Algorithm as Manager: How Algorithmic Judge-Case Assignment Influences Court Performance

Rui Zuo*

October, 2024

Preliminary and Incomplete.
Please do not cite or circulate.
Click here for the latest version.

Abstract

Algorithms are increasingly used in high-stakes sectors to manage complex decisions, yet their performance in managerial roles remains understudied. This paper examines the impact of algorithmic case assignment in the Chinese court system, where algorithms have been adopted to enhance fairness and impartiality. Leveraging a natural experiment from 2014 to 2020, during which courts transitioned from human-led assignments to either random or machine learning (ML)-based assignment systems, I analyze 66 million unstructured case documents to assess the effects on assignment patterns, court performance, and judicial effort. Using a generalized Difference-in-Differences approach, I find that random assignment weakens the link between judge experience and case complexity, leading to modest declines in performance and effort. In contrast, ML-based assignment preserves assignment patterns and court performance levels comparable to manual systems. These findings provide novel insights into the role of algorithms in high-stakes, managerial decision-making, extending the literature on both judicial reforms and algorithmic performance in the workplace.

*University of Texas at Austin, ruizuo11@utexas.edu.

Algorithms are being increasingly applied across high-stakes sectors to manage complex decisions. Yet, the performance of algorithms in high-stakes managerial roles is not well understood. In binary decision-making tasks in high-stakes settings—such as predicting recidivism, assessing creditworthiness, and diagnosing disease—algorithms have demonstrated their potential due to their consistency and impartiality. By following structured guidelines, algorithms often outperform humans, reducing the discretion and noise inherent in human decisions (Mullainathan and Obermeyer, 2022; Angelova et al., 2023; Kleinberg et al., 2018; Suhadolnik et al., 2023). However, when algorithms take on the role of work allocation—deciding 'who gets what'—they must not only be impartial but also assign tasks to the most suitable worker based on nuanced context, a task traditionally reliant on human insight and discretion. Whether the strengths of an algorithm's impartiality extend to managerial roles, or whether its lack of sensitivity to nuance limits its effectiveness in high-stakes task distribution, remains an open question.

The court system provides an ideal lens into the use of algorithms, as courts—with their profound impact on society—increasingly rely on these tools to handle complex functions in pursuit of greater impartiality. This paper estimates the impact of two types of algorithmic case assignment systems in courts. Specifically, it examines the shift from human-led case assignments to (1) random assignment, a rule-based algorithm, and (2) machine learning-based assignment, a data-driven algorithm, assessing how this shift influences assignment patterns ("who gets what"), court performance (output), and judicial effort (input). Each approach has distinct strengths and limitations. Manual assignment relies on a senior judge or staff member selecting the most suitable judge based on availability and expertise, accessing detailed case- and judge-specific information but allowing human discretion. Random assignment, by contrast, selects judges based on predefined rules, reducing discretion but disregarding judge and case characteristics, which can lead to mismatches. ML-based assignment adapts to case and judge information, optimizing based on historical case outcomes. However, it may inherit biases from its training data and lack access to the full range of information available to human decision-makers. Given these trade-offs, this study seeks to understand how each assignment system shapes assignment patterns, court performance, and judicial effort.

To address these questions, I leverage the quasi-random rollout of algorithmic case assignment systems across Chinese courts. In 2014, the Supreme People' s Court (SPC) initiated a transition away from human-led assignments to enhance judicial fairness, allowing each provincial court to decide on the timing and type of algorithm. By 2020, most provinces had adopted either a random assignment or ML-based system. Based on provincial court procurement contracts, cross-referenced with over 2,000 news and policy documents, I identify eight provinces that implemented random assignment and 18 that adopted ML-based assignment from 2014 to 2020.

Because all courts are eventually treated, and courts choose either random or ML, I treat the switch to random and ML-based systems as separate treatments and use the staggered Difference-in-Differences (DiD), treating not-yet-adopted provinces as the control group. This approach compares court outcomes in courts of the provinces that adopted random or ML assignment with those that had not yet implemented these systems at various time points. The key identification assumption is that adoption timing is exogenous, which is reasonable since the decision was driven by budget constraints, procurement processes, and provincial leadership preferences rather than pre-existing trends in assignment patterns, performance, or effort. First, to assess assignment patterns change, I examine changes in the correlation between judge characteristics and case complexity post-adoption. Then, I measure the performance and effort change due to the adoption of the algorithm, using a generalized DiD approach following Callaway and Sant' Anna (2021).

To quantify assignment pattern, court performance, and effort, I structuralize 66 million unstructured case documents text data from 2014 to 2021, capturing case characteristics, judge characteristics, verdict quality, and case outcomes. These case documents, hosted by China Judgment Online (CJO), represent the universe of publicly disclosed cases in the designated provinces. To measure assignment patterns, I construct a summary index of case complexity by synthesizing several pre-assignment variables, such as litigation reasons, money involved, and litigant information, following Anderson (2008). To measure the impact of the court's output from cases and input into hearing cases, I create two summary indices, the performance index and effort index, with similar approach. The performance index includes metrics such as the number of cases processed, appeal rates, and reversal rates as they are performance metrics in the court's official reports. The effort index, used by economists and legal scholars to measure effort in legal reasoning, is based on verdict quality, including verdict length, the number of legal provisions cited, and the depth of reasoning.

According to the case assignment patterns, the shift to random assignment prevents experienced judges from handling more complex cases, whereas the ML-based assignment continues to align complex cases with experienced judges, albeit with a weaker correlation. In the manual assignment era, both random and ML courts exhibited similar patterns, where experienced judges were slightly more likely to take on complex cases. After the adoption of random assignment, however, judge experience no longer correlated with case complexity, potentially contributing to a decline in performance. In contrast, the ML system preserved some correlation between experience and complexity, though the strength of this relationship diminished, resulting in more ambiguous effects on performance.

The main results suggest that random assignment reduces both court performance and judicial effort, while ML-based assignment shows no significant change in either. The random courts see a 0.054 decrease in the standardized performance index, primarily driven by higher appeal rates, and a 0.114 decrease in the effort index, linked to shorter verdicts and fewer legal provisions cited. These declines are consistent across court divisions, where the same level of judicial effort translates into weaker performance. This pattern aligns with the disruption of the pre-policy correlation between judge experience and case complexity, as less experienced judges may be ill-suited to handle more complex cases. In contrast, no significant changes were observed in ML courts for either input or output, or across divisions. This lack of change reflects the ML-based system's ability to maintain pre-existing judge-case correlations, with only minor adjustments. The ML system likely replicates manual assignment patterns from its training data, preserving the original relationship between judge experience and case complexity.

These findings are robust to multiple checks, including alternative sample constructions (e.g., excluding courts with less public case disclosure), alternative weighting methods for the index components, and the inclusion of additional controls for judge and court characteristics.

;This paper contributes to multiple strands of literature. First, it adds to the literature on algorithm performance in the workplace, providing insights into algorithms' managerial role in high-stakes settings. Much of the existing research in high-stakes settings focuses on the performance of algorithms versus humans, particularly in areas such as the judiciary, healthcare, and domestic violence prevention (Rittenhouse et al., 2023; Cheng and Chouldechova, 2022; Berk et al., 2016; Agarwal et al., 2023; Mullainathan and Obermeyer, 2022; Berk, 2017; Angelova et al., 2023; Kleinberg et al., 2018). In these contexts, algorithms are typically tasked with binary decisions, such as determining bail or parole outcomes (Berk, 2017; Angelova et al., 2023; Kleinberg et al., 2018), and the literature shows that algorithms often outperform human decision-makers in these tasks. However, there is limited research on algorithms in a managerial role, specifically in complex matching tasks. The algorithm manager (AM) literature primarily consists of qualitative studies that focus on gig platforms, such as food delivery or ride-sharing, where algorithms match workers to tasks in low-skill settings without human assignment as a control group (Becker et al., 2023; Lee et al., 2015; Cram et al., 2020; Jarrahi et al., 2023). One related study by Hoffman et al. (2018) shows that algorithms can make better hiring decisions than humans, but this is also based on a low-skill setting. Research on algorithmic performance in high-stakes, matching-based managerial roles remains scarce, and my paper is the first to quantitatively assess this in the context of court case assignment. My paper extends this literature by evaluating algorithms in a more complex managerial function—case assignment in courts—where the algorithm determines which judge is assigned to which case. I find that, in this context, the algorithm does not strictly outperform humans; instead, it largely mirrors the patterns of manual assignment used in its training data.

Second, this paper contributes to the literature on case assignment systems in the judicial process by quantitatively assessing changes in assignment patterns and suggesting how the assignment system itself can influence case outcomes. The extensive literature on judge fixed effects demonstrates how factors such as ideology, personality, gender, and race influence judicial rulings (Eren and Mocan, 2018; Shayo and Zussman, 2011; Sunstein et al., 2004; Boyd et al., 2010; Harris and Sen, 2019; Glynn and Sen, 2015), often using random case assignment as a quasi-random tool to match judges with varying characteristics to similar types of cases. A smaller body of literature examines case assignment itself, primarily through qualitative case studies that compare random assignment guidelines across U.S. courts or internationally (Gramckow et al., 2016; Fabri and Langbroek, 2007; Macfarlane, 2023; Jin, 2020). An exception is (Chilton and Levy, 2015), which analyzes over 10,000 case assignment outcomes to assess assignment patterns in U.S. federal courts. Building on this literature, my paper quantitatively evaluates the effects of changes in case assignment systems on judicial performance and effort, using a comprehensive dataset of 66 million case documents over six years, in a developing country context. I find that random case assignment disrupts the relationship between judge experience and case complexity but leads to a modest reduction in both court performance and judge effort. Furthermore, I extend this analysis by investigating ML-based case assignment, a newer approach. My results show that this system largely retains prior assignment patterns while maintaining performance and judicial effort at levels comparable to those seen under manual assignment.

Third, this paper contributes to the literature on judicial reforms and their broader economic and legal impacts, particularly in developing countries. Existing studies, such as Liu et al. (2022), demonstrate how reforms that enhance judicial independence reduce local protectionism and promote economic integration in China. Similarly, Mehmood (2022) shows that judicial independence reforms in Pakistan reduce pro-government rulings, while Helmke and Rosenbluth (2009) provide a comparative analysis of regimes and the rule of law. My paper extends this body of work by examining how internal judicial reforms—specifically, case assignment systems—affect judicial performance and effort in a high-stakes institutional setting. By exploring how random and ML-based case assignment systems influence judicial outcomes, this study provides new insights into how court management reforms can enhance legal and economic efficiency, especially in developing contexts.

The remainder of the paper is structured as follows: Section 1 provides background, Section 3 outlines the data, Section 2 details the empirical strategy, Section 4 describes the case assignment systems, Section 5 presents the results and robustness checks and Section 6 concludes.

1 Institutional Background

The case assignment system plays a crucial role in China's legal system. Unlike judicial systems with active jury participation or legally binding precedents, Chinese courts rely solely on judges for decision-making. This highlights the critical importance of who gets assigned to what case, as mismatches could impair court performance. Historically, Chinese courts relied on manual assignment. However, in September 2015, the Supreme People's Court (SPC) published "Several Opinions on Improving Judicial Accountability System of the People's Courts"[1]. This initiative aimed to promote automated case assignment systems to improve judicial fairness by reducing judge shopping. Subsequently, courts gradually transitioned to either random or ML-based assignment systems. All three assignment systems have their advantages and limitations, making the overall impact of the transition unclear. The following subsections outline the case assignment procedures, the three different sys-

[1]Supreme People's Court. "Several Opinions on Improving Judicial Accountability System of the People's Courts." http://gongbao.court.gov.cn/Details/58f02f7ad96f8dcb0e75b8c7e08999.html

tems (manual, random, and ML-based), and the timing and rationale behind their adoption.

1.1 Case Assignment Procedure

Each judge works at only one division and only handles cases within this division. After a case is filed, the case-filing division forwards it to the relevant division based on the type of the case (e.g., family law cases go to the civil division). Within a division, case assignment is carried out manually (old) or via automated systems (new). Courts within each province follow a top-down approach: the provincial court signs one case assignment system procurement contract with one company for all courts within the province. As a result, all divisions within a court, and all courts within a province, use the same assignment system.

1.2 Three Assignment Systems

There are several methods for assigning cases, each with its own advantages and disadvantages. This subsection discusses the decision-making logic behind the three assignment systems and highlights the countries implementing each system outside of China.

Manual In manual assignment, a court staff or the division head manually assigns cases based on perceived judge suitability. This process allows the decision-makers to consider extensive information about judges (e.g., performance, communication skills), which might lead to better matches (positive). However, the reliance on human discretion potentially leads to issues such as judge shopping (negative), as it allows inappropriate factors (e.g., personal connections, gender) to influence assignments. This concern was highlighted in a commentary published by a judge from the Shanghai Second High Court[2]. In 2014, manual assignment was still the predominant case assignment method in China. Internationally, manual assignment remained common in the U.S. until the mid-1990s and was still in practice in some OECD countries as of 2007 (Fabri and Langbroek, 2007).

Random Random assignment eliminates human discretion by assigning cases to judges based on predefined rules, without considering case or judge characteristics. For example, an incoming case is assigned to a random judge with the lowest current caseload. While this system increases transparency (positive), it may lead to poor

[2] Wang Zhigang. Exploration and Practice of Case Assignment System Reform. People's Court Daily, 2016-03-02.

matches (negative), such as assigning a family law judge to a small loan case or a junior judge to a highly complicated case. Random assignment in China began around 2014, following the Supreme People' s Court's guidance to move away from manual assignment. Internationally, the U.S. implemented random case assignment as early as 1995 to promote fairness and minimize manipulation risks, setting a trend that spread globally. By 2016, random assignment was used in most countries, though with many variations (see Figure A1).

ML ML-based assignment combines the advantages and disadvantages of both manual and random systems. An ML algorithm selects judges based on a limited set of characteristics and attempts to optimize case outcomes (positive), but it lacks access to the full set of judge characteristics considered in the manual assignment (negative)[3]. ML also eliminates direct human discretion in the match (positive), but the system may still reflect discretions hidden in the training data (negative). The ML-based assignment system is implemented exclusively in China, introduced over the past decade.

1.3 Timing and Choice of Assignment System

As shown in Figure A3, of the 26 provinces sampled, 8 adopted random assignment while 18 implemented ML-based systems[4]. The geographical proximity of the provinces to algorithm developers likely influenced the choice: northern provinces, for instance, tended to adopt ML-based systems due to the presence of an ML company in Beijing. By 2021, most Chinese courts had completed the transition from manual to either random or ML-based assignment systems[5]. The timing of adoption, as shown in Figure A4, varies at the provincial level[6]. This variation was influenced by multiple factors, including budget constraints, government procurement processes, and provincial leadership preferences. There is no clear correlation between provincial wealth and timing (e.g., Beijing adopted the system relatively late). The data section will provide further details on the definitions and timing of random and ML

[3] While software companies provide details about the algorithms used (e.g., Random Forest, XGBoost), the exact parameters are proprietary. This paper focuses on the outcomes rather than replicating the algorithms.

[4] Five provinces did not disclose their assignment systems and were therefore excluded from the sample.

[5] This study focuses on provinces that implemented new systems between 2014 and 2020, based on data accessibility.

[6] Although anecdotal evidence suggests variation in timing could trickle down to the prefecture level due to technical challenges in installing systems, provincial-level variation is used for consistency.

adoption.

2 Empirical Strategy

The identification strategy utilizes the plausibly exogenous variation in the timing of the implementation of automated case assignment systems across Chinese courts from 2014 to 2020[7]. I use a staggered Difference-in-Differences (DiD) framework by comparing outcomes between treated and control groups before vs. after the treated court divisions' adoption of the random or ML-based assignment. Because all court divisions are eventually treated, I use not-yet-treated court divisions as the control group for already treated court divisions at each point in time. Here, random and ML-based systems serve as separate treatments, to avoid selection bias into assignment types. To be specific, I compare court divisions within each treatment type—not-yet-random court divisions to already random court divisions, and not-yet-ML court divisions to already ML court divisions[8]. The following subsection provides a detailed explanation of my DiD framework. I will first discuss why my setting does not satisfy the assumptions required for unbiased estimation using the traditional Two-Way Fixed Effects (TWFE) method. Next, I will justify the choice of Callaway and Sant' Anna (2021) approach and demonstrate how my setting meets the necessary assumptions. Finally, I will outline the procedure for estimating average treatment effects and event studies using CSDiD, tailored to the context of this study.

2.1 Choice of Generalized DiD

The standard solution for analyzing staggered adoption is the Two-Way Fixed Effects (TWFE) approach. However, TWFE relies on two key assumptions to provide unbiased results—static treatment effects and homogeneous treatment effects

[7]As discussed in Section 1, the timing of policy adoption was influenced by a range of factors, including uncertainties in procurement processes, and budget constraints. Currently, there is no evidence suggesting that these factors correlate with pre-policy outcome variables. An ongoing robustness check is to estimate if provincial characteristics can predict the timing of policy adoption. This analysis will be presented in later versions of the paper.

[8]There is limited evidence showing selection into treatment. Pre-treatment case assignment patterns between random and ML court divisions are similar (see Section 4). The decision to adopt random or ML is partially influenced by factors like geographical proximity to assignment system developers—e.g., ML court divisions are often concentrated in northern China, where the only ML assignment developer is based. To be cautious, I limit the control group to not-yet-treated court divisions within the same treatment type. A pooled control group with not-yet-treated groups across types will be presented in future versions of the paper.

across groups—as highlighted in recent literature (Callaway and Sant' Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021)[9]. In my setting, these assumptions are likely not met. First, TWFE's use of already-treated groups as controls assumes static treatment effects—the treatment impact must remain constant over time. This is a challenge in the context of case assignment, where judges gain experience as they accumulate cases under the new system[10]. Second, TWFE requires homogenous treatment effects across different groups, which is unlikely given that court divisions in various provinces handle cases with distinct characteristics and have diverse judicial personnel, leading to heterogenous effects[11]. Differences in random and ML assignment systems calibration across court divisions can further introduce heterogeneity. To address these concerns, I use the Difference-in-Differences (DiD) method developed by Callaway and Sant' Anna (2021), which accommodates multiple time periods and variation in treatment timing.

The reliability of CSDiD depends on several key assumptions:

Irreversibility of Treatment Once a province adopts a case assignment reform, it remains treated. This is consistent with the context of case assignment systems in China, where provinces rarely revert to manual assignment after transitioning to random or ML-based systems[12].

Random Sampling Each observation should be randomly drawn from the population of interest. My dataset consists of panel data covering court performance, judicial effort, and case characteristics for every division within all court divisions over the specified timeframe (6 quarters pre-policy and 8 quarters post-policy). This dataset includes the universe of publicly disclosed cases, ensuring representative coverage.

[9] Sun and Abraham (2021) demonstrate that heterogeneous treatment effects can create pretrends, complicating interpretation. Additionally, Goodman-Bacon (2021) shows that TWFE only assigns positive weights when treatment effects remain constant; varying effects can skew results towards groups with larger sample sizes, especially those treated mid-study.

[10] If new case assignment systems improve performance, their benefits could build as judges grow more adept. Conversely, if these systems hinder performance, the effects might lessen as judges adjust or worsen if mismatches increase over time.

[11] In staggered designs involving multiple groups and time periods, assuming homogenous treatment effects is often unrealistic (De Chaisemartin and d' Haultfoeuille, 2020).

[12] Even if reversals occur, they are limited to specific small divisions within court divisions and not on a large scale. However, provinces may upgrade or switch to new assignment systems after the expiration of the current contracts, which typically last for two years. During my study's post-policy period (eight quarters), the treatment is stable.

Limited Treatment Anticipation The assumption is that provinces do not alter their assignment patterns in anticipation of reform implementation. This assumption appears reasonable, as adoption timing was driven by a mix of factors—budget limitations, procurement logistics, and preferences of provincial leaders—rather than pre-existing trends in performance or effort.

Conditional Parallel Trends The identification strategy requires that, conditional on observable covariates, treated and not-yet-treated court divisions would have followed similar trends in the absence of the reform. The event studies presented in Section **??** shows the absence of pre-trends. Not-yet-treated court divisions serve as the control group. Although displaying raw parallel trend plots for all court divisions would result in 38 distinct plots for random and ML court divisions, I provide examples of trends for court divisions treated in 2018 Q4 for both random and ML systems in Figures B2 and B3.

Common Support There should be sufficient overlap in characteristics between treated and control groups, ensuring that treated and control court divisions have comparable profiles. This condition is met, as the control group consists of court divisions in not-yet-treated provinces. Given the substantial variability in court performance, judge attributes, and case features within a single province, the earlier treated court divisions are likely to find suitable comparisons among the later treated court divisions.

In the not-yet-treated as comparison group, the group time average treatment effect (ATT) is:

$$ATT(g,t) = \mathbb{E}\left[Y_t - Y_{g-1} \mid G_g = 1\right] - \mathbb{E}\left[Y_t - Y_{g-1} \mid D_t = 0, G_g = 0\right] \tag{1}$$

where $g \in (2014\text{ Q3}, 2018\text{ Q4})$ is the quarter the court divisions was first treated (for example, 2016 Q2 corresponds to the court divisions first treated in second quarter of 2016, and so on), $t \in (2014\text{ Q1}, 2018\text{ Q4})$ is a quarter, G_g is a binary variable that equals to one if a unit is first treated in quarter g and zero otherwise. D_t is a binary variable that equals to one if a unit is treated in quarter t. Here, the first term shows the difference of outcome variable of the treated court divisions between period t and the period before first-treated, $g-1$. The second term shows the difference between the court divisions not-yet-treated by quarter t and themselves one period before the treatment group's first period. The ATT is a weighted average of the group time

ATT:

$$\theta = \frac{1}{\kappa} \sum_{g \in G} \sum_{t} \nVdash\{t \geq g\} ATT(g,t) P(G = g \mid G \leq 2020 \text{ Q4}) \tag{2}$$

where $ATT(g,t)$ is defined in Equation , and $\kappa = \sum_{g \in G} \sum_t \nVdash\{t \geq g\} P(G = g \mid G \leq 2020 \text{ Q4})$. κ ensures that the weights on $ATT(g,t)$ in the second term sum up to one. This weighting method ensures positive weights and gives larger weight to group with more court divisions. The event study specification is:

$$\delta_e = \sum_{g} \sum_{t} \nVdash\{t - g = e\} P(G = g \mid t - g = e) ATT(g,t), \tag{3}$$

where δ_e represents the average treatment effect e periods after adoption, calculated across all groups that participated in the treatment for exactly e periods.

3 Data

This study relies on government procurement data to identify treatment timing and type and 66 million case documents to analyze case assignment systems, judge characteristics, and case outcomes. Below, I detail the sources of data, data processing methods, and key summary statistics.

3.1 Case Assignment System Adoption Data: China Government Procurement Online (中国政府采购网)

I collected contracts between provincial high courts and technology firms to identify when and what type of automated case assignment systems were adopted. The adoption date is set as the start date of the contract, rounded to the first day of the corresponding quarter for standardization. This timeline is cross-referenced with local policy documents and news reports to ensure accuracy. Notably, all divisions within a court and courts within a province follow the same adoption schedule.

To determine the type of assignment system, I examined the technical specifications from contracted firms' websites. Based on the algorithms listed, I classified each court into random or ML-based assignment categories. Figure A4 shows the quarterly distribution of courts adopting random and ML systems, indicating a relatively even spread over time.

3.2 Case Documents: China Judgements Online (中国裁判文书网)

China Judgements Online (CJO) is a centralized platform, launched by the SPC in 2013, that hosts publicly accessible judicial documents. According to SPC guidelines, all judgment documents must be uploaded unless exempt for legal reasons[13][14]. As of January 2023, CJO contained 120 million case records.

3.2.1 Sample Selection

I focus on 26 provinces where the timing and type of automated system adoption are confirmed. Some provinces, while known to have adopted automation by 2020, lack precise adoption details and are therefore excluded from the analysis. The study period spans from January 1, 2014, to December 31, 2020, covering the full rollout of automation. My sample includes civil, business, criminal, and administrative cases, comprising 85% of the total caseload—resulting in a final dataset of 66 million cases[15].

3.2.2 Raw Case Document

Each CJO document consists of four sections: title, facts, verdict, and closing. These sections provide details on case characteristics, judge characteristics, and case outcomes, which are extracted as follows:

Title The title includes the case name, unique case ID, and court ID. From this, I extract the court name, case ID, filing year, and case type (e.g., civil, business).

Fact The facts section covers litigant details and case background. I extract litigant names, the number of litigants, their identity (individual or corporate), the monetary amount in dispute, cause of action (COA), and related case IDs.

[13] This document can be found at http://gongbao.court.gov.cn/Details/d0e837bbafb75a8863b4d4c407d694.html

[14] Some cases are restricted from public access, such as divorce or juvenile cases, but the platform still displays basic case details and reasons for non-disclosure. This allows calculation of transparency rates for different courts. For robustness, I use courts with disclosure rates above 85% and 90% as alternative samples, and results remain consistent.

[15] Law enforcement cases, accounting for 15% of total caseloads, are excluded as they do not involve judge decision-making but rather execution of court rulings, which falls outside the scope of traditional judicial reasoning.

Verdict The verdict section contains judicial reasoning and the final decision. Key extracted elements are verdict word count, legal provisions cited, and provisions written in the decision.

Closing This section records judge(s) and clerk(s)' names and the decision date. I use judge names, case type and court ID to construct judge-level panel data and court-division-quarter panel data, enabling analyses of both assignment patterns and case outcomes[16][17].

3.2.3 Variables for Case Assignment Patterns and Case Outcomes

To assess case assignment patterns (Section 4) and case outcomes (Section 5), I use summary indices for case complexity, judge/court-division performance, and judge/court-division effort, reducing dimensionality and offering a clearer overview. Following Anderson (2008), summary indices are beneficial for minimizing over-testing risks and providing a comprehensive view of general effects. I aggregate variables by 1) standardizing signs so that higher values indicate better outcomes, 2) normalizing based on pre-treatment standard deviations, and 3) weighting by the inverse of the covariance matrix[18].

The case complexity index, inspired by Gramckow et al. (2016), includes the number of litigants, the percentage of corporate litigants, disputed monetary amounts, and the number of COAs, indicating the workload involved[19]. For performance index, I use number of case processed, appeal rate, and reversal rate. This selection of performance metrics follows SPC's 2011 official guideline on assessing performance and Kahn and Li (2019)[20]. For judicial effort index, I draw from verdict word count, provisions cited from existing law, and provisions written in the decision, guided by

[16]Judges rarely move between courts in China, and the combination of judge name and court ID serves as a reliable identifier due to the unique nature of Chinese names.

[17]Court divisions are not explicitly specified in raw documents. I proxy divisions using case types and COAs, as case processing standards vary across divisions, e.g., civil vs. anti-trust. This granularity allows for normalization of outcome measures within divisions.

[18]This method down-weights highly correlated variables and emphasizes variables with unique information. I also conducted robustness checks using equal weights, yielding consistent results.

[19]Admittedly, the complexity index cannot fully capture the real workload. Some cases might appear straightforward initially but become more intricate during the hearing process. The aim of the complexity index is to reflect the court's initial assessment of case difficulty, which influences assignment decisions. Therefore, I limit the index to pre-assignment variables, avoiding any bias introduced by judges' characteristics that might alter the perceived complexity after the case is assigned.

[20]This document can be found at http://www.court.gov.cn/zixun-xiangqing-2298.html.

the SPC’ s standards and (Liu et al., 2022; Liu, 2018)[21].

3.2.4 Summary Statistics

Table A1 and A2 compares pre-treatment averages for courts using random and ML systems. Columns (1) and (2) present averages for random and ML courts, respectively, while column (3) reports the differences. Random courts show higher caseloads and slightly lower appeal and reversal rates, implying better performance. They also produce longer and more complex verdicts. Panel C reveals higher caseload inputs and fewer judges in random courts, which handle more civil cases than ML courts. These differences hints potential selection bias, reinforcing the need to treat not-yet-treated courts within the same category as controls for each treatment type.

4 Who Gets What Case?

Before examining how the change in case assignment patterns impacts court performance and effort, it’ s helpful to start with a descriptive analysis of judge-case assignments in the current system. I run a regression of case complexity as the outcome and judge's characteristics as explanatory variables, with an interaction term of the policy. This analysis provides insight into how the judge-case correlation shifts post-policy and hints at the expected direction of changes in court performance and effort. The findings are as follows: (1) The correlation between judge characteristics and case complexity is small but statistically significant, with all correlations below 0.1 on the standardized complexity index. (2) Before the policy, both random and ML court divisions display similar case distribution patterns. (3) During the policy, random assignments reduce the correlation between judge experience and case complexity while negatively correlating historical judge performance with complexity. In ML courts, there is a decrease in the correlation between experience and case complexity, with a tendency to assign more complex cases to female judges. (4) After the policy, random court divisions show no correlation between judge experience and case complexity, though female judges are more likely to receive complex cases. In ML court divisions, the directions of judge-case correlations remain consistent, with positive, negative, and insignificant relationships unchanged. The following subsections will go into the methods and detailed findings associated with these observations.

[21]This document can be found at http://gongbao.court.gov.cn/Details/25a9b4684d384ea16f78e276f14f13.html.

4.1 How Are Cases Assigned to Judges?

The case assignment process for judges in China remains a black box. Existing studies typically rely on fieldwork within individual courts to understand assignment patterns, focusing on variations across courts and potential harm from human's discretion (Gramckow et al., 2016; Fabri and Langbroek, 2007; Macfarlane, 2023; Jin, 2020). In a manual assignment system, decisions are generally made by court staff or division heads, who aim to match cases with judges based on perceived suitability and current workload. However, there is no comprehensive quantitative analysis of judge-case assignments in China, leaving questions about what constitutes an "appropriate" assignment largely unanswered.

The introduction of random and machine learning (ML)-based assignment systems is expected to influence judge-case correlations in different ways. A random assignment system ideally eliminates correlations between judge characteristics and case complexity. This randomness disrupts established connections between characteristics like experience, historical performance, and effort, and the types of cases judges handle. Conversely, an ML-based system operates based on its training data, aiming to optimize assignments. This approach may preserve or even strengthen historical patterns by assigning cases algorithmically, or it may make slight adjustments intended to improve judgment quality based on predictive modeling.

To analyze these potential changes, I estimated the following regression model on a 10% random sample of cases:

$$\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt} \quad (4)$$

In this model, case_complexity_i represents the complexity index for case i, calculated from pre-assignment factors such as the cause of action, monetary stakes, and litigant composition.[22] The binary variable Post_t indicates whether quarter t occurs after the policy's implementation, distinguishing pre- and post-policy periods. The vector judge_char_{jt} includes characteristics of judge j at time t, such as gender, experience (normalized), historical performance, and historical effort.[23] The interac-

[22] The complexity index reduces the dimensionality of multiple indicators into a single measure, based on Anderson (2008). All variables are determined before assignment to ensure they are exogenous to the judge's influence. Alternative dimensionality reduction methods like PCA yielded principal components with less than 30% explanatory power.

[23] Historical performance and effort indices are calculated as averages over prior quarters, following the method in Anderson (2008).

tion term $(\text{judge\_char} \times \text{Post})_{jt}$ captures any shifts in the relationship between judge characteristics and case complexity due to the policy. Standard error is clustered at the provincial level, as the main regression.

The key coefficients are β_2 and β_3. The coefficient β_2 represents the baseline correlation between judge characteristics and case complexity under the previous system. In contrast, β_3 reflects changes in this correlation following the implementation of random or ML-based assignments. By analyzing these coefficients, we can determine the extent to which each system changes, preserves, or disrupts existing assignment patterns.

It is worth noting potential limitations of this difference-in-differences (DiD) approach. As explained in the method section, staggered policy implementation across provinces may introduce bias in traditional DiD estimates. Although generalized DiD methods (e.g., Callaway & Sant' Anna, 2021) address staggered adoption, they focus on β_3 and do not provide information on β_2, which is essential for understanding baseline correlations. Furthermore, generalized DiD methods often require binary variables, limiting the precision of continuous judge characteristics. While CSDiD methods may offer additional insights, they require substantial data preprocessing, which risks perceptions of data manipulation. For this analysis, the combination of binary (gender) and normalized continuous variables (indices and experience) is expected to capture judge-case correlation changes effectively.

4.2 Pooled Regression from All Court-divisions

Pre-policy Judge-Case Correlation Figure A6 shows the baseline correlations between judge characteristics and case complexity before random and ML-based assignment, represented by β_2. These correlations are small but significant: a one standard deviation (SD) increase in historical effort or performance is associated with less than a 0.1 change in case complexity, indicating a modest link between judge attributes and the types of cases assigned. Correlations are consistent across both court systems.

Specifically, judges with higher historical effort are more likely to handle complex cases, while those with better historical performance are assigned simpler cases. Experience is positively correlated with case complexity, suggesting that more experienced judges are given more challenging cases. Gender (Predicted Female) shows no significant correlation with case complexity, indicating a gender-neutral case assignments across both systems. These pre-policy correlations establish a baseline to assess how each assignment system affects judge-case patterns post-policy.

Change in Judge-Case Correlation Figure A7 illustrates the changes in judge-case correlations after policy implementation, represented by β_3. The effects are modest, with changes under 0.05 in standardized case complexity, but distinct across the systems. In the random court, experience is no longer correlated with complexity, meaning experienced judges are no longer assigned more complex cases. Historical performance remains negatively associated with complexity, suggesting high-performing judges continue to receive simpler cases. In the ML court, the experience-complexity correlation also weakens, reflecting a partial disruption of the pre-policy pattern. Additionally, a significant positive change in the female-complexity correlation indicates that female judges are now more likely to be assigned complex cases.

Post-Policy Judge-Case Correlation Figure A8 presents the combined post-policy correlations, $\beta_2 + \beta_3$, for both systems. In the random court, the experience-complexity correlation is now insignificant, showing that experience no longer influences case complexity assignments. In contrast, the ML court maintains a weakened, but positive, experience-complexity correlation, suggesting partial retention of pre-policy patterns.

Overall, while both systems modify judge-case correlations, random assignment disrupts traditional patterns more strongly, particularly by removing the experience-complexity link, while ML assignment tends to preserve existing relationships with some adjustments. This supports the expectation that random assignment breaks down prior patterns more thoroughly, whereas ML-based assignment retains elements of the initial structure.

4.3 Judge-Case Correlation by Division

The judge-case correlations vary across divisions, reflecting the unique characteristics and types of cases handled by each. In crime divisions, where decisions are typically evidence-based and judge discretion is less emphasized (particularly in civil law countries), there is no significant pre-policy correlation between judge characteristics and case complexity. Following the adoption of random and ML-based assignments, these divisions exhibit minimal changes in these correlations, indicating that the assignment systems have little impact on how cases are allocated based on judge attributes in both ordinary and economics-related crime divisions.

The civil division, which primarily deals with high-volume, lower-stakes cases like family and small loan cases, shows different patterns. Before the policy change, more experienced and female judges were more likely to be assigned complex cases. However, the introduction of random and ML systems disrupts this pattern, eliminating the correlation between judge experience and case complexity. This suggests that the new assignment systems neutralize the previous bias toward assigning complex cases to judges based on experience or gender.

In business and civilian vs. government divisions, which handle high-stakes cases such as intellectual property disputes and government-related housing demolition cases, the pre-policy period shows that judges with lower historical performance are assigned more complex cases. After the implementation of the new systems, these divisions experience a reduced correlation between experience and case complexity, indicating that both random and ML-based systems weaken the tendency to assign challenging cases to more experienced judges.

Overall, the impact of the assignment systems differs across divisions, with crime divisions showing little change, while civil, business, and civilian vs. government divisions experience a breakdown of traditional judge-case assignment patterns, particularly regarding experience and case complexity.

5 Result

The adoption of random and machine learning (ML) judge-case assignment systems in courts yields distinct outcomes regarding court's performance and effort. In court-divisions that switched to random assignment, both performance and effort experience a measurable decline, with performance decreasing by 0.054 and effort by 0.114 on the standardized index. By contrast, court-divisions adopting ML-based assignment systems do not exhibit statistically significant changes in either performance or effort. The effects of random and ML assignment remain consistent across all types of divisions. The following subsections will provide detailed discussions of each outcome.

5.1 Random Courts

Main Results The event study diagrams in Figure A9 show no significant pre-treatment trends, indicating stability in performance and effort indices prior to the policy change. The left panel of Figure A1 illustrates the estimated effects on the performance index, while the right panel displays the results for the effort index. Both diagrams show that the estimates for the five pre-treatment periods are close to zero and statistically insignificant.

After the implementation of random case assignment, the performance index exhibits a delayed but significant decline starting in the third quarter, with this negative trend persisting in subsequent periods.[24] Table A3, Panel A, provides the point estimates for the performance index (column 1), as well as the normalized (columns 2–4) and raw outcome variables (columns 5–7). After the policy intervention, the performance index declines by 0.054, statistically significant at the 1% level. This decline is primarily driven by the normalized non-appeal rate, which decreases by 0.110 standard deviations, significant at the 10% level, corresponding to an increase of approximately six additional appeals per court-division per quarter, significant at the 5% level.[25]

In contrast, the effort index shows an immediate decline post-policy, which stabilizes around zero in subsequent quarters.[26] This index, measuring judicial input, reflects the adaptation phase judges undergo. Figures C2–C4 display the event study diagrams for each component of the effort index: normalized verdict word count, normalized number of legal provisions cited, and normalized provisions written. The absence of significant pre-treatment trends for these variables supports the conclusion that the observed declines are attributable to the policy change.

Finally, Table A3, Panel B, quantifies the decline in the effort index, showing a reduction of 0.114 standard deviations, significant at the 1% level, primarily due to decreases in verdict word count and provisions cited. Specifically, the normalized verdict word count decreases by 0.097, equivalent to approximately 46.533 fewer characters per case per quarter, significant at the 1% level. Additionally, the normalized number of provisions cited decreases by 0.097 standard deviations, corresponding to

[24]This delayed effect is expected because the performance index comprises variables such as processed caseloads, appeal rates, and reversal rates. Changes in appeal and reversal rates tend to occur gradually; litigants typically need time to consider the costs and potential benefits of appealing a decision. Similarly, a reversal by a higher court implies that substantial errors are identified in the original judgment.

[25]While the performance index provides a general overview of court output, it is important to recognize that appeal and reversal rates can be influenced by multiple factors. For instance, litigants might be more inclined to appeal if they perceive higher courts as less corrupt post-policy implementation. Alternatively, if litigants perceive the new judges as less experienced or less professional, they may be more inclined to file appeals. Here, the performance index is used as a pragmatic measure of court performance, somewhat analogous to a company metric, where a higher appeal rate suggests increased cost and resources needed to resolve cases.

[26]The initial decrease in effort is consistent with a learning curve, as judges adapt to handling cases assigned under the new random system. Initially, they may produce shorter verdicts as they get acquainted with new case types, but over time, they develop familiarity and adjust accordingly.

0.096 fewer provisions per case per quarter, also significant at the 1% level.

Heterogeneity by Divisions Figure A11 indicates a reduction in efficiency across divisions: those with stable effort levels exhibit lower performance, while divisions with increased effort do not show corresponding performance improvements. In Figure A11, the left panel presents the heterogeneous effects of random assignment on the performance index across five divisions, with most effects being either negative or not significantly different from the pre-period. Notably, the civil division, which handles the highest caseload, experiences a significant decline of 0.13 in the standardized performance index. The right panel depicts the effects on effort by division, showing that while the economic-related crimes division (handling high-stakes cases) records a slight increase of 0.1 in the standardized effort index, other divisions exhibit no significant change in effort. Together, these results suggest that the random assignment system generally leads to decreased efficiency across divisions, with increased effort failing to yield corresponding gains in performance.

5.2 ML Courts

Main Results The machine learning (ML) case assignment system shows no significant impact on either judicial performance or effort, suggesting stability across both indices post-implementation. As illustrated in the event study diagrams (Figure A10), the left panel displays the effects on the performance index, while the right panel shows the effects on the effort index. Both diagrams indicate that there are no significant deviations from zero across all quarters, with no discernible pre-treatment trends. This stability implies that ML assignment does not disrupt the performance and effort levels in the same way that random assignment does.

In further detail, Table A4, Panel A, provides point estimates for the performance index and its component variables. The overall performance index increases slightly by 0.025, but this effect is statistically insignificant. The normalized variables—such as the number of cases processed, non-appeal rates, and reversal rates—also show no significant changes, supporting the conclusion that ML assignment has a neutral effect on court output.

Similarly, the effort index results, as shown in Table A4, Panel B, indicate an insignificant change of -0.010, suggesting that judicial input remains largely unaffected by the policy. The table also shows minor and statistically insignificant fluctuations in the normalized variables that compose the effort index, such as verdict word count and the number of provisions cited, which further confirms the consistency in judicial effort under ML assignment.

Heterogeneity by Divisions The heterogeneous analysis by division, presented in Figure A12, underscores the absence of significant effects across different types of cases. Both the performance and effort indices remain close to zero across all divisions—including economic-related crime, ordinary crime, business, civil, and civilian vs. government cases—highlighting the uniform impact (or lack thereof) of ML assignment. These results collectively suggest that ML-based case assignment does not substantially change the efficiency within each division, as the input and output level stays the same.

In summary, the implementation of random and ML assignment has yielded divergent impacts on court operations. Random court-divisions shows a slight decrease in the performance index, driven by the increase in appeals and decrease in verdict length and legal provisions cited. Heterogeneity analysis among divisions shows a decline in efficiency: the same effort (input) is related to worse performance (output), while more effort (input) is related to non-increase in performance (output). ML court divisions experience no significant change in performance or effort. None of the component variables see a significant change compared to pre-policy period. This absence of change is consistent among the divisions.

5.3 Robustness Checks

I conduct several sensitivity analyses to confirm the robustness of the baseline results. The main results remain stable across different sets of controls, alternative sample restrictions, and variations in outcome variable weights.[27]

First, I examine the effect of varying the weights assigned to outcome variables. The performance and effort indices are weighted averages of normalized measures, initially constructed using the inverse of the covariance matrix to give higher weight to less correlated variables. This weighting approach reduces the influence of redundant information, making it sensitive to the specific data structure. I construct indices by assigning equal weight to each component variable to assess robustness. As shown in Figure **??** and **??**, the effect sizes on both performance and effort indices remain consistent across random and ML court divisions under this alternative weighting scheme.

[27] In future versions, I intend to incorporate the alternative empirical method from **?**.

Second, I apply alternative sample restrictions by limiting the analysis to courts with case disclosure rates of at least 85% and 90%, respectively. As discussed in the data section, certain case types (e.g., juvenile or divorce cases) are legally withheld from public disclosure, potentially impacting courts with lower disclosure rates. Excluding courts with lower disclosure rates helps ensure that incomplete data does not bias the results. Figure B6 and B7 demonstrate that results for both performance and effort indices are robust to these alternative disclosure thresholds in both random and ML court divisions.

Lastly, I add additional controls for case and judge characteristics to verify that the results are not driven by shifts in case composition or judge profiles. As shown in Table C1 and C2, these additional controls have minimal impact on the estimated effects, underscoring the robustness of the findings.

6 Conclusion

Building on the existing literature on algorithms in high-stakes settings, this paper investigates the impact of algorithm-based case assignment systems on court performance and judicial effort, focusing on the shift from manual to automated random and ML-based assignments in the Chinese judicial system. Leveraging the staggered adoption of these assignment systems across courts, I use a generalized Difference-in-Differences (DiD) approach to estimate their effects on judicial performance and effort. For this analysis, I use government procurement data to identify the timing of each court's transition to automated case assignment and classify the type of algorithm adopted. Additionally, I utilize 66 million case documents to construct detailed measures of case characteristics, judge attributes, and case outcomes, enabling a comprehensive evaluation of the impacts on assignment quality and court effectiveness.

The findings indicate that random assignment disrupts the link between judge experience and case complexity, resulting in a slight decrease in court performance (0.054 SD) and judicial effort (0.11 SD). By contrast, ML-based assignment maintains a partial match between complex cases and experienced judges, showing no significant changes in performance or effort across divisions. These results are consistent under alternative court performance and effort measures, high-disclosure court samples, and additional case and judge controls.

This study contributes to the literature on algorithmic decision-making in high-stakes settings, illustrating that automated assignment systems influence judicial outcomes through the assignment pattern itself. Contrary to the assumption that algorithmic assignments are inherently efficient, this research demonstrates that random assignment may reduce effectiveness due to its inflexibility, while ML-based assignment, by retaining certain judge-case correlations, can achieve comparable results to human judgment in court settings. Furthermore, this study extends judge fixed-effect literature by showing that not only individual judge characteristics but also the method of judge-case assignment can affect case outcomes.

The key takeaway for policy implication is to recognize the influence assignment systems have on case outcomes. This underscores the importance of understanding, refining, and transparently disclosing these systems. My findings urge caution against assuming that random assignment is always preferable and suggest the potential of ML systems to balance impartiality with assignment quality, reducing human discretion. Transparent guidelines for algorithmic operations could enhance public trust in judicial independence, while careful calibration of algorithmic assignments could help courts achieve optimal results.

While this study provides valuable insights into the impact of algorithms on case assignment patterns and case outcomes, several avenues for future research remain. First, the specifics of the case assignment algorithms are worth closer examination. Even nuanced variations within random assignment rules may yield distinct outcomes. Advanced text analysis could capture further dimensions of judge-case matching, enriching our understanding of assignment patterns. Second, research could explore the long-term impacts of algorithmic case assignment on society, such as effects on crime or business practices. Lastly, examining the externalities of algorithmic assignment may offer a more comprehensive evaluation of algorithmic managers' broader value in the legal context. For instance, the shift from human decision-makers to algorithm-driven assignments could affect litigants' perceptions of judicial independence, potentially changing litigation strategies, such as judge-shopping practices. These future research directions can help refine the ideal case assignment approach and its impact on society.

References

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, "Combining human expertise with artificial intelligence: Experimental evidence from radiology," Technical Report, National Bureau of Economic Research 2023.

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484), 1481–1495.

Angelova, Victoria, Will S Dobbie, and Crystal Yang, "Algorithmic recommendations and human discretion," Technical Report, National Bureau of Economic Research 2023.

Becker, Luc, Bastian Wurm, and Thomas Hess, "Will Algorithms Replace Managers? A Systematic Literature Review on Algorithmic Management," 2023.

Berk, Richard, "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism," *Journal of Experimental Criminology*, 2017, *13*, 193–216.

Berk, Richard A, Susan B Sorenson, and Geoffrey Barnes, "Forecasting domestic violence: A machine learning approach to help inform arraignment decisions," *Journal of empirical legal studies*, 2016, *13* (1), 94–115.

Boyd, Christina L, Lee Epstein, and Andrew D Martin, "Untangling the causal effects of sex on judging," *American journal of political science*, 2010, *54* (2), 389–411.

Callaway, Brantly and Pedro HC Sant' Anna, "Difference-in-differences with multiple time periods," *Journal of econometrics*, 2021, *225* (2), 200–230.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American economic review*, 2020, *110* (9), 2964–2996.

Cheng, Lingwei and Alexandra Chouldechova, "Heterogeneity in algorithm-assisted decision-making: A case study in child abuse hotline screening," *Proceedings of the ACM on Human-Computer Interaction*, 2022, *6* (CSCW2), 1–33.

Chilton, Adam S and Marin K Levy, "Challenging the randomness of panel assignment in the Federal Courts of Appeals," *Cornell L. Rev.*, 2015, *101*, 1.

Cram, W Alec, Martin Wiener, Monideepa Tarafdar, Alexander Benlian et al., "Algorithmic Controls and their Implications for Gig Worker Well-being and Behavior.," in "ICIS," Vol. 2020 2020, pp. 1–17.

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

Fabri, Marco and Philip M Langbroek, "Is There a Right Judge for Each Case-A Comparative Study of Case Assignment in Six European Countries," *Eur. J. Legal Stud.*, 2007, *1*, 292.

Glynn, Adam N and Maya Sen, "Identifying judicial empathy: Does having daughters cause judges to rule for women's issues?," *American Journal of Political Science*, 2015, *59* (1), 37–54.

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

Gramckow, Heike P., Omniah Ebeid, Erica Bosio, and Jorge Luis Silva Mendez, "Good practices for courts report: Helpful elements for good court performance and the World Bank ' s quality of judicial process indicators - Key elements, lessons learned, and good practice examples," 2016.

Harris, Allison P and Maya Sen, "Bias and judging," *Annual Review of Political Science*, 2019, *22* (1), 241–259.

Helmke, Gretchen and Frances Rosenbluth, "Regimes and the rule of law: Judicial independence in comparative perspective," *Annual Review of Political Science*, 2009, *12* (1), 345–366.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li, "Discretion in hiring," *The Quarterly Journal of Economics*, 2018, *133* (2), 765–800.

Jarrahi, Mohammad Hossein, Mareike Möhlmann, and Min Kyung Lee, "Algorithmic management: The role of AI in managing workforces," *MIT Sloan Management Review*, 2023.

Jin, Changwei, "Analysis on the Mechanism of AI Case Division," *Journal of CUPL*, 2020, *2020* (02). China University of Political Science and Law.

Kahn, Matthew E and Pei Li, "The effect of pollution and heat on high skill public sector worker productivity in China," Technical Report, National Bureau of Economic Research 2019.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human decisions and machine predictions," *The quarterly journal of economics*, 2018, *133* (1), 237–293.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, and Laura Dabbish, "Working with machines: The impact of algorithmic and data-driven management on human workers," in "Proceedings of the 33rd annual ACM conference on human factors in computing systems" 2015, pp. 1603–1612.

Liu, Ernest, Yi Lu, Wenwei Peng, and Shaoda Wang, "Judicial independence, local protectionism, and economic integration: Evidence from China," Technical Report, National Bureau of Economic Research 2022.

Liu, Zhuang, "Does reason writing reduce decision bias? Experimental evidence from judges in China," *The Journal of Legal Studies*, 2018, *47* (1), 83–118.

Macfarlane, Katherine A, "Constitutional Case Assignment," *NCL Rev.*, 2023, *102*, 977.

Mehmood, Sultan, "The impact of Presidential appointment of judges: Montesquieu or the Federalists?," *American Economic Journal: Applied Economics*, 2022, *14* (4), 411–445.

Mullainathan, Sendhil and Ziad Obermeyer, "Diagnosing physician error: A machine learning approach to low-value health care," *The Quarterly Journal of Economics*, 2022, *137* (2), 679–727.

Rittenhouse, Katherine, Emily Putnam-Hornstein, and Rhema Vaithianathan, "Algorithms, Humans and Racial Disparities in Child Protective Systems: Evidence from the Allegheny Family Screening Tool," *Unpublished Working Paper*, 2023.

Shayo, Moses and Asaf Zussman, "Judicial ingroup bias in the shadow of terrorism," *The Quarterly journal of economics*, 2011, *126* (3), 1447–1484.

Suhadolnik, Nicolas, Jo Ueyama, and Sergio Da Silva, "Machine learning for enhanced credit risk assessment: An empirical approach," *Journal of Risk and Financial Management*, 2023, *16* (12), 496.

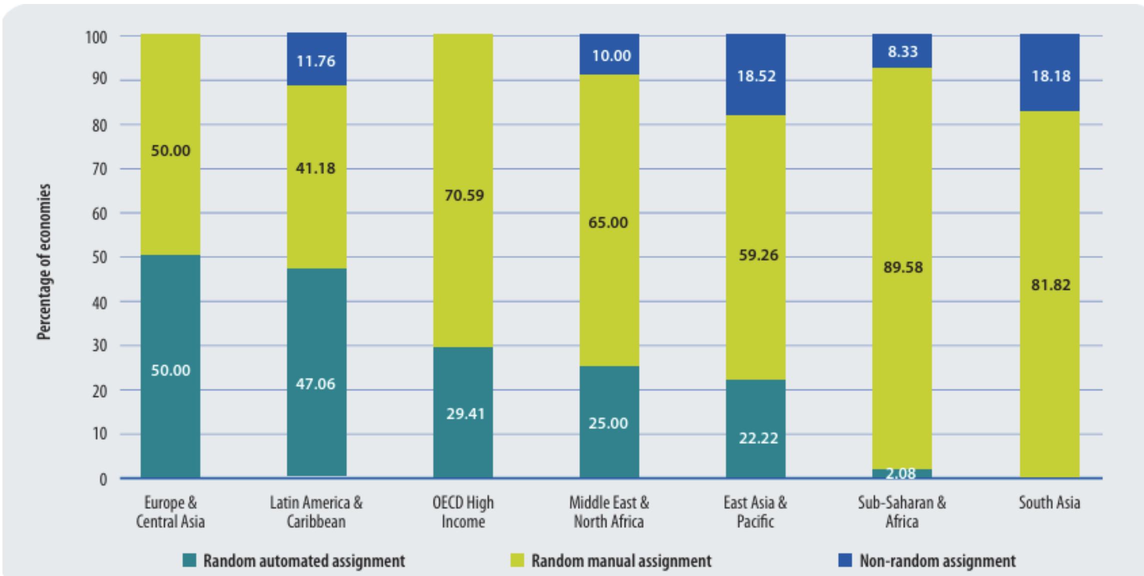
Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.

Sunstein, Cass R, David Schkade, and Lisa Michelle Ellman, "Ideological voting on federal courts of appeals: A preliminary investigation," *Virginia Law Review*, 2004, pp. 301–354.

A Tables and Figures

A.1 Figures

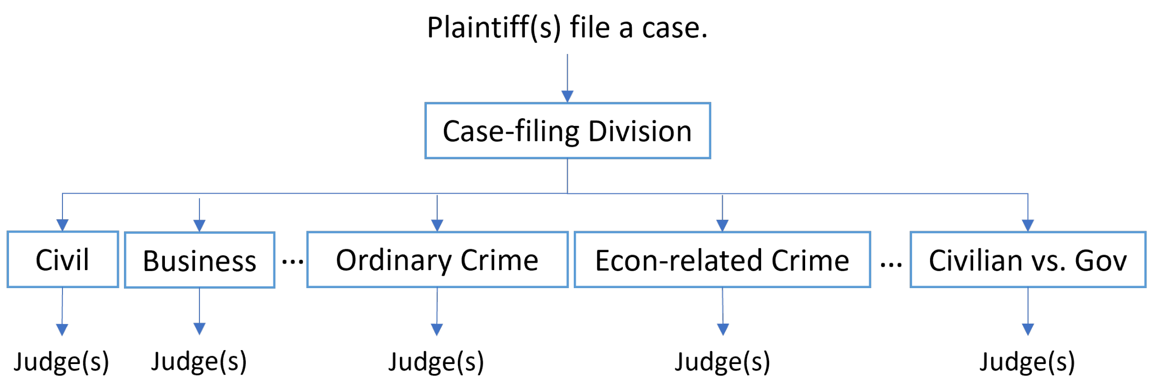
Appendix Figure A1. Case Assignment System Worldwide by 2016



Notes: This figure illustrates the case assignment system worldwide by 2016. Data from at least one city from each country is collected. The percentages shown in the figure are based on data for 189 economies, though for economies in which Doing Business collects data for two cities, the data for the two cities are considered separately.

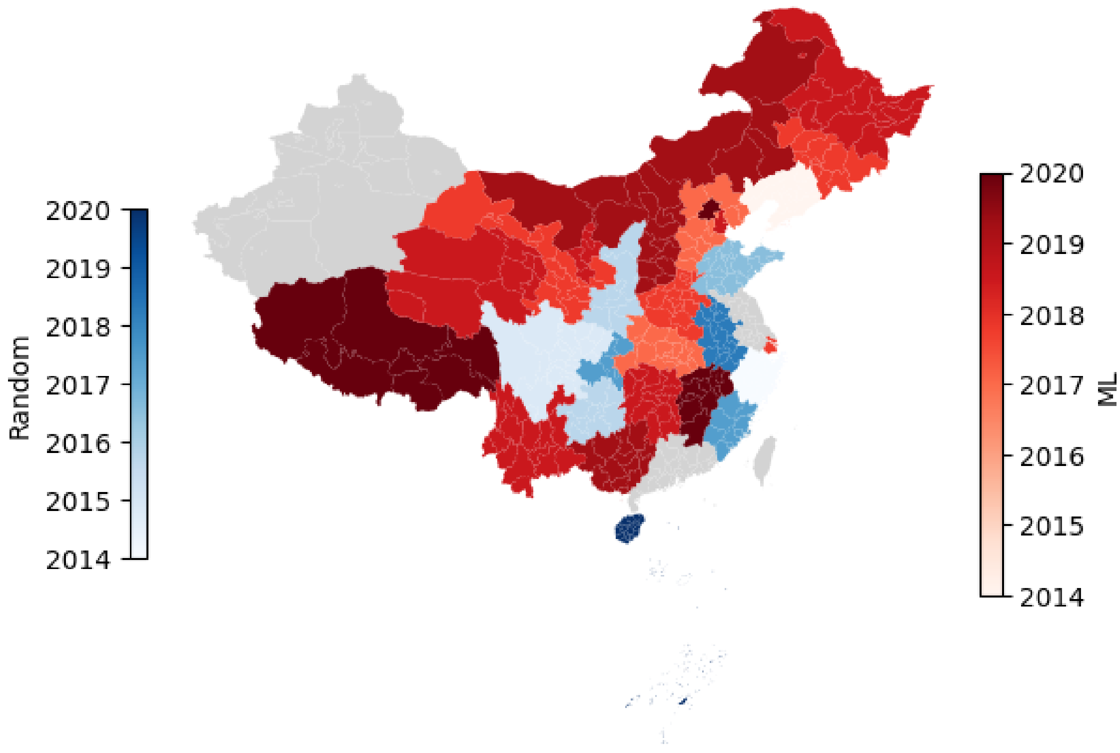
Source: Doing Business database, http://www.doingbusiness.org/data.

Appendix Figure A2. Flow Chart of Judicial Process



Notes: This figure illustrates the case assignment process in China. Plaintiffs file the case at the case-filling division. Then, the case-filing division distribute to each division according to the case type. Then, each division assign the case to the judge(s) by either a human decision maker, or an algorithm. Judge(s) only hear case within their division. All divisions within a court, and all courts within a province uses the same type of case assignment approach.

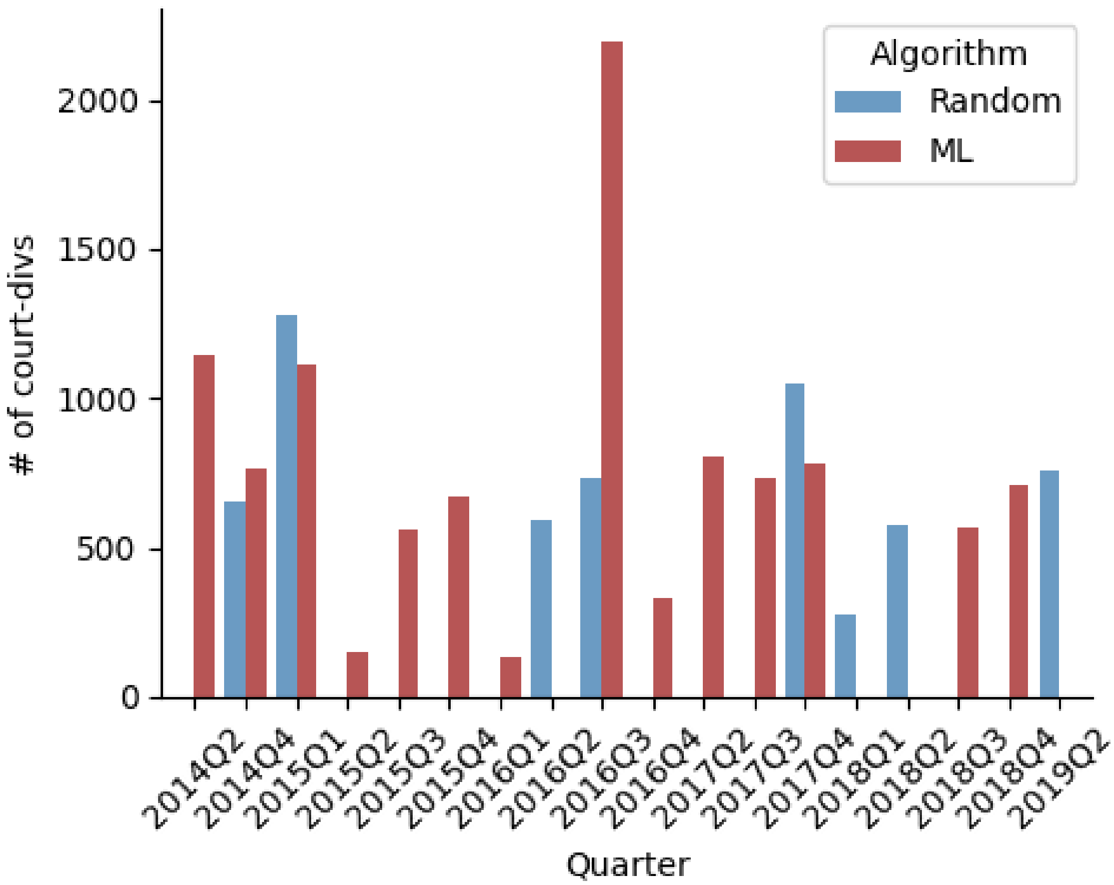
Appendix Figure A3. Algorithmic Case Assignment System Adoption by Prefecture



Notes: This figure illustrates the timing and type of case assignment systems in China' s prefectures. The blue (red) gradient scale indicates prefectures using random (ML-based) assignment systems, with darker shades showing later adoption. Gray areas represent prefectures with unknown system details, which are excluded from analysis. Thin white lines denote prefecture borders. According to the data source, prefectures in the same province generally adopt the same system simultaneously. While prefectural variations likely exist, additional validation is needed. This analysis uses uniform timing per province. The future versions will incorporate prefectural variation.

Source: China Government Procurement Online.

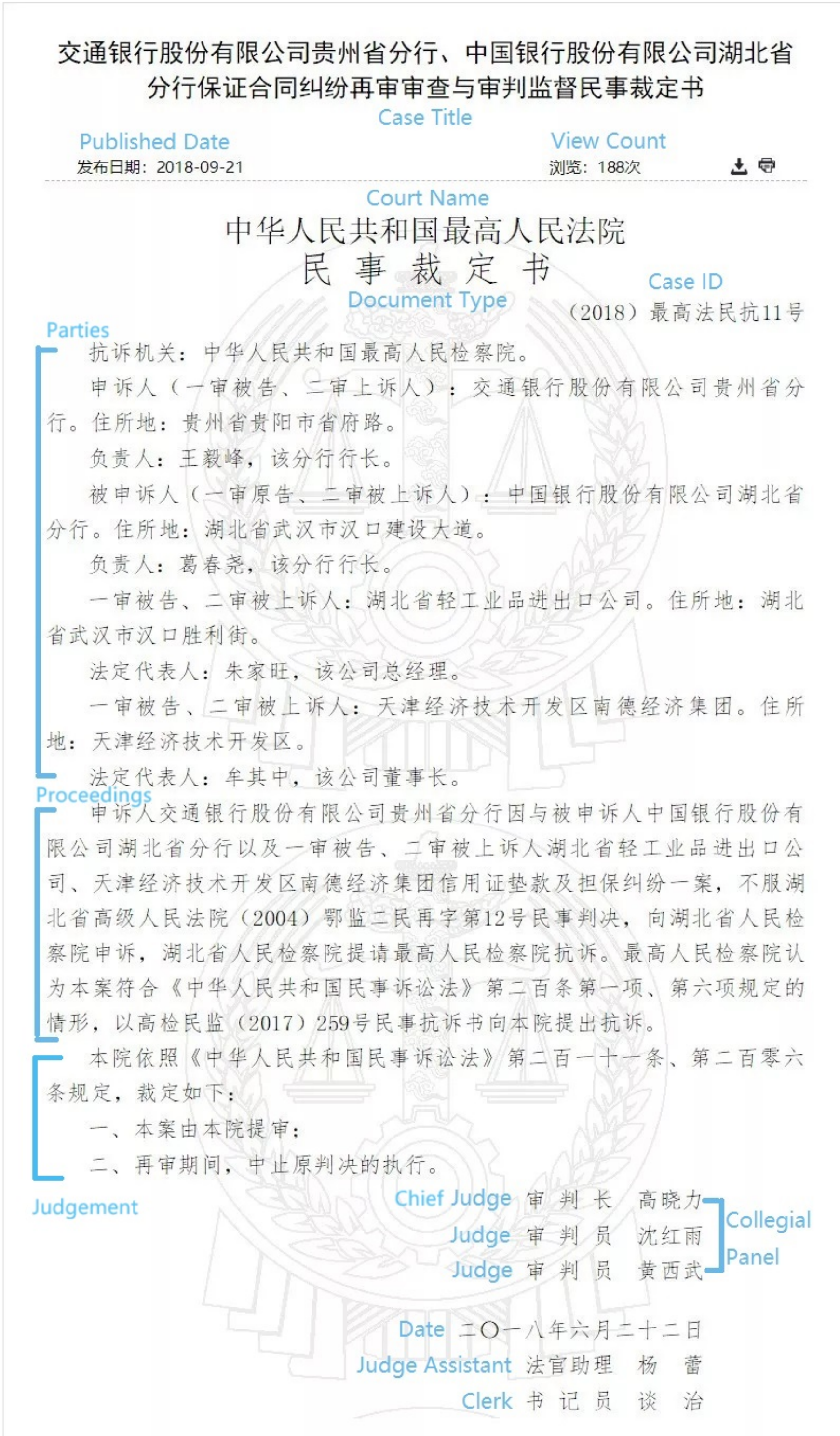
Appendix Figure A4. Algorithmic Case Assignment System Adoption by Courts



Notes: This figure illustrates the timing and type of case assignment systems by court. The blue (red) bar indicates courts using random (ML-based) assignment systems. From this figure, we can see the rollout of the algorithmic case assignment from 2014 to 2020. According to the data source, courts in the same province generally adopt the same system simultaneously. While prefectural variations likely exist, additional validation is needed. This analysis uses uniform timing per province. The future versions will incorporate prefectural variation.

Source: China Government Procurement Online.

Appendix Figure A5. Case Document Example



交通银行股份有限公司贵州省分行、中国银行股份有限公司湖北省分行保证合同纠纷再审审查与审判监督民事裁定书

发布日期：2018-09-21　浏览：188次

中华人民共和国最高人民法院

民 事 裁 定 书

（2018）最高法民抗11号

抗诉机关：中华人民共和国最高人民检察院。

申诉人（一审被告、二审上诉人）：交通银行股份有限公司贵州省分行。住所地：贵州省贵阳市省府路。

负责人：王毅峰，该分行行长。

被申诉人（一审原告、二审被上诉人）：中国银行股份有限公司湖北省分行。住所地：湖北省武汉市汉口建设大道。

负责人：葛春尧，该分行行长。

一审被告、二审被上诉人：湖北省轻工业品进出口公司。住所地：湖北省武汉市汉口胜利街。

法定代表人：朱家旺，该公司总经理。

一审被告、二审被上诉人：天津经济技术开发区南德经济集团。住所地：天津经济技术开发区。

法定代表人：牟其中，该公司董事长。

申诉人交通银行股份有限公司贵州省分行因与被申诉人中国银行股份有限公司湖北省分行以及一审被告、二审被上诉人湖北省轻工业品进出口公司、天津经济技术开发区南德经济集团信用证垫款及担保纠纷一案，不服湖北省高级人民法院（2004）鄂监二民再字第12号民事判决，向湖北省人民检察院申诉，湖北省人民检察院提请最高人民检察院抗诉。最高人民检察院认为本案符合《中华人民共和国民事诉讼法》第二百条第一项、第六项规定的情形，以高检民监（2017）259号民事抗诉书向本院提出抗诉。

本院依照《中华人民共和国民事诉讼法》第二百一十一条、第二百零六条规定，裁定如下：

一、本案由本院提审；

二、再审期间，中止原判决的执行。

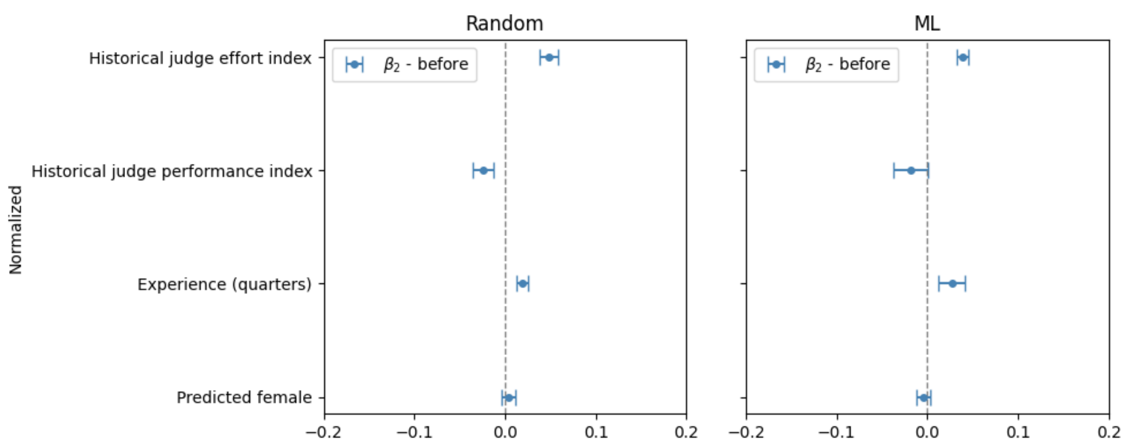
审 判 长　高晓力
审 判 员　沈红雨
审 判 员　黄西武

二〇一八年六月二十二日
法官助理　杨　蕾
书 记 员　谈　治

Notes: This figure shows an example of the case document. A case document has of four parts: title, fact, verdict and closing. The title contains case title, case type, court name, document type and case ID. The fact contains litigant(s) name, their information, facts, evidence, previous proceeding and related case IDs. The verdict contains the judgment and the reasoning. The closing contains the judge(s) and clerk(s) names and the date of judgment.

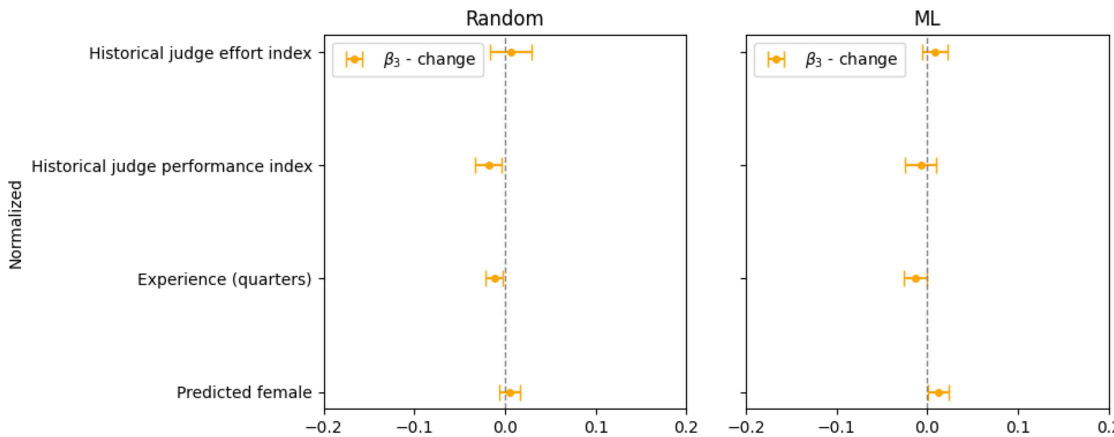
Source: China Judgment Online.

Appendix Figure A6. Case Assignment Pattern Relative to Case Complexity Index Before Random/ML



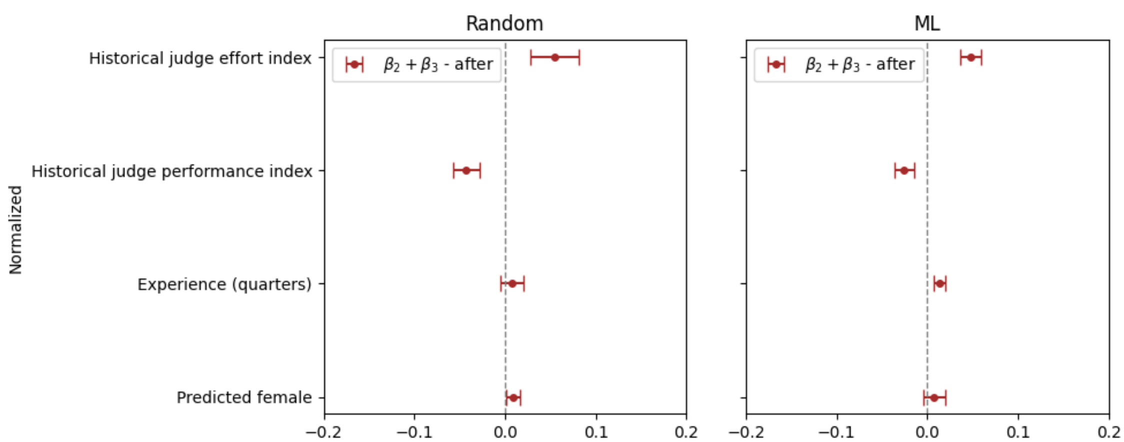
Notes: This figure illustrates case assignment patterns in random and ML-based courts during the manual era, displaying the β_2 coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3(\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how judge characteristics correlate with case complexity before the assignment system change. The analysis uses a 10% representative case-level sample. Historical judge effort and performance indices are standardized indices from weighted averages of multiple variables, normalized to pre-treatment observations. Experience is similarly normalized. Predicted female is a binary variable indicating female judges based on name prediction. The left panel shows judge-case correlations for courts that transitioned to random assignment, while the right panel shows those that transitioned to ML-based assignment.

Appendix Figure A7. Case Assignment Pattern Relative to Case Complexity Index Change Due to Random/ML



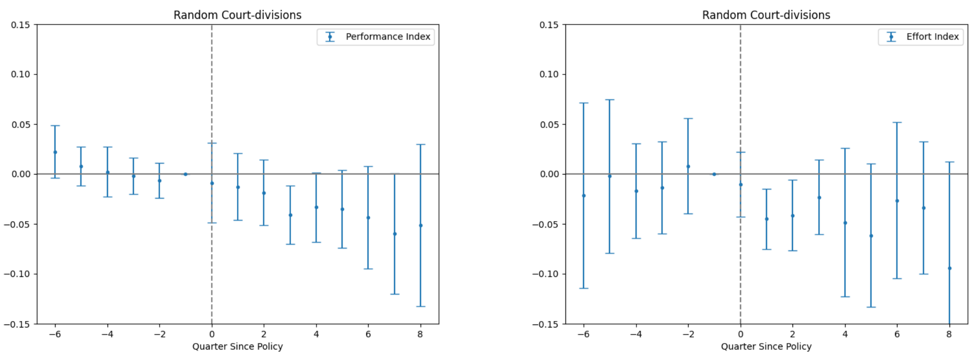
Notes: This figure illustrates the change in case assignment patterns in random and ML-based courts due to the algorithmic systems, displaying the β_3 coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how differently judge characteristics correlate with case complexity since the assignment system change. The analysis uses a 10% representative case-level sample. Historical judge effort and performance indices are standardized indices from weighted averages of multiple variables, normalized to pre-treatment observations. Experience is similarly normalized. Predicted female is a binary variable indicating female judges based on name prediction. The left panel shows judge-case correlations for courts that transitioned to random assignment, while the right panel shows those that transitioned to ML-based assignment.

Appendix Figure A8. Case Assignment Pattern Relative to Case Complexity Index After Random/ML



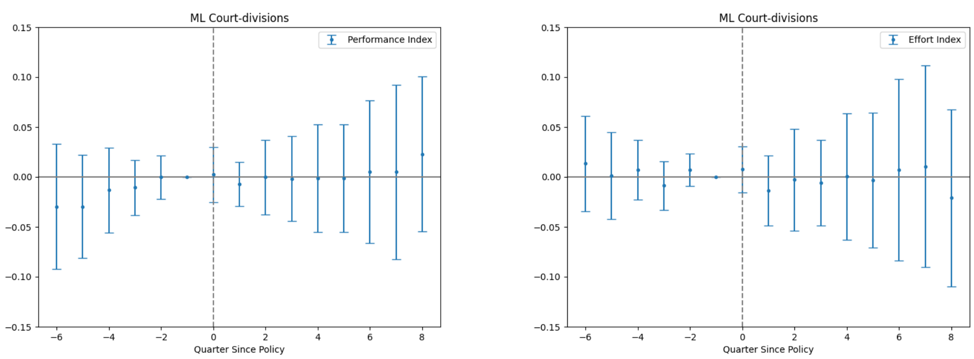
Notes: This figure illustrates case assignment patterns in random and ML-based courts after the algorithmic system adoption, displaying the $\beta_2 + \beta_3$ coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how judge characteristics correlate with case complexity after the assignment system change. The analysis uses a 10% representative case-level sample. Historical judge effort and performance indices are standardized indices from weighted averages of multiple variables, normalized to pre-treatment observations. Experience is similarly normalized. Predicted female is a binary variable indicating female judges based on name prediction. The left panel shows judge-case correlations for courts that transitioned to random assignment, while the right panel shows those that transitioned to ML-based assignment.

Appendix Figure A9. Event Study of Random Assignment on Performance Index and Efforts Index



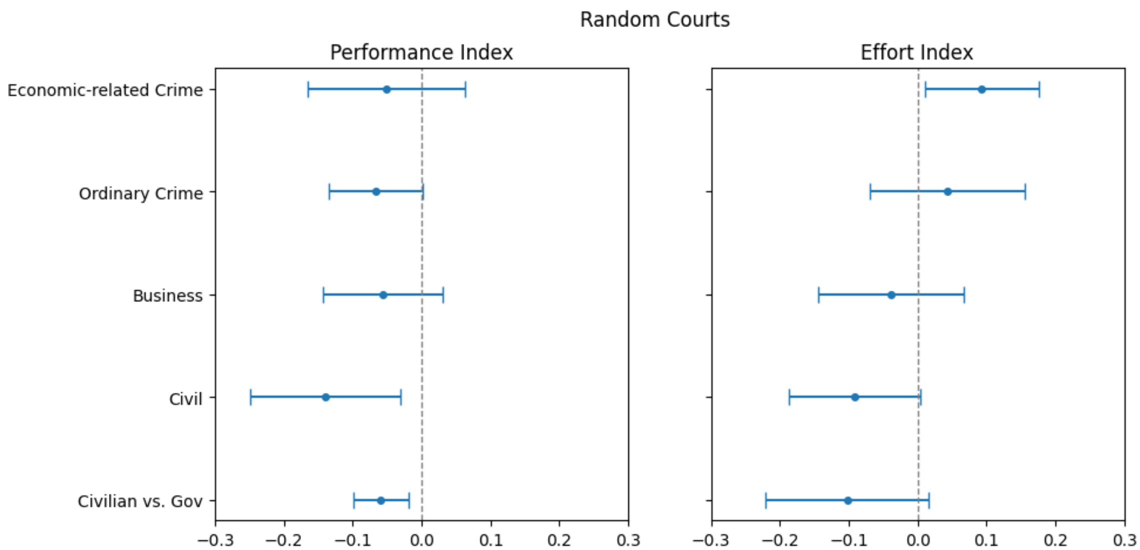
Notes: The figure presents event-study estimates of the effects of a random case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure A10. Event Study of ML-Based Assignment on Performance Index and Efforts Index



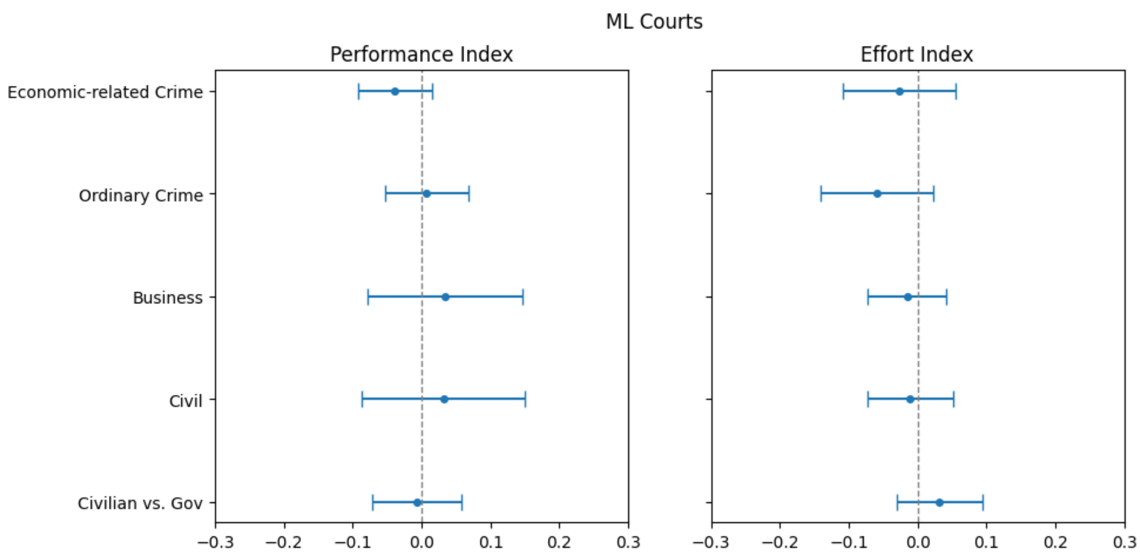
Notes: The figure presents event-study estimates of the effects of a ML-based case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure A11. ATT of Random Assignment on Performance Index and Efforts Index across Divisions



Notes: The figure shows the average treatment effects of an random case assignment system on performance (output) and effort (input) indices across five court divisions: civil vs. government, civil, business, ordinary crime, and economic-related crime. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment effects with 95% confidence intervals, derived from the CSDiD corresponding to Equation 2. Each coefficient represents a separate regression. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure A12. ATT of ML-Based Assignment on Performance Index and Efforts Index across Divisions



Notes: The figure shows the average treatment effects of an ML-based case assignment system on performance (output) and effort (input) indices across five court divisions: civil vs. government, civil, business, ordinary crime, and economic-related crime. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment effects with 95% confidence intervals, derived from the CSDiD corresponding to Equation 2. Each coefficient represents a separate regression. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

A.2 Tables

Appendix Table A1. Summary Statistics of the Random and ML-Based Court-Divisions Before Treatment

Variable	Random	ML	Difference
Panel A: Case Characteristics Before Algo			
# of Cause of Actions (COA)	0.99	1.00	-0.01***
	(0.17)	(0.17)	(0.00)
$ involved (Yuan)	4572965.36	4889624.26	-316658.90
	(30535612.29)	(34873296.30)	(90646.96)
# of litigants	2.43	2.40	0.03**
	(1.22)	(1.40)	(0.00)
Ratio of corporate litigants	0.30	0.27	0.03
	(0.30)	(0.30)	(0.00)
Ratio of female litigants	0.08	0.09	-0.01
	(0.11)	(0.13)	(0.00)
Panel B: Judge Characteristics Before Algo			
Predicted gender	0.29	0.29	-0.002***
	(0.24)	(0.25)	(0.00)
Experience (quarters)	9.84	7.05	2.79***
	(7.18)	(5.71)	(0.02)
Historical # cases processed/quarter	49.43	33.57	15.87
	(57.46)	(55.61)	(0.16)
Historical NOT appeal rate	0.79	0.76	0.03
	(0.16)	(0.17)	(0.00)
Historical NOT reversal rate	0.99	0.98	0.01
	(0.03)	(0.05)	(0.00)
Historical verdict word count	535.67	575.16	-39.48
	(227.00)	(251.71)	(0.66)
Historical # of provisions cited	1.73	1.82	-0.10
	(0.70)	(0.74)	(0.00)
Historical # of provisions written	0.86	0.88	-0.02
	(0.57)	(0.62)	(0.00)

Notes: This table shows the case and judge characteristics of the random and ML-based court-divisions before the algorithm adoption. Column (1) shows the pre-treatment average of random court-divisions, column (2) shows the pre-treatment average of the ML-based divisions, and column (3) shows the difference between two groups. Panel A presents the case characteristics, in which I use the first four to calculate case complexity. Panel B presents the judge characteristics. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Appendix Table A2. Summary Statistics of the Random and ML-Based Court-Divisions Before Treatment

Variable	Random	ML	Difference
Panel A: Outcomes before Algo - Performance			
Performance Index	0.06	-0.05	0.12***
	(0.58)	(0.74)	(0.00)
# of cases processed/quarter	66.81	44.91	21.89***
	(161.42)	(167.59)	(0.36)
NOT appeal rate	0.79	0.77	0.02***
	(0.26)	(0.28)	(0.00)
NOT reversal rate	0.98	0.97	0.01***
	(0.08)	(0.10)	(0.00)
Panel B: Outcomes before Algo - Effort			
Effort Index	0.05	-0.04	0.09***
	(0.75)	(0.80)	(0.00)
Verdict word count	502.87	484.31	18.57***
	(403.94)	(416.85)	(0.89)
# of provisions cited	1.62	1.54	0.08***
	(1.05)	(1.09)	(0.00)
# of provisions written	0.78	0.71	0.07***
	(0.97)	(0.94)	(0.00)

Notes: This table shows the performance and effort of the random and ML-based court-divisions before the algorithm adoption. Column (1) shows the pre-treatment average of random court-divisions, column (2) shows the pre-treatment average of the ML-based divisions, and column (3) shows the difference between two groups. Panel A and panel B presents the performance index and compositing variables and effort index and compsiting variables, respectively. Panel B presents the judge characteristics. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Appendix Table A3. ATT of Random Assignment on Performance Index and Efforts Index

			Panel A: Outcome Variables - Performance			
Index	**Normalized Variables**			**Raw Variables**		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Performance	# Cases Processed	Not Appealed Rate	Not Reversal Rate	# Cases Processed	Appeals	Reversals
-0.054***	0.051	-0.110*	-0.060	8.056	6.066**	0.209
(0.02)	(0.04)	(0.06)	(0.04)	(5.00)	(2.44)	(0.17)
			Panel B: Outcome Variables - Effort Index			
Index	**Normalized Variables**			**Raw Variables**		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effort	Verdict Word Count	Provisions Cited	Provisions Written	Verdict Word Count	# Provisions Cited	# Provisions Written
-0.114***	-0.097***	-0.097***	0.012	-46.533***	-0.096***	-0.012
(0.02)	(0.04)	(0.03)	(0.03)	(12.57)	(0.02)	(0.02)
Unit			Court-division			
Cluster			Province			
Obs			350,552			

Notes: This table presents the ATT of random assignment system on performance index, effort index, the normalized and raw elements composing the index. In Panel A, Column (1) shows the ATT on the performance index. Columns (2)-(4) display normalized variables (# cases processed, not appealed rate, and not reversal rate) that compose the index, while Columns (5)-(7) show their raw counterparts (# cases processed, appeals, and reversals). In Panel B, Column (1) shows the ATT on the effort index. Columns (2)-(4) include normalized variables (verdict word count, provisions cited, and provisions written), and Columns (5)-(7) show the raw counterparts for these components. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

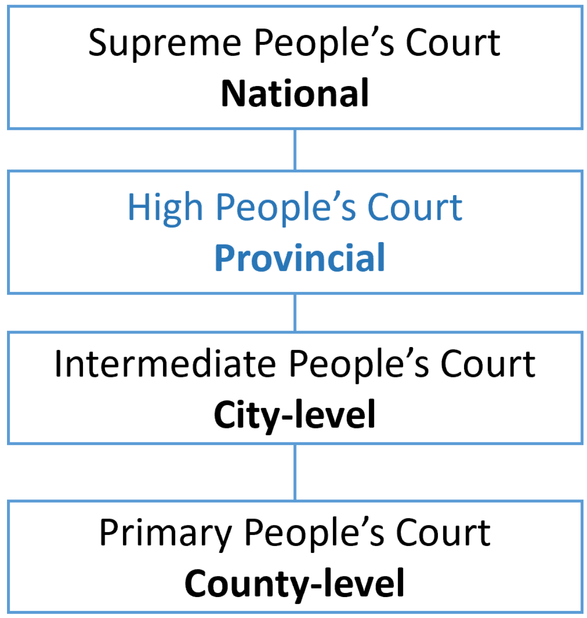
Appendix Table A4. ATT of ML Assignment on Performance Index and Efforts Index

	Panel A: Outcome Variables - Performance					
Index	**Normalized Variables**			**Raw Variables**		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Performance	# Cases Processed	Not Appealed Rate	Not Reversal Rate	# Cases Processed	Appeals	Reversals
0.025	0.001	0.042	0.023	-3.289	-0.889	-0.155
(0.03)	(0.04)	(0.06)	(0.03)	(5.25)	(1.46)	(0.12)
	Panel B: Outcome Variables - Effort Index					
Index	**Normalized Variables**			**Raw Variables**		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effort	Verdict Word Count	Provisions Cited	Provisions Written	Verdict Word Count	# Provisions Cited	# Provisions Written
-0.010	0.021	0.002	-0.031	13.150	0.018	-0.009
(0.03)	(0.04)	(0.04)	(0.03)	(7.21)	(0.03)	(0.01)
Unit	Court-division					
Cluster	Province					
Obs	534,601					

Notes: This table presents the ATT of ML-based assignment system on performance index, effort index, the normalized and raw elements composing the index. In Panel A, Column (1) shows the ATT on the performance index. Columns (2)-(4) display normalized variables (# cases processed, not appealed rate, and not reversal rate) that compose the index, while Columns (5)-(7) show their raw counterparts (# cases processed, appeals, and reversals). In Panel B, Column (1) shows the ATT on the effort index. Columns (2)-(4) include normalized variables (verdict word count, provisions cited, and provisions written), and Columns (5)-(7) show the raw counterparts for these components. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

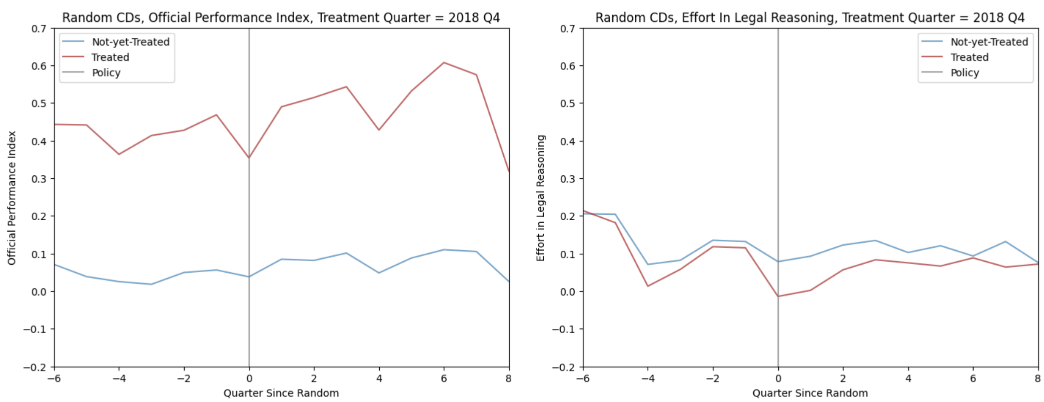
B Additional Graphs

Appendix Figure B1. Court Structure of China



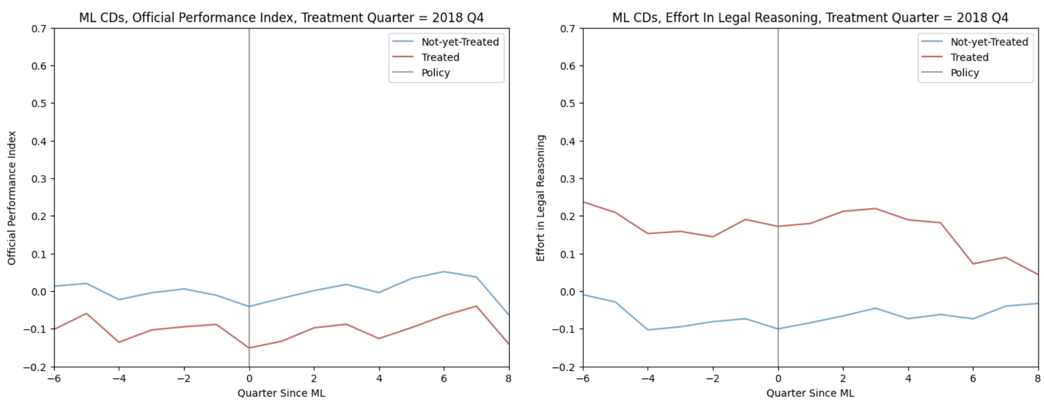
Notes: This figure illustrates the court structure in China. The Chinese court system is not a homogeneous entity. Policy documents and commands are issued top-down, with instructions cascading from the Supreme People's Court (SPC) to the primary courts. China's judicial system operates hierarchically with four levels that correspond to different levels of government: the SPC, 33 high people's courts, 416 intermediate people's courts, and 3087 primary people's courts. In the absence of appellate courts, the SPC and the high and intermediate courts have appellate jurisdiction over the courts one level below them. Primary courts handle exclusively first-instance cases, the SPC hears only appeals, and the remaining courts manage a mix of both first-instance cases and appeals. This structure is complicated by specialized courts, such as military, maritime, forest, and railway courts, which have jurisdiction limited to specific topics. These courts are not representative of the broader population and, due to their specialized nature and complicated organizational affiliations, this paper focuses exclusively on the ordinary courts.

Appendix Figure B2. Example Parallel Trend of Random Assignment on Performance Index and Efforts Index, 2018 Q4



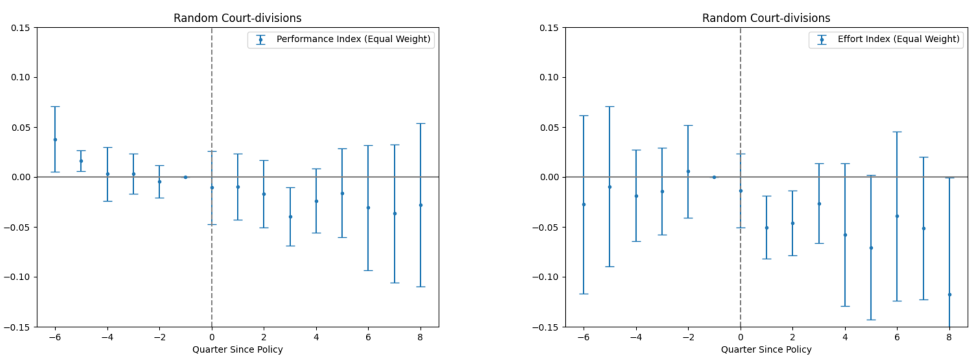
Notes: The figure illustrates descriptive evidence on the impact of random assignment on court-division outcomes by plotting performance (left panel) and effort (right panel) trends in both the control and treatment groups over time, relative to the period of random assignment adoption. This example treatment group is 2018 Q4. This plot does not include any controls. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions.

Appendix Figure B3. Example Parallel Trend of ML-Based Assignment on Performance Index and Efforts Index, 2018 Q4



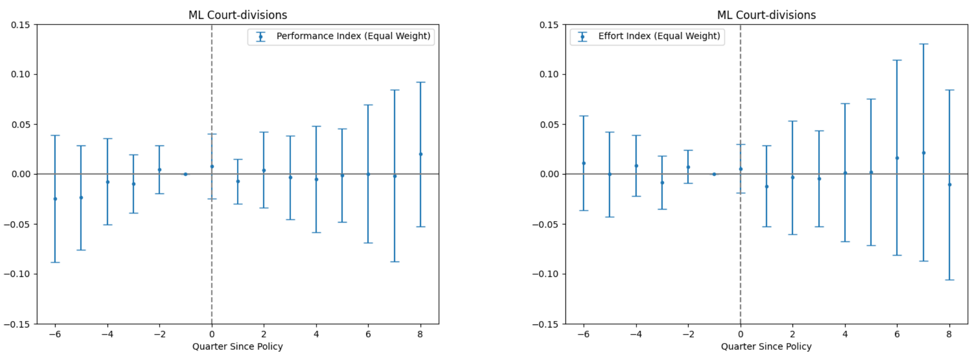
Notes: The figure illustrates descriptive evidence on the impact of ML-based assignment on court-division outcomes by plotting performance (left panel) and effort (right panel) trends in both the control and treatment groups over time, relative to the period of ML-based assignment adoption. This example treatment group is 2018 Q4. This plot does not include any controls. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions.

Appendix Figure B4. Event Study of Random Assignment on Performance Index and Efforts Index, Equal Weight



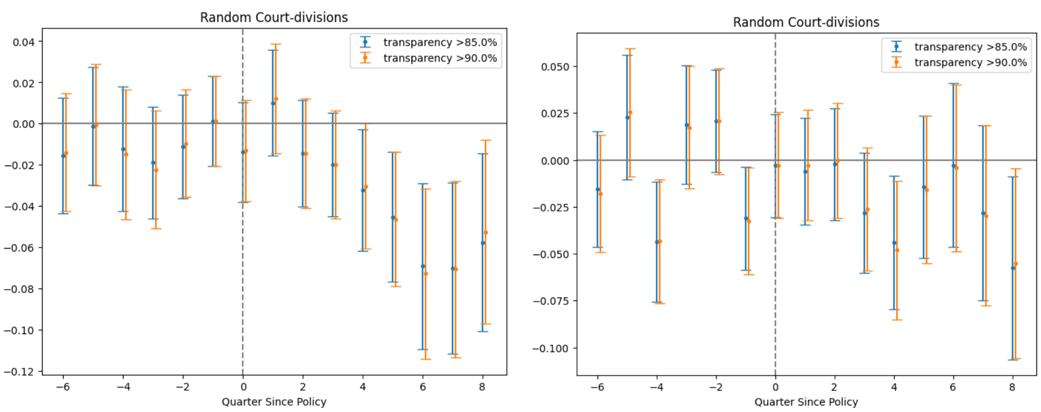
Notes: The figure presents event-study estimates of the effects of a random case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, an equal weighted average of normalized # of case processed, not appeal rate and not reversal rate. The right panel uses the effort index, an equal weighted average of normalized verdict word count, # of provisions cited and # of provisions written. Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure B5. Event Study of ML-Based Assignment on Performance Index and Efforts Index, Equal Weight



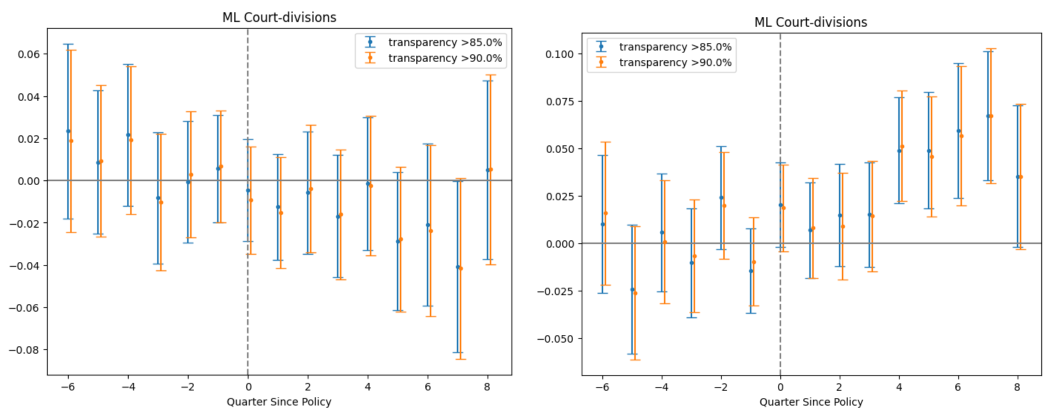
Notes: The figure presents event-study estimates of the effects of a ML-based case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, an equal weighted average of normalized # of case processed, not appeal rate and not reversal rate. The right panel uses the effort index, an equal weighted average of normalized verdict word count, # of provisions cited and # of provisions written. Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure B6. Event Study of Random Assignment on Performance Index and Efforts Index, High Transparency Rate Courts



Notes: The figure presents event-study estimates of the effects of a random case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The blue and orange bars are the estimates of the sample restricted to courts with 85% and 90% case document disclosure rate, respectively. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure B7. Event Study of ML-Based Assignment on Performance Index and Efforts Index, High Transparency Rate Courts



Notes: The figure presents event-study estimates of the effects of a ML-based case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal rate according to Anderson (2008). The right panel uses the effort index, a weighted average of normalized verdict word count, # of provisions cited and # of provisions written according to Anderson (2008). Both panels display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 2.1. The blue and orange bars are the estimates of the sample restricted to courts with 85% and 90% case document disclosure rate, respectively. The unit of analysis is court-division and the unit of time is quarter. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

C Additional Tables

Appendix Table C1. ATT of Random Assignment on Performance and Efforts Index With Additional Controls

	Panel A: Outcome Variable - Performance			
	(1)	(2)	(3)	(4)
Random	-0.054***	-0.053***	-0.055***	-0.055***
	(0.03)	(0.02)	(0.02)	(0.02)
	Panel B: Outcome Variable - Effort Index			
	(1)	(2)	(3)	(4)
Random	-0.114***	-0.042**	-0.041**	-0.046**
	(0.03)	(0.02)	(0.02)	(0.02)
Control		Judge Characteristics	Level of Court	Document Type
Unit	Court-division	Court-division	Court-division	Court-division
Cluster	Province	Province	Province	Province
Obs	350,552	350,552	350,552	350,552

Notes: This table presents the ATT of random assignment system on performance and effort index with additional controls. In Panel A, Column (1)-(4) shows the ATT on the performance index with different controls. In Panel B, Column (1)-(4) shows the ATT on the effort index with different controls. The treatment group includes court-divisions in provinces already implementing random assignment, while the control group comprises not-yet-implemented random assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Appendix Table C2. ATT of ML Assignment on Performance and Efforts Index With Additional Controls

	Panel A: Outcome Variable - Performance			
	(1)	(2)	(3)	(4)
ML	0.025	0.022	0.023	0.021
	(0.03)	(0.03)	(0.03)	(0.03)
	Panel B: Outcome Variable - Effort Index			
	(1)	(2)	(3)	(4)
ML	-0.010	-0.020	-0.018	-0.010
	(0.03)	(0.03)	(0.03)	(0.03)
Control		Judge Characteristics	Level of Court	Document Type
Unit	Court-division	Court-division	Court-division	Court-division
Cluster	Province	Province	Province	Province
Obs	534,601	534,601	534,601	534,601

Notes: This table presents the ATT of ML-based assignment system on performance and effort index with additional controls. In Panel A, Column (1)-(4) shows the ATT on the performance index with different controls. In Panel B, Column (1)-(4) shows the ATT on the effort index with different controls. The treatment group includes court-divisions in provinces already implementing ML-based assignment, while the control group comprises not-yet-implemented ML-based assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.