sections: the basic information, claims of litigants, facts recognized by the court, legal principles applied and outcomes of the litigation.*