Fewer Opportunities, Larger Gender Promotion Gap?

Alex Eble and Feng Hu*

March 2025

Abstract

This paper studies how the reduction of opportunities for professional advancement widens the gender promotion gap. We leverage a personnel reform that significantly cut the number of available positions at higher ranks in China's court system. We find that the promotion gap favoring men widens after the reform, despite women performing at least as well as their male counterparts. We provide suggestive evidence that this widened disadvantage for women stems from male preference in promotion evaluations. A large female presence in court leadership teams mitigates the gender promotion gap. Our findings highlight how fewer chances along the ladder, coupled with male preference, contribute to the "leaky pipeline" of talented women.

*Eble: Columbia University and National Bureau of Economic Research (NBER). Email: eble@tc.columbia.edu, Hu: University of Science and Technology Beijing. Email: feng3hu@gmail.com. Eble and Hu are co-first authors and contributed equally; authors are listed alphabetically by surname. Keywords: gender, labor market, inequality, promotion, leadership, personnel, public service, judicial reform. JEL codes: J16, L14, J62, J71 I23, O30

1 Introduction

In most white-collar work, women advance up the career ladder at lower rates than men (Alper, 1993; Ginther and Kahn, 2004; Bertrand, Goldin and Katz, 2010; Ceci and Williams, 2011; Buckles, 2019). In this study, we ask whether women's chances of success diminish relative to men's when there are fewer higher-ranked positions available. To do so, we leverage a personnel reform that abruptly reduced the number of available higher-ranked positions in the Chinese judiciary.

We find that a reduction in higher-ranked judge positions widens the gender promotion gap. We document that this occurs despite the fact that women's performance on a series of objective measures, putatively used to decide on promotion, is as good or better than that of their male colleagues. We provide suggestive evidence that women's disadvantages in promotion stem largely from male preference exerted during the evaluation process. Finally, we show that when there is a greater proportion of females among the court leadership team, the negative effects of the policy we measure are much smaller in magnitude. Surprisingly, however, the gender of the head of a given court, the court president, has far less of an impact.

The policy we study was announced in 2014 and it aimed to optimize the allocation of judicial resources in China's court system. Issued by China's Supreme People's Court, it mandated that judges comprise no more than 39% of formal judicial personnel in local courts. The reform was rolled out across provinces in 2015 and 2016, and the quotas it put in place have remained largely unchanged ever since. The main impact of the policy was to reduce the total number of judges employed in courts across China by 40% from their pre-policy levels. Furthermore, in almost all cases, judges employed prior to the policy had to reapply for their jobs, a process which involved undergoing the normal set of civil service exams and evaluations. This policy provides the variation in opportunities for career advancement that we use in our empirical analysis.

Our main dataset contains from administrative data on 85 million judgment files, linked to 335,000 judges covering the period from 2013 to 2021. Linking each document to personnel records, we can track officially recognized measures of work performance as well as promotion

records for these individuals over time. We also collect personnel data covering judges hired after the quota went into place. Our final dataset contains information on the leadership teams of over 1,800 courts, comprising roughly half of all courts in China during the reform period. This provides us with information on the gender of all members of each court's leadership team over this period.

We first show that the reduction in the number of senior positions decreases women's likelihood of promotion by six percent, compared to similarly aged and qualified men. Our results remain robust when we restrict our analysis to only judges first hired prior to the reform, and separately when we use only judges first hired after the quota went into place. We also find that the gender promotion gap intensifies when there is a larger number of candidates competing for a given position.

The widened gender promotion gap we find following the reform is unlikely to be explained by gender differences in work performance or gender differentials in professional exit from the court system. We show that women participate in more case verdicts, and their verdicts are of similar or better quality than those of male judges. We also show that women are no more likely than men to exit the court system. Together, this evidence indicates that the negative impact of the reform on women's career advancement is likely to stem from male preference in the evaluation process. This process includes two parts – a competence assessment and a subjective evaluation component – which, we argue, both confer additional favor to the selection of male candidates for these career advancement opportunities.

We then investigate whether the gender composition of a given court's leadership team changes the magnitude or sign of the effects we measure. Each court has a team of several (more than three and less than ten) individuals, comprising a president, vice president, and members of the court's CPC (Communist Party of China) committee. We find that when a greater proportion of this team is female, the negative impacts of the policy lessen in magnitude. Indeed, the negative impact almost disappears when women comprise the majority of court leadership teams. In contrast, we find no significant differences in our estimates between courts with a female, as opposed to a male president.

We push out on the frontier of three active literatures in labor and applied microeconomics. First, our paper advances understanding of the reasons why females are less likely than males to reach senior white-collar roles (Alper, 1993; Ginther and Kahn, 2004; Bertrand, Goldin and Katz, 2010; Ceci and Williams, 2011; Buckles, 2019). We advance on this literature by focusing on a novel mechanism - greater scarcity in opportunities for professional advancement, for which both men and women must compete. We estimate the empirical relevance and magnitude of this mechanism. Furthermore, our findings highlight the role of male preference in widening this gap in this context. In this regard, we also contribute to the literature on how deepening resource scarcity leads to rationing based on group identity in settings with underlying discrimination (Akbarpour et al., 2024; Singh and Venkataramani, 2024).

Second, our paper adds to the recent body of work studying the ways that female managers do or do not address gender inequality in the workplace (Matsa and Miller, 2011; Joshi, Son and Roh, 2015; Kunze and Miller, 2017; Bertrand et al., 2019; Lawson et al., 2022; Cullen and Perez-Truglia, 2023). We push out on this literature in several ways. First, we show that female leadership can help mitigate the negative effects of the mechanism we study – a global contraction in opportunities for advancement. This diverges from prio work, which has primarily focused on how female leadership directly promotes female subordinates. Second, we show that having a large proportion of women in a given leadership team – as opposed to simply having a female head – holds potential to address inequality by gender in the likelihood of promotion. This finding contributes to the ongoing debate on the effectiveness of female leadership in supporting women's career advancement.

Third, our paper contributes to the literature on the consequences of management personnel policies (Biasi and Sarsons, 2022; Hospido, Laeven and Lamo, 2022; Hoffman and Stanton, 2024). We advance this literature in the following ways. First, we focus on the unintended consequences that a management personnel policy with no explicit mention of gender has on gender inequality in the labor market. Our results show that this ostensibly gender-neutral labor force adjustment policy widens the gender gap in career development. This sheds evidence on the economic and gender implications of reductions in force and "right-sizing" efforts common today in both government

and industry. Our findings also highlight the mechanisms through which such policy does so. Second, we focus on civil court judges, a group of professionals whose personnel management has clear and vital links to fairness and integrity in broader society, and which, to date, has received little attention in the literature.

The rest of the paper proceeds as follows: Section 2 provides background on China's judge quota reform and Section 3 describes the data. Section 4 examines how the reform-induced reduction of promotion opportunities affects the gender promotion gap. Section 5 explores why women are disadvantaged in promotions as overall opportunities decrease. Section 6 studies whether female leadership helps mitigate the gender promotion gap. The final section concludes.

2 The Judge Quota Reform in China

2.1 China's court system reform

While China's court system differs from its Western counterparts in many aspects, it has looked to western models of courts in recent years and has substantial autonomy in most local and torts decisions (Sun and Fu, 2022). Relatedly, in the past few decades, China has implemented various reforms aimed at enhancing judges' competence and the professionalism of its court system (Liebman, 2017; Sun and Fu, 2022). These reforms require judges to adjudicate cases based on the law, free from external interference in court decision-making.

To further optimize the allocation of judicial resources and enhance judges' autonomy, the judge quota reform (Yuan'ezhi Gaige, in Chinese) aims to streamline the court system based on a formula established by the Central Committee of Chinese Communist Party (Sun and Fu, 2022). The guiding principle is to select qualified judges and empower them to make independent judicial decisions while holding them accountable for those decisions. Following the reform, only "quota judges" are authorized to issue independent verdicts, with the assistance of judge assistants and court clerks.[1] Initiated by the Chinese Supreme People's Court in 2014, this top-down re-

[1] Before the judge quota reform, judges and assistant judges were both considered judges who could issue independent verdicts.

form began with seven pilot provinces: Shanghai, Guangdong, Jilin, Hubei, Hainan, Qinghai, and Guizhou. From January to March 2015, four local pilot courts in Shanghai conducted the inaugural batch of assessments for quota judge recruitment, marking the formal commencement of the judge quota reform.[2] The other six pilot provinces launched the reform later in 2015, while the remaining provinces formally initiated the reform in 2016.[3]

According to the "Opinions on Related Issues of Judicial System Pilot Reform" issued by the Central Political and Legal Committee in 2014, a maximum of 39% of the central political-legal special staffing in the court system could be reappointed. The new judge position following the reform was known officially as serving as a "quota judge" (Sun and Fu, 2022).[4] Typically, each city or province usually starts with a reappointment rate below 39% in the first round of the reform, creating opportunities for future quota judges.

The reform was nearly completed in January 2017. Judges of the Chinese Supreme People's Court were reconfirmed as quota judges on July 3, 2017, marking the official conclusion of the judge quota reform.[5] As shown in Figure 1, about 120,000 previous judges were reconfirmed as quota judges after the reform, while about 90,000 judges were released from their former positions. The number of judges has remained stable since 2017, reflecting the binding nature of the "39 percent" rule.

Figure 1 shows that while men still comprise the majority of judges in China, the number of women entering the judiciary has been increasing over time. The share of female judges rose rapidly from about one-quarter in 2011 to one-third in 2016 before the reform. However, it experienced a slight decline during the reform year, followed by a slower rate in the second year after the reform. This suggests that the reform may have had an unintended adverse effect on women's career progression.

[2]Source: `http://www.chinatrial.net.cn/news/8171.html`.

[3]While the second batch of 11 pilot provinces (Shanxi, Inner Mongolia, Heilongjiang, Jiangsu, Zhejiang, Anhui, Fujian, Shandong, Chongqing, Yunnan, Ningxia) were selected in May 2015, these provinces formally started the reform in 2016. Source: `http://xinhuanet.com/politics/2016-01/22/c\_128655621.htm`.

[4]The central political-legal special staffing is a targeted and centrally controlled staffing allocation to a court, including verdict-related personnel (i.e. judges, assistant judges, judge assistants, and court clerks) and judicial administrative personnel.

[5]Source: `https://www.chinacourt.org/article/detail/2017/07/id/2908362.shtml`.

Figure 1. Number of judges/share of female judges over time

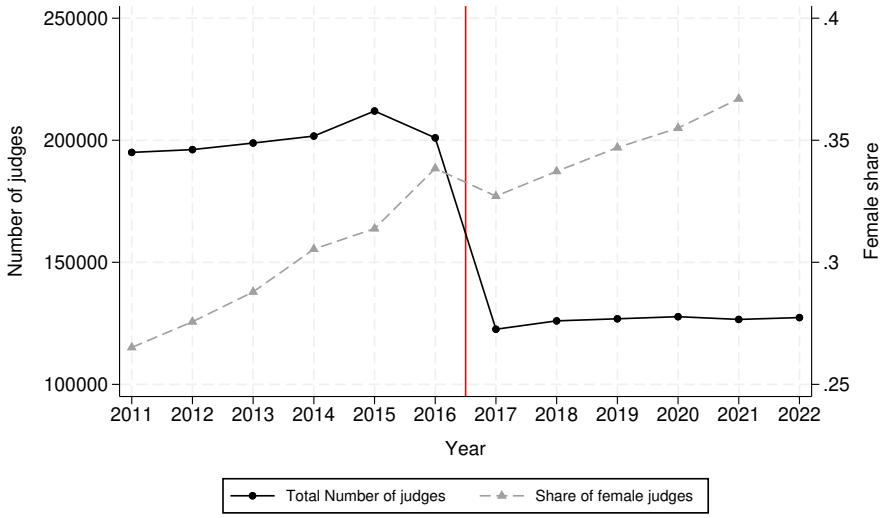


Figure 1 notes: this figure shows the total number of judges and share of female judges during the 2011-2022 period. Data sources: 2014-2022 *Statistics on the Status of Chinese Women and Children* and 2019-2023 *Work Reports of the Supreme People's Court*.

2.2 Selection of quota judges

Next, we explain how former court employees at the judge, assistant judge, and judge assistant levels can become a quota judge.[6] With the exception of senior court leaders, most applicants for post-policy judge (aka "quota judge") positions must pass both a written exam and the court's evaluation to qualify (Zuo, 2020). The written exam assesses applicants' legal knowledge, their ability to apply the knowledge in judicial decisions, and their skills in drafting judicial documents. The court's evaluation consists of competence indicators such as the quantity and quality of judicial decisions, as well as subjective assessments from court leaders and colleagues.[7] Based on the scores of the written exam and the court evaluation, a judicial selection committee ("Committee for the Selection of Judges and Procurators") decides on whether to recommend an applicant for the position of quota judge.

[6]Assistant judges could issue independent verdicts before the judge quota reform, while judge assistants provided support to judges and could not issue independent verdicts.

[7]Source: `https://gxfy.gxcourt.gov.cn/article/detail/2021/11/id/6364514.shtml`.

While provincial high people's courts oversee the assessment of quota judge applicants, the selection of quota judges is largely decided by local courts. Typically, the standard written exam score is weighted less than the evaluation score assigned by local courts. Unlike civil service exams, the assessment of quota judge applicants includes subjective evaluations of court leaders and colleagues, which usually accounts for more than one-quarter of the total assessment score (Zuo, 2020). The evaluation from court leaders carries a much larger weight in this subjective assessment, highlighting their dominant role in deciding quota judge qualifications.

Where did disqualified judges go? According to the Chinese Supreme People's Court [8] and interviews conducted by Sun and Fu (2022), many disqualified judges, particularly those with prior administrative roles, chose to transition to political or administrative departments. This aligns with the reform's goal of separating the judicial role from the political and administrative arms of the court system.[9] Other disqualified judges were downgraded to the rank of judge assistants, chose to retire early, or exited the court system. With limited prospects of becoming quota judges, many turned to private-sector employment such as positions in law firms or civil servant roles in other systems.

3 Data

We use the following two primary datasets in our paper. First, we exploit an administrative dataset comprising 85 million judgment files to track judges' ranks and work performance. Second, we manually compiled a unique dataset of quota judges, representing half of all quota judges in China. This quota judge dataset serves as a complementary analysis to verify the accuracy of judge ranks derived from the judgment files.

[8] Source: `https://www.chinacourt.org/article/detail/2017/07/id/2908362.shtml`.

[9] In China's hierarchical court system, frontline judges who handle actual cases are typically lower in rank and status. Before the judge quota reform, they often sought a career path to a management or administrative position that involves little judging but still holds a judge title.

3.1 Judgment files

To enhance judicial transparency and fairness, China's Supreme People's Court has mandated that all local courts publicize both contemporary and historical verdicts on its official website (China Judgments Online, https://wenshu.court.gov.cn/) since 2014. The judgment files contain the names and ranks of judges responsible for the verdicts, which form the basis of our main judge sample. We retrieved data on 340,000 judges from the 85 million judgment files available on the official website, covering the period from 2013 and 2021.[10]

We infer the gender of judges using their full Chinese names and a machine-learning approach.[11] Our dataset includes 98,092 female judges, representing 29% of the total sample. We track changes in judge ranks over time, resulting in 1.37 million observations at the judge-year level. As shown in Figure A1, we find a similar pattern of female judge shares documented in Figure 1 which is based on national statistical yearbooks, suggesting that our analysis sample is representative of the judge population.

We construct indicators of judges' work performance at the judge-year level based on judgment files, ensuring our metrics align with those used for promotions in the Chinese court system.[12] These comprise five measures. The first is the number of cases a judge handled in a given year, which reflects their workload. The second is the average time taken to close a case in a given year, serving as a rough measure of the efficiency with which the judge processed cases. The third is the number of words of reasoning in the judgment file issued by the judge, which is a rough measure of the quality of a judicial decision (Li, 2008; Cao, Liu and Zhou, 2023). We also construct the following two distinct measures of the quality of a decision: the proportion of first-instance cases (i.e., new cases, rather than appeals) in a given year that were appealed; and the proportion of

[10] Due to time lags between the judgment and its disclosure, our sample period begins in 2013, a year before the website's launch (Cao, Liu and Zhou, 2023). We limit the end year of our sample period to 2021, as the official website has published significantly fewer judgment files after that year.

[11] Specifically, we infer judge gender by using a Python package "ngender," which employs a Naive Bayes classifier based on more than 20 million Chinese names.

[12] These indicators are informed by the *Guiding Opinions on Strengthening and Improving the Assessment of Judges* issued by the Supreme People's Court, which emphasizes both the "quantity" of cases and the "quality" of judicial decisions. Source: `https://www.court.gov.cn/zixun-xiangqing-330041.html`.

first-instance cases in which the judge's original decision was reversed by a higher court (Cao, Liu and Zhou, 2023; Liu et al., 2023).

3.2 Quota judges

We scraped administrative data on judges from the judicial process information publicly released by China's Supreme People's Court (China Judicial Process Information Online, `https://splcgk.court.gov.cn/gzfwww/`). To supplement this data, we further manually gathered information on quota judges from local court websites when the quota judge information was missing in the administrative data. Overall, we have successfully collected information on about 60,000 judges, representing half of all judges in China.[13]

Because the judgment files contain no information on whether a judge is a quota judge, we match judges in the administrative data with performance indicators derived from judgment files for robustness checks. The information on quota judges contains judges' names, gender, affiliated courts, and administrative positions. As shown in Figure A2, we observe a clear pattern of a leaky pipeline for women: they represent the highest proportion of the lowest position, "judge assistant," but the lowest proportion of the highest position, "court president." This information also allows us to construct the gender composition of leadership at the court level, which we will analyze later to see how the gender composition of leadership team mediates the reform effect on female judges.

4 Court Reform and Gender Promotion Gap

In this section, we estimate the impact of the personnel reform on gender disparities in judicial appointments. Our analysis proceeds in two steps. First, we present the key estimates regarding how the reform differentially affects men's and women's opportunities for appointment as quota judges. Second, we explore the heterogeneity effects in terms of the intensity of the shock induced by the reform.

[13]The quota judge information comes from 2861 courts, accounting for 83% of all courts in China.

Figure 2. Likelihood of being a judge or a judge assistant by gender and over time

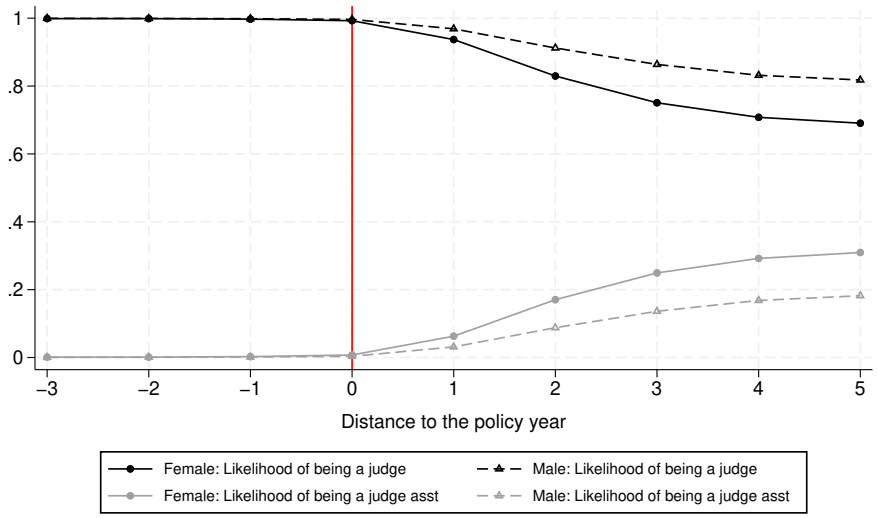


Figure 2 notes: this figure shows annual data on the transition from one judge role to another. We start with the set of male and female judges observed in 2011, three years prior to the policy. We then plot the likelihood, by gender, that these judges remain in their roles, and that they transition to "judge assistant" positions.

4.1 Main estimates

As illustrated in Figure 2, a significant portion of both male and female former judges were not reappointed following the reform, with the likelihood of non-reappointment for women nearly doubling that for men. In addition, women are more likely than men to transition into judge assistant roles after the reform, indicating that the exogenous reduction in available judge positions may disproportionately hinder women's career advancement.[14]

Next, we conduct an event study analysis to causally estimate the dynamic effect of the personnel policy on the likelihood of mean and women becoming quota judges. We estimate the following equation:

$$Judge_{it} = \alpha_0 + \sum_{p=-3}^{5} \alpha_{1\#p} * Female_i * p_{it} + \rho_i + \lambda_t + \varepsilon_{it} \quad (1)$$

[14]Figure A3 presents a similar pattern when we exclude judge assistants who joined the judicial personnel team after the reform.

$Judge_{it}$ is a dummy variable indicating whether individual i holds a judge position in year t. An individual is defined as a judge if they served as a judge (or assistant judge) before the reform or as a quota judge after the reform. The time distance from the reform year is captured by period variables p_{it}, with $p = 0$ set to the reform year. Fixed effects for individuals (ρ_i) and years (λ_t) are included to account for individual and time-specific factors, respectively. We, thus, leverage only within-individual comparisons, relying on changes in position status for the same individual over time resulting from the implementation of the judge quota reform.

Our main coefficients of interest are the set of $\alpha_{1\#p}$ across periods p_{it}, which estimate how the gender gap in the likelihood of being a judge evolves over time. These estimates are measured relative to one year prior to the reform year in the province where the individual worked. In our main analysis, we restrict our data for analysis to observations for each individual within a window of three years before and five years after the reform year. Standard errors are clustered at the individual level.

As Panel (a) of Figure 3 shows, women are less likely than men to hold judge positions following the judge quota reform, with the gender gap peaking two to three years after the reform. To account for potential compositional changes after the reform, we restrict the analysis to those who served as judges before the reform. We find a smaller yet still significant post-reform gender gap in judge appointments.[15]

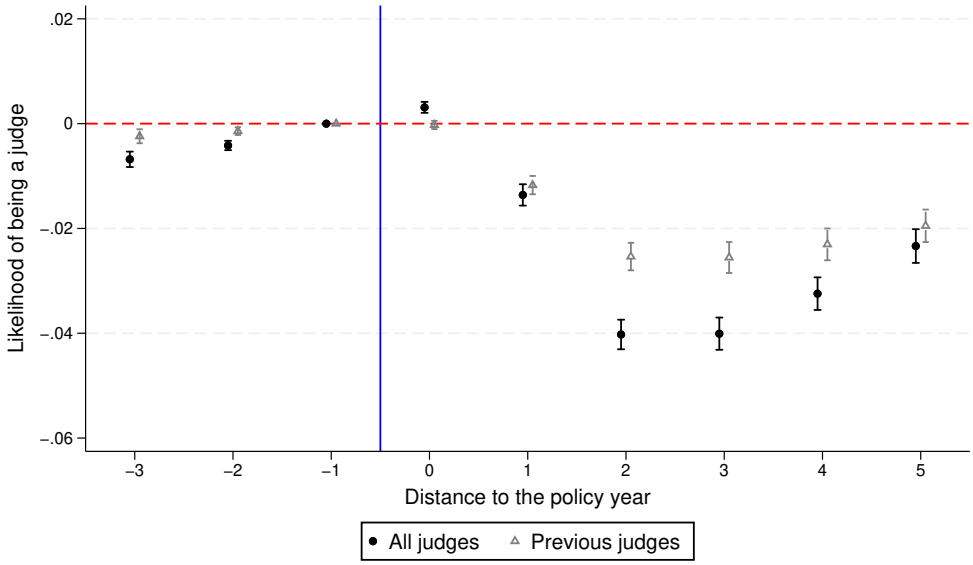
Following the reform, some non-quota judges (including assistant judges) are permitted to serve temporarily as judges due to the heavy workload in the court system.[16] Since judgment files do not distinguish between those "temporary" judges and quota judges, a higher proportion of female "temporary" judges could lead to an underestimation of the gender gap in judge appointments. To address this concern, we restrict the sample to "actual" quota judges, whose information is collected from government websites. As Panel (b) of Figure 3 shows, we find an even stronger negative reform effect on women's likelihood of being appointed as judges.

[15] As shown in Section B, we find similar results when limiting the analysis to judges selected through the civil servant exam.

[16] Source: `https://www.chinacourt.org/article/detail/2017/05/id/2853535.shtml`.

Figure 3. The reform effect on women's likelihood of being a judge

(a) Judges identified from judgment files



(b) Actual quota judges

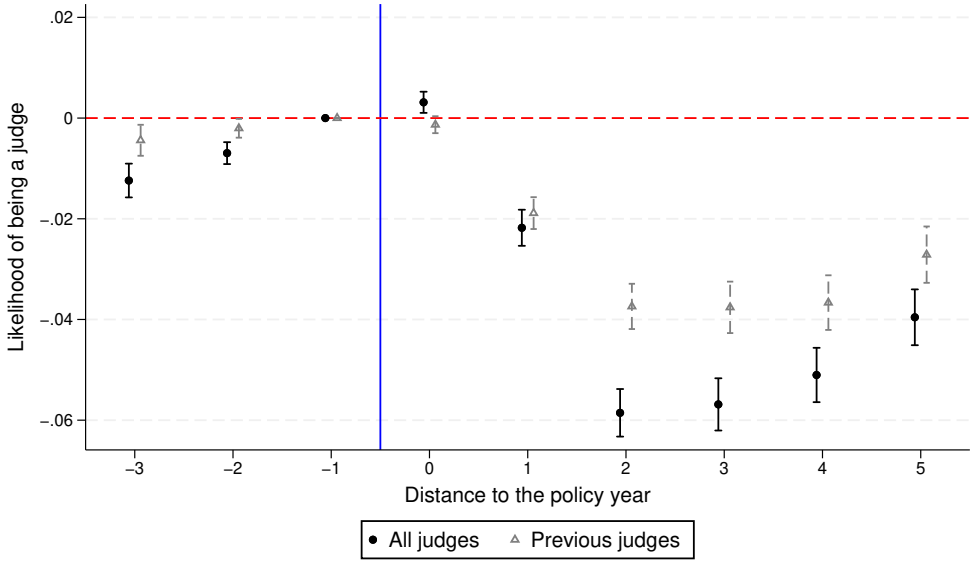


Figure 3 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether an individual is a judge in a given year. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Within each panel, we show estimates for both all court employees and those who served as judges before the reform. Standard errors are clustered at the judge level. We show the tabular analogue to these results in Table A1.

4.2 Heterogeneity in terms of shock intensity

In this subsection, we further examine the impact of the reform on the gender promotion gap by analyzing the heterogeneity in shock intensity resulting from the reform. The central government mandates that the proportion of quota judges must remain within 39% of the total judicial personnel in a court. Consequently, it becomes more challenging for a court employee to attain the status of a quota judge when a greater number of former judges or assistant judges are competing for the limited quota within a court.

As Panel (a) of Figure 4 shows, the share of judges among court employees exceeded the pre-specified red line "39%" in most courts before the reform, ranging from only slightly above "39%" to nearly "70%." However, after the reform, the judge share drops below "39%" in most courts (as shown in Panel (b)), suggesting much greater promotion pressure among court employees in those courts with a higher proportion of judges before the reform.[17]

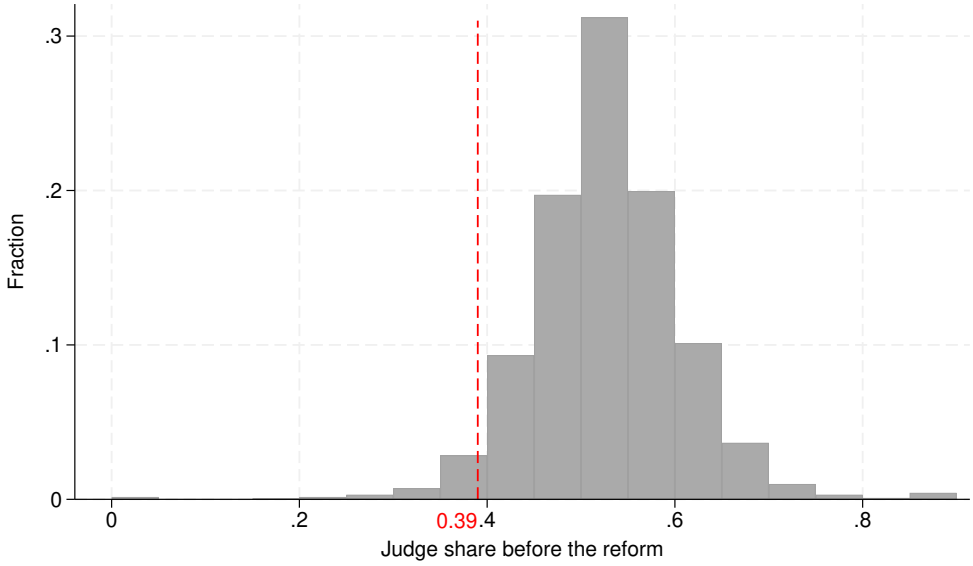
When more people compete for the limited resources ("judge quota"), the resource allocation ("promotion") may increasingly be influenced by group identity in settings with underlying discrimination (Singh and Venkataramani, 2024). As the availability of the judge quota diminishes, women's chances for promotion are likely to decline due to gender bias in promotion decisions. As illustrated in Figure 5, the gender promotion gap is more pronounced when a female employee works in a court with a higher proportion of judges prior to the reform and thus experiences a larger reform shock. We find an even larger gender promotion gap when analyzing the sample of actual quota judges instead of those identified from judgment files.[18] This suggests that women face greater disadvantages in promotions as competition for the judge quota (scarce resources) intensifies.

[17]Figure A4 shows the judge share among all court employees over time

[18]Figure A5 presents similar findings when we restrict the analysis to those who were judges before the reform.

Figure 4. Judge share among court employees at the court level before/after the reform

(a) Judge share before the reform



(b) Judge share after the reform

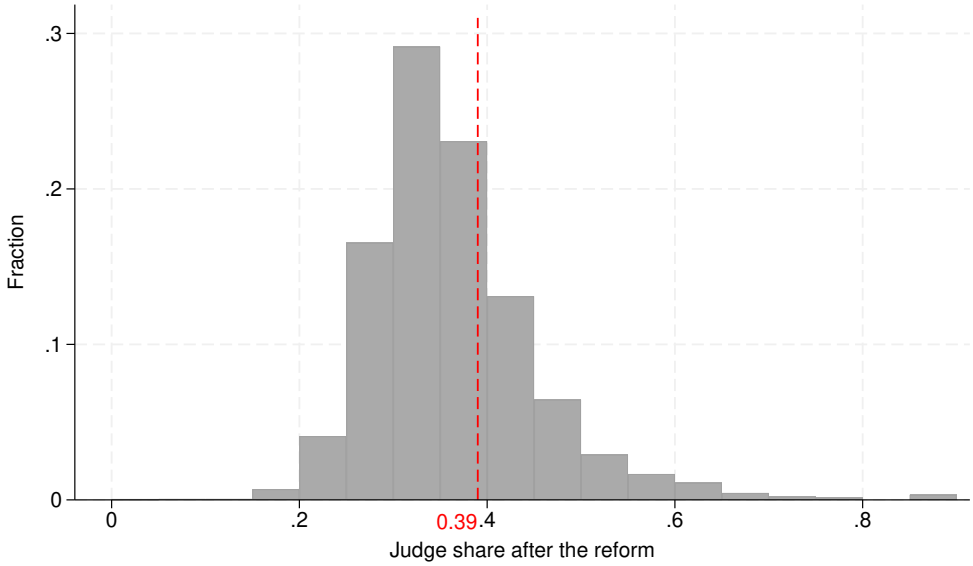


Figure 4 notes: this figure presents the distribution of the judge share among court employees at the court level before and after the reform.

Figure 5. The reform effect on women's likelihood of being a judge: Heterogeneity in terms of shock intensity

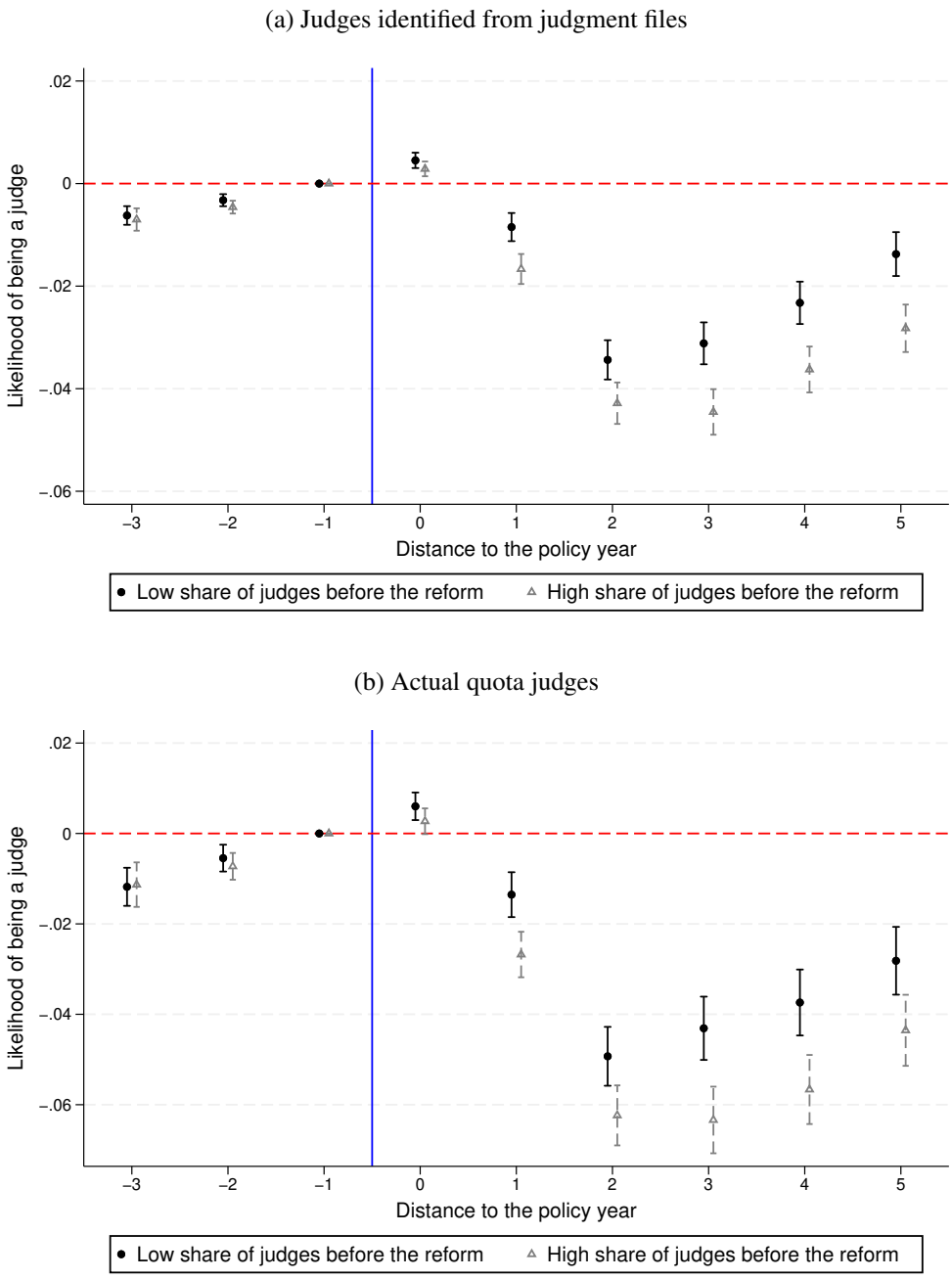


Figure 5 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. We compare the reform effect estimates between those who experienced a small reform shock (working in a court with a low judge share before the reform) and those who experienced a large reform shock (working in a court with a high judge share before the reform). The dependent variable is a dummy variable indicating whether an individual is a judge in a given year. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level.

5 Why are Women Disadvantaged in Promotions?

In this section, we explore why women are disadvantaged in the promotion to higher-ranked judge positions following the reform. First, we assess whether female judges perform worse than their male counterparts after the reform. Second, we investigate whether capable women are more likely than men to leave their judge positions. Finally, we discuss possible gender discrimination against women in the selection process for quota judges.

5.1 Do female judges perform worse than males?

In this subsection, we study whether female quota judges perform worse than their male counterparts after the reform. We use the following five performance measures that are officially recommended for evaluating judges in China's court system.[19] First, we measure the "quantity" of judge performance using the annual number of case verdicts. Second, we use the following four indicators to measure the "quality" of judge performance: (1) the average time taken to close a case in a given year, which is a rough measure of the efficiency with which the judge processes cases; (2) the number of words of reasoning in the judgment file issued by the judge, which reflects the quality of a judicial decision (Li, 2008; Cao, Liu and Zhou, 2023); (3) the proportion of first-instance cases that were appealed in a given year; and (4) the proportion of first-instance cases in which the judge's original decision was overturned by a higher court (Cao, Liu and Zhou, 2023; Liu et al., 2023).

Our results in Figure 6 show that women participate in more case verdicts and the quality of their verdicts is at least comparable to that of men's.[20] As shown in Panel (a) of Figure 6, women handle a greater workload by completing more case verdicts, though the number of case verdicts is not comparable across years due to variations in the annual number of online case verdicts.[21]

[19] We refer to the *Guiding Opinions on Strengthening and Improving the Assessment of Judges* issued by the Supreme People's Court, which emphasizes both "quantity" of cases and "quality" of judicial decisions. Source: `https://www.court.gov.cn/zixun-xiangqing-330041.html`.

[20] As shown in Section B, we find similar results when limiting the analysis to judges selected through the civil servant exam.

[21] The number of case verdicts varies across years in the analysis sample due to the availability of online judgment files, with the year 2020 having the highest number. The availability of these files is unlikely to be correlated with

As shown in Panel (b) of Figure 6, women demonstrate slightly greater efficiency in handling cases, as they spend less time closing a case. Panel (c) shows that women provide more detailed judicial reasoning in their judgment files compared to men. Panels (d) and (e) reveal that both women and men have nearly identical appeal rates and reversal rates of first-instance decisions – two important metrics used for judge evaluations in China's court system. These findings suggest that the quality of women's judicial decisions is at least on par with that of men.[22]

Overall, our results show that women perform at least as well as men, suggesting that differences in judge performance are unlikely to be the cause of the widening gender promotion gap after the reform. Furthermore, as shown in Figure A6, we observe a large gender gap in promotions to judges even after controlling for the five performance indicators discussed in this subsection.

5.2 Are capable women more likely to leave the court system?

An alternative explanation for women's disadvantages could be a higher propensity among capable women to leave the court system post-reform. If this is the case, it stands to reason that women are less likely than men to be promoted to judge positions after the reform.

To investigate this explanation, we limit our analysis to those who were judges or assistant judges before the quota judge reform. We define a dropout as an individual with at least one year of missing case verdict information following their last recorded year in the judgment files.[23]

First, we show that those who exit the court system (referred to as "dropouts") are less capable than those who remain (referred to as "stayers") after the reform. As shown in Figure 7, dropouts exhibit much higher appeal rates and reversal rates of first-instance decisions – two key metrics for assessing judge performance in China's court system. Further, female "stayers" perform at least as well as their male counterparts, and female "dropouts" also perform at least as well as male

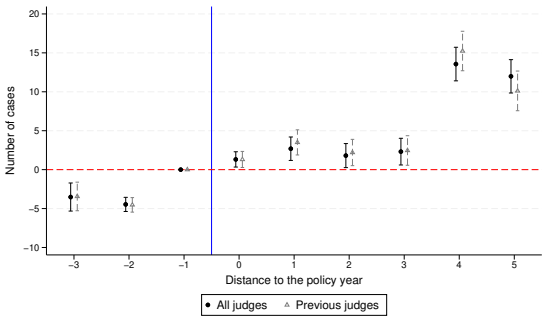
judge gender, suggesting that the sample selection issue does not affect our interpretation of results.

[22] One concern is the possibility of gender-specific case assignments. If female judges are disproportionately assigned easier cases, the measured performance could be biased against male judges. However, this concern is less relevant in our context, as we observe consistent patterns even after 2017, when the Supreme People's Court of China mandated the random assignment of cases to judges. Source: `http://gongbao.court.gov.cn/Details/a4bafde7d2e94a72f12f97204cd1be.html`.

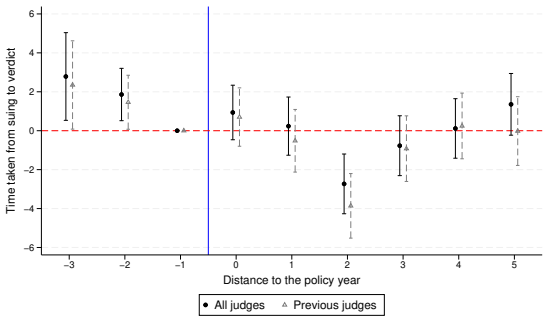
[23] We also redefine a dropout as someone with "at least two years of missing case verdict information" and find similar results.

Figure 6. Gender differences in judge performance

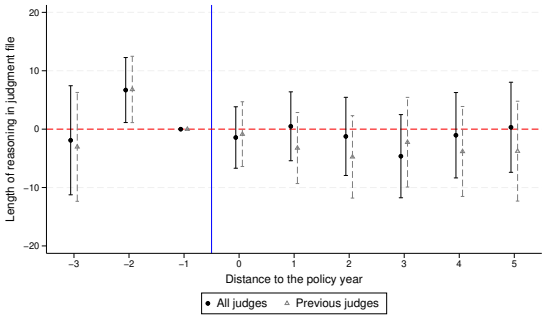
(a) Number of case verdicts



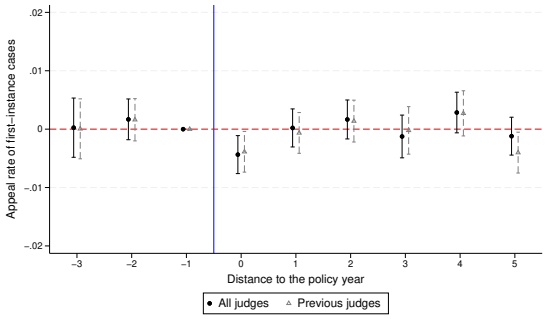
(b) Time for closing a case



(c) Length of reasoning



(d) Appeal rate



(e) Reversal rate

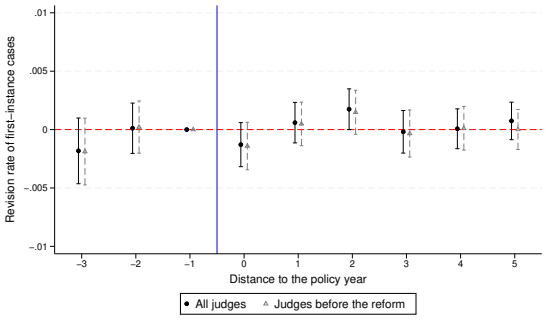


Figure 6 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variables are judge performance indicators, including (a) number of case verdicts, (b) time for closing a case, (c) average length of reasoning in judgment files, (d) appeal rate of first-instance decisions, and (e) reversal rate of first-instance decisions. We show the results both with the all-judge sample and with the sample who were judges before the reform. Standard errors are clustered at the judge level.

"dropouts."

Second, as shown in Figure 8, we find almost no gender differences in the likelihood of exiting the court system among high-ability judges.[24] Together, these findings suggest that the concern about selective exit is minimal in our context, indicating that the gender promotion gap among judges is unlikely to be attributed to women's selective exit from the court system.

Finally, to further alleviate the above concern, we limit the analysis to those who have remained in the court system after the reform. In Figure 9, our results indicate that we continue to observe a substantial gender gap in the likelihood of being a judge after the reform.[25] As shown in Panel (a), women are about 4% less likely than men to hold judge positions after the reform, though the estimates are slightly smaller than those obtained from the full sample, as shown in Figure 3. Overall, our results suggest that women's disadvantages in promotions to judge positions are unlikely to stem from changes in the gender composition of capable judges after the reform.

5.3 Evidence of discrimination in the selection process

Our previous results indicate that the gender promotion gap is unlikely to be attributed to gender differences in judicial performance or to more capable women leaving the court system after the reform. In this subsection, we propose that male preference in promotion decisions may contribute to the gender promotion gap after the reform, which reduced available judge positions.

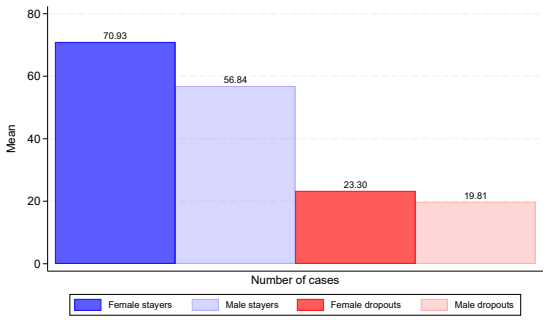
As discussed in Section 2.2, the selection of quota judges is based on the weighted scores of three components: a written exam administrated by provincial high people's courts, a competence evaluation conducted by local courts, and a subjective evaluation known as "democratic appraisal," which is rated by court leaders and colleagues. The typical weights assigned to these components are 0.4 for the written exam, 0.3 for the competence evaluation, and 0.3 for the democratic appraisal. Further, the competence evaluation score is itself a weighted composite of several factors: work performance evaluation (weight: 0.17), seniority (weight: 0.06), which is based on years of work experience and position rank, and other criteria (weight: 0.07), including education

[24] We show the regression estimates in Figure A8. Among non-high-ability judges, we find small gender differences in the likelihood of exiting the court system, as shown in Figure A9

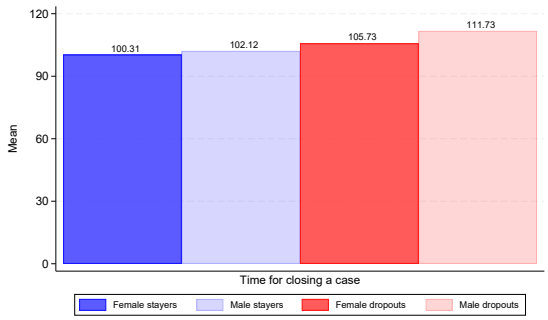
[25] Figure A10 also reveals a gender promotion gap among those who ever exited the court system.

Figure 7. Judge performance by gender and by dropout status

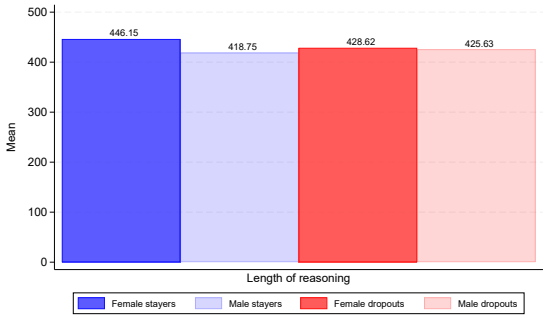
(a) Number of case verdicts



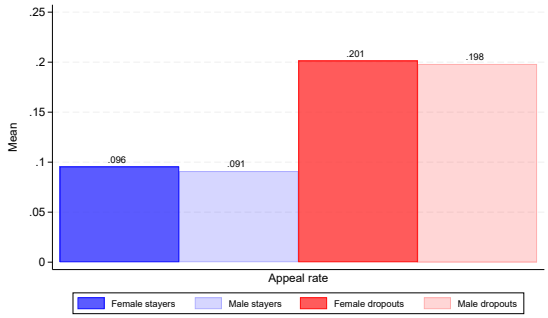
(b) Time for closing a case



(c) Length of reasoning



(d) Appeal rate



(e) Reversal rate

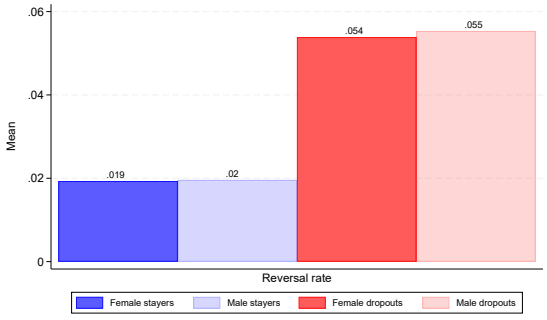
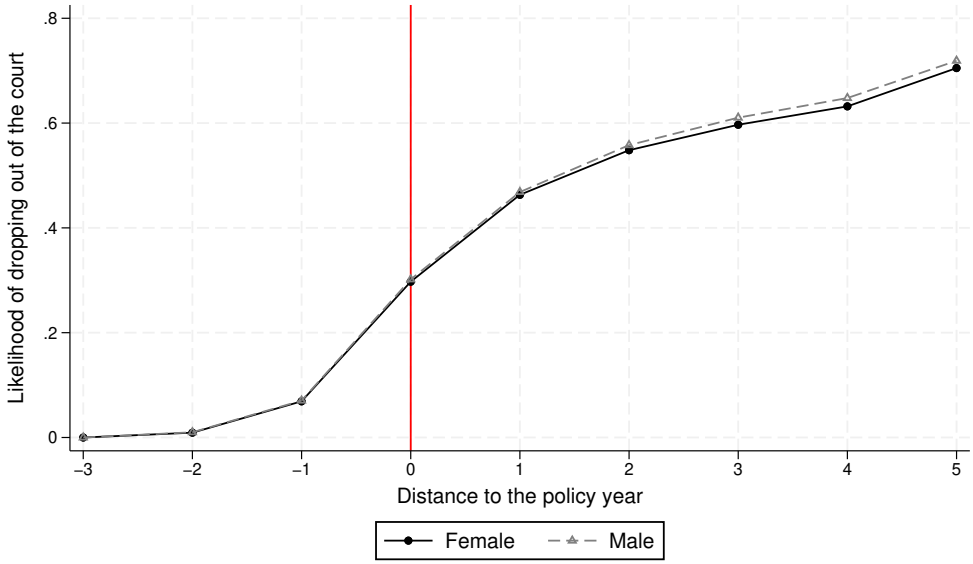


Figure 7 notes: this figure compares judge performance by gender and by dropout status. We use the following five performance metrics: (a) number of case verdicts, (b) time for closing a case, (c) average length of reasoning in judgment files, (d) appeal rate of first-instance decisions, and (e) reversal rate of first-instance decisions.

Figure 8. Gender differences in exiting the court system among high-ability judges

(a) Judges with zero reversal rates



(b) Judges with zero appeal rates

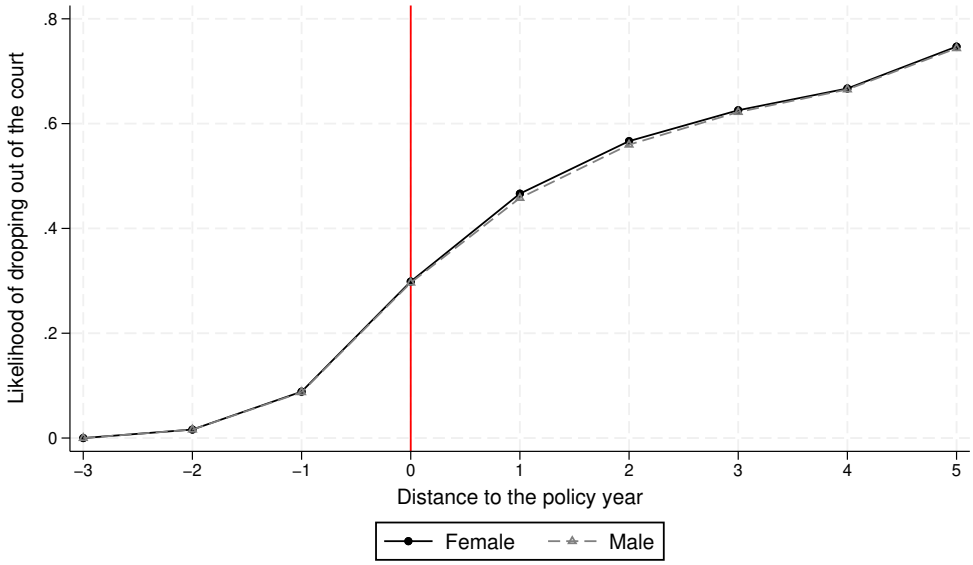
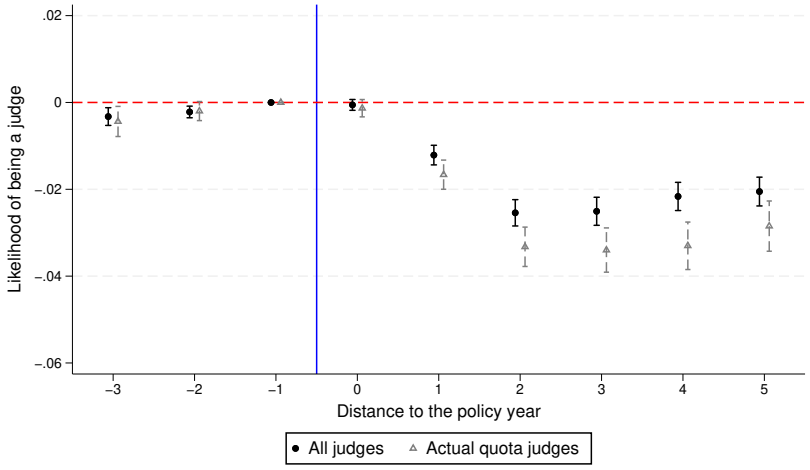


Figure 8 notes: this figure compares the dropout likelihood by gender and over time. We focus on two high-ability groups: judges with zero reversal rates of first-instance case verdicts, and judges with zero appeal rates of first-instance case verdicts.

Figure 9. Gender difference in the likelihood of being a judge among stayers

(a) Main results



(b) Excluding judges with two missing years of data

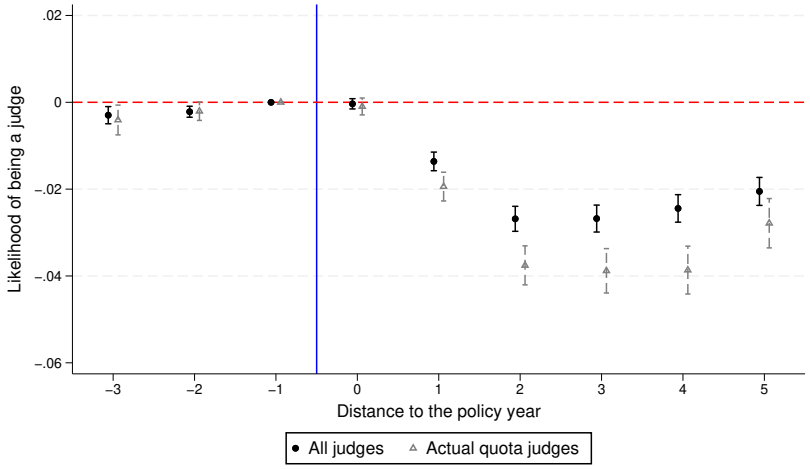


Figure 9 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is the likelihood of being a judge in a given year. In Panel (a), we present estimates using a dataset where we drop all judges with more than one missing year of data (that is, in one of our study years, they appear in no judgment files). In Panel (b) we present analog results, relaxing this to excluding only judges with more than two missing years of data. We define judges both based on online judgment files and on the information on quota judges collected from government websites. Standard errors are clustered at the judge level.

background, awards, and extra scores for factors such as judicial integrity evaluation and political quality evaluation (Zuo, 2020).

As shown in Figure 10, while female candidates outperform male candidates on the objective written exam, they receive lower scores in both the competence assessment and the subjective evaluation. Although women's work performance is at least on par with men's, as shown in Section 5.1, they generally have fewer years of work experience, are more likely to hold junior positions, and are less likely to receive awards or to earn extra points, which are largely subjective. This indicates that women face disadvantages in the competence assessment.

Further, even if the gender differences in the written exam and competence evaluation cancel each other out, the subjective evaluation ("democratic appraisal") still favors men over women. Discrimination against disadvantaged groups such as women is particularly likely in subjective evaluations (Benson, Li and Shue, 2024; Shukla, 2024). This male preference may be exacerbated by the significant weight of court leaders' evaluations in the subjective assessment, with most court leaders being male.[26] Such male preference may become even more pronounced when the competition for judge positions intensifies, as documented in Section 4.2.

Overall, the gender promotion gap is likely attributed to male preference in the evaluation process, which includes a competence assessment designed to favor men and a subjective evaluation component that disproportionately prioritizes male leaders.

6 Can Female Leadership Help Alleviate Women's Disadvantages in Promotion?

As documented in Section 5, judge promotion decisions are largely shaped by subjective evaluations of court leadership. Motivated by this fact, in this section, we explore how female court leadership may help alleviate women's disadvantages in promotions to higher-ranked judges. Because the gender of a court president ("boss") and the gender composition of leadership teams may have different impacts on women's career advancement (Kunze and Miller, 2017; Karpowitz et al.,

[26]For example, the typical weights are 0.5 for court leaders and 0.5 for other colleagues.

Figure 10. Gender differences in promotion evaluation scores

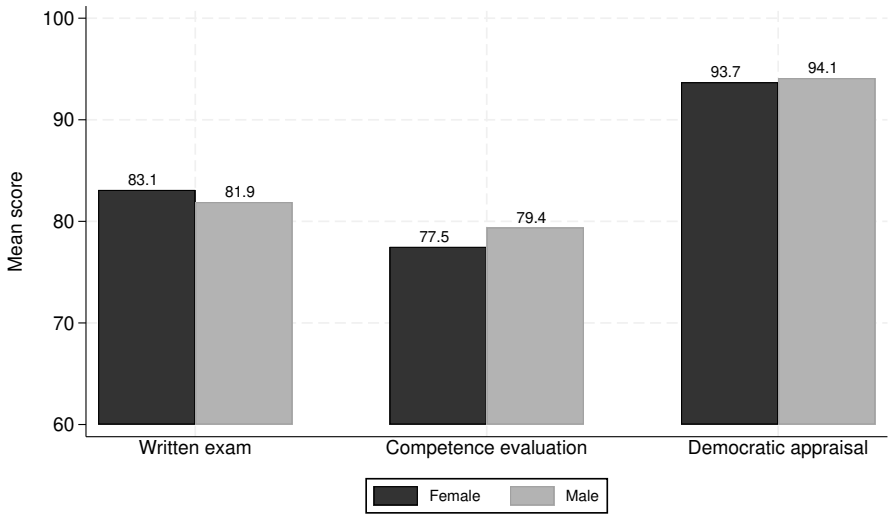


Figure 10 notes: this figure shows gender differences in the three components of the judge promotion evaluation: written exam scores, competence evaluation scores, and subjective evaluation ("democratic appraisal") scores. The summary statistics are from (Zuo, 2020).

2024), we study their roles in women's promotions separately.

6.1 The role of a female court president

In this subsection, we explore whether having a female court president can help reduce the gender promotion gap following the judge quota reform. Figure 11 shows that the gender promotion gap is only slightly smaller when a female court president is in place compared to a male president after the reform.[27] The results are robust when we limit the analysis to actual quota judges or to those who were judges before the reform.[28]

The findings suggest that having a female president ("boss") may not necessarily reverse the gender promotion gap (Alan et al., 2024). This is likely due to the principle of collective

[27] One concern is that the gender of court presidents may change after the reform. For instance, a female court president may be replaced by a male after the reform, potentially underestimating the role of a female court presidency in reducing the gender promotion gap. This concern, however, is minor in our setting because our court leadership sample is limited to those who were in position during the reform period. The presidency is unlikely to change significantly within a short timeframe (i.e. the 2016-2021 period in our setting).

[28] Figure A11 shows the estimation results by limiting the analysis to those who were judges before the judge quota reform.

leadership in China's civil service system, where important decisions such as promotions should be made collectively by a group of court leaders, rather than solely by the court president (Thornton, 2021).[29] To further explore the role of female leadership in mitigating women's disadvantages in promotions, we will next focus on the gender composition of court leadership teams.

6.2 Gender composition of court leadership teams

In this subsection, we further explore how the gender composition of court leadership teams affects women's promotions to higher-ranked judge positions. We divide the analysis sample in terms of whether the female share of court leadership members is equal to or larger than 0.5.[30]

In Figure 12, we find that although women still face disadvantages in promotions, the gender promotion gap is much smaller when women make up at least half of the court leadership. The gender promotion gap nearly disappears when we limit the analysis to "actual" quota judges.[31]

We conduct the following two robustness checks. First, we limit the analysis to those who were judges before the judge quota reform to account for potential compositional changes after the reform. Second, we exclude the sample of court leaders from the analysis because almost all of them were reappointed as quota judges after the reform (Zuo, 2020). Figures A13 and A14 show that the main estimates remain robust using alternative analysis samples.

6.3 Why does female leadership matter?

As documented in Section 2.2, promotion to a quota judge position is determined by a candidate's performance on a written exam organized by provincial high people's courts, a competence evaluation conducted by local courts, and a subjective evaluation by court leaders and colleagues. While

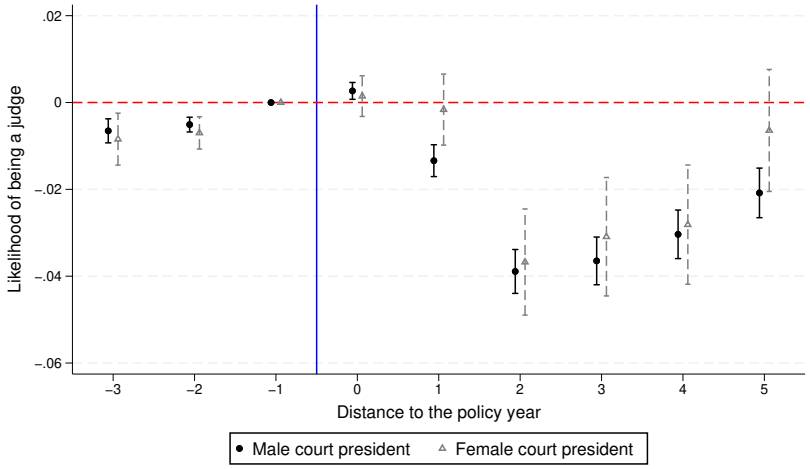
[29] Collective leadership in China refers to a governance principle within China's civil service system, where the Chinese Communist Party (CCP) plays a central role. This principle requires that decision-making authority be exercised collectively by a group of leaders. It is designed to ensure that no single leader holds unchecked power and that decisions are made through consultation and consensus among leadership teams or committees.

[30] In our analysis sample, only 7.44% of courts have a female share of leadership members being equal to or larger than 0.5, while 40% of courts have no female leaders.

[31] We further divide the analysis sample into three subgroups: court employees with a female share of leadership members [0, 0.33], court employees with a female share of leadership members (0.33, 0.66], and court employees with a female share of leadership members (0.67, 1]. As shown in Figure A12, a larger female share of court leadership members is associated with a smaller gender promotion gap.

Figure 11. Gender difference in likelihood of being a judge: Male vs. female court presidents

(a) Judges identified from judgment files



(b) Actual quota judges

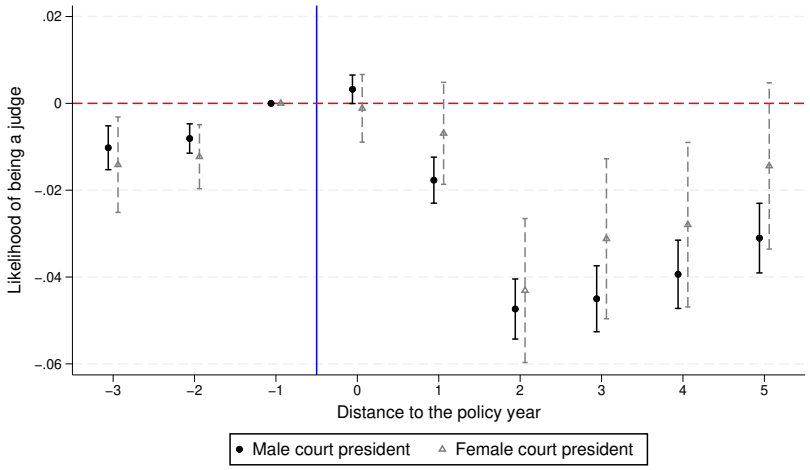
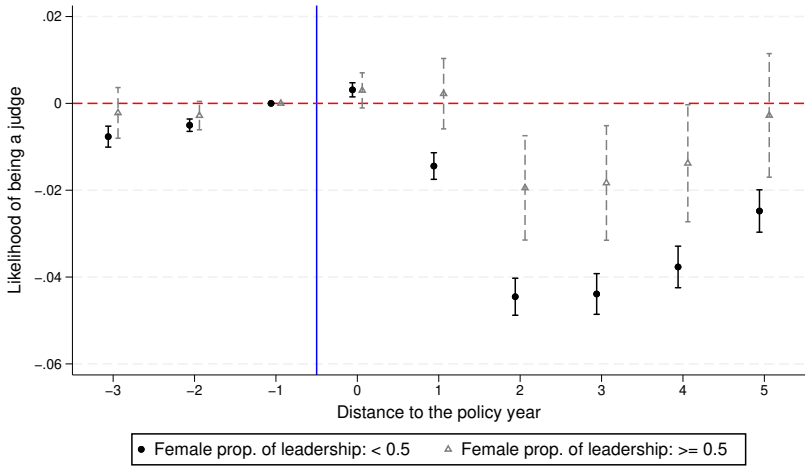


Figure 11 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is the likelihood of being a judge in a given year. We compare the gender gap in the likelihood of being a judge between courts with a male president and courts with a female president. Figure **??** shows the results with the sample of judges who first appeared in judgment files before the reform. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level. We show the tabular analogue to these results in the first two columns of Table A2.

Figure 12. Gender difference in likelihood of being a judge: A low vs. high female share of court leaders

(a) Judges identified from judgment files



(b) Actual quota judges

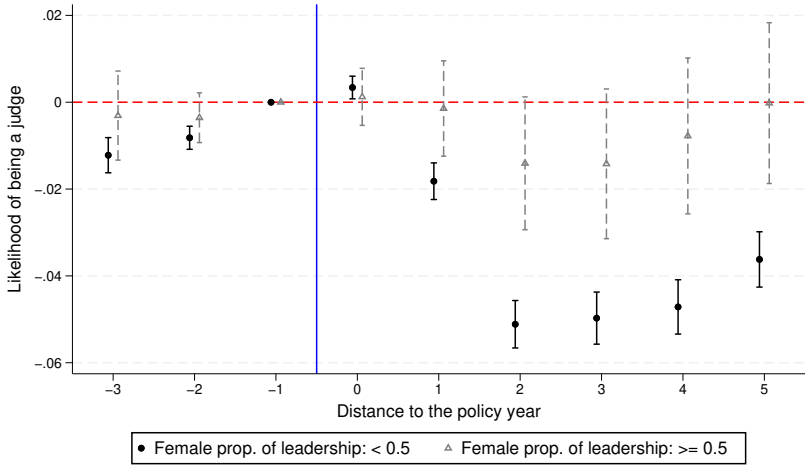


Figure 12 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is the likelihood of being a judge in a given year. We compare the gender gap in the likelihood of being a judge between courts with a high female share (>= 0.5) of court leadership members and courts with a low female share (< 0.5). We use the sample of judges who first appeared in judgment files both before and after the reform. Figure A13 shows the results with the sample of judges who first appeared in judgment files before the reform. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level. We show the tabular analogue to these results in the first two columns of Table A3.

female leaders are unlikely to help improve women's written exam scores – which assess legal knowledge and its application – we focus on the role of female leadership in enhancing women's competence such as work performance and subjective evaluations.

First, we examine gender differences in judges' work performance in courts with a high female share of leadership members compared to those with a low female share. As shown in Figure A15, we find that the gender gap favoring women in the number of judicial verdicts is more pronounced in courts with a higher proportion of female leaders. We, however, observe almost no differences in the gender gap across the other four work performance indicators with respect to the share of female leadership. Overall, our results suggest that women's work performance benefits to some extent from a higher proportion of female leadership compared to lower female representation.

Second, we explore the role of female leadership in subjective evaluations during the application process for quota judge positions. While we lack data on applicants' subjective evaluation scores, the structure of the quota judge application evaluations indicates that court leaders have a dominant influence, as their assessments carry significantly more weight in the subjective evaluation component of the overall evaluation process. Because female leaders tend to favor female subordinates or are better at interpreting women's job quality (Joshi, Son and Roh, 2015; Kunze and Miller, 2017; Flabbi et al., 2019), women are expected to receive higher subjective evaluation scores in courts with a higher proportion of female leadership compared to those with lower female representation.

In summary, female leadership enhances women's work performance or leads to higher subjective evaluation scores, thereby helping to mitigate the widening gender promotion gap that emerged after higher-ranked judge positions became scarcer due to the judge quota reform.[32]

[32] As shown in Table A4, female leadership is not significantly correlated with regional characteristics including economic development level, population size, education levels, and gender norms (measured by sex ratio at birth). This suggests that the role of female leadership is unlikely to be confounded by local factors.

7 Conclusion

In this paper, we study how the gender promotion gap evolves when promotion opportunities become fewer due to a personnel reform in China's court system. We find that the scarcity of higher-ranked judge positions exacerbates the promotion gap in favor of men, despite women performing at least as well as their male counterparts. The widened gender promotion gap cannot be attributed to gender-specific shifts in the composition of the personnel. Instead, we provide suggestive evidence that male preference in subjective evaluations is a significant driver of women's disadvantages in promotions. Moreover, we show that a large female presence in court leadership teams helps mitigate the gender promotion gap, while the gender of court presidents plays a less prominent role.

These findings highlight how reduced opportunities, coupled with male-biased evaluation processes, contribute to the "leaky pipeline" phenomenon, where talented women are underrepresented in higher-ranked positions. The persistence of this gap suggests that systemic biases within promotion frameworks can have lasting implications for gender equity in professional settings. Our research underscores the need for a critical examination of evaluation processes to ensure they are fair and equitable, particularly in environments where competition for promotions is intensified.

Importantly, our results offer an optimistic pathway forward: increasing the representation of female leaders at the top. This aligns with ongoing discussions in both academic and industry contexts regarding the importance of diverse leadership. Our evidence suggests that female leaders may possess a unique ability to better assess and support women's productivity. Furthermore, we find that the effectiveness of female leadership in promoting gender equity is amplified when their representation is substantial rather than minimal. This is particularly crucial given that many promotion decisions rely on collective judgments shaped by subjective evaluations. Thus, fostering a more equitable leadership landscape could play a pivotal role in addressing the gender promotion gap and ensuring that talented women are not left behind in their professional advancement.

References

Akbarpour, Mohammad, Eric Budish, Piotr Dworczak, and Scott Duke Kominers. 2024. "An economic framework for vaccine prioritization." Quarterly Journal of Economics, 139(1): 359–417.

Alan, Sule, Gozde Corekcioglu, Mustafa Kaba, and Matthias Sutter. 2024. "Female leadership and workplace climate." Working paper.

Alper, Joe. 1993. "The pipeline is leaking women all the way along." Science, 260(5106): 409–411.

Benson, Alan M., Danielle Li, and Kelly Shue. 2024. "'Potential" and the gender promotion gap." American Economic Review.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." American economic journal: applied economics, 2(3): 228–255.

Bertrand, Marianne, Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway." Review of Economic Studies, 86(1): 191–239.

Biasi, Barbara, and Heather Sarsons. 2022. "Flexible wages, bargaining, and the gender gap." The Quarterly Journal of Economics, 137(1): 215–266.

Buckles, Kasey. 2019. "Fixing the leaky pipeline: Strategies for making economics work for women at every stage." Journal of economic perspectives, 33(1): 43–60.

Cao, Guangyu, Chenran Liu, and Li-An Zhou. 2023. "Suing the government under weak rule of law: Evidence from administrative litigation reform in China." Journal of Public Economics, 222: 104895.

Ceci, Stephen J, and Wendy M Williams. 2011. "Understanding current causes of women's underrepresentation in science." Proceedings of the National academy of sciences, 108(8): 3157–3162.

Cullen, Zoe, and Ricardo Perez-Truglia. 2023. "The old boys' club: Schmoozing and the gender gap." American Economic Review, 113(7): 1703–1740.

Flabbi, Luca, Mario Macis, Andrea Moro, and Fabiano Schivardi. 2019. "Do female executives make a difference? The impact of female leadership on gender gaps and firm performance." Economic Journal, 129(622): 2390–2423.

Ginther, Donna K, and Shulamit Kahn. 2004. "Women in economics: moving up or falling off the academic career ladder?" Journal of Economic perspectives, 18(3): 193–214.

Hoffman, Mitchell, and Christopher T Stanton. 2024. "People, Practices, and Productivity: A Review of New Advances in Personnel Economics."

Hospido, Laura, Luc Laeven, and Ana Lamo. 2022. "The gender promotion gap: evidence from central banking." Review of Economics and Statistics, 104(5): 981–996.

Joshi, Aparna, Jooyeon Son, and Hyuntak Roh. 2015. "When can women close the gap? A meta-analytic test of sex differences in performance and rewards." Academy of Management Journal, 58(5): 1516–1545.

Karpowitz, Christopher, Stephen D O'Connell, Jessica Preece, and Olga Stoddard. 2024. "Strength in numbers? Gender composition, leadership, and women's influence in teams." Journal of Political Economy, 132(9): 3077—-3114.

Kunze, Astrid, and Amalia R Miller. 2017. "Women helping women? Evidence from private sector data on workplace hierarchies." Review of Economics and Statistics, 99(5): 769–775.

Lawson, M Asher, Ashley E Martin, Imrul Huda, and Sandra C Matz. 2022. "Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language." Proceedings of the National Academy of Sciences, 119(9): e2026443119.

Liebman, Benjamin L. 2017. "China's courts: Restricted reform." In Law and society in East Asia. 237–280. Routledge.

Li, Feng. 2008. "Annual report readability, current earnings, and earnings persistence." Journal of Accounting and Economics, 45(2-3): 221–247.

Liu, Ernest, Yi Lu, Wenwei Peng, and Shaoda Wang. 2023. "Judicial independence, local protectionism, and economic integration: Evidence from China." National Bureau of Economic Research.

Matsa, David A, and Amalia R Miller. 2011. "Chipping away at the glass ceiling: Gender spillovers in corporate leadership." American Economic Review, 101(3): 635–639.

Shukla, Soumitra. 2024. "Making the elite: Discrimination at top firms." Working paper.

Singh, Manasvini, and Atheendar Venkataramani. 2024. "Rationing by race." American Economic Review, Revise and Resubmit.

Sun, Ying, and Hualing Fu. 2022. "Of judge quota and judicial autonomy: an enduring professionalization project in China." China Quarterly, 251: 866–887.

Thornton, Patricia M. 2021. "Of constitutions, campaigns and commissions: a century of democratic centralism under the CCP." China Quarterly, 248(S1): 52–72.

Zuo, Weimin. 2020. "Empirical research on the reform of mechanisms for the selection of quota judges: The case of Province A." China Legal Science, 37(4): 261–282.

Appendices

A Appendix figures and tables

Figure A1. Number of judges/share of female judges over time: The analysis sample

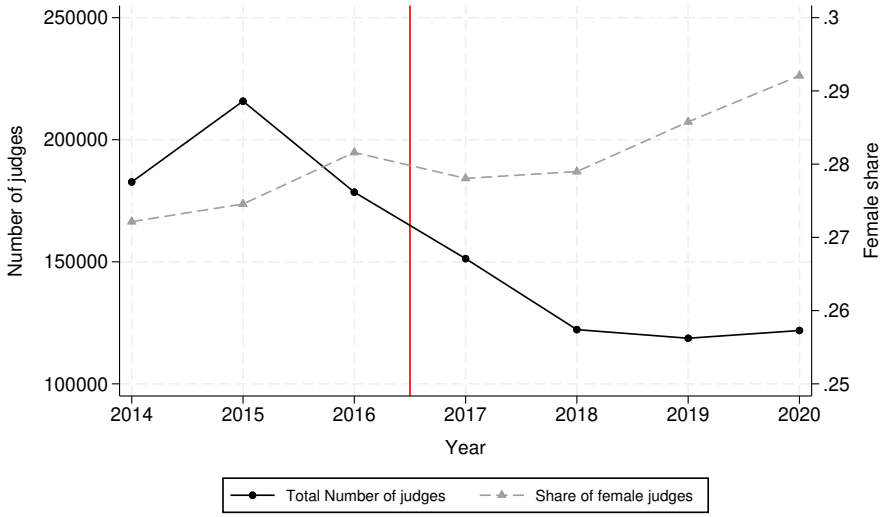


Figure A1 notes: this figure shows the total number of judges and share of female judges during the 2014-2020 period using the analysis sample. We drop observations in the year 2013/2021 from the analysis sample due to many missing observations in the two years.

Figure A2. A leaky pipeline for women in China's court system

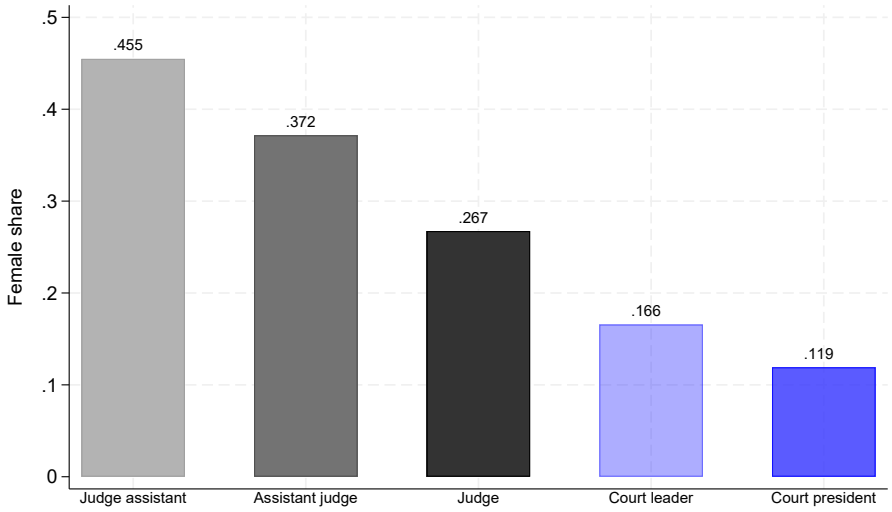


Figure A2 notes: this figure shows female shares across positions in China's court system, ranging from the lowest "judge assistant" and the highest "court president." Court leaders include court presidents, vice presidents, and other members of the Communist Party leadership group of the court.

Figure A3. Likelihood of being a judge or a judge assistant by gender and over time: previous judges

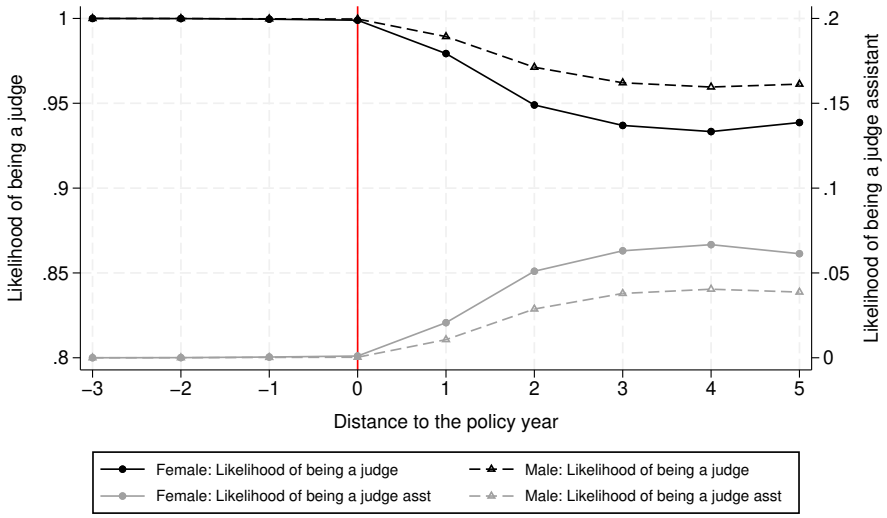


Figure A3 notes: this figure shows women's and men's likelihood of being a judge or a judge assistant over time by limiting the sample to those who were judges before the reform.

Figure A4. Judge share among all court employees over time

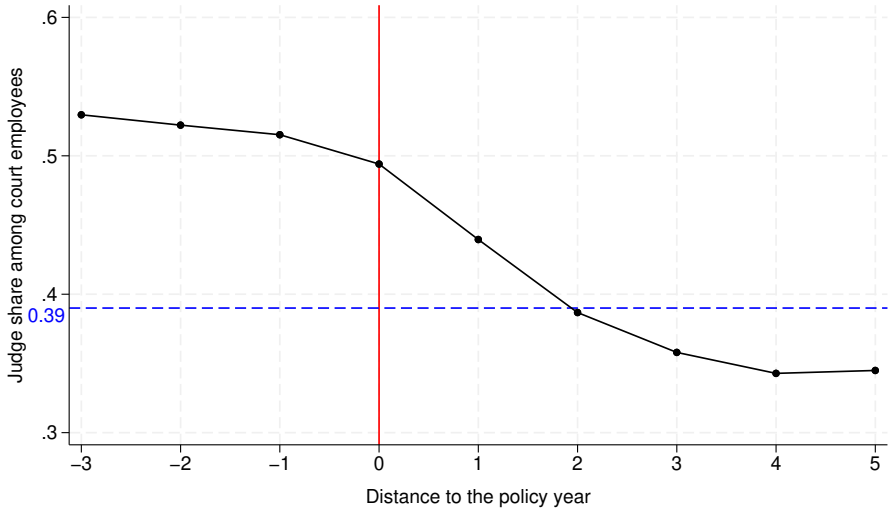
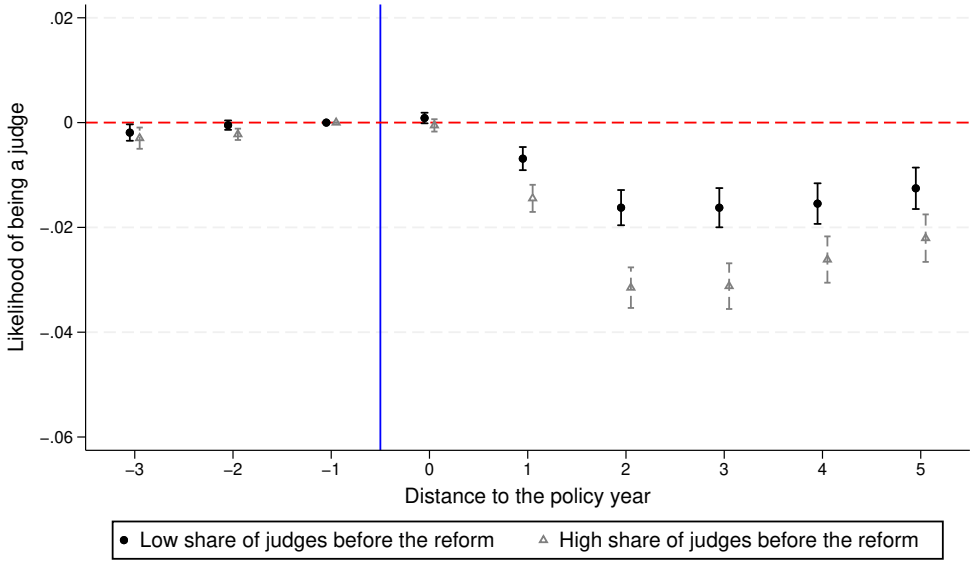


Figure A4 notes: this figure presents the judge share among all court employees over time.

Figure A5. The reform effect on women's likelihood of being a judge among previous judges: Heterogeneity in terms of shock intensity

(a) Judges identified from judgment files



(b) Actual quota judges

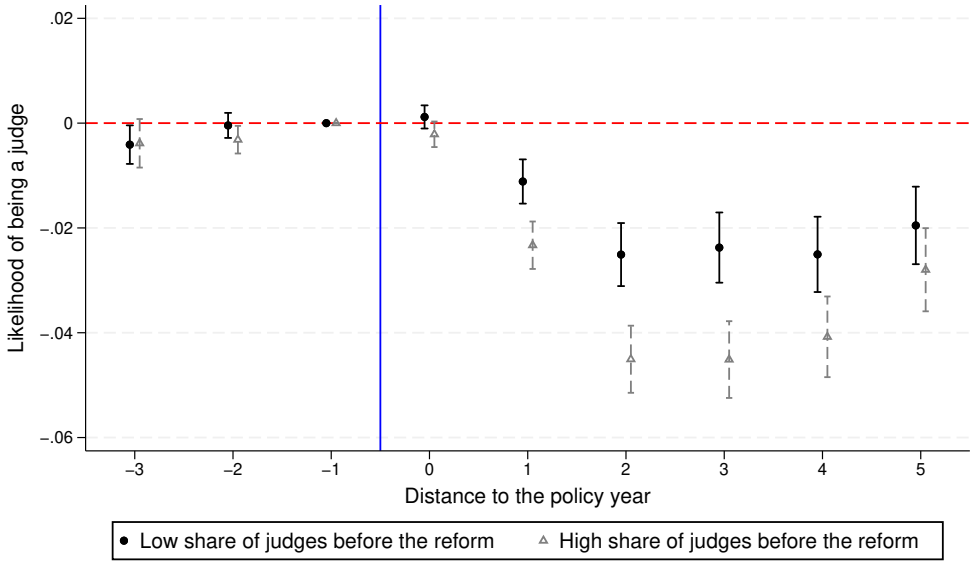
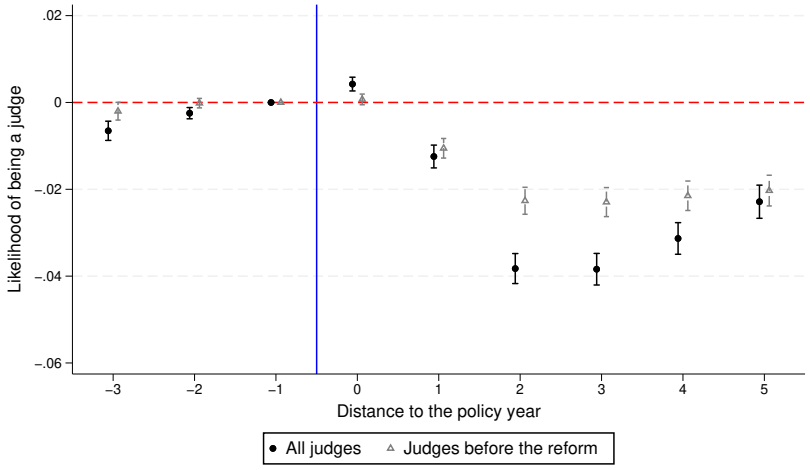


Figure A5 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. We compare the reform effect estimates between those who were judges before the reform and experienced a small reform shock (working in a court with a low judge share before the reform) and those who were judges before the reform and experienced a large reform shock (working in a court with a high judge share before the reform). The dependent variable is a dummy variable indicating whether to be a judge in a given year. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level.

Figure A6. Gender difference in likelihood of being a judge: controlling for performance

(a) All judges



(b) Actual quota judges

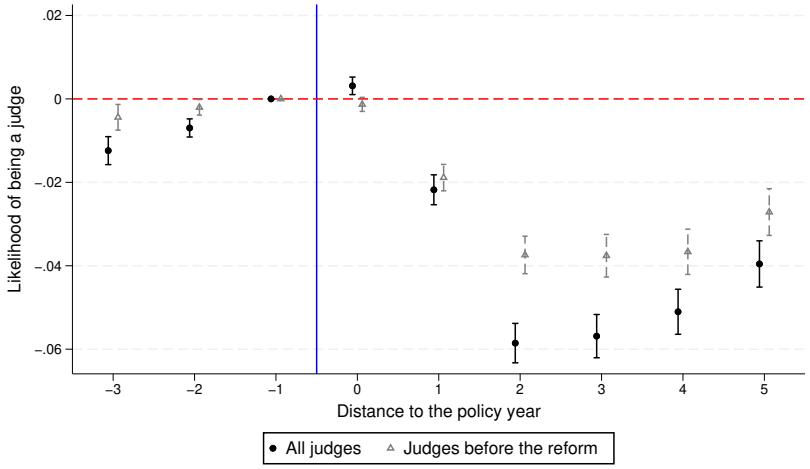
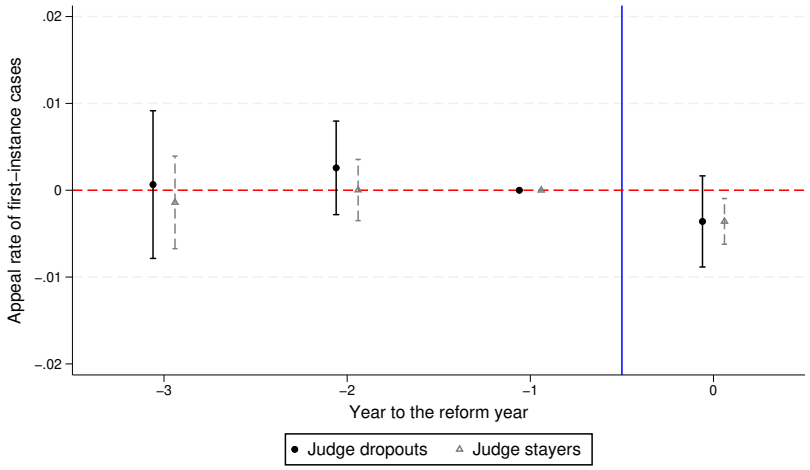


Figure A6 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1 by further controlling for all the five performance indicators in the model: number of case verdicts, time for closing a case, average length of reasoning in judgment files, appeal rate of first-instance decisions, and reversal rate of first-instance decisions. The dependent variable is likelihood of being a judge in a given year. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level.

Figure A7. Gender difference in performance among dropouts and stayers

(a) All judges



(b) Real quota judges

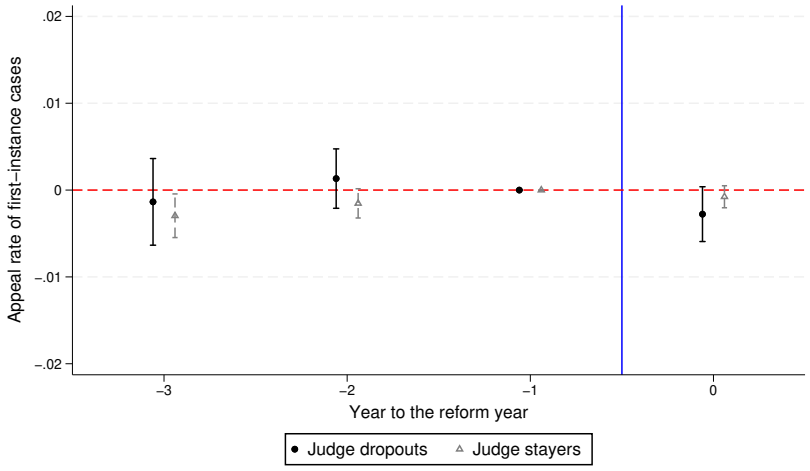


Figure A7 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is appeal rate of first-instances cases for a judge in a given year in Panel (a), and reversal rate of first-instances cases for a judge in a given year in Panel (b). We compare the performance indicators between those who exit the court system ("dropouts") and those who have been staying ("stayers"). Standard errors are clustered at the judge level.

Figure A8. Gender difference in likelihood of exiting the court system

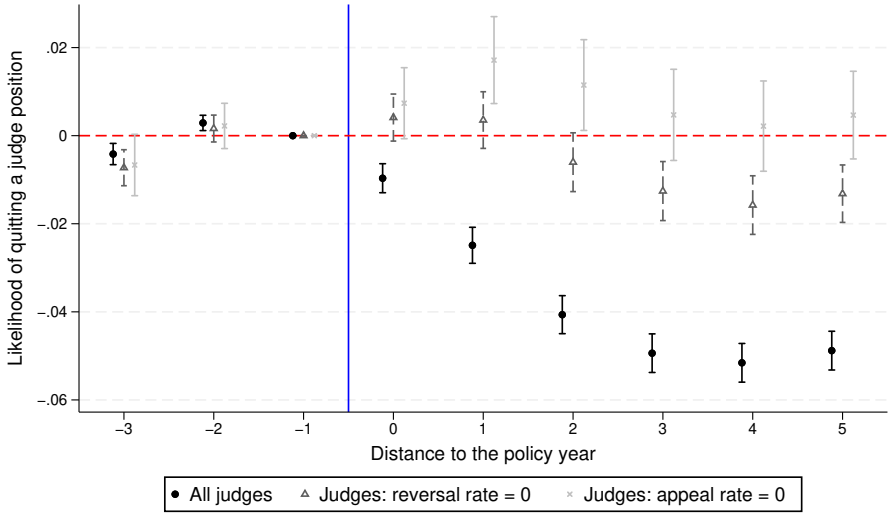
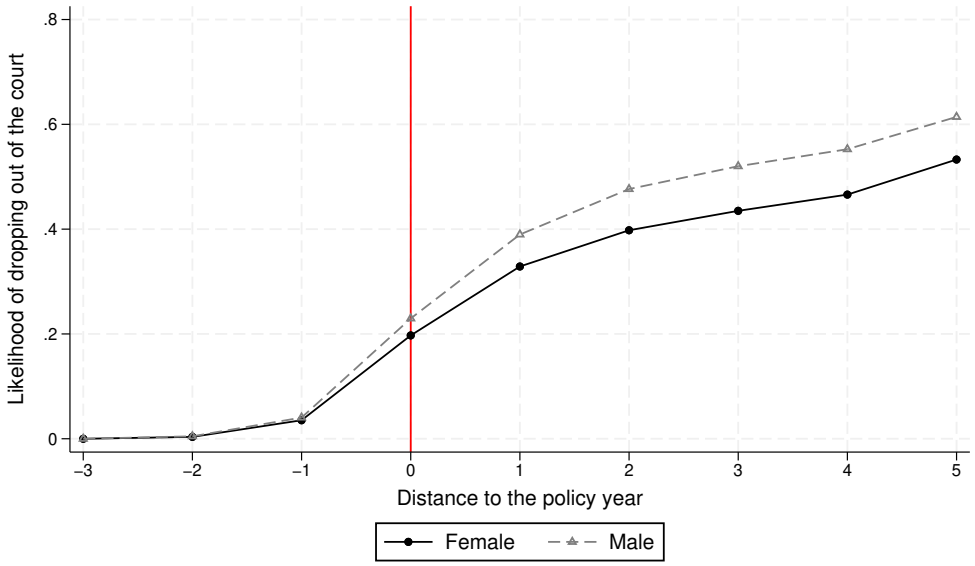


Figure A8 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether a judge exits from the court system in a given year. Except for the estimates using the all-judge sample, we also show the estimates by limiting the analysis sample to two high-ability groups: judges with zero reversal rates of first-instance case verdicts, and judges with zero appeal rates of first-instance case verdicts. Standard errors are clustered at the judge level.

Figure A9. Gender differences in exiting the court system among high-ability judges

(a) Judges with non-zero reversal rates



(b) Judges with non-zero appeal rates

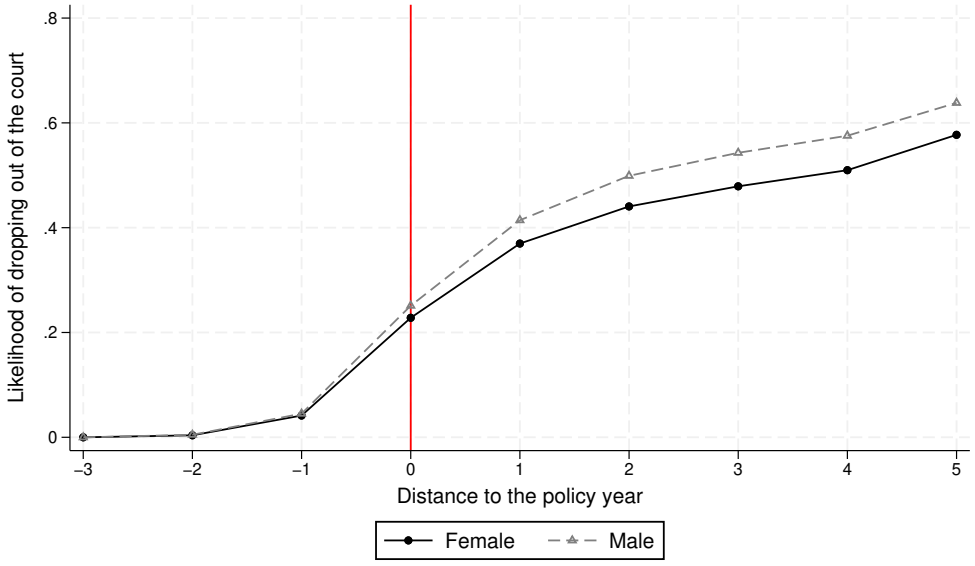
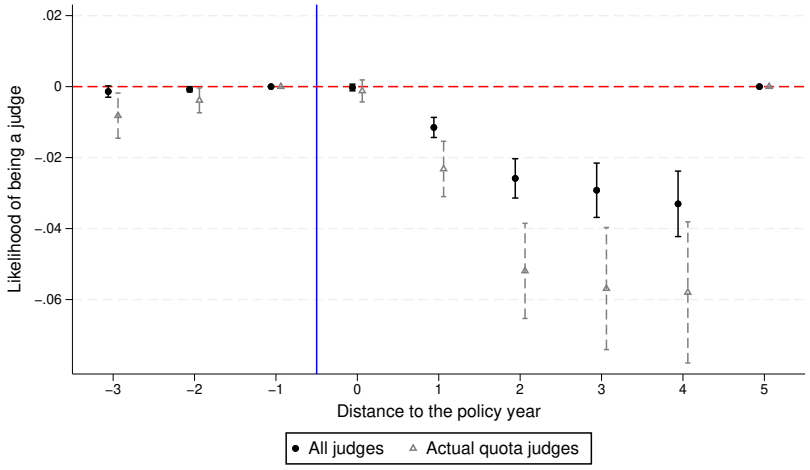


Figure A9 notes: this figure compares the dropout likelihood by gender and over time. We focus on two non-high-ability groups: judges with non-zero reversal rates of first-instance case verdicts, and judges with non-zero appeal rates of first-instance case verdicts.

Figure A10. Gender difference in likelihood of being a judge among stayers

(a) Judges with one missing year



(b) Judges with two missing years

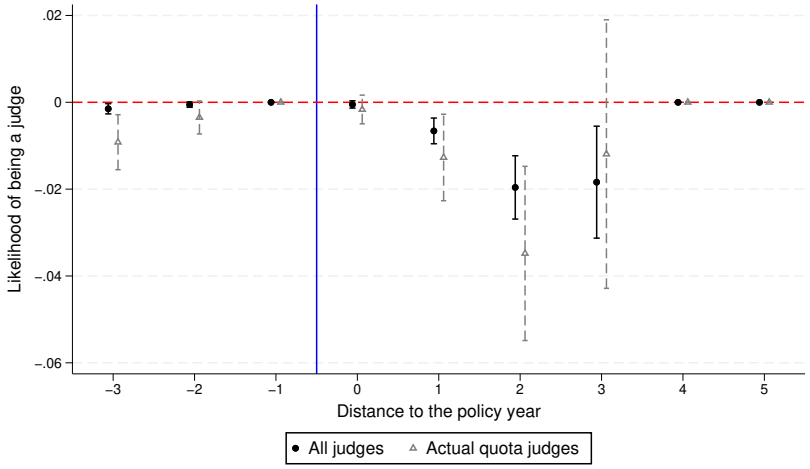
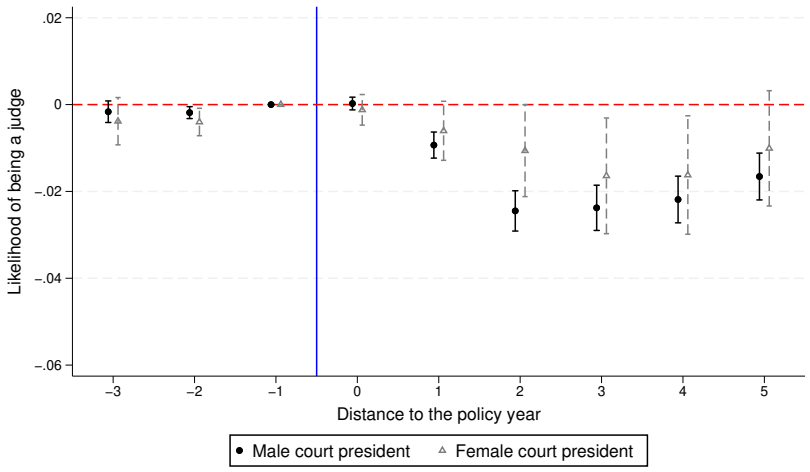


Figure A10 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is likelihood of being a judge in a given year. As indicated in the panel titles, the estimates in Panel (a) were generated by using judges with more than one missing year of appearing in judgment files, and those in Panel (b) were generated by using judges with more than two missing years of appearing in judgment files. We define judges both based on online judgment files and on the information on quota judges collected from government websites. Standard errors are clustered at the judge level.

Figure A11. Gender difference in likelihood of being a quota judge among previous judges: Male vs. female court presidents

(a) Judges identified from judgment files



(b) Actual quota judges

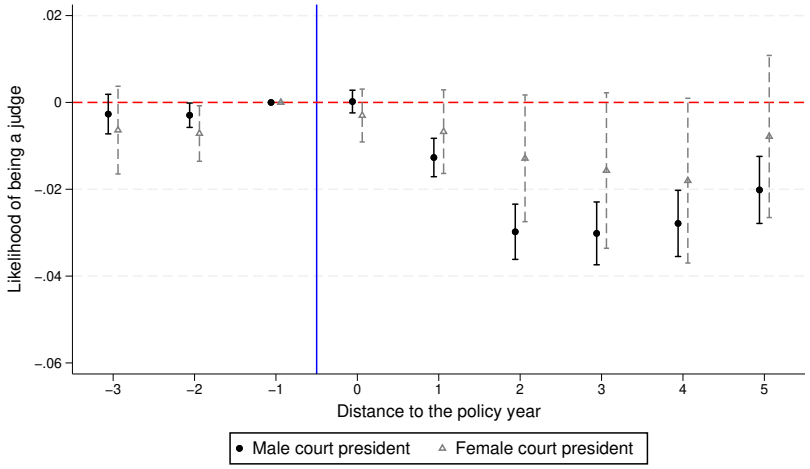
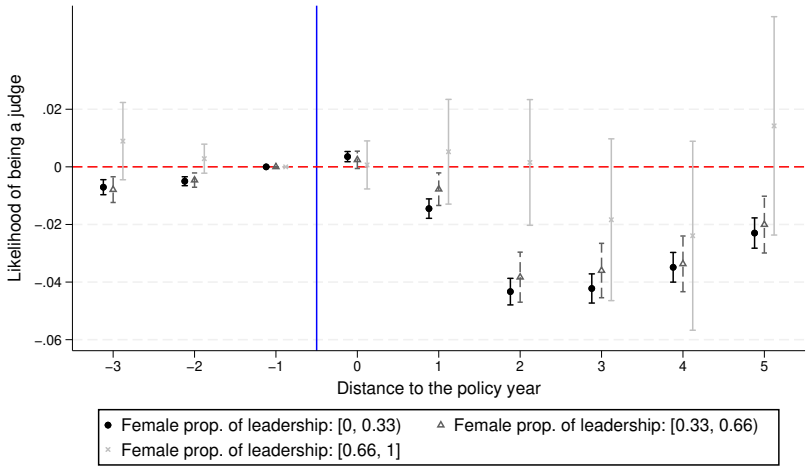


Figure A11 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is likelihood of being a judge in a given year. We compare the gender gap in likelihood of being a judge between courts with a male president and courts with a female president. We use the sample of judges who first appeared in judgment files before the reform. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level. We show the tabular analogue to these results in the last two columns of Table A2.

Figure A12. Gender difference in likelihood of being a judge: heterogeneity in terms of female share of court leaders

(a) Judges identified from judgment files



(b) Actual quota judges

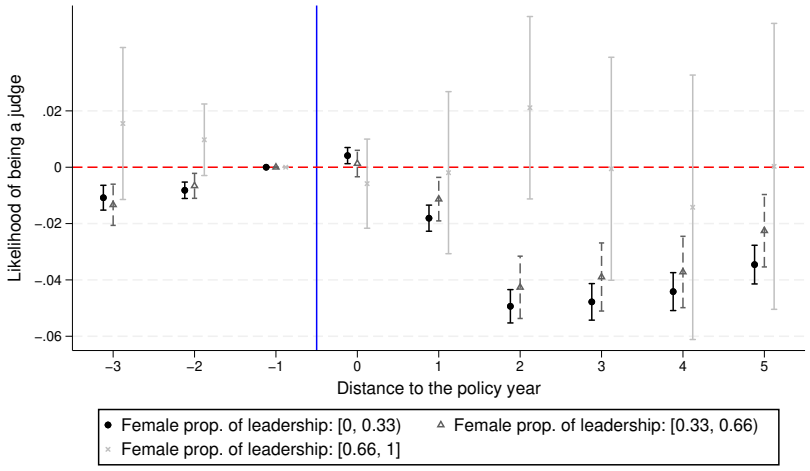
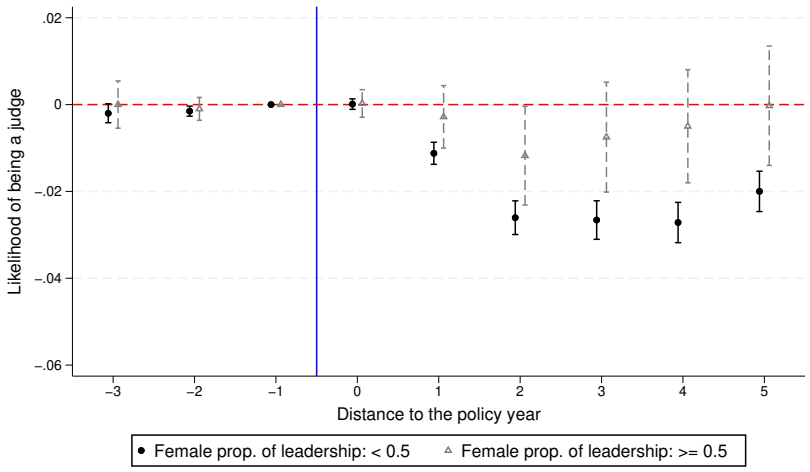


Figure A12 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is likelihood of being a judge in a given year. We show the heterogeneity results in terms of female share of court leadership members. We use the sample of judges who first appeared in judgment files both before and after the reform. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level.

Figure A13. Gender difference in likelihood of being a quota judge among previous judges: A low vs. high female share of court leaders

(a) Judges identified from judgment files



(b) Actual quota judges

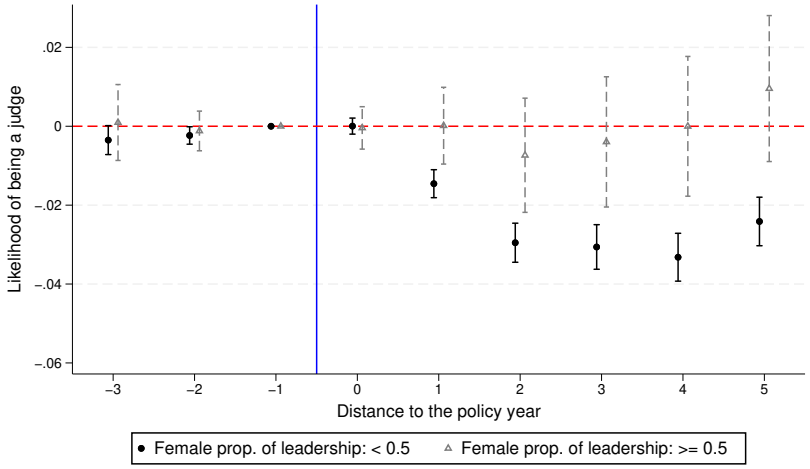
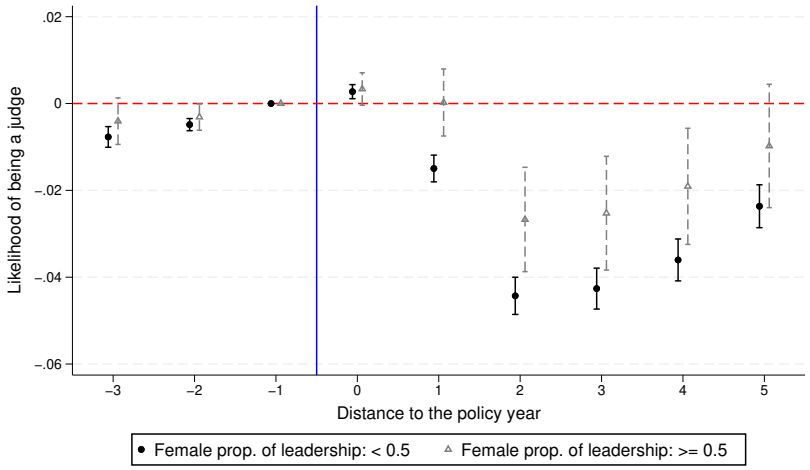


Figure A13 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is likelihood of being a judge in a given year. We compare the gender gap in likelihood of being a judge between courts with a high female share of court leadership members and courts with a low female share. We use the sample of judges who first appeared in judgment files before the reform. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level. We show the tabular analogue to these results in the last two columns of Table A3.

Figure A14. Gender difference in likelihood of being a judge by excluding leaders: A low vs. high female share of court leaders

(a) Judges identified from judgment files



(b) Actual quota judges

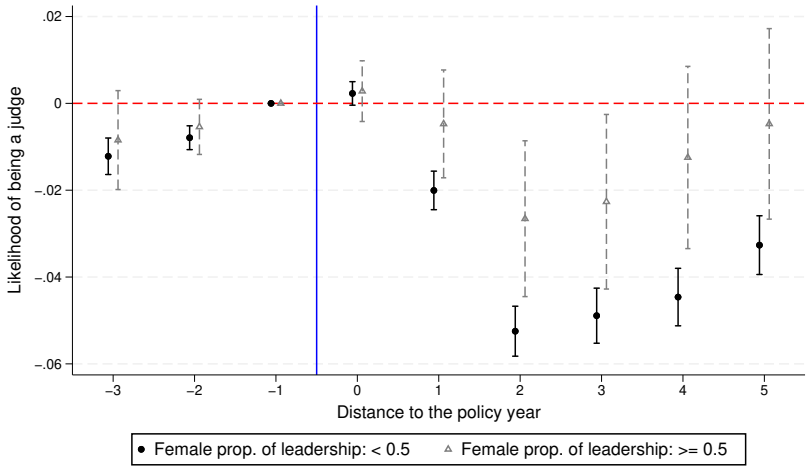
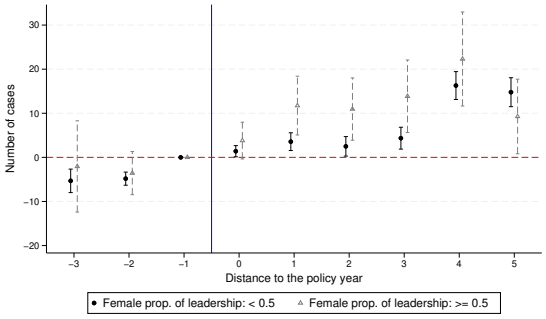


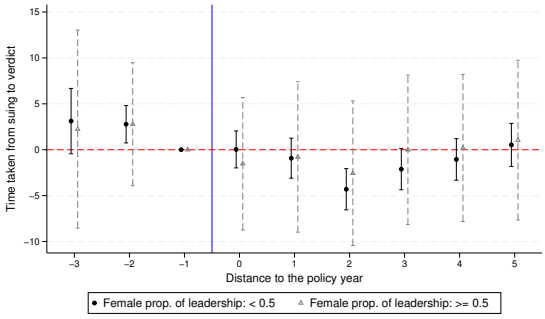
Figure A14 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is likelihood of being a judge in a given year. We compare the gender gap in likelihood of being a judge between courts with a high female share of court leadership members and courts with a low female share. We exclude the sample of court leaders from the analysis. As indicated in the panel titles, the estimates in Panel (a) are based on judge data identified from judgment files, while those in Panel (b) rely on the information on data regarding quota judges collected from government websites. Standard errors are clustered at the judge level.

Figure A15. Gender differences in judge performance: A low vs. high female share of court leaders

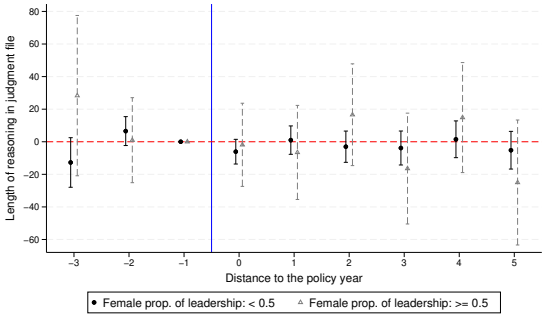
(a) Number of case verdicts



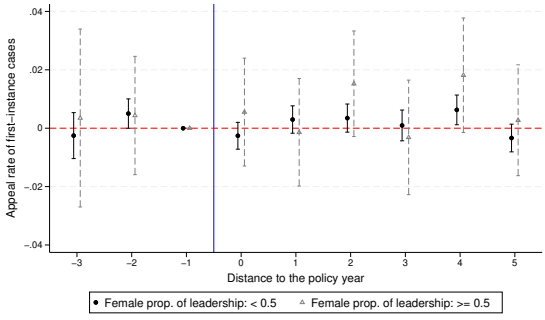
(b) Time for closing a case



(c) Length of reasoning



(d) Appeal rate



(e) Reversal rate

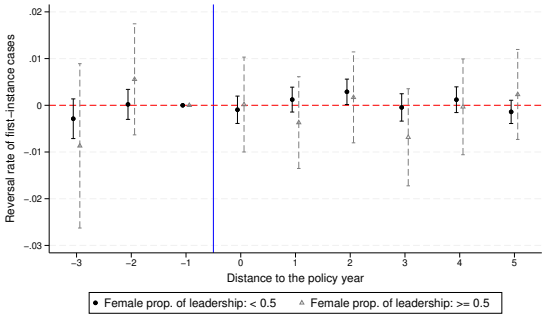


Figure A15 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variables are judge performance indicators, including (a) number of case verdicts, (b) time for closing a case, (c) average length of reasoning in judgment files, (d) appeal rate of first-instance decisions, and (e) reversal rate of first-instance decisions. We compare the gender gap in likelihood of being a judge between courts with a high female share ($>= 0.5$) of court leadership members and courts with a low female share (< 0.5). Standard errors are clustered at the judge level.

Table A1. The reform effect on women's promotion to a judge position

	Judges identified from files			Actual quota judges		
	(1)	(2)	(3)	(4)	(5)	(6)
Female * Post	-0.016***	-0.013***	-0.014***	-0.025***	-0.020***	-0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1313630	1087295	899132	590310	446269	429742
R^2	0.811	0.454	0.452	0.835	0.485	0.484
Sample	All judges	Previous judges	Retained judges	All judges	Previous judges	Retained judges

Table A1 notes: this table presents the estimates of the reform effect on women's likelihood of being a judge. We control for both person and year fixed effects in the regressions. We define a "judge" based on judgment files in the first three columns and define a "judge" based on actual quota judge information collected from government websites. We limit the analysis sample to those who were judges before the reform in columns (2) and (5), and to previous judges who are retained after the reform in columns (3) and (6). Standard errors are clustered at the person level. *** < 0.01, ** < 0.05, * < 0.1.

Table A2. Gender difference in likelihood of being a judge: Male vs. female court presidents

	(1) Judges identified from files	(2) Actual quota judges	(3) Previous judges	(4) Actual + previous judges
Panel A: Male presidents				
Female * Post	-0.015***	-0.020***	-0.012***	-0.016***
	(0.001)	(0.002)	(0.001)	(0.002)
N	426342	240674	349702	185955
R^2	0.824	0.846	0.481	0.507
Panel B: Female president				
Female * Post	-0.010***	-0.010**	-0.006**	-0.005
	(0.003)	(0.004)	(0.003)	(0.004)
N	59519	34141	48820	26572
R^2	0.812	0.837	0.489	0.514

Table A2 notes: this table presents the estimates of the reform effect on women's likelihood of being a judge in terms of the gender of a court president. We control for both person and year fixed effects in the regressions. We define a "judge" based on judgment files in columns (1) and (3) and define a "judge" based on actual quota judge information collected from government websites in columns (2) and (4). We limit the analysis sample to those who were judges before the reform in the last two columns. Standard errors are clustered at the person level. *** < 0.01, ** < 0.05, * < 0.1.

Table A3. Gender difference in likelihood of being a judge: A low vs. high female share of court leaders

	(1) Judges identified from files	(2) Actual quota judges	(3) Previous judges	(4) Actual + previous judges
Panel A: Female prop. of court leaders < 0.5				
Female * Post	-0.017***	-0.022***	-0.013***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)
N	612707	372060	502591	291779
R^2	0.817	0.841	0.473	0.497
Panel B: Female prop. of court leaders >= 0.5				
Female * Post	-0.004	-0.001	-0.003	0.001
	(0.003)	(0.004)	(0.003)	(0.004)
N	43488	24269	35890	19005
R^2	0.861	0.881	0.485	0.509

Table A3 notes: this table presents the estimates of the reform effect on women's likelihood of being a judge in terms of the gender of a court president. We control for both person and year fixed effects in the regressions. We define a "judge" based on judgment files in columns (1) and (3) and define a "judge" based on actual quota judge information collected from government websites in columns (2) and (4). We limit the analysis sample to those who were judges before the reform in the last two columns. Standard errors are clustered at the person level. *** < 0.01, ** < 0.05, * < 0.1.

Table A4. Factors associated with female court leadership

	Being a female president		% of female leaders >= 0.5	
	(1)	(2)	(3)	(4)
GDP per capita (10,000 RMB)	-0.002	-0.002	0.004	0.004
	(0.005)	(0.005)	(0.003)	(0.003)
% of the second industry	0.027	0.021	-0.090	-0.086
	(0.147)	(0.146)	(0.117)	(0.117)
% of the third industry	0.054	0.045	0.135	0.141
	(0.188)	(0.187)	(0.148)	(0.149)
Population (million)	0.002	0.002	-0.004	-0.004
	(0.004)	(0.004)	(0.003)	(0.003)
% of college education	0.442	0.478	-0.147	-0.177
	(0.338)	(0.344)	(0.204)	(0.205)
Municipality	-0.089	-0.090	0.052	0.051
	(0.099)	(0.099)	(0.073)	(0.073)
Sex ratio at birth		0.133		-0.120
		(0.254)		(0.167)
N	1028	1028	1402	1402
R^2	0.003	0.003	0.007	0.008

Table A4 notes: this table presents the correlations between female leadership and local characteristics including GDP per capita, industrial structure (measured by proportions of the second and tertiary sectors in GDP), population size, education levels, and gender norms (measured by sex ratio at birth). The dependent variable in the first two columns is a dummy for having a female court president, while in the last two columns, it is a dummy for courts where more than half of the leaders are female. The analysis is conducted at the court level, and the results indicate no significant correlations between female leadership and local characteristics.

B Analysis of judges selected through the civil servant exam

In this Appendix, we complement the main estimates using the sample of judges who competed for the same positions from the same courts in China's civil servant exam. The judges, therefore, are more comparable than the sample used in the main analysis of this paper. We first describe the data, then show how the judge quota reform affects the gender promotion gap among this group of judges, and finally explore whether gender differences in work performance can explain the widened gender promotion gap.

B.1 Data and summary statistics

Our first dataset comes from the civil servant exam data, which should be made publicly available according to China's Civil Servant Law. We hand-collected these from both official government websites and the website of private companies which collate these data.[33] This dataset contains information on roughly 250,000 final-round job candidates who applied to the same positions and were hired between 2010 and 2018 in 21 of China's 31 provinces. The information includes individuals' traits (name, gender, and education), civil servant exam scores, job positions, and associated institutions.

Our second dataset is the judge dataset used in the main analysis of this paper. We use the judge names and associated institutions to match the two datasets.[34] This yields a dataset containing promotion history and work performance for 5,100 civil servant judges working in about 800 courts, with 56% of them being female.

As shown in Figure B1, at time of hiring, women have higher civil service test scores and are more likely to have graduated from an elite university (as defined by the 211-project). Women also complete more court cases and the quality of their judicial decisions is at least as good as that of men's, as measured by the length of reasoning in the judgments, time for closing a case, the appeal and reversal rate of first-instance cases. However, women are less likely than men to be promoted to higher-rank positions and are more likely to work as judge assistants, a position providing support to judges.

B.2 Court reform and gender promotion gap among judges selected via the civil servant exam

As shown in Figure B2, the likelihood of a civil servant being appointed as a judge increases over time before the reform and women are even slightly more likely than men to become a judge. But some previous judges of both genders are not reappointed as quota judges after the reform, and the likelihood of leaving the position is larger for women than for men. In contrast, women are more likely than men to become judge assistants after the reform, suggesting that the exogenous reduction of available judge positions may hurt women's career development.

Next we use an event study analysis to causally estimate the dynamic effect of the personnel policy on men's and women's promotion likelihood. We estimate an equation similar to 1.

As Panel (a) of Figure B3 shows, women are less likely than men to be judges after the judge quota reform, with the largest gender gap appearing two years after the reform, which is consistent

[33] These include Zhonggong Education Co. (www.offcn.com), Huatu Education Co. (www.huatu.com), and others.

[34] We drop those observations who have the same names and the same institutions, which account for only 1% of the total sample.

Figure B1. Gender differences in court judges' characteristics, performance, and promotion

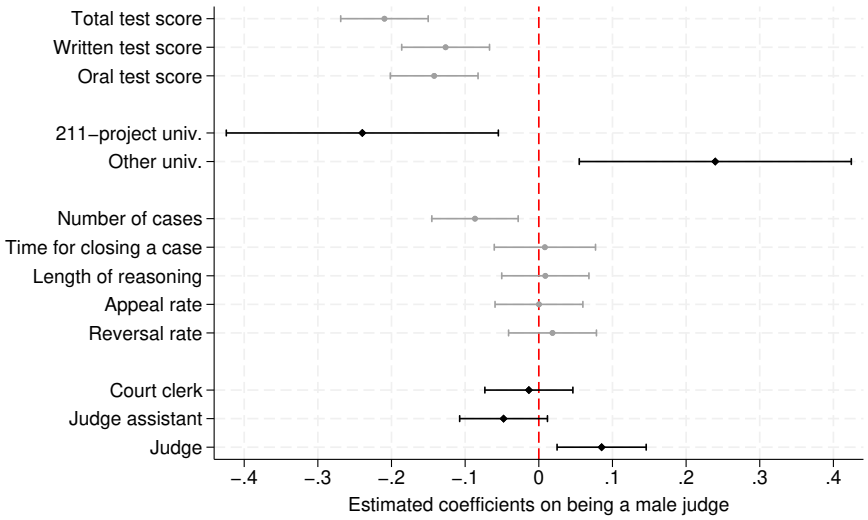


Figure B1 notes: this figure shows gender differences in court judges' background information, work performance, and their promotion. The coefficient estimates are scaled by the standard deviation of outcome variables.

Figure B2. Likelihood of being a court clerk, a judge assistant, and a judge by gender and over time

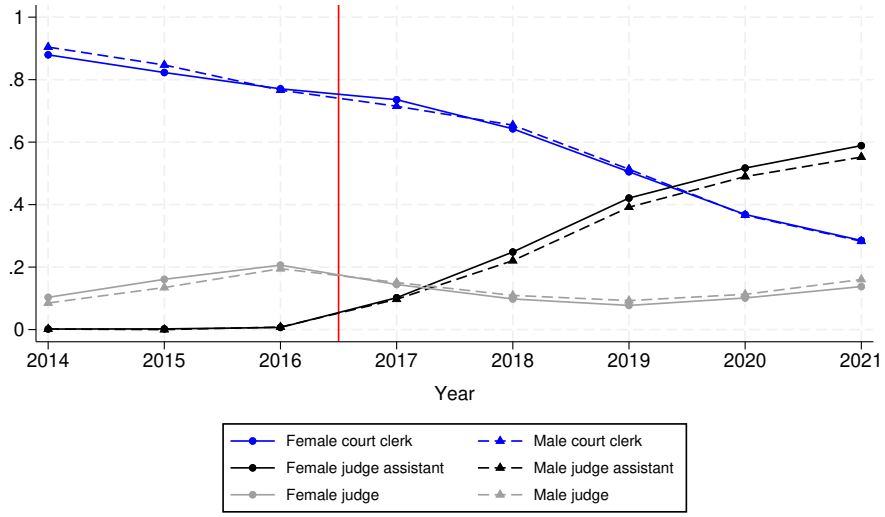


Figure B2 notes: this figure illustrates the likelihood of women and men being appointed as court clerks (entry level), judge assistants, and judges (the highest level) over time.

with our main findings shown in Figure 3. To account for possible compositional changes after the reform, we limit the analysis sample to those who worked as judges before the reform.[35] We still find a larger but imprecisely estimated gender gap in being judges after the reform.

Panel (b) shows how court civil servants are appointed as judge assistants after the reform. As expected, women are more likely than men to be judge assistants after the reform, and the gender gap is even larger when limiting the analysis to those who worked as judges before the reform. Overall, women are disadvantaged in promotions due to the judge quota reform because they are more likely than men to be appointed as judge assistants over judges after the reform.

We further explore the heterogeneity in the reform effect in terms of shock intensity induced by the reform, which requires that the proportion of quota judges should be kept within 39% of the formal judicial personnel of a court. As expected, the gender promotion gap is larger when a female employee works in a court with a higher judge share before the reform and thus experiences a larger reform shock. This suggests that women are more disadvantaged in promotions when more people compete for the limited resources ("judge quota").

B.3 Gender differences in work performance among judges selected via the civil servant exam

In this subsection, we explore whether the widened gender promotion gap is explained by gender differences in work performance among judges selected via the civil servant exam.[36]

As shown in Figure B5, we find that women participate in more case verdicts and the quality of their verdicts is at least as good as that of men's, which is consistent with the findings of our main analysis shown by Figure 6. For instance, as shown in Panel (a) of Figure B5, women have more workload by completing more case verdicts for all three positions: court clerks, judge assistants, and judges. As shown in Panels (d) and (e), women have slightly lower appeal rates and reversal rates of first-instance decisions, the two important metrics used for judge evaluations by China's court system. Therefore, the gender promotion gap is unlikely to be explained by possible gender differences in work performance among judges selected via the civil servant exam.

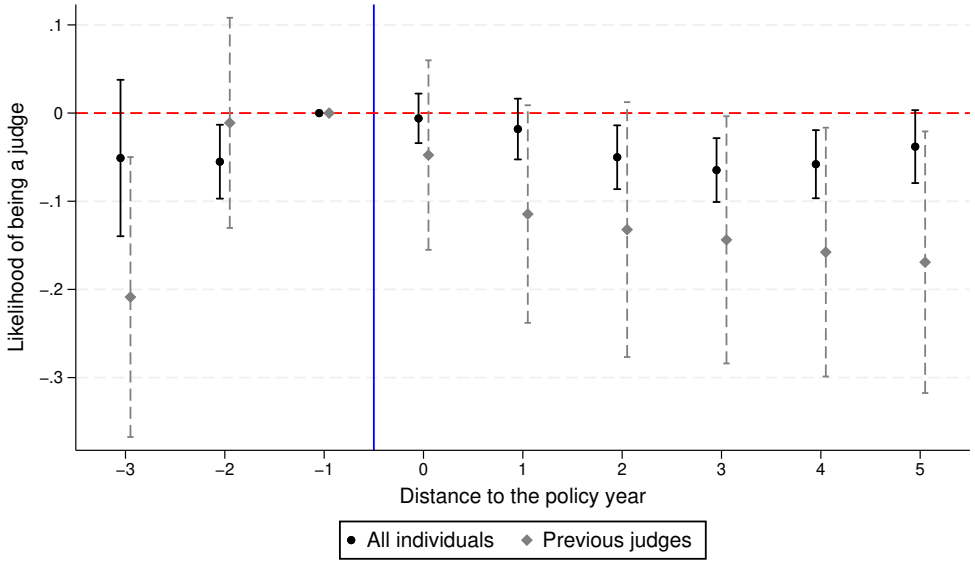
Overall, the estimates from the sample of judges selected through the civil servant exam are consistent with our main results from the much larger sample of judges identified from judgment files, indicating the robustness of the paper's main findings.

[35] Only 335 court employees are left when we limit the analysis to previous judges.

[36] Due to the small sample size of judges selected via the civil servant exam, we do not explore other explanations documented in Section 5 and the role of female leadership in mitigating the gender promotion gap shown in Section 6.

Figure B3. The reform effect on women's promotion likelihood among judges selected via the civil servant exam

(a) Promotion to judge positions



(b) Promotion to judge assistant positions

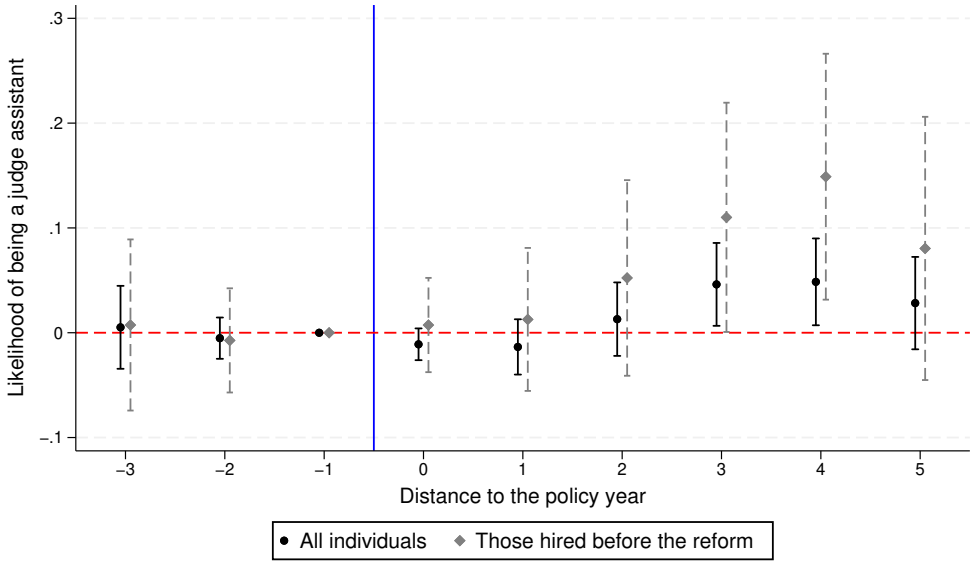


Figure B3 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variable is a dummy variable indicating whether to be a judge or a judge assistant in a given year. Within each panel, we show estimates using both all court employees and those who were judges before the reform. Standard errors are clustered at the judge level.

Figure B4. The reform effect on women's likelihood of being a judge: Heterogeneity in terms of shock intensity

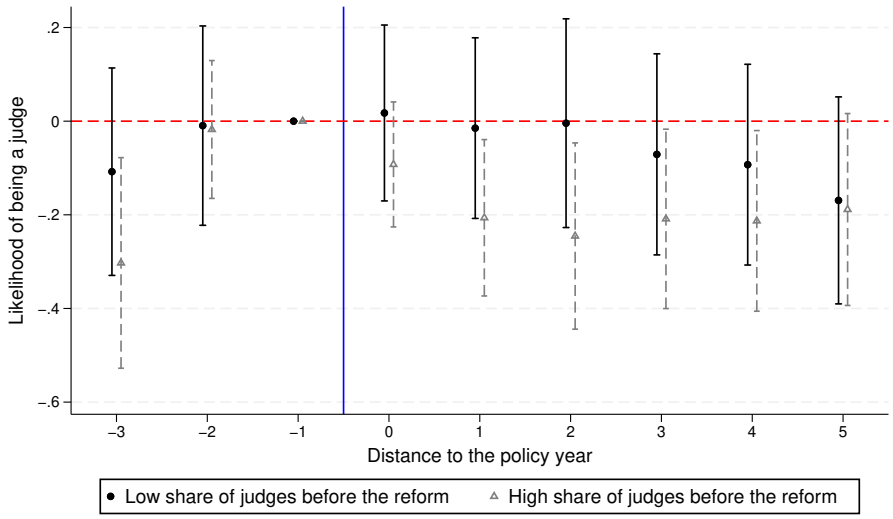
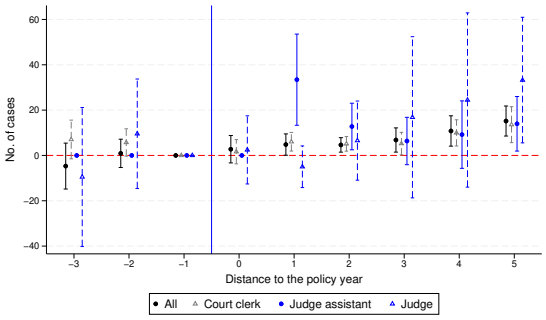


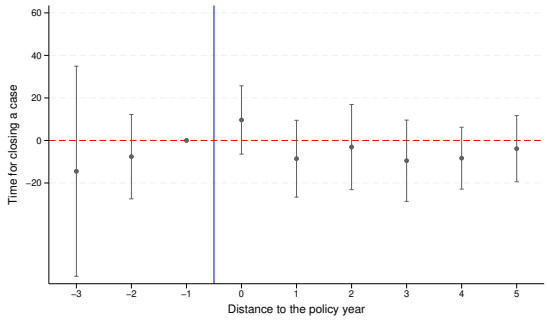
Figure B4 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. We compare the reform effect estimates between those who experienced a small reform shock (working in a court with a low judge share before the reform) and those who experienced a large reform shock (working in a court with a high judge share before the reform). The dependent variable is a dummy variable indicating whether to be a judge in a given year. Standard errors are clustered at the judge level.

Figure B5. Gender differences in judge performance

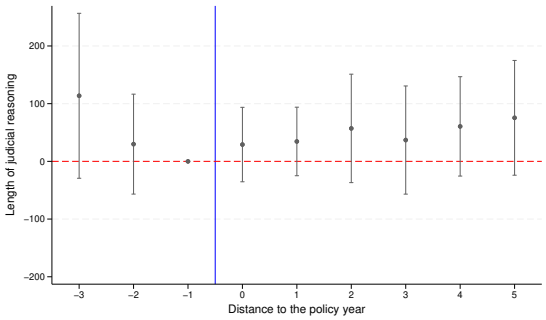
(a) Number of case verdicts



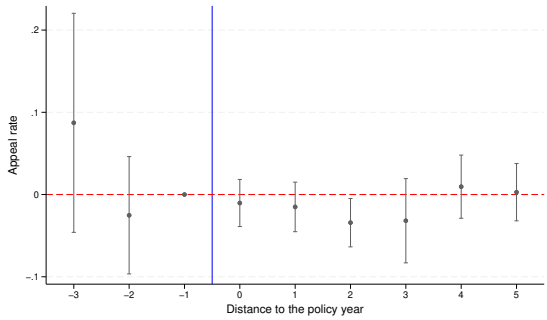
(b) Time for closing a case



(c) Length of reasoning



(d) Appeal rate



(e) Reversal rate

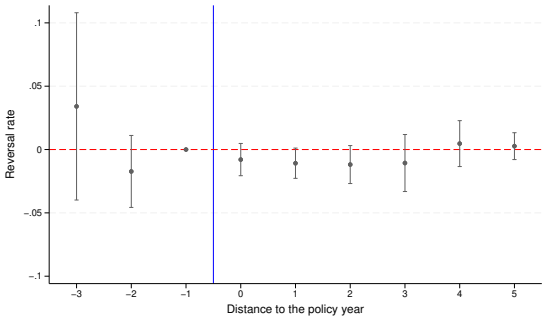


Figure B5 notes: this figure presents estimates of $\alpha_{1\#p}$ from Equation 1, along with 95 percent confidence intervals. The dependent variables are judge performance indicators, including (a) number of case verdicts, (b) time for closing a case, (c) average length of reasoning in judgment files, (d) appeal rate of first-instance decisions, and (e) reversal rate of first-instance decisions. Standard errors are clustered at the judge level.