Women in the Courtroom: Technology and Justice*

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

December 5, 2024

Abstract. Our study analyzes 6 million civil judgments in China from 2014 to 2018, documenting gender disparities that disfavor female litigants. We investigate the impact of an open justice reform that mandated courts to broadcast legal proceedings live on a centralized online platform. By exploiting variations in its implementation across courts and over time and employing both difference-in-differences and Bartik IV approaches, we find that gender disparities in chances of winning decrease as broadcast intensity increases. Analysis of the textual content of judicial decisions provides further evidence that these changes in judicial outcomes stem from altered judge behaviors (i.e., attention and effort) under enhanced judicial transparency. Our results demonstrate how information technology shapes judges' conduct, underscoring its broader potential to improve accountability in public institutions.

Keywords: 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: Guanghua School of Management, Peking University, Email: chenyuyu@gsm.pku.edu.cn; and Qingxu YANG: The Faculty of Business and Economics, The University of Hong Kong, Email: qxyang@hku.hk. 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 202111159164). Yuyu Chen acknowledges financial support by the National Natural Science Foundation of China (Grant No. 72394393).

1. Introduction

Gender disparities are pervasive across societal domains - from labor markets to political spheres - in countries at all stages of economic development. While these disparities have been extensively documented, a critical gap remains in our understanding of gender disparities within civil litigation, particularly in developing countries. Gender bias in civil litigation may have significant economic consequences, such as discouraging women from pursuing entrepreneurial activities or entering into judicially enforceable contracts, thereby contributing to a misallocation of talent and economic resources. This economic significance is underscored by prior studies that have documented substantial gains from reducing gender-based barriers in various areas (Duflo 2012; Falk and Hermle 2018; Doepke and Tertilt 2019; Hsieh et al. 2019).

Despite the clear significance of this issue, the scarcity of rigorous evidence can be attributed to two primary challenges. First, there exists a severe limitation in the availability of civil litigation data, particularly in less developed economies. Second, there are inherent issues within the judicial systems that impede convincing identification: while researchers often leverage random assignment of cases as an identification tool in the context of developed countries, such a principle may not be perfectly enforced in developing country contexts due to institutional weaknesses.

To address these challenges, we examine gender bias in civil litigation within the context of China. In July 2013, the Supreme People's Court (SPC) of China launched the China Judgements Online (CJO) website, mandating that all courts publish their rulings on this online platform. This initiative aimed to create an extensive and readily accessible digital repository of legal decisions for public scrutiny. We obtain a comprehensive dataset from the CJO, comprising approximately 6 million civil court judgments involving only individual litigants between 2014 and 2018. These judicial documents offer a wealth of detailed information on civil litigation, including litigant characteristics, event specifics, area of the disputes, decision deliberations, and litigation outcomes. This rich corpus not only enables us to investigate gender bias in civil litigation at scale but also allows for an in-depth examination of the underlying mechanisms through analysis of judges' decision-making processes as encoded in the textual data, overcoming the typical data limitations in such contexts.

To identify gender bias in judicial decisions, we leverage a 2016 nationwide open justice reform in China. This reform introduced surveillance technologies in courtrooms and mandated live trial broadcasts on a centralized platform (China Court Trial Online). Court proceedings became accessible through both live streaming and preserved recordings, enhancing judicial transparency and public scrutiny. The reform's staggered implementation across courts provides quasi-experimental variation to test

whether increased oversight reduces gender disparities in judicial decisions–a pattern that would indicate pre-existing discriminatory practices.

This open justice reform provides an attractive setting for two reasons. First, its primary objective is to enhance judicial transparency and quality, rather than targeting gender disparities, which mitigates policy endogeneity concerns, as the reform's implementation was not driven by gender disparity. Second, this setting allows us to both provide evidence on judicial bias against women and gain insights into how information technology adoption can reshape court decisions and enhance public sector monitoring mechanisms in general.

Building upon this setting, we begin our analysis by investigating judicial decisions in the pre-reform period (2014Q1-2016Q2). We focus on the extent to which judges support plaintiffs' claims, as plaintiffs in civil litigation typically initiate proceedings, while defendants play a more passive role. Our findings reveal significant gender-based disparities in litigation outcomes: female plaintiffs are 5 percentage points less likely to win than male plaintiffs. This result holds after controlling for numerous case-level characteristics, including defendant gender, legal costs, and region-level variables, as well as incorporating court-area (i.e., each area within a court), year-quarter, and judge fixed effects.

Acknowledging that unobserved factors related to both plaintiff gender and case outcomes likely drive this finding, we leverage the quasi-experimental feature of the open justice reform to test its effect on gender differences in litigation outcomes. Our empirical design exploits variations in policy intensity over time and across courts. Specifically, we construct an intensity measure by calculating the ratio of broadcast cases to the total number of cases at the court-area by year-quarter level. We prefer this strategy over leveraging case-level treatment, i.e., whether an individual case was broadcast or not, because decisions about which cases to broadcast can be selected by courts handling them. Our approach of leveraging aggregate level policy intensity can avoid this issue: if the broadcasting mechanism had no impact on litigation outcomes, our approach would not identify any effects, despite the selective broadcasting of certain cases by the courts.

The primary empirical specification is the generalized difference-in-differences (DID) with continuous treatment, controlling for court-area and year-quarter fixed effects. We find that female plaintiffs are at less of a disadvantage after the reform is introduced and when the broadcasting intensity increases. Specifically, when the intensity of live broadcasting increases from 0 to 10 percent (i.e., the mean value of the intensity at the court level since the reform started), the gap between female and male plaintiffs' chances of winning declines by more than half a percentage point.

A potential threat to identification is that the aggregate reform intensity may be endogenous. For instance, it could be contaminated by unobserved confounding factors that vary across court-area and over time. To address this concern, we strengthen our analysis by implementing a Bartik-like instrumental variable (IV) strategy. Our Bartik instrument exploits two sources of variation: (1) the distribution of cases across areas in each court in 2014 (before the reform), which is exogenous to the subsequent policy change, and (2) the time-varying average broadcasting intensity for each area and each prefecture (excluding the focal court). The IV estimates are only slightly larger than the corresponding baseline DID estimates, suggesting that potential endogeneity concerns are limited. While the Bartik IV and DID strategies rely on distinct identifying assumptions, our results remain remarkably consistent across both approaches.

The final set of results in this paper focuses on how judges' behavior changes when their decision-making process is subject to public scrutiny. This analysis sheds light on the mechanism through which surveillance technology impacts judicial decisions, thereby lending more credibility to the established impact of the reform.

First, gender disparities in civil litigation may arise because judges are more dismissive and less attentive towards female litigants' claims compared to male litigants'. Motivated by this conjecture, we proxy judicial attentiveness by developing a novel text similarity index that quantifies the extent to which judges incorporate plaintiffs' claims into their decisions. Our analysis reveals that before the reform, judges were less likely to acknowledge and integrate factual claims from female plaintiffs compared to male plaintiffs and that this gender disparity diminished after the reform.

Second, the reform may take effect because judges may behave more professionally and exert more effort under scrutiny. We assess changes in judicial effort through two metrics: the number of legal articles cited and an entropy-based index of writing style. As broadcast intensity increased, judges cited more legal articles in their decisions, suggesting greater effort, and this change was primarily driven by cases with female plaintiffs. Furthermore, judicial decisions became less standardized, containing more nuanced deliberations and factual details following the reform. These more elaborate deliberations, suggesting higher quality decisions, predict smaller gender disparities in outcomes. The documented behavioral changes align with the observed post-reform reductions in gender disparities.

Third, given that the changes in judicial outcomes were likely driven by shifts in judges' behavior, we examine whether the reform affects female and male judges differently. We find that judges of both genders respond to the surveillance technology, with a larger response from male judges, suggesting the reform may be particularly effective in disciplining them.

Finally, we examine alternative channels through which the surveillance technology might influence judicial outcomes, focusing on potential changes in litigant or lawyer behavior. By analyzing subsamples of cases with no litigants or lawyers present during proceedings, we find that changes in their behavior are unlikely to be primary drivers of the observed effects. We also present evidence that the reform did not significantly alter litigants' case filing decisions. This set of evidence combined suggests that changes in judges' behavior are the primary contributors to the observed effects of the reform.

Our paper contributes to a few strands of literature. First, our work provides new findings about gender disparities in civil litigation. Previous studies have predominantly focused on criminal cases and documented gender gaps favoring female defendants. For example, Bindler and Hjalmarsson (2020) find that female defendants tend to receive more lenient treatment compared to male defendants in the UK. The same pattern has been documented in the US (Sorensen, Sarnikar, and Oaxaca 2012; Starr 2015; and Butcher, Park, and Piehl 2017). However, it is unclear whether these findings can be generalized to civil cases or to the context of developing countries. We investigate the gender effects of litigants in civil cases between individuals in China and reveal a distinct pattern: female plaintiffs in civil courts face a disadvantage.[1]

Second, this study advances the literature on how information technology plays a role in public affairs (Pierce, Snow, and McAfee 2015; Mastrobuoni 2020). The introduction of broadcasting cameras into courtrooms is analogous to equipping police officers with body-worn cameras (BWCs), both serving as monitoring devices to enhance transparency and accountability. Research on BWCs has shown that their presence can lead to behavioral changes in both officers and civilians, but it is challenging to distinguish the effects of BWCs on officers vs. civilians (Cubukcu et al. 2021; Zamoff, Greenwood, and Burtch 2021; Ferrazares 2023). Our study evaluates the impact of surveillance cameras in courtrooms. Using a large corpus of judicial decision texts, we show that judges alter their behavior in response to this technology, affecting judicial outcomes. This finding further supports the disciplinary effects of recording cameras in public institutions.[2]

Finally, our paper contributes new evidence on transparency's impact on public sector decision-making and governance (Kosack and Fung 2014). While Hansen, McMahon, and Prat (2018) show how public access to monetary policy deliberations

[1]This finding also enriches the literature that investigates judicial bias against minorities (Shayo and Zussman 2011; Abrams, Bertrand, and Mullainathan 2012; Bar and Zussman 2012; Kastellec 2013; Alesina and La Ferrara 2014; Arnold, Dobbie, and Yang 2018; Anwar, Bayer, and Hjalmarsson 2019; Hou and Truex 2019; Bielen, Marneffe, and Mocan 2021 and Chen and Ornaghi 2023).

[2]While such an intervention may help address biases against racial minorities or women in the public sector, it may have a limited impact on gender bias in the private sector (e.g., Kuhn and Shen 2013; Blau and Kahn 2017; Charles, Guryan, and Pan 2018; Bertrand et al. 2021; and Adams-Prassl et al. 2024).

affects policymakers, we demonstrate that real-time scrutiny prompts judges to act more professionally and deliberate more thoroughly. While existing literature focuses on the impact of enhanced transparency from institutional changes, we examine the effect of higher transparency resulting from the adoption of new technology, a novel contribution to this line of research.

2. Background

2.1. Courts and Judges in China

China's court system consists of four hierarchical levels: local courts, intermediate courts, high courts, and the Supreme Court. Local courts handle general cases at the county level, while intermediate courts oversee larger cases and appeals from local courts at the prefecture level. High courts manage cases appealed from intermediate courts and significant public interest cases at the provincial level. As of January 2021, the Chinese court system consisted of the Supreme People's Court, 33 High Courts (typically one per province), 416 Intermediate Courts (one per prefecture), and 3,087 Local Courts (one per county). In this hierarchy, higher courts supervise and monitor the courts below them in their jurisdictions.[3] The Supreme Court, as the country's highest judicial authority, is the primary driver of the judicial reforms discussed in this paper.

Courts typically comprise several sub-courts, each staffed with specialized judges. Our paper focuses on civil cases, which are categorized into nine major areas: (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. Cases that do not fit these categories are classified as "others." Specific sub-courts handle groups of these areas, with their responsibilities varying across jurisdictions based on population and economic development.

In this judicial system, judges wield considerable power in adjudication with limited supervision during proceedings (Peerenboom 2002). Their attitudes and preferences strongly shape court decisions and litigation outcomes. For "clear facts and few disputes" cases, typically one judge presides using summary procedures, with their opinion dominating the outcome. In "complex cases" (e.g., those with more litigants) following general procedures, 3, 5, or 7 judges are involved, but the chief judge's influence remains dominant in determining litigation outcomes.

[3]For example, local courts are guided and supervised by the corresponding upper intermediate court within the same prefecture. Their personnel decisions and performance evaluations are significantly impacted by these intermediate courts.

2.2. Judicial Reforms

The substantial discretionary power afforded to judges, coupled with limited external oversight effectively turned judicial decision-making into a "black box." Recognizing these issues, China's Supreme Court has implemented a series of initiatives since 2014 aimed at promoting transparency in the judicial system. As part of this reform, four major information disclosure platforms were created to enhance transparency in judicial proceedings. These platforms specialize in publishing legal documents, live broadcasting of trials, providing information about litigation procedures, and displaying blacklists of defaulters.[4]

The key focus of this paper is the impact of the open justice reform, particularly the live broadcasting of trials. The Supreme Court mandated that courts at all levels broadcast trials live on the China Court Trial Online website, with the long-term goal of broadcasting every trial (excluding exceptional cases). The China Court Trial Online website, officially launched in September 2016, allows the public to observe trials in real time or view videotaped proceedings later. By the end of 2021, more than 16 million cases had been broadcast. The platform has attracted considerable attention from the public, including citizens, journalists, and legal practitioners.[5]

While the reform swept through courts across the whole judicial system within a relatively short period (roughly one and a half years), there was considerable variation in the timing of connecting to the website across courts. By September 2016, 383 courts (10.89 percent of all courts) were connected; by January 2017, another 762 courts were connected; and by December 2017, all 3,517 courts were connected.[6] In addition to the variation in timing, there are also large differences in the proportion of cases broadcast across courts.

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 share the same ambition or possess the technical and financial capacities to implement this reform, even though their cooperation was essential to its success. The solution adopted is characteristically Chinese: higher-level courts quan-

[4]The China Judgements Online website, launched in July 2013, is one of the key initiatives. The Supreme Court mandates the publication of all court documents (with some exceptions) on this platform, which had amassed over 120 million documents by December 2021. This reform aims to improve the quality of court rulings by providing easy access to past decisions and to curb judicial misconduct by facilitating the identification of outlier decisions (Liebman et al. 2020). This website serves as our primary source for judicial documents.

[5]For instance, when the Shenzhen Intellectual Property Court broadcast Huawei's patent infringement cases against Samsung on January 11, 2018, over 1 million users watched live. This level of engagement is not uncommon; the Supreme Court reports that more than 10 broadcast cases have each exceeded 10 million views.

[6]Institute of Law of the Chinese Academy of Social Sciences, "Third-Party Evaluation Report on Open Trial of the People's Court 2018."

tify the desired progress and include it in the rubric used to assess the performance of subordinate courts. Although practices vary across localities, intermediate courts typically specify a quota of broadcasted trials for local courts and incorporate it into their evaluation system.[7]

The head of local courts determines the broadcast quota based on superior court requirements and allocates assignments to subcourts. Subcourt head judges then select specific cases for broadcast and submit their plan to the court head. In principle, the court head reviews the plan and makes a final decision; in practice, however, approval is typically automatic.

3. Data Construction

3.1. Sample Construction

We obtained legal documents published on the China Judgements Online website, with assistance from a commercial data company. In this paper, we restrict our sample to civil litigations involving individual litigants rather than institutions or companies. Our data covers January 2014 to December 2018, a period bisected by the gradual introduction of the live broadcasting reform. While courts are responsible for uploading legal documents, publication can lag by several months. Having acquired these documents in 2020, our dataset should closely approximate the complete set of cases up to 2018. The full sample comprises 6,424,324 civil judgments.

The full sample contains cases with a single litigant on both sides as well as those with multiple litigants on at least one side. In our empirical explorations, we consistently analyze both the main sample with all litigants and the subsample with only two litigants (hereafter, the two-litigant subsample). The results obtained using the latter should be particularly informative, especially when we focus on the magnitude of the estimated effects, because the litigants' gender is clearly defined and easily measured in this subsample.

Data on whether a case was broadcast live were acquired from China Court Trial Online. As of April 2021, a total of 11,016,416 cases had been broadcast on China Court Trial Online. We have obtained all of these broadcast records and matched them with our sample of judgments.

[7] The China Court Trial Online website annually selects and publishes lists of 'Excellent Courts' and 'Excellent Judges' from courts at all levels, based on their court broadcasting performance. According to Article 15 of the "Guidelines of the Supreme People's Court for Advancing the Judicial Accountability System," trial broadcasts are a key evaluation criterion for courts and judges. Local courts are assessed by their superior intermediate courts and can receive honors and rewards for exceptional broadcasting performance.

3.2. The Structure and Content of Judicial Decisions

A standard judgment mainly consists of five sections: basic information, claims of litigants, facts recognized by the court, legal principles applied, and outcomes of the litigation. Figure A1 in the Appendix provides an example.

The basic information section includes the title of the document, the case number, the corresponding 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 outlines the causes of the litigation, including the plaintiffs' claims and reasons as well as the defendants' arguments and defenses. The facts section contains the facts and evidence recognized by the court and clarifies the controversial aspects of the case. The subsequent section elucidates the legal principles applied and the judges' justifications.

The final section presents the court decisions and litigation outcomes, such as the extent to which the plaintiff's claims are supported and how the legal costs are shared. The judge's signature concludes the document.

3.3. Constructing Variables

The semi-fixed structure of these judgments allows us to efficiently extract case information. For each judicial decision, we extract a host of variables from the basic information section. For example, we acquire the case numbers, which help us merge the data with other datasets. The *instance* indicates whether a case is in the first instance (heard in the original jurisdiction) or the second instance (rehearing an appeal). The litigant information reveals the names of litigants, which we use to identify whether the case involves individuals, organizations, or enterprises. We can also obtain the number of plaintiffs and defendants, litigant gender (if available), and whether the plaintiffs and defendants appear in court. The data contain the names of lawyers on both sides and the number of lawyers for the plaintiffs and the defendants. Additionally, from the signature located at the bottom of the legal document, we obtain the names of the judge(s).[8]

Among these extracted variables, litigant gender is particularly crucial for our analysis. Of our full sample, 72% of judgments include litigants' genders, while 26% and 27% omit plaintiff's and defendant's genders, respectively. Excluding cases with missing gender information yields our main sample of 4,601,718 cases. For multiple plain-

[8]We identify individual judges by their name and the court they work for. Judges who share the same name but work for different courts simultaneously are identified as distinct individuals.

tiffs or defendants, we code gender based on the first listed party, as civil litigation documents typically order litigants by their relevance to the case. Unlike litigant gender, judges' genders are not disclosed in legal documents. We infer them using *Ngender*, an algorithm widely used in industry to predict Chinese individuals' profiles.[9]

Beyond gender information, we also collect information on the area to which each case belongs from the judicial decisions. However, recognizing potential significant heterogeneity within each area, we introduce a more refined measure called "case type." We construct this measure based on the combination of laws applied to each case, categorizing cases as similar if they apply similar laws. To ensure an objective categorization process, we employ an unsupervised topic model that classifies all 6 million cases into 50 distinct types. Appendix B provides details on this process. All of our analyses include case type as an additional control variable, where applicable.

Having obtained this set of case-level characteristics, we now turn to our key outcome variable: the litigation outcome. According to civil procedure law, the extent to which the court supports one party in a lawsuit is inversely proportional to the share of the legal costs that he or she pays. For instance, if the judge rules that the plaintiff prevails entirely, then the defendant must pay 100 percent of the legal costs. If the judge supports the plaintiff's claim only partially, such as 80 percent, then the plaintiff pays 20 percent of the legal costs. Therefore, we define the plaintiffs' chances of winning as the share of the legal costs borne by the defendants. We extract this cost allocation information from the outcome section of each case.

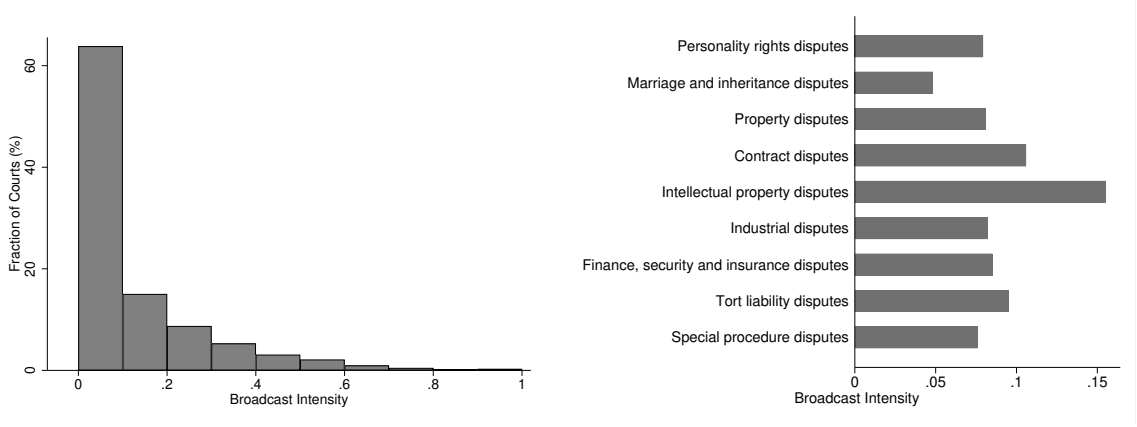
In addition to case-specific variables, we also consider measures of policy implementation. Using broadcast information, we construct broadcast *intensity*–the ratio of broadcast cases to total cases at a given aggregate level. The default level is court-area $\times$ year-quarter.

We supplement our judicial document data with region-level information from the *China City Statistical Yearbook*. This includes each prefecture's annual GDP per capita, population, and internet penetration rate (households with internet access divided by total households). These variables provide basic characteristics of the prefectures in our analysis.

3.4. Summary Statistics

Table A1 in Appendix A reports the summary statistics for our judicial decision data. Column (1) shows means and standard deviations in the main sample. The average plaintiff winning rate is 74%, representing the extent to which plaintiffs' claims are

[9]We validate the *Ngender* algorithm against a database of nearly 2,000 judges, compiled from the annual "Merit Judges" lists issued by the Supreme Court from 2000 to 2020. This external validation shows 92% accuracy in predicting judges' genders from their names.



(a) Courts Distribution (b) By Areas

Figure 1. *Broadcast intensity across courts and areas. The left panel displays the distribution of courts by broadcast intensity post-reform. The right panel shows the broadcast intensity across nine civil litigation areas.*

supported. The average number of plaintiffs per case is 1.13, indicating that most civil cases involving individuals have only one plaintiff, while the number of defendants is higher at 1.72. Nearly half of plaintiffs (49%) appear in the courtroom during trials, compared to only 8% of defendants, suggesting a more passive role for the latter. Plaintiffs have an average of 0.48 lawyers per case, while defendants have even fewer at 0.28, which implies that the majority of litigants in Chinese civil cases do not employ legal representation.

Females are less present in courtrooms. On average, 31% of cases have female plaintiffs, while only 19% have female defendants. Cases adjudicated by female judges account for 29% of the total, which aligns with the proportion of female judges in the judicial population. The average number of judges per case is 1.43, indicating that most cases are adjudicated by a single judge. The vast majority of cases in our sample (89%) are first instance cases. The legal costs, which are proportional to the size of the case (or the amount of money involved), are presented in logarithmic form.

We also present summary statistics for the aforementioned variables in a subsample with two litigants (i.e., one plaintiff and one defendant). Column (2) of Table A1 summarizes the means and standard deviations of relevant variables in this subsample. The two-litigant subsample exhibits similar characteristics to the main sample, except that both the number of lawyers and the average legal costs are lower. This difference is expected, as the size of cases may correlate with the number of litigants.

Court broadcast intensity varies across courts and areas. Figure 1(a) shows the distribution of courts by average broadcast intensity from September 2016 to December 2018. Over 60% of courts broadcast fewer than 10% of their cases, while less than 20%

broadcast more than 20%. Figure 1(b) shows the average intensity for each area during the same period. The overall average intensity is approximately 10%, with modest variations across different areas. Notably, marriage, family, and inheritance disputes have a lower average intensity (around 5%), while intellectual property disputes have a higher intensity (around 15%).

4. Gender, Reform and Litigation Outcomes

4.1. Pre-reform Gender Disparities

We begin our investigation by examining the gender disparity in civil courts during the per-reform period of our sample, specifically from 2014Q1 to 2016Q2. We estimate the following equation:

$$y_{ijktl} = \beta_0 + \beta \text{Female}_i + X_i\zeta + \omega_{j\times k} + \lambda_t + \delta_l + \varepsilon_{ijktl}, \tag{1}$$

where y_{ijktl} denotes the plaintiff's chances of winning in case i adjudicated by judge l, area j, court k, and year-quarter t; Female_i is a dummy variable that takes a value of 1 if the plaintiff is female in case i, and 0 otherwise. To control for area-, court-, and time-specific factors that affect the plaintiff's chances of winning, we include the court-area fixed effect (i.e., $\omega_{j\times k}$) and year-quarter fixed effects (i.e., λ_t), respectively. We also include chief judge fixed effects (i.e., δ_l) to control for individual judges' ruling patterns. All standard errors are clustered at the court level. Our main coefficient of interest is β, which captures the gender difference in the chances of winning.

We also include a set of control variables (i.e., X_i) at the case and prefecture levels. At the case level, we include the defendant's gender as it may affect the plaintiff's chances of winning. We also consider the number of parties involved (plaintiffs and defendants) and the number of presiding judges, which correlates with the complexity of the dispute. In addition, we examine whether the plaintiffs and defendants appear in court, as their presence or absence may affect the outcome. The number of lawyers representing the plaintiff and defendant is used to proxy for the legal resources available to each side. We also include a variable for legal fees (log) as a measure of the size of the case. Furthermore, we include a dummy variable for the instance of the case (first or second), as the characteristics of cases may differ between the two instances. We include case type constructed using topic modeling (as discussed in section 3.3) to control for the nuanced characteristics of the cases in our analysis. At the prefecture level, we include GDP per capita and population to proxy for the region's level of development and size, respectively. We also incorporate the internet penetration rate to proxy for local accessibility to Internet which could be relevant for the reform.

Table 1. *Plaintiff Gender and Litigation Outcomes: The Pre-reform Period*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	All Areas	Excl Marriage	All Areas	Excl Marriage
	(1)	(2)	(3)	(4)
Female	-0.0356***	-0.0226***	-0.0530***	-0.0344***
	(0.00156)	(0.00159)	(0.00229)	(0.00233)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.636	0.730	0.534	0.680
N	1,070,802	883,115	614,414	436,647
R^2	0.581	0.486	0.657	0.609

Notes: This table presents the results of analyzing gender differentials in the plaintiff's chances of winning for civil cases adjudicated during the pre-reform period from 2014Q1 to 2016Q2. Columns (1) and (2) report the results with main sample and columns (3) and (4) present the results with two-litigant subsample. In columns (2) and (4), we further exclude cases in the area of marriage and family disputes. Case-level control variables include the defendant gender, 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 case type. 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.

We first estimate Equation (1) using our full sample (i.e., all litigants), in which each side can have more than one litigant. Table 1 presents the regression results. Column (1) reports the estimated coefficient on the dummy *Female*, which is negative and statistically significant. For comparison, we repeat our estimation with the two-litigant subsample with only one litigant on either side (i.e., column (3) of Table 1). After controlling for a set of control variables and fixed effects, we find that the chances of winning for female plaintiffs are approximately 3.56 percentage points lower than for male plaintiffs (equivalent to 5.6% of the sample mean). This gender difference is even more pronounced in the two-litigant scenarios, where it is approximately 5.3 percentage points (equivalent to 10% of the sample mean). This difference is expected, as our full-sample estimations only consider the first litigant's gender, and gender mixing on either side likely biases estimates toward zero.

To ensure that the association is not driven only by gender-salient issues, such as litigation related to marriage, family and inheritance, we drop cases in this area and repeat our estimations with both samples. The results are reported in columns (2) and (4) of Table 1. Both estimated coefficients are smaller in size but still significant, suggesting that such an association exists in areas that are not gender salient.

4.2. Gender Disparities and Reform Implementation

The findings on pre-reform gender disparities set the stage for our analysis of the reform's impact. The reform's rollout introduced significant variations in public scrutiny across courts and over time, providing a unique opportunity to examine how increased transparency affects gender disparities in judicial decisions. Notably, the reform's primary objective was to enhance the overall transparency and quality of judicial proceedings, rather than to specifically address gender disparities (as detailed in section 2.2). This feature allows us to interpret any post-reform changes in these disparities as evidence of gender bias.

While the top-down implementation of the reform at the aggregate level is likely exogenous, its application within courts–specifically, the selection of individual cases for broadcasting–may not be random. For instance, it is plausible that the head judges of subcourts select straightforward cases with clear paths to fair adjudication for broadcasting. If this case selection is correlated with both the plaintiff's gender and judicial outcomes, the empirical design could be susceptible to omitted variable bias.

To examine this possibility, we follow the approach of Bhuller et al. (2020) and conduct balancing tests to investigate the determinants of live broadcasting at the case level (Table A2). First, we regress the plaintiff's chances of winning on a set of case characteristics. The results in column (1) show strong correlations between all case characteristics and judicial decisions. Next, we examine differences in these characteristics between broadcast and non-broadcast cases by regressing a dummy variable indicating live broadcast on the same set of control variables (column (2)). Our analysis reveals that several observable attributes, including plaintiff gender, can predict live broadcasting, suggesting a non-random case allocation process.

This selection mechanism motivates us to utilize variations in the reform implementation at the aggregate level. In the following subsection, we employ court-area level broadcast intensity as our treatment variable, leveraging its substantial variations across both courts and areas (see Figure 1(a) and 1(b)). By applying a generalized difference-in-differences (DID) design with this treatment, we can effectively circumvent the case-level selection issue: if the reform has no impact on litigation outcomes, any reshuffling of cases between broadcast and non-broadcast categories would not produce any effects in our analysis.

4.3. The Difference-in-differences Design and Results

In this section, we leverage the broadcast reform to examine gender disparity in judicial decisions. Using the full sample (2014-2018), we estimate a generalized (continu-

Table 2. *Difference-in-Differences Estimation: Impacts of Live Broadcasting*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	All Areas	Excl Marriage	All Areas	Excl Marriage
	(1)	(2)	(3)	(4)
Female × Intensity	0.0394***	0.0249***	0.0536***	0.0302***
	(0.00255)	(0.00247)	(0.00336)	(0.00306)
Female	-0.0116***	-0.00455***	-0.0199***	-0.00943***
	(0.000847)	(0.000822)	(0.00108)	(0.00101)
Intensity	-0.0177***	-0.0117***	-0.0178**	-0.00993
	(0.00441)	(0.00450)	(0.00727)	(0.00756)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.744	0.787	0.709	0.788
N	3,974,316	3,696,048	2,063,379	1,810,190
R^2	0.464	0.386	0.584	0.495

Notes: This table presents the main results for the DID specification, using the sample of cases adjudicated between 2014 and 2018 (both the pre- and post-reform periods). Columns (1) and (2) report the results with main sample and columns (3) and (4) report the results with the two-litigant subsample. In columns (2) and (4), we exclude cases in the area of marriage and family disputes. Case-level control variables include the defendant gender, 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 case type. 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.

ous) DID model by controlling for court-area and year-quarter fixed effects as follows:

$$y_{ijktl} = \beta_0 + \beta \text{Female}_i + \gamma \text{intensity}_{jkt} + \theta \text{Female}_i * \text{intensity}_{jkt} + X_i\zeta + \omega_{j\times k} + \lambda_t + \delta_l + \varepsilon_{ijktl}, \tag{2}$$

where intensity_{jkt} is the ratio of live-broadcast cases to total cases in area j, court k, and year-quarter t. We also control for the same set of control variables X_i and judge fixed effects (i.e., δ_l), as in Equation (1). Note that in this DID model, a continuous treatment is used, given the reform intensity is continuous and time varying.

The coefficient on the interaction term is our main interest; it captures how live broadcasting affects gender differences in the chances of winning, conditional on fixed differences across court-areas and year-quarters. First, the inclusion of court-area and year-quarter fixed effects ensures that our estimated effects are not driven by area or time specific factors. Second, the interaction term effectively differences out potential fixed effects correlated with gender. This approach allows us to distinguish the effect of increased transparency from other confounding factors that might influence gender differences in judicial outcomes.

Results from our estimations are presented in Table 2. Columns (1) presents the estimation results for the main sample. Columns (3) report the results for the two-litigant

subsample for comparison. Columns (2) and (4) further exclude cases related to marriage, family, and inheritance and conduct the same analyses on both samples. All specifications yield consistent outcomes: the gender disparity against female plaintiffs diminishes when a larger fraction of cases is broadcast live. The magnitude of this effect is fairly large: based on the estimate in column (3) of Table 2, an increase in broadcasting intensity from 0 percent to 10 percent (the mean value of court-level intensity since the reform's introduction) removes approximately one-tenth of female plaintiffs' pre-reform disadvantage (i.e., 5.4×10 percent ≈ 0.54 percentage points). The narrowing of the gender gap as litigation becomes more transparent provides evidence of the existence of gender discrimination.[10]

Dynamics Effects Beyond examining the reform's average effects, we also investigate its dynamic impact over time. As broadcast intensity gradually increased after the reform's introduction, we expect to observe a progressive strengthening of the reform's impact, rather than an abrupt, discontinuous change.

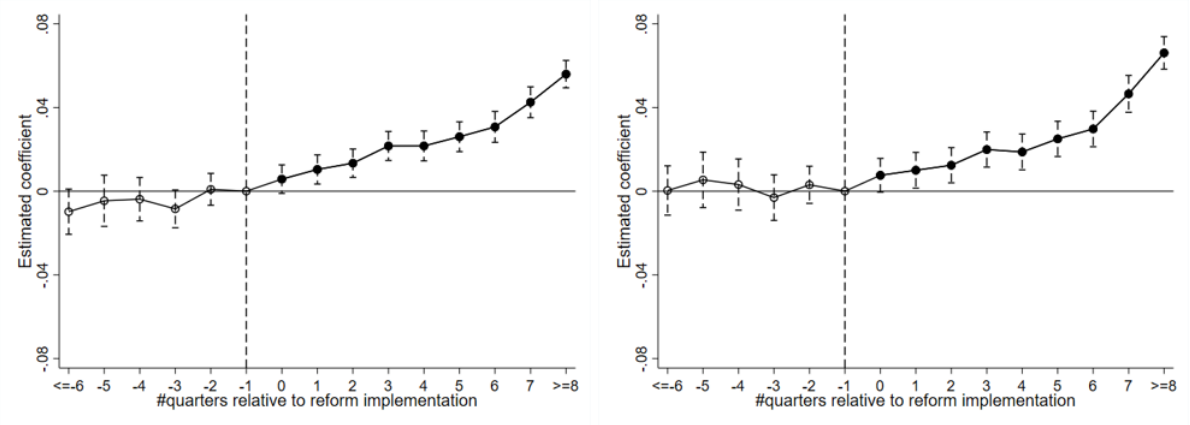
To investigate, we employ an event study model and exploit the variation in the timing of the reform across prefectures (i.e., the extensive margin), the administrative level at which the reform was decided, financed and implemented. We construct a dummy variable $\text{Reform}_{p,t}$, which takes a value of 1 if prefecture p (where court k is located) started the reform in its jurisdiction in quarter t, and 0 otherwise. To estimate the dynamic effects, we employ the following specification:

$$\begin{aligned} y_{ijktl} = \beta_0 + \beta \text{Female}_i + \sum_{\tau=-6}^{8} \gamma_\tau \text{Reform}_{p,t}^{\tau} + \sum_{\tau=-6}^{8} \theta_\tau \text{Female}_i * \text{Reform}_{p,t}^{\tau} \\ + X_i\zeta + \omega_{j\times k} + \lambda_t + \delta_l + \varepsilon_{ijktl}, \end{aligned} \tag{3}$$

where $\text{Reform}_{p,t}^{\tau}$ for τ = -6,...,8 is a sequence of dummy variables, indicating that quarter t is τ quarters away from the introduction of the reform in prefecture p. To absorb the effects outside of the estimation window, we additionally set $\tau = -6$ for $\tau \leq -6$ and $\tau = 8$ for $\tau \geq 8$, respectively.[11] The quarter prior to the reform (i.e., $\tau = -1$) is dropped as the reference period. The same control variables and fixed effects as in our baseline specification are also included. Our main interest is the estimated coefficients on interaction terms between Female_i and $\text{Reform}_{p,t}^{\tau}$, which capture the dynamic impact of reform on the gender differential. Figure 2 displays the estimation results. The

[10] In addition to the interaction term, the sum of the main effect of the female dummy and the interaction term is also of interest. We perform a joint significance test to determine whether this sum is significantly different from zero when the broadcasting intensity is set to its mean (10%). The F-test yields an F-statistic of 170 (p-value < 0.01), indicating that the reform partially mitigates the disadvantage for female plaintiffs, but it does not completely eliminate or reverse the gender disparity.

[11] We select 6 quarters for the pre-reform period and 8 quarters for the post-reform period to ensure balanced sample sizes at both ends of the estimation window. This approach accounts for the gradual increase in the number of civil cases over time.



(a) All Litigants (b) Two Litigants

Figure 2. *Event Study. The figure illustrates the reform's estimated effects on the gender differential in winning chances over time, spanning from 6 quarters before its introduction ($\tau = -6, ..., -1, 0$) to 8 quarters and beyond afterwards ($\tau = 1, ..., 8$). Dots represent the estimated coefficients for each period, with bars indicating the corresponding 95 percent confidence intervals.*

positive and increasing post-reform coefficients suggest that the reform's impact on gender disparities strengthens over time, aligning with our earlier expectation and boosting our confidence in the credibility of our findings.[12]

Control for linear trend and alternative fixed effects. We further conduct a series of exercises to check the robustness of our results. It is possible that some unobserved time-varying variables at the court level simultaneously influence both the reform intensity and the trends of gender disparities. To account for potential linear trends, we estimate a specification that includes court-specific time trends in Equation (2). This approach captures smooth, court-specific changes over time. Furthermore, it is also likely that other court-specific shocks may be correlated with both the reform implementation and changes in judicial outcomes (e.g., court leadership turnover and other judicial reforms implemented during the same period). To alleviate this concern, we estimate a specification that incorporates court $\times$ year-quarter fixed effects into Equation (2). Table A3 reports the results of these alternative specifications, demonstrating the robustness of our main findings.

[12]Recent literature has raised concerns about the reliability of event study analyses under staggered treatment timing (Borusyak, Jaravel, and Spiess 2021; Roth et al. 2023). While a number of robust estimators have been proposed as general solutions, these methods cannot be directly applied to our specification because our primary interest lies in the interaction terms (i.e., *Female* $\times$ *Intensity*). To our knowledge, existing frameworks for robust estimators have not yet been extended to accommodate interaction terms. To address this potential challenge, we divided the sample based on the plaintiff's gender and conducted separate event study regressions. Both robust and TWFE estimators demonstrate consistent patterns, revealing no significant pre-existing trends. Furthermore, the winning chances of male plaintiffs held steady following the reform, whereas female plaintiffs experienced a substantial improvement in their success rate. This suggests the policy primarily improved litigation outcomes for women. These results are available upon request.

Robustness of TWFE A recent strand of literature highlights potential issues with the two-way fixed effects (TWFE) estimator that can lead to unreliable estimates. Callaway, Goodman-Bacon, and Sant'Anna (2021) demonstrate that in settings with continuous treatment, the validity of TWFE estimation relies on the assumption that units receiving a stronger treatment would exhibit the same treatment effects as those receiving a weaker treatment if they were subjected to the same weaker treatment. Cook et al. (2023) suggest that this assumption is more likely to hold when treatment strength is uncorrelated with observable characteristics. Following this line of literature, we address this concern in Appendix C, showing that these issues are unlikely to be problematic in our context.

Alternative measures and samples. To ensure the robustness of our results across various measures and sample choices, we conduct several additional analyses. First, we employ an alternative treatment variable, using overall court-level intensity instead of court-area-level intensity. Second, we test our findings against alternative outcome measures, including a coarser definition of litigation outcome: a dummy variable *Win* that equals 1 when the plaintiffs' chances of winning exceed 50 percent, and 0 otherwise. Third, to address potential inaccuracies in intensity measurement due to insufficient case numbers, we exclude court-area-quarter cells with fewer than 20 cases from our analysis. Lastly, we address missing gender information for litigants by using a gender imputation algorithm (Ngender) to predict litigants' gender in these cases and incorporate them into our main sample. We re-estimate Equation (2) using these alternative measures and samples, reporting the results in Table A4. The magnitude and significance of these estimated coefficients closely align with our baseline results.

4.4. The Bartik Instrument Design and Results

While our analyses in the previous section support the validity of our DID design, a potential threat to identification is that broadcast intensity may be endogenous due to unobserved factors varying across courts and over time. To address this concern, we complement our DID analysis with a Bartik IV approach.

In our context, the Bartik IV effectively addresses endogeneity concerns by exploiting two sources of variation: pre-determined shares of areas in each court, and area-specific broadcast intensity at the prefecture level. This approach follows the literature developed by Goldsmith-Pinkham, Sorkin, and Swift (2020) and Borusyak, Hull, and Jaravel (2022). They show that Bartik instruments can address time-varying confounders through the interaction of local shares and aggregate shocks.

Our investigation using the Bartik IV approach proceeds in two steps. First, we describe the construction of the Bartik instrument and present the estimation results of the IV regressions. Second, we examine the identification assumptions underlying

our approach and provide evidence supporting their validity.

Bartik instrument and results. The Bartik instrument we construct exploits two sources of variation: (1) Share—the distribution of cases across area in each court before the reform (specifically, in 2014) and (2) Shift—the time-varying average broadcasting intensity for each area and prefecture (excluding the focal court). The predicted intensity is calculated as:

$$Z_{kt} = \sum_{j \in \text{area}} \alpha_{jk}^{2014} \, \text{intensity}_{jp_{-k}t}, \tag{4}$$

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} denotes the share of area j for court k in 2014, i.e., 2 years before the reform's implementation. The term $\text{intensity}_{jp_{-k}t}$ is the average intensity in area j and year-quarter t of prefecture p (where court k's is located), excluding court k.

Table 3 reports our IV estimates using both samples. For the main sample, columns (1) and (2) of Table 3 present the first-stage results and column (3) displays 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). The coefficients in the first stage are highly significant, implying that the instrument has high predictive power. As shown in columns (1) and (4), each percentage increase in predicted broadcasting intensity is associated with a 0.688 and 0.703 percentage increase in actual intensity in the main sample and two-litigant sample, respectively. 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 significant. We observe that the IV estimates are only slightly larger than the corresponding estimates in the baseline model (reported in columns (1) and (3) of Table 2). The similarity in effect magnitudes across specifications with and without instruments suggests that endogeneity is less of a concern in our analysis.

Exogenous share assumption. In our study, the validity of the Bartik instrument depends on the exogeneity of early shares in areas as of 2014.[13] That is, the exclusive restriction condition requires that, conditional on court-area and year-quarter fixed effects, third factors affecting the litigation outcome after the reform should not be si-

[13]The validity of Bartik instruments can arise from either the exogeneity of initial shares or common shocks (Goldsmith-Pinkham, Sorkin, and Swift 2020; Borusyak, Hull, and Jaravel 2022). Examining our Bartik setting through the lens of the "share view" (Goldsmith-Pinkham, Sorkin, and Swift 2020), the identifying assumption is that the differential exposure to common shocks is exogenous. Alternatively, Borusyak, Hull, and Jaravel (2022) show that exogenous independent shocks to many sectors cause the Bartik estimator to be consistent, even when the shares are not exogenous. In our setting, we do not rely on the number of areas to be large enough to ensure consistency.

Table 3. *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			0.0726***			0.0889***
			(0.00599)			(0.00699)
Female	0.000375	0.0224***	-0.0138***	0.000503*	0.0217***	-0.0224***
	(0.000265)	(0.00163)	(0.000903)	(0.000293)	(0.00168)	(0.00115)
Intensity			-0.0207***			-0.0198***
			(0.00677)			(0.00724)
Predicted Intensity	0.687***	-0.0799***		0.703***	-0.0745***	
	(0.0451)	(0.00796)		(0.0447)	(0.00767)	
Female × Predicted Intensity	-0.0150***	0.960***		-0.0161***	0.983***	
	(0.00550)	(0.0360)		(0.00556)	(0.0347)	
Controls	Y	Y	Y	Y	Y	Y
Judge FE	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 Value			137.923			122.984
K-P Wald Value			114.858			122.643
N	3,974,316	3,974,316	3,974,316	2,063,379	2,063,379	2,063,379
R^2	0.739	0.565	0.017	0.765	0.592	0.026

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 areas in each court in 2014 and the time-varying prefecture-level average broadcasting intensity for each 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, 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 case type. 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.

multaneously correlated with both the early area share in 2014 and the post-reform prefecture-level average intensity by area . Following Goldsmith-Pinkham, Sorkin, and Swift (2020), we conduct three complementary tests to corroborate the validity of this assumption in our context. This set of tests has been widely used in the Bartik IV literature (e.g., Acemoglu and Restrepo 2020 and Fouka, Mazumder, and Tabellini 2022).

First, one threat to the validity of our instrument is that the pre-reform legal share might be correlated with unobserved time-varying confounding factors at regional level. The regional characteristics that determined the pre-reform share distribution might be correlated with long-run confounding factors affecting both the reform implementation patterns and trends in judicial outcomes. To examine this possibility, we conduct a placebo test for the "pre-trends". If there were long-run factors that contaminated the predicted intensity we construct, we would expect the judicial outcomes *before* the reform to be correlated with the predicted intensity *after* the reform.

Specifically, we examine whether gender disparities in litigation outcomes in 2014,

2015 and the first 2 quarters of 2016 (prior to the reform) respond to the predicted intensity in 2018. We estimate Equation (2) using litigation outcomes from the 2014, 2015 and 2016 (first two quarters) subsamples and the predicted intensity constructed for the subsample for 2018. Table A5 reports the results. In this table, the estimated coefficients on the interaction terms are negligible in magnitude and statistically insignificant, which increases our confidence in the validity of this Bartik instrument.

Second, our identifying assumption could be violated if the initial characteristics that determined pre-reform legal share had dynamic effects on the changes in judicial outcomes over time. To address it, we conduct two robustness checks. We begin by incorporating interaction terms of time dummies with the 2014 regional characteristics, including internet penetration rate, GDP per capita, and population, into our Bartik specification. This refined specification allows us to account for the dynamic effects of these regional characteristics over time. The results, displayed in Table A6, indicate that the estimates remain consistent and stable across all specifications, with *almost no changes* in the magnitude of the coefficients, suggesting that these regional characteristics do not exhibit any systematic correlation with trends of gender disparity in litigation outcomes.

Furthermore, we directly examine whether the early shares are correlated with the initial conditions, i.e., the regional characteristics in 2014. If these regional characteristics are not correlated with the early shares at the initial stage, it is less likely that these initial conditions would dynamically correlate with the time-varying predicted intensity and bias our estimates. Following the suggestions made by Goldsmith-Pinkham, Sorkin, and Swift (2020), we calculate the Rotemberg weight for each area and examine whether the key areas driving the bulk of the variation in the instrument are correlated with the three aforementioned regional characteristics: Internet penetration rate, GDP per capita and population. For the four most significant (based on their Rotemberg weights) areas —contract, tort, marriage, and property—we regress the court-level initial shares in 2014 on the three prefecture-level control variables and present the results in Table A7. We find that the correlations between each share and all three control variables remain consistently small and statistically insignificant.

Alternative instrument. To ensure our results are not sensitive to the choice of instruments, we construct an additional instrumental variable. Specifically, we use a weighted average of the broadcasting intensity of non-j areas in non-k courts in the same prefecture p. This approach excludes not only own-court cases but also own-area cases from the construction of the IV. The average intensity of other courts and areas within the same prefecture should be positively correlated with the broadcasting intensity of area j in court k, due to the peer effects (such as competition for higher rankings) among local courts during the reform implementation (as discussed in sec-

tion 2.2). The identifying assumption is that, conditionally on a set of control variables and fixed effects, the average intensity of other courts in other areas should be unrelated to the litigation outcomes of court k, except through its influence on the intensity of court-area $k \times j$. Table A8 reports the estimation results, which are similar to our Bartik IV estimation in section 4.4.

In conclusion, our two empirical strategies rest on different sets of identifying assumptions. Nevertheless, they deliver qualitatively similar and quantitatively comparable results, bolstering our confidence that the reform indeed had a significant impact on judicial outcomes.

5. Judges and the Reform

After documenting the reform's effect on gender disparities in litigation outcomes, we investigate the underlying mechanisms. While direct measures of courtroom conduct (e.g., time spent with female versus male plaintiffs) are limited, judicial decisions' textual content reveals key aspects of judicial behavior. Through textual analysis, we examine the reform's impact on judges' acknowledgment of plaintiffs' facts across gender and their effort levels in proceedings. By examining these shifts in judicial conduct, we aim to uncover the pathways through which transparency influences judicial decisions in the courtroom.

5.1. Judicial Attention Under Enhanced Scrutiny

Gender disparities in civil litigation may arise from judges giving less attention and weight to female litigants' claims compared to those of male litigants. This differential treatment likely stems from gender stereotypes that portray women as more emotional and less rational, leading judges to perceive female plaintiffs' statements as less credible and informative. As a result, judges may be less likely to incorporate female litigants' factual statements into their deliberations, potentially contributing to women's lower success rates in court. This mechanism aligns with Patton and Smith (2017) finding that female lawyers receive less speaking time and face earlier interruptions in U.S. courts. The reform, by exposing judicial conduct to public scrutiny, may compel judges to counter such bias, behave more professionally, and afford equal consideration to female litigants.

To provide suggestive evidence of this judicial behavior, we evaluate how judges acknowledge litigants' statements by comparing the content of litigants' and judges' statements. Inspired by Kelly et al. (2021), we develop a novel text similarity measure to quantify how judges incorporate litigants' statements into their decisions.[14]

[14]Kelly et al. (2021) identified influential patents through textual similarity to previous and subse-

Table 4. *Text Similarity, Plaintiff Gender and the Reform*

	Outcome Variable: Δ Similarity			
	Pre-reform Period		Full Period	
	All Litigants	Two Litigants	All Litigants	Two Litigants
	(1)	(2)	(3)	(4)
Female	-0.0116***	-0.0195***	-0.00851***	-0.0134***
	(0.000718)	(0.00106)	(0.000491)	(0.000703)
Intensity			-0.00177	-0.00595
			(0.00658)	(0.00750)
Female × Intensity			0.00612***	0.0159***
			(0.00226)	(0.00283)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.154	0.204	0.143	0.219
N	1,065,884	612,174	3,957,595	2,054,867
R^2	0.528	0.534	0.525	0.554

Notes: This table presents the regression results analyzing the relationship between text similarity, plaintiff gender and the reform. Δ similarity is measured as the difference between the similarity of the plaintiff's claim and the judge's claim, and the similarity of the defendant's claim and the judge's claim. The outcome variable measures the degree to which the judge acknowledges the plaintiff's statements compared to the defendant's statements. In column (1), we examine the correlation between Δ similarity and plaintiff gender, using the pre-reform subsample. In column (3), we further include broadcast intensity and the interaction term of female dummy and intensity. Consistent results are found in the two-litigant subsample, as presented in columns (2) and (4). 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 case type. 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.

For each judgment, we analyze three sections: facts presented by the plaintiff, facts presented by the defendant, and facts recognized by the judge. We calculate textual similarity between judges' acknowledged facts and those of each party, generating measures of "plaintiff-similarity" and "defendant-similarity."[15]

The assumption underlying this measure is that higher text similarity between a litigant's statements and a judge's acknowledged facts indicates greater judicial attention to that litigant's account. Specifically, when judges devote more attention to considering a litigant's statements, their recognized facts should more closely reflect that litigant's account, all else being equal.

We introduce a measure called Δ *similarity*, representing the difference between

quent work.

[15]To compute the textual similarity of any two texts i and j, we follow a two-step process. First, using CoSENT, we produce two vectors (embeddings), each representing the semantic meaning of the input text, denoted as $\mathbf{V}_i$ and $\mathbf{V}_j$. The CoSENT model is a variant of Sentence-BERT, a large language model pretrained on billions of data using the BERT methodology. These high-dimensional vectors can capture deep semantic information from the text. Second, we compute the cosine similarity between vectors i and j: $d_{ij} \equiv \frac{\mathbf{V}_i \cdot \mathbf{V}_j}{\|\mathbf{V}_i\| \|\mathbf{V}_j\|}$, where $d_{ij} \in [-1, 1]$. The larger the value of d_{ij}, the more similar texts i and j are.

plaintiff-similarity and defendant-similarity. This measure captures the relative attention judges allocate to plaintiffs versus defendants in each case. To validate this measure, we examine its correlation with plaintiffs' chances of winning. A positive correlation would suggest it contains meaningful information about judicial attention. Figure A2 demonstrates a robust positive correlation between winning probability and Δ similarity in both the all-litigant sample and the two-litigant subsample.

Building on this measure, we first investigate whether female plaintiffs were less likely than male plaintiffs to have their statements acknowledged by judges, focusing on the pre-reform subsample. To do so, we estimate Equation (1) by replacing the dependent variable with Δsimilarity. The results are presented in columns (1) and (2) of Table 4. In both the full sample and the two-litigant subsample, the estimated coefficient is negative and statistically significant, implying that in the pre-reform period, judges were less likely to recognize a female plaintiff's statement (in relation to the defendant's statement) compared to one submitted by a male plaintiff.

To assess the reform's impact on judicial acknowledgment of litigants' statements, we estimate Equation (2) using Δsimilarity as the dependent variable. The results are presented in columns (3) and (4) of Table 4. In both samples, the coefficients associated with intensity_{jkt} are not statistically significant, indicating that the reform does not significantly affect judges' attitudes toward plaintiffs overall. However, the estimated coefficients on the interaction term are statistically significant and positive, indicating that as the intensity of live broadcasting increases, the gender disparity in how judges acknowledge and integrate plaintiffs' statements diminishes.

These findings suggest that judges paid less attention to female plaintiffs compared to male plaintiffs before the reform, when scrutiny was low. The reform's enhanced transparency promoted more professional judicial conduct, leading judges to pay more attention to female litigants and reducing gender disparities in judicial attention during proceedings. These behavioral changes align with the observed post-reform reductions in gender disparities in judicial outcomes.

5.2. Judicial Effort Under Enhanced Scrutiny

The second aspect we investigate is judicial effort, a key indicator of diligence and commitment to thorough case examination that directly influences decision quality. Enhanced transparency may motivate judges to dedicate more effort to proceedings, potentially improving judicial quality. We assess judicial effort using two proxies from written decisions: the number of law articles cited and the writing style of judgments.

First, we analyze the number of legal articles cited in judgments as an indicator of judicial effort, a metric widely used in law and economics to assess judges' perfor-

Table 5. *Number of Law Articles Cited and Impacts of Trials Broadcasting*

	Outcome Variable: Number of Law Articles Cited					
	All Litigants			Sub-sample: Two Litigants		
	Plaintiff's Gender:					
	All	Female	Male	All	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Intensity	0.170***	0.296***	0.114*	0.170***	0.283***	0.119**
	(0.0629)	(0.0959)	(0.0677)	(0.0554)	(0.0828)	(0.0597)
Controls	Y	Y	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Mean of outcome	8.663	8.628	8.690	6.536	6.437	6.582
N	3,974,316	1,144,191	2,826,878	2,063,379	583,122	1,476,902
R^2	0.678	0.696	0.676	0.524	0.545	0.527

Notes: This table presents results for the analysis of the number of law articles cited in the judgments. We examine the relationship between the number of law articles cited in the judgement and the reform intensity. The corresponding results are shown in column (1) for full sample, column (2) for female plaintiff sub-sample and column (3) for male plaintiff sub-sample. We also present the respective results with the two-litigant subsample in columns (4),(5) and (6). 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 case type. 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.

mance and document quality (e.g., Chen et al. 2022). Judges exerting minimal effort may simply cite generic, commonly applied articles. In contrast, more diligent judges may cite both common and less frequently used articles to address novel case aspects.

To this end, we extract the number of law articles cited in each judgment and regress this measure on the court-area-level intensity of broadcasting, incorporating the same set of fixed effects and controls used in Equation (2). The results are presented in Table 5. As shown in column (1), an increase in the intensity of live broadcasting leads to an increase in the number of law articles cited. This pattern also holds for the two-litigant subsample, as reported in column (4). These findings suggest that the reform has prompted judges to invest more effort in their judicial deliberations.

Furthermore, we explore whether this pattern varies across cases with different plaintiff genders. We divide the full sample into two subsamples: a female plaintiff sub-sample and a male plaintiff sub-sample, and conduct the analysis separately. The results are presented in Table 5, with columns (2) and (5) showing findings for female plaintiffs, and columns (3) and (6) for male plaintiffs. The estimated coefficients are significant and larger in magnitude for the female plaintiff subsamples but smaller and less significant for the male plaintiff subsamples. This suggests that the effects of the reform on judicial effort, as measured by law article citations, are primarily driven by cases involving female plaintiffs. This finding indicates that increased judicial effort

in cases with female plaintiffs may be a mechanism through which the reform reduces gender disparity in judicial decisions.

Second, beyond citation patterns, we analyze the writing style of legal documents that may correlate with judges' effort in evidence collection and legal deliberation. We hypothesize that when a judge invests less time and effort in deliberating a case, they are more likely to rely on standardized templates, resulting in judgments that are in a cookie-cutter style and less informative. In such instances, the language tends to follow predictable patterns. Conversely, when a judge engages in more thorough evidence gathering and nuanced deliberation, the judgment's content becomes more informative and less predictable.

To quantify this aspect of judgments, we use an entropy-based measure drawing from computational legal studies literature on text predictability. Following Friedrich, Luzzatto, and Ash (2020), we construct the Lempel-Ziv Compression measure to generate a global entropy metric, calculating the proportion of the compressed file relative to the original. Larger fractions indicate less compressible information, implying content beyond a cryptic, pseudocode-like style. Conversely, smaller fractions suggest higher text predictability and lower entropy, indicating a more standardized or templated approach to judgment writing. This metric thus quantifies the diversity and complexity of language in judgments.

We compute the global entropy measure for judgments at the court-area $\times$ year-quarter level. This measure is constructed by aggregating all relevant parts of judgment documents within the corresponding court-area and year-quarter. The raw text is then compressed, and the ratio of the compressed file to the original file is calculated. We follow this procedure to construct measures for three aspects of the judgments: factual descriptions, deliberations, and a combined measure incorporating both parts.

We estimate the change in writing style in response to the reform using the following specification:

$$\text{Entropy}_{jkt} = \beta_0 + \beta \text{intensity}_{jkt} + \omega_{j\times k} + \lambda_t + \varepsilon_{jkt} \tag{5}$$

where the outcome variable Entropy_{jkt} is the global entropy of legal documents for area j, court k, and year-quarter t. We include court-area and year-quarter fixed effects to be consistent with the previous specifications, but the judge fixed effect is not applicable. We also control for the total volume (equivalent to word count) of the original texts, as well as the set of regional-level control variables as discussed in section 4.1. The results are presented in Table 6. As indicated in column (1), the increased intensity of live broadcasting leads to an overall increase in entropy. Examining the factual and deliberation parts separately, the findings remain consistent, as shown in columns (2)

Table 6. *Writing Style, Reform and Litigation Outcomes*

	Outcome Variable:					
	Entropy for			Chances of Winning		
	All	Facts	Deliberation	c=All	c=Facts	c=Deliberation
	(1)	(2)	(3)	(4)	(5)	(6)
Intensity	0.00495***	0.00829*	0.00872*			
	(0.00146)	(0.00463)	(0.00475)			
Female × Entropy$_c$				0.0663***	0.0221***	0.0497***
				(0.0145)	(0.00760)	(0.00816)
Female				-0.0250***	-0.0147***	-0.0199***
				(0.00364)	(0.00223)	(0.00195)
Entropy$_c$				-0.0329***	0.00141	0.00103
				(0.0119)	(0.00429)	(0.00671)
Controls	Y	Y	Y	Y	Y	Y
Judge FE	N	N	N	Y	Y	Y
Court-Area FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
Mean of outcome	0.314	0.358	0.363	0.744	0.744	0.744
N	258,995	258,995	258,995	3,965,940	3,965,940	3,965,940
R^2	0.422	0.212	0.359	0.462	0.462	0.462

Notes: This table presents results for the analysis of global entropy of judgements. We examine the correlation between entropy and broadcast intensity at the court-area-year-quarter level. We use three different entropy measures: Fact (factual descriptions), Deliberation (descriptions of deliberation), and ALL (combined measure of both parts). The corresponding results are displayed in columns (1), (2), and (3) for All, Facts, and Deliberation, respectively. In addition, we investigate whether the entropy level at the court-area-year-quarter level can predict gender disparity in litigation outcomes. To do so, we estimate Equation (2) by substituting the intensity measure with the entropy measure. The results for the three types of entropy measures are presented in column (4) for All, column (5) for Fact, and column (6) for Deliberation, respectively. 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 case type. In all regressions, 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.

and (3), respectively.

Furthermore, if the entropy measure is a valid indicator of judges' effort and judicial quality, we expect it to predict gender disparities in judicial outcomes. To test this, we modify Equation (2) by replacing the intensity measure with our entropy measure. The results, presented in column (4) of Table 6, support our conjecture. They reveal a significant correlation: as judgments become less standardized and exhibit higher entropy (i.e., deviating more from a cookie-cutter style), the gender disadvantage experienced by female plaintiffs relative to male plaintiffs decreases. This pattern holds when we examine the factual descriptions (column 5) and deliberation sections (column 6) separately.

5.3. Do Judges of Both Genders Respond to the Reform?

The previous sections provide suggestive evidence that changes in judicial behavior in response to the reform could account for changes in judicial outcomes. We therefore

Table 7. *Responses across Judge Genders*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	Chief Judge Gender:			
	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
Female × Intensity	0.0314***	0.0424***	0.0404***	0.0588***
	(0.00398)	(0.00293)	(0.00492)	(0.00395)
Female	-0.00871***	-0.0126***	-0.0164***	-0.0213***
	(0.00106)	(0.000921)	(0.00139)	(0.00120)
Intensity	-0.0199***	-0.0167***	-0.0237***	-0.0154*
	(0.00512)	(0.00525)	(0.00610)	(0.00932)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.747	0.743	0.721	0.705
N	1,144,191	2,826,878	583,122	1,476,902
R^2	0.454	0.474	0.572	0.595

Notes: This table shows the differential responses of female judges and male judges to the reform. We subdivide the sample into separate samples for female and male judges and re-estimate the baseline regression, as shown in columns (1) and (2) with the main sample and columns (3) and (4) with the two-litigant subsample. Case-level control variables include the defendant gender, 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 case type. 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.

examine whether this response varies by judge gender or is common across both male and female judges.

To examine it, we extract the names of chief judges from the texts of judicial decisions and inferred their gender information. We divide our sample into two subsets: cases presided over by female judges and those by male judges. We then applied our main specification (Equation (2)) to each subsample separately. Table 7 presents these results. Column (1) shows findings for the female judges subsample, while column (2) displays results for male judges. We display estimation results in columns (3) and (4) for the two-litigant sample. Our analysis reveals significant positive coefficients on the interaction term for both male and female judge subsamples, indicating that judges of both genders respond to the surveillance technology.

Another observation is that the coefficient on the female plaintiff dummy is larger for the male judges sample, suggesting a more pronounced gender disparity in cases presided over by male judges than female judges. Correspondingly, the reform's impact is larger for male judges than female judges across both the all-litigant and two-litigant samples, suggesting that the reform has a stronger disciplinary effect on male

judges.[16]

5.4. Alternative Channels: Judges or Litigants?

Explorations in the previous sections suggest that the reform disciplined judges' judicial conduct, leading to changes in judicial outcomes. However, an alternative interpretation could be that *litigants* may have altered their behaviors in and out of the courtroom, resulting in the observed effects. Two potential avenues of litigant behavioral change warrant exploration: alterations in during-trial conduct and shifts in pre-trial case filing decisions.

We first explore the possibility that litigant behavioral changes in response to the reform lead to the observed effects. Litigants might adjust their conduct when aware of broadcasting, and judges might adapt their decisions based on these changes in litigant behavior rather than the presence of cameras themselves. To address this concern, we implement an empirical strategy to mitigate potential confounding effects from changes in litigant behavior. Our dataset includes appearance records of plaintiffs and defendants, allowing us to create a subsample of cases where no litigants were present in court. This subsample is particularly useful because changes in litigant behavior in response to the broadcasting mechanism would not be a factor in judicial decisions.[17]

We construct this subsample from our main sample and use it to estimate Equation (2). We also re-estimate the equation using the two-litigant subsample. The results are reported in columns (1) and (3) of Table 8, respectively. Both estimated coefficients remain positive and highly significant, yet are approximately half the size of those in the baseline case (see Table 2).

In a similar vein, one might argue that the live broadcasting mechanism encouraged lawyers to change their behavior in court, which in turn led to changes in judges' decisions. However, this hypothesis is even less plausible in our context, given the limited presence of lawyers in Chinese civil lawsuits: in 69 percent of civil cases, neither side has legal representation. We build a subsample of cases in which none of the

[16]This pattern likely reflects greater initial discrimination against female litigants by male judges, suggesting they had more room for improving gender equality in judicial conduct when the reform was implemented. The disciplinary effects of the reform also vary across regions, likely due to differences in local gender norms and cultural attitudes toward gender equality. We have conducted detailed analyses of both the gender-specific discrimination patterns and regional variations in reform effects; these results are available upon request. Our findings enrich the literature on judge gender's role in judicial decisions (Boyd, Epstein, and Martin 2010; Lim, Snyder Jr, and Stromberg 2015; Knepper 2018; Philippe and Ouss 2018; Chen and Ornaghi 2023).

[17]In China, most civil case litigants are not required to appear in court and do not have to inform the court in advance if they will attend. They are required to appear in person only for cases involving custody in the area of marriage, family and inheritance disputes. Therefore, judges do not know in advance whether litigants will appear.

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

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	Parties that do not appear:			
	Litigants	Lawyers	Litigants	Lawyers
	(1)	(2)	(3)	(4)
Female × Intensity	0.0193***	0.0461***	0.0337***	0.0529***
	(0.00316)	(0.00318)	(0.00442)	(0.00385)
Female	-0.00129	-0.0205***	-0.0128***	-0.0231***
	(0.000960)	(0.00102)	(0.00135)	(0.00125)
Intensity	-0.0178***	-0.0133**	-0.0297***	-0.00736
	(0.00502)	(0.00623)	(0.00681)	(0.0101)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.748	0.771	0.737	0.738
N	1,935,756	2,044,952	782,414	1,251,355
R^2	0.415	0.554	0.550	0.651

Notes: This table presents DID estimation results with a subsample that none of litigant or lawyer appears in court. Both estimated coefficients remain significant with similar magnitude to those in the baseline case (see Table 2). Case-level control variables include the defendant gender, 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 case type. 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.

plaintiffs or defendants hire lawyers from the main sample. We estimate Equation (2) using this subsample and report the respective results in columns (2) and (4) of Table 8. These estimated coefficients are still significant and close to those estimated using the main sample (see Table 2).

Next, we examine whether the reform influenced plaintiffs' case filing behaviors. The introduction of the broadcasting mechanism may have discouraged litigants with weaker cases from filing claims. If female plaintiffs, on average, had a lower threshold for case strength when filing – meaning they tended to file weaker cases compared to male plaintiff – this could explain their lower winning chances before the reform. Consequently, the reform might mechanically lead to an increase in female plaintiffs' chances of winning, as the post-reform pool of female plaintiffs would predominantly consist of stronger cases.

While our Bartik IV approach in section 4.4 could mitigate this concern, we provide two additional tests. First, we examine whether the reform affected the proportion of female plaintiffs. We aggregate the data into the court-area-year-quarter level, and assess whether the proportion of female plaintiff responds to the broadcast intensity. Results in Table A9 show no significant changes in gender composition for all cases

(column (1)) or two-litigant cases (column (2)).

Second, we conduct a placebo test to examine changes in case strength. While the strength of a case is not directly observable, it can be correlated with its characteristics. We construct predicted winning chances for each case based on these pre-determined attributes and test whether these predicted chances of winning changed in response to the reform. Specifically, we estimate the DID specification using the predicted chances of winning as the dependent variable. Reassuringly, the estimated coefficients on the interaction terms are not significant across all specifications (see columns (3) and (4) of Table A9). In other words, neither the gender composition nor the case strength likely drive our results.

5.5. Effectiveness and Limitations of the Open Justice Reform

The evidence presented in sections 5.1 and 5.2 indicates that judges modify their judicial behaviors, i.e. exhibiting greater impartiality during deliberation and exerting more effort, when subjected to additional scrutiny mechanisms. These findings provide empirical support for the mechanism underlying the reform's impact, substantiating the link between increased public scrutiny and changes in judicial behavior.

To further interpret and contextualize these results, we can draw upon a set of behavioral theories from the existing literature. First, increased judicial visibility can heighten "image concerns" (Ariely, Bracha, and Meier 2009) and incentivize judges to act more professionally. When judges know they are being observed by stakeholders, legal professionals, and the public, they may adhere more strictly to court rules, scrutinize claims more carefully, and draft legal documents with greater deliberation. This enhanced professionalism may, in turn, result in more impartial decisions.

Second, higher judicial visibility likely raises the cost of discrimination when judges act on biased preferences. Parsons et al. (2011) documented this mechanism in baseball games, where umpires' racial bias decreased under camera monitoring. Similarly, Ferrazares (2023) shows that body-worn cameras can reduce complaints from black civilians against white officers by regulating officer behavior. These findings suggest that when under camera surveillance, decision-makers - whether judges, umpires, or officers - likely become more vigilant about their behavior and consciously correct their biases against minorities.

In addition, factors specific to China's judicial system may amplify these effects. The hierarchical structure of the Chinese judiciary, comprising 12 levels, requires judges to undergo consistent performance assessments for career advancement and higher incomes. The average age of Chinese judges is 39 (compared to 51 for US district judges), making frequent performance evaluations an effective tool for incentivizing

judges. Consequently, the availability of surveillance records to hold judges accountable provides strong incentives for professional and diligent behavior.

While our results demonstrate positive effects of the reform, courtroom surveillance is not a panacea. Our findings and prior studies suggest two critical conditions for effectiveness: (1) the discriminated group's identity must be readily observable, and (2) the discrimination must be preference-based. This approach may be ineffective when litigants share the same identity but one has an undisclosed connection to the judge, potentially leading to intentional bias. Thus, while increased transparency can mitigate certain forms of discrimination, it cannot address all judicial biases.

6. Concluding Remarks

Our findings provide evidence on how information technology can play a role in monitoring court proceedings in general and how increased judicial visibility can reduce gender disparity in civil litigation in particular. These insights contribute to both the literature on judicial bias and the broader research on the relationship between technology and public affairs.

There are two promising avenues to further our discussion. It is crucial to acknowledge that this reform may generate hidden social costs extending beyond the courtroom. For instance, Adams, Adams-Prassl, and Adams-Prassl (2022) raise concerns about privacy abuses resulting from publishing judicial decisions online in the UK. This privacy concern can be further exacerbated when litigants' images and videos are publicly accessible at no cost. To address this, future reforms should consider balancing transparency with privacy protection.

Finally, our study focuses on the impacts of live broadcasting on judicial outcomes around the time of the reform's introduction. However, the long-term influence on judicial behavior remains an intriguing area for research. Prolonged exposure to heightened scrutiny may permanently reshape judicial conduct, fostering enduring professionalism even without ongoing monitoring. These long-term effects warrant further investigation to understand the lasting impact of increased judicial transparency.

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.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy 128*(6), 2188–2244.

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 (2024). Violence against women at work. *The Quarterly Journal of Economics 139*(2), 937–991.

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

Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.

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

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–555.

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

Bar, T. and A. Zussman (2012). Partisan grading. *American Economic Journal: Applied Economics 4*(1), 30–48.

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.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020). Incarceration, recidivism, and employment. *Journal of Political Economy 128*(4), 1269–1324.

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.

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

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

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.

Butcher, K. F., K. H. Park, and A. M. Piehl (2017). Comparing apples to oranges: Differences in women's and men's incarceration and sentencing outcomes. *Journal of Labor Economics 35*(S1), S201–S234.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

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.

Chen, D. L., J. Graham, M. Ramos Maqueda, and S. Singh (2022). Do judges favor their own ethnicity and gender?

Chen, D. L. and A. Ornaghi (2023). Gender attitudes in the judiciary: Evidence from US circuit courts. *American Economic Journal: Applied Economics*.

Cook, L. D., M. E. Jones, T. D. Logan, and D. Rosé (2023). The evolution of access to public accommodations in the United States. *The Quarterly Journal of Economics 138*(1), 37–102.

Cubukcu, S., N. M. Sahin, E. Tekin, and V. Topalli (2021). Body-worn cameras and adjudication of citizen complaints of police misconduct. Technical report, National Bureau of Economic Research.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

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

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

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

Ferrazares, T. (2023). Monitoring police with body-worn cameras: Evidence from Chicago. *Journal of Urban Economics*, 103539.

Fouka, V., S. Mazumder, and M. Tabellini (2022). From immigrants to Americans: Race and assimilation during the Great Migration. *The Review of Economic Studies 89*(2), 811–842.

Friedrich, R., M. Luzzatto, and E. Ash (2020). Entropy in legal language. In *NLLP 2020 Natural Legal Language Processing Workshop 2020. Proceedings of the Natural Legal Language Processing Workshop 2020 co-located with the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2020)*, Volume 2645, pp. 25–30. CEUR-WS.

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

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

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.

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

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–320.

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.

Kosack, S. and A. Fung (2014). Does transparency improve governance? *Annual Review of Political Science 17*(1), 65–87.

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

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., 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), 103–135.

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

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics 94*(1), 20–36.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review 101*(4), 1410–1435.

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.

Peerenboom, R. (2002). *China's long march toward rule of law*. Cambridge University Press.

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.

Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth 9*, 131–165.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.

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

Sorensen, T., S. Sarnikar, and R. L. Oaxaca (2012). Race and gender differences under federal sentencing guidelines. *American Economic Review: Papers & Proceedings 102*(3), 256–260.

Starr, S. B. (2015). Estimating gender disparities in federal criminal cases. *American Law and Economics Review 17*(1), 127–159.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

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.

Online Appendix

(Not intended for publication)

A. Supplementary Figures and Tables

A.1. Figures

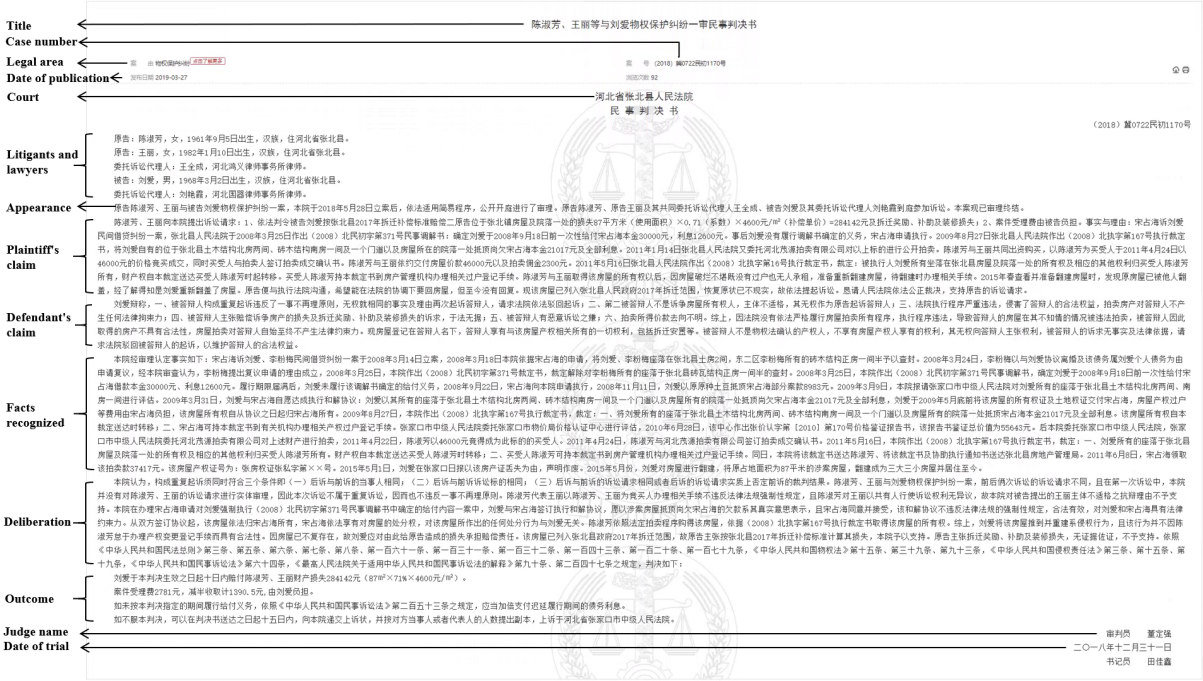


Figure A1. *An example of case information extracted from a legal document. This is a sample of civil judgement that includes five main sections: the basic information, claims of litigants, facts recognized by the court, legal principles applied and outcomes of the litigation.*

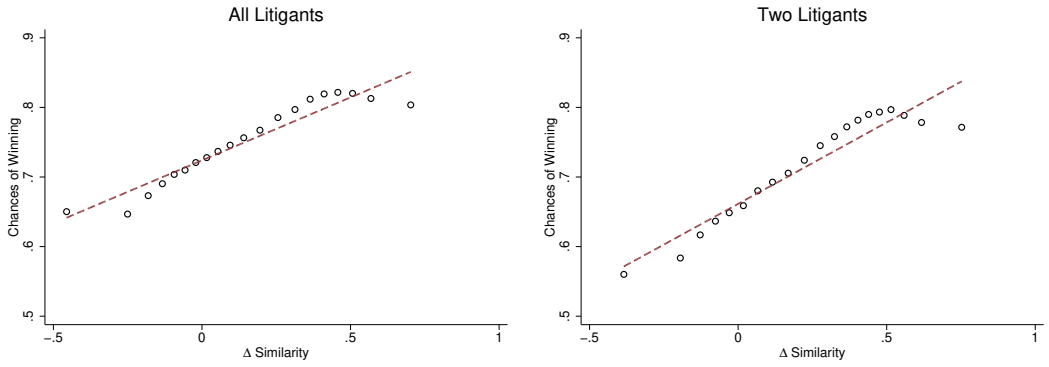


Figure A2. *Bin-scatter plot for the correlation between a plaintiff's likelihood of winning and Δ similarity. Δ similarity gauges the disparity between the similarity of the plaintiff's statements with those acknowledged by the judge and the similarity of the defendant's statements with the judge's acknowledgments. Controls for case characteristics and court fixed effects are included. In both the all-litigant sample and the subsample of two litigants, we observe a robust and positive correlation between plaintiff winning chances and Δ similarity.*

A.2. Tables

Table A1. *Summary Statistics: Case Level Variables*

	Main sample: All Litigants	Sub-sample: Two Litigants
	mean/sd	mean/sd
	(1)	(2)
Chances of winning for plaintiff	0.74	0.71
	(0.39)	(0.42)
Number of plaintiffs	1.13	1.00
	(0.71)	(0.00)
Number of defendants	1.72	1.00
	(1.24)	(0.00)
Plaintiff's appearance (1=Yes)	0.49	0.60
	(0.50)	(0.49)
Defendant's appearance (1=Yes)	0.08	0.09
	(0.27)	(0.29)
Number of lawyers for plaintiff	0.48	0.41
	(0.67)	(0.62)
Number of lawyers for defendant	0.28	0.15
	(0.63)	(0.41)
Plaintiff's gender (1=female)	0.31	0.32
	(0.46)	(0.47)
Defendant's gender (1=female)	0.19	0.20
	(0.39)	(0.40)
Chief judge female (1=Yes)	0.29	0.28
	(0.45)	(0.45)
Number of judges	1.43	1.38
	(0.74)	(0.70)
Instance (1=first instance)	0.89	0.91
	(0.31)	(0.29)
Legal cost in log	6.45	5.99
	(2.07)	(1.96)
Observation	4,516,547	2,333,715

Notes: This paper summarizes the statistical property of the case characteristics of our sample. Columns (1) and (2) report the means and standard deviations of a number of case-level variables in the main sample and two-litigant subsample, respectively.

Table A2. *Testing for Random Assignment of Cases to Broadcast*

	Outcome Variable	
	Chances of Winning	Whether Broadcast Online (Yes=1)
	(1)	(2)
Plaintiff's gender (1=female)	-0.0204***	-0.00115***
	(0.00123)	(0.000311)
Defendant's gender (1=female)	-0.0286***	-0.00000698
	(0.000992)	(0.000338)
Number of plaintiffs	-0.00864***	-0.000318
	(0.000641)	(0.000223)
Number of defendants	0.00564***	-0.00162***
	(0.000458)	(0.000169)
Plaintiff's appearance (1=Yes)	-0.00869***	0.00299***
	(0.00108)	(0.000542)
Defendant's appearance (1=Yes)	-0.0661***	0.00162**
	(0.00136)	(0.000653)
Number of lawyers for plaintiff	-0.0102***	0.000467
	(0.000657)	(0.000323)
Number of lawyers for defendant	-0.0464***	-0.00142***
	(0.00132)	(0.000307)
Instance (1=first instance)	0.146***	0.0114***
	(0.0195)	(0.00395)
Number of Judges	-0.0103***	-0.00926***
	(0.00116)	(0.000892)
Legal cost in log	0.0195***	0.00132***
	(0.000511)	(0.000176)
F-statistics for joint test	475.72	30.99
p-value	0.000	0.000

Notes: This table presents the test for random assignment of civil cases to be broadcast online, with cases adjudicated during the full period from 2014 to 2018. All specifications include the court-area and year-quarter fixed effects. Reported F-statistic refers to a joint test of the null hypothesis for all variables. Standard errors in parentheses clustered at court level; * p<0.1, ** p<0.05, *** p<0.01.

Table A3. *Specifications with Time Trend and Alternative Fixed Effects*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	Court Time Trend	Court×Time FE	Court Time Trend	Court×Time FE
	(2)	(1)	(4)	(3)
Female × Intensity	0.0393***	0.0387***	0.0538***	0.0524***
	(0.00253)	(0.00250)	(0.00334)	(0.00334)
Female	-0.0116***	-0.0114***	-0.0200***	-0.0195***
	(0.000838)	(0.000831)	(0.00106)	(0.00105)
Intensity	-0.0228***	-0.0331***	-0.0281***	-0.0415***
	(0.00368)	(0.00575)	(0.00446)	(0.00894)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Court×Year-Quarter FE	N	Y	N	Y
Court×Time Trend	Y	N	Y	N
Mean of outcome	0.744	0.744	0.709	0.709
N	3,974,316	3,970,468	2,063,379	2,059,196
R^2	0.466	0.476	0.587	0.600

Notes: To rule out confounding factors such as time-varying shocks at the court level, we estimate two additional specifications and add court-specific time trend and court × year-quarter fixed effects 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, 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 case type. 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 A4. *Robustness*

Outcome Variable	Chances of Winning	Win	Chances of Winning	
	Overall Inten.	Intensity	Drop Cells	Supplement
All Litigants	(1)	(2)	(3)	(4)
Female × Overall Intensity	0.0459***			
	(0.00295)			
Female × Intensity		0.0384***	0.0368***	0.0390***
		(0.00290)	(0.00279)	(0.00254)
Female	-0.0143***	-0.0123***	-0.0112***	-0.0115***
	(0.000874)	(0.000897)	(0.000939)	(0.000842)
Overall Intensity	-0.0141**			
	(0.00551)			
Intensity		-0.0152**	-0.0152***	-0.0175***
		(0.00690)	(0.00530)	(0.00436)
Mean of outcome	0.744	0.744	0.774	0.750
N	3,976,383	3,974,316	3,345,727	4,031,953
R^2	0.444	0.433	0.477	0.462
Two Litigants	(5)	(6)	(7)	(8)
Female × Overall Intensity	0.0609***			
	(0.00378)			
Female × Intensity		0.0530***	0.0433***	0.0530***
		(0.00381)	(0.00344)	(0.00336)
Female	-0.0226***	-0.0208***	-0.0180***	-0.0197***
	(0.00114)	(0.00116)	(0.00115)	(0.00107)
Overall Intensity	-0.0144*			
	(0.00825)			
Intensity		-0.0127	-0.0116	-0.0176**
		(0.0126)	(0.00878)	(0.00719)
Mean of outcome	0.709	0.701	0.741	0.722
N	2,065,978	2,063,379	1,744,528	2,090,624
R^2	0.564	0.551	0.609	0.583
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court FE	Y	N	N	N
Area FE	Y	N	N	N
Court-Area FE	N	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Notes: This table presents the robustness check using alternative variables and sample. In columns (1) and (5), we use the court level intensity instead of court-area level intensity. In columns (2) and (6), we use the dummy variable Win as the dependent variable. In columns (3) and (7), to address the concern that insufficient cases in a court-area-quarter cell may lead to inaccurate measurement of broadcast intensity, we exclude cells with fewer than 20 cases from our analysis. In columns (4) and (8), we supplement our sample by using a name-based gender deduction method to interpolate missing gender information. Case-level control variables include the defendant gender, 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 case type. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors are in the parentheses beneath coefficients, clustered at the court level; * p<0.1, ** p<0.05, *** p<0.01.

Table A5. *Placebo Test: Early Outcome Against Later Predicted Intensity*

	Outcome Variable: Chances of Winning					
	All Litigants			Sub-sample: Two Litigants		
	2014	2015	2016Q1∼2	2014	2015	2016Q1∼2
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Predicted Intensity	0.0301	-0.00380	0.0785	0.00147	0.0307	0.130
	(0.0392)	(0.0266)	(0.113)	(0.0612)	(0.0371)	(0.127)
Female	-0.0395***	-0.0386***	-0.0185***	-0.0605***	-0.0584***	-0.0247***
	(0.00218)	(0.00192)	(0.00165)	(0.00337)	(0.00282)	(0.00241)
Predicted Intensity	-0.000946	-0.0114	-0.0327	-0.0110	-0.0351**	-0.0655
	(0.0142)	(0.0132)	(0.0389)	(0.0199)	(0.0165)	(0.0445)
Controls	Y	Y	Y	Y	Y	Y
Judge FE	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	338,347	447,424	256,635	199,479	253,251	136,756
R^2	0.615	0.590	0.656	0.685	0.669	0.745

Note: This table presents placebo test for pre-trends of our Bartik instrument. Using subsamples in 2014, 2015 and the first 2 quarters of 2016, we regress chances of winning on female dummy and predicted broadcasting intensity in 2018 (according to court and year-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 on the interaction terms are insignificant. Case-level control variables include the defendant gender, 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 case type. 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 A6. *Robustness Checks by Adding 2014 Regional Control Variables*

	Outcome Variable: Chances of Winning			
All Litigants	(1)	(2)	(3)	(4)
Female × Intensity	0.0696***	0.0693***	0.0694***	0.0697***
	(0.00550)	(0.00549)	(0.00549)	(0.00550)
Female	-0.0136***	-0.0136***	-0.0136***	-0.0136***
	(0.000875)	(0.000874)	(0.000874)	(0.000875)
Intensity	0.0265	0.0326	0.0368	0.0325
	(0.0357)	(0.0394)	(0.0390)	(0.0394)
Mean of outcome	0.744	0.744	0.744	0.744
N	4,253,908	4,256,360	4,256,363	4,253,861
R^2	0.017	0.017	0.017	0.017
Two Litigants	(5)	(6)	(7)	(8)
Female × Intensity	0.0853***	0.0851***	0.0851***	0.0854***
	(0.00643)	(0.00642)	(0.00642)	(0.00644)
Female	-0.0220***	-0.0220***	-0.0220***	-0.0220***
	(0.00112)	(0.00112)	(0.00112)	(0.00112)
Intensity	0.0433	0.0572	0.0594	0.0587
	(0.0519)	(0.0567)	(0.0568)	(0.0568)
Mean of outcome	0.709	0.709	0.709	0.709
N	2,209,214	2,210,432	2,210,429	2,209,144
R^2	0.025	0.025	0.025	0.025
Internet Penetration × Year-Quarter FE	Y	N	N	Y
GDP per capita × Year-Quarter FE	N	Y	N	Y
Population × Year-Quarter FE	N	N	Y	Y
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Notes: This table presents the robustness check for Bartik IV regression by adding interactions of 2014 regional controls and time fixed effects. We control for the interaction terms of internet penetration rate, GDP per capita and population in turn, as well as all three characteristics with year-quarter fixed effects. The main effects remain unchanged in both the full sample and the two-litigants subsample. Case-level control variables include the defendant gender, 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 case type. Prefecture-level control variables include internet penetration rate, GDP per capita (log) and population (log). Standard errors are in the parentheses beneath coefficients, clustered at the court level; * p<0.1, ** p<0.05, *** p<0.01.

Table A7. *Initial Share of Legal Areas*

	Outcome Variable: Initial Share of Legal Area			
	Contract	Tort	Marriage	Property
	(1)	(2)	(3)	(4)
Internet Penetration	-0.0134	0.00895	0.0183	0.00410
	(0.0313)	(0.0202)	(0.0285)	(0.00949)
N	2,845	2,348	2,737	2,720
GDP per capita	0.0143	0.00340	0.0130	-0.00292
	(0.0105)	(0.00689)	(0.00859)	(0.00241)
N	2,852	2,351	2,755	2,730
Population	0.0117	0.00127	0.0121	-0.00141
	(0.00858)	(0.00644)	(0.00763)	(0.00205)
N	2,852	2,351	2,755	2,730

Notes: This table presents the regression results of the 2014 share of the top four legal areas (according to their Rotemberg weights) for each court, against three key prefecture-level characteristics including internet penetration rate, GDP per capita, and population in 2014. Standard errors are clustered at the province level and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A8. *Estimations with IV Constructed from Other Courts and Other Areas*

	Alternative IV 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			0.0737***			0.0911***
			(0.00513)			(0.00601)
Female	0.0000188	0.0207***	-0.0132***	0.0000862	0.0204***	-0.0231***
	(0.000356)	(0.00188)	(0.000906)	(0.000463)	(0.00191)	(0.00114)
Intensity			-0.0147**			-0.0165**
			(0.00639)			(0.00717)
Predicted Intensity	0.830***	-0.0352***		0.838***	-0.0375***	
	(0.0326)	(0.00488)		(0.0328)	(0.00491)	
Female × Predicted Intensity	-0.0167***	0.937***		-0.0230***	0.954***	
	(0.00371)	(0.0285)		(0.00439)	(0.0285)	
Controls	Y	Y	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y	Y	Y
Court FE	Y	Y	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y	Y	Y
K-P LM Value			239.072			219.832
K-P Wald Value			323.305			323.846
N	3,969,909	3,969,909	3,969,909	2,060,948	2,060,948	2,060,948
R^2	0.769	0.639	0.017	0.791	0.656	0.026

Notes: This table presents the first stage and second stage results for the IV regressions. We construct another instrument from the weighted average of the broadcast intensity among other issue areas in other courts within same prefecture. 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). The estimates shown in columns (3) and (6) remain positive and highly significant, similar to those obtained in our Bartik IV estimation. Case-level control variables include the defendant gender, 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 case type. 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 A9. *Testing for Changes in Plaintiff Composition and Case Filing*

	Outcome Variable:			
	Proportion of Female Plaintiffs		Predicted Chances of Winning	
	All Litigants	Two Litigants	All Litigants	Two Litigants
	(1)	(2)	(3)	(4)
Female × Intensity			-0.0000691	-0.0000474
			(0.0000621)	(0.0000430)
Female			-0.0803***	-0.0803***
			(0.0000155)	(0.0000174)
Intensity	0.00141	-0.0142	0.000163	0.000408**
	(0.00881)	(0.0120)	(0.000184)	(0.000194)
Control	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Court×Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.355	0.363	0.649	0.628
N	225,022	172,328	3,987,661	2,075,461
R^2	0.463	0.517	0.988	0.990

Notes: This table presents the tests for changes in plaintiffs' composition. In columns (1) and (2), we examine whether more women initiate litigation in response to the reform by regressing the proportion of female plaintiffs at the court-area-year-quarter level on the broadcasting intensity, controlling for court-area and court × year-quarter fixed effects to absorb the court-specific shocks (such as other judicial reforms during the study period). In columns (3) and (4), we conduct a placebo test by examining whether changes in judicial outcomes can be attributed to changes in case characteristics. In all specifications, we include prefecture-level control variables including 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. Construction of Case Type

In our baseline regressions, we control for a text-based indicator called case type to capture nuanced case information. In this appendix, we describe our construction approach. Our primary approach employs a topic modeling algorithm–an unsupervised learning technique–to automatically classify cases into refined categories, thereby eliminating subjective assessment through automated processing.

The rationale for our classification stems from judicial practice, where judges cite specific combinations of law articles to substantiate their decisions. Cases can thus be considered similar when they draw upon similar legal provisions. Following this logic, we group cases based on the similarity of applied laws.

For implementation, we employ the Top2vec model, an unsupervised topic modeling algorithm developed by Angelov (2020). The algorithm maps documents and words into a shared semantic vector space, where semantically similar documents form clusters. Each cluster represents a distinct topic, enabling classification based on semantic relationships.

In our context, each document represents the set of laws cited in a judicial decision. To classify cases based on their legal frameworks, we retain only the names of cited laws while excluding specific clause numbers. For instance, if a decision cites Article 10 of the Forestry Law and Article 20 of the Economic Law, the document is coded as "Forestry Law, Economic Law" for classification purposes.

The Top2vec algorithm maps these legal citations into a unified vector space and identifies dense regions of similar documents. These regions represent shared legal frameworks among cases, capturing thematic patterns in judicial citations. The algorithm organizes all documents into 50 distinct clusters. We validate this classification by comparing it with a human-assisted dictionary approach, finding consistent results across both methods.

C. Reliability of Two Way Fixed Effect Model

The reliability of the TWFE model has been questioned by a large body of recent literature, which demonstrates that it only provides consistent estimates under strong assumptions of homogeneous treatment effects (De Chaisemartin and d'Haultfoeuille 2020; Goodman-Bacon 2021; Borusyak, Jaravel, and Spiess 2021; Callaway, Goodman-Bacon, and Sant'Anna 2021; Callaway and Sant'Anna 2021; Sun and Abraham 2021; Roth et al. 2023). The TWFE model generates a weighted average of all 2x2 difference-in-differences comparisons across treatment timing groups. While consistent for homogeneous treatment effects, the model can produce negative weights and biased estimates under treatment effect heterogeneity.

In the context of continuous treatments (dosing DID), Callaway, Goodman-Bacon, and Sant'Anna (2021) show that strengthening the parallel trends assumption–a modification not applicable to binary treatments–can ensure positive weights in the decomposition of TWFE estimator. However, they identify an additional "selection bias" in TWFE model for continuous treatments. Specifically, units with weaker treatment serve as counterfactuals for those with stronger treatment. The strong parallel trends assumption requires that stronger treatment units would experience the same treatment effects as weaker ones if assigned the same treatment.

In applied works, Cook et al. (2023) suggest that this assumption is likely to hold if the treatment strength is uncorrelated with observable factors. In our context, we implement this strategy by using propensity score matching (PSM) to construct a sample where treatment intensity is orthogonal to observable court characteristics. We classify courts into high-dose and low-dose groups based on whether their broadcasting intensity falls above or below the median.

Our matching incorporates both time-varying socioeconomic factors and time-invariant local characteristics. Time-varying indicators include population, GDP per capita, and internet penetration, which reflect local economic development and constrain reform implementation through their effects on public expenditure. We also include time-invariant geographic and environmental characteristics that shape long-term economic development and social norms. Economic factors comprise proximity to ports and rivers (in logarithms), location within the Yangtze Economic Zone, and area size, which influence trade opportunities, economic integration, and modernization. Factors influencing social norms include altitude, slope, waviness, and elevation. These topographical features historically determined early agricultural practices and household division of labor, potentially exerting lasting impacts on gender roles and attitudes (Nunn and Puga 2012 and Rodrik, Subramanian, and Trebbi 2004).

Using 1:1 nearest-neighbor matching without replacement, we construct a balanced sample of 2,124 courts (1,062 high-dose and 1,062 low-dose), with comparable observable characteristics across groups. Table A10 presents the balancing tests.

Re-estimating our baseline DID specification with this matched sample yields results similar in magnitude and significance to our baseline findings (Table A11), despite the reduced sample size. This consistency suggests that selection issues in the dosing DID design do not substantially affect our results.

Table A10. *Balancing Test for high-dose and low-dose adopters*

Variables	Before Matching		After Matching	
	Diff.	t-statistics	Diff.	t-statistics
Population (log)	-0.0518*	(-2.40)	-0.00358	(-0.14)
GDP per capita (log)	-0.0792***	(-4.42)	0.0127	(0.65)
Internet Penetration	0.00811	(1.72)	-0.00341	(-0.65)
Area (log)	0.000638	(0.02)	-0.000982	(-0.02)
Altitude	158.7***	(5.55)	9.125	(0.30)
Slope	1.815***	(8.29)	-0.278	(-1.14)
Waviness	166.0***	(4.86)	-0.146	(-0.00)
Elevation	0.570**	(3.17)	-0.0482	(-0.24)
Distance to port (log)	-0.0590	(-1.09)	0.00211	(0.03)
Distance to river (log)	-0.239***	(-4.33)	0.0827	(1.35)
Distance to coast (log)	0.123***	(6.82)	-0.0207	(-0.99)
Yangtze Economic Zone	0.302***	(7.09)	-0.0300	(-0.67)

Notes: This table summarizes the balancing test between high-dose courts and low-dose courts before and after the matching procedure. We find no significant difference among all control variables between the two groups after matching. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A11. *Matched Difference-in-Differences Estimation: High-dose and Low-dose Adopters*

	Outcome Variable: Chances of Winning			
	All Litigants		Sub-sample: Two Litigants	
	All Areas	Excl Marriage	All Areas	Excl Marriage
	(1)	(2)	(3)	(4)
Female × Intensity	0.0426***	0.0264***	0.0595***	0.0340***
	(0.00318)	(0.00299)	(0.00429)	(0.00384)
Female	-0.0131***	-0.00500***	-0.0223***	-0.0103***
	(0.00108)	(0.00104)	(0.00136)	(0.00130)
Intensity	-0.0257***	-0.0186***	-0.0289***	-0.0201***
	(0.00424)	(0.00435)	(0.00486)	(0.00506)
Controls	Y	Y	Y	Y
Judge FE	Y	Y	Y	Y
Court-Area FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y
Mean of outcome	0.745	0.791	0.703	0.788
N	2,611,040	2,420,525	1,354,270	1,180,477
R^2	0.469	0.381	0.595	0.499

Notes: This table presents the main results for the DID specification, using the sample cases from matched courts with comparable average broadcast intensity between 2014 and 2018. Columns (1) and (2) report the results with main sample and columns (3) and (4) report the results with the two-litigant subsample. In Columns (2) and (4), we exclude cases in the area of marriage and family disputes. Case-level control variables include the defendant gender, 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 case type. 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.