The Case Assignment System as a Technology: Algorithmic Allocation and Its Impact on Court

Rui Zuo*

June 23, 2025

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Please do not cite or circulate.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

Abstract

Courts, as public institutions central to dispute resolution and the rule of law, operate with internal administrative systems. Yet, little is known about how these systems specifically impact judicial performance and public trust. This paper studies the impact of case assignment systems, exploiting a nationwide reform in China that replaced manual assignment with algorithmic systems—either random or machine learning (ML)-based. Using 66 million court documents from 2014 to 2021, I estimate the effects of these systems on judge-case matching, case-processing performance, and public opinion. I find that random assignment reduces path dependence but weakens expertise matching, leading to modest declines in performance and verdict quality, yet increasing public trust. ML-based assignment improves matching and reduces path dependence but does not significantly change performance or trust. These findings demonstrate that seemingly neutral administrative technologies embed consequential trade-offs that require explicit policy choices balancing efficiency and public legitimacy.

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

Court decisions permeate nearly all aspects of society and daily life, from family matters and social norms to the enforcement of the rule of law, making court performance a fundamental determinant of societal wellbeing. Understanding what drives variation in court decisions and performance has therefore attracted sustained interest across fields. Increasing evidence suggests that judges' characteristics—their ideology, identity, experience, and training—interact with case features to create measurable variation in judicial outcomes[1]. If this judge-case matching generates differences in court performance, a natural follow-up question arises: what about the systems that create these matches? Every court must decide who hears which case, yet the case assignment process is often treated as a neutral infrastructure—fixed, procedural, and outside the scope of analysis. However, if judge-case assignment influences court performance, as a growing body of research suggests, then the system generating these assignments may itself influence performance.

This paper asks: do case assignment systems affect case assignment patterns, court performance, and public trust? I approach this question by studying a nationwide reform in China, where courts transitioned from manual assignment to two types of algorithmic systems: either random assignment or machine learning-based assignment. Using 66 million court documents from 2014-2021, I estimate how different assignment technologies on (1) assignment patterns—measured by judge-case expertise matching and the alignment between judge characteristics and case complexity; (2) court performance—captured through case processing efficiency and verdict quality; and (3) public trust—reflected in attitudes toward the courts, judicial fairness, and the rule of law.

Each case assignment approach—manual, random, and ML-based—presents distinct strengths and weaknesses. Manual assignment, for example, allows for nuanced, human-observed case- and judge-specific information, potentially leading to tailored pairings. However, this reliance on human is susceptible to inconsistency or discretion. Random assignment, as a rule-based system, effectively minimizes such discretion. Yet, by its design, it disregards specific judge and case characteristics, which can result in inappropriate or inefficient judge-case pairings (e.g., assigning family law cases to judges who primarily handle patent cases). ML-based assignment, a data-driven alternative, seeks to optimize pairings by learning from historical out-

[1] An extensive literature examines how judge characteristics—including ideology, gender, race, personality traits, and even external factors like home football team performance—influence judicial decisions (Eren and Mocan, 2018; Shayo and Zussman, 2011; Boyd et al., 2010; Harris and Sen, 2019; Glynn and Sen, 2015).

come data. This approach, however, might inherit discretions present in its training data and fail to account for unquantified information that human decision-makers might consider. Given this coexistence of strength and weaknesses, the net effect of each system on assignment patterns, court performance, and public trust is not self-evident.

This paper overcomes a key empirical challenge in studying assignment system effects by finding a setting that enables large-scale observation of both institutional change and court outcomes. The challenge has two dimensions. First, identifying when and how assignment systems change: OECD countries with comprehensive court records have used random assignment for decades, providing no variation to study, while developing countries with flexible assignment systems often lack documentation of when and how changes occur. I solve this by leveraging China's nationwide court reform where I can pinpoint the exact timing and type of assignment system changes through government procurement contracts. Second, observing assignment outcomes and court performance: I overcome the typical data limitations by using 66 million publicly available court documents that reveal both who gets assigned to which cases (from judge names) and detailed case outcomes (from written facts and verdicts).

I exploit the staggered rollout of algorithmic assignment systems across Chinese provinces from 2014-2020. Eight provinces adopted random assignment while 18 adopted ML-based assignment, creating plausibly exogenous variation in treatment timing driven by budget and bidding process rather than court performance. Each system has distinct characteristics: manual assignment relies on senior judges selecting suitable matches using detailed information but allowing discretion; random assignment eliminates discretion but disregards judge-case characteristics; ML-based assignment optimizes using historical data but may inherit training biases. Using generalized difference-in-differences following Callaway and Sant'Anna (2021), I compare outcomes in treated provinces with not-yet-treated provinces to identify causal effects of these different assignment technologies.

This study shows that case assignment systems are far from a neutral infrastructure; they impacts case assignment patterns, court performance metrics, and public trust in the justice system. The transition from manual to random case assignment significantly impacted case assignment patterns, court performance, and public trust, with public perception notably diverging from internal court metrics. Specifically, random assignment reduced path dependence: the complexity of cases assigned to a judge became less dependent on the complexity of their past caseload. However, this procedural change did not improve the matching of judge expertise to case type (e.g., a family law judge was not more likely to be assigned a family law case). This outcome is anticipated, as random assignment, by its rule-based nature, disregards judge-specific expertise and case characteristics.

This shift in assignment patterns correlated with a slight reduction in court performance. We observe a decrease in the official performance index (-0.054 standard deviations) and in a measure of verdict quality (-0.114 standard deviations). These declines appear driven by an increase in appeal rates (approximately six additional appealed cases per court-division per quarter) and decreases in verdict length (around 47 fewer words) and the number of legal provisions citing (0.096 less). Counterintuitively, despite these performance declines, public trust in courts increased (by 0.067 standard deviations), primarily driven by improved perceptions of fairness in law execution and rule of law. The approximate two-year lag before this increase in trust became statistically significant suggests that the public's updated belief does not come from immediate reactions to the assignment system change itself, but rather from accumulating observations of court over time.

The transition from manual to machine learning (ML)-based case assignment yielded distinct changes in assignment patterns. Similar to random assignment, ML systems reduced path dependence in case complexity. Additionally, under ML assignment, case complexity no longer correlated with judge experience. This procedural change, however, increases the expertise-matching ratio for case assignments (i.e., cases were more likely to be assigned to judges with relevant specialization).

Despite these modifications to assignment patterns—including improved expertise matching—this did not translate into detectable gains in overall court performance. Empirical evidence indicates no significant change in either the official performance indicators or the verdict quality index. Furthermore, no significant changes were observed in the components of these indices. Consequently, public trust in courts also showed no significant change following the adoption of ML-based assignment. This lack of shift in public perception may due to the absence of discernible changes in overall court performance, or perhaps a lag in public awareness of the subtle changes in assignment mechanisms under the ML system.

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 the empirical literature on judicial decision-making by examining how case assignment systems influence court outcomes. A large body of research demonstrates that heterogeneity in judicial decisions (such as bailing, paroling and sentencing) creates systematic variation in case outcomes, with much of this work relying on random case assignment as a source of exogenous variation to identify causal effects of different judges (Kling, 2006; Agan et al., 2023; Aizer and Doyle Jr, 2015; Bhuller et al., 2020; Eren and Mocan, 2021)[2]. This approach treats assignment systems as neutral administrative infrastructure that generates useful variation without independently affecting outcomes. However, a smaller literature has begun to examine assignment systems themselves, primarily through qualitative comparisons of assignment practices across courts (Gramckow et al., 2016; Fabri and Langbroek, 2007; Macfarlane, 2023; Jin, 2020; Chilton and Levy, 2015). My paper advances this emerging area by providing the first quantitative comparison of different assignment technologies—manual, random, and machine learning-based systems—and their effects on court performance and public trust.

Second, this study contributes to the growing literature on algorithmic governance in public sector settings. As government entities increasingly use algorithms for decision-making, research has examined both custom-built and existing systems, typically comparing algorithmic performance to human decision-makers. The closest context to my setting is algorithmic use in judicial processes, where algorithms primarily assist with binary decisions such as bail, parole, and sentencing guidelines (Arnold et al., 2025; Angelova et al., 2023, 2024; Berk, 2017; Berk et al., 2016; Kleinberg et al., 2018; Stevenson and Doleac, 2024). These studies generally find that algorithms often outperform humans in consistency and impartiality (Ludwig et al., 2024). Similar findings appear in other algorithmic governance applications, including policing misconduct, tax auditing, and child protection (Stoddard et al., 2024; Kleinberg et al., 2015; Battaglini et al., 2024; Rittenhouse et al., 2023). My paper advances this literature by examining algorithms in a different task: rather than making binary decisions ("yes or no"), I study algorithms in work allocation tasks ("who gets what"). This represents a shift from algorithmic decision-making to algorithmic management that requires allocating different tasks to different workers. I find that well-designed algorithms can perform at least as effectively as human assigners, while different algorithmic approaches produce different outcomes.

Third, this paper contributes to the algorithmic management literature, where

[2]For comprehensive reviews of this literature, see Frandsen et al. (2023) and Chyn et al. (2024).

algorithms take on managerial responsibilities by assigning tasks to human workers. The algorithmic management literature primarily consists of qualitative studies focusing 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 (Rosenblat and Stark, 2016; 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 in a low-skill setting. My paper contributes quantitative evidence using causal inference to identify the effects of algorithmic versus human assignment in a high-stakes professional environment. This work extends algorithmic management beyond routine task to complex work allocation.

These findings offer crucial insights for court reform efforts globally. As judicial systems worldwide increasingly transition towards algorithmic case assignment, policymakers must recognize that these are not neutral administrative tools, but technologies with complex effects on court operations, performance metrics, and public perception. The choice of system requires careful consideration of inherent trade-offs. For instance, our evidence suggests that while random assignment can enhance public trust, it may not optimize certain aspects of measured court performance. Conversely, more sophisticated ML-based assignment systems, despite potentially improving specific case allocation, did not necessarily translate into significant gains in overall court performance or public trust in our context. The broader lesson is that seemingly neutral administrative technologies embody value trade-offs that require explicit policy choices rather than purely technical solutions. As institutions increasingly rely on algorithms for resource allocation, understanding these trade-offs becomes essential for designing systems that balance efficiency, fairness, and public acceptance.

The remainder of the paper is structured as follows: Section 1 provides institutional background on Chinese courts and the assignment reform. Section 3 describes the data and variable construction. Section 2 outlines the empirical strategy. Section 4 describes results on assignment patterns. Section 5 presents results on court performance and robustness checks. Section 6 reports results on public trust in court. Section 7 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"[3]. 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 systems (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.

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

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[4]. 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 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)[5]. 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.

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

[5] 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.

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[6]. 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[7]. The timing of adoption, as shown in Figure A4, varies at the provincial level[8]. 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 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[9]. 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-

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

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

[8]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.

[9]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.

ML court divisions to already ML court divisions[10]. 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 across groups—as highlighted in recent literature (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021)[11]. 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[12]. 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[13]. 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

[10]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.

[11]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.

[12]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.

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

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[14].

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.

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

[14]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.

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:

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

where $ATT(g,t)$ is defined in Equation 1, and $\kappa = \sum_{g \in G} \sum_t \mathbb{1}\{t \geq g\} P(G = g \mid G \leq 2020\ \text{Q4})$. 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:

$$\delta_e = \sum_{g} \sum_{t} \mathbb{1}\{t - g = e\} P(G = g \mid t - g = e) ATT(g,t), \tag{3}$$

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[15][16]. 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[17].

---

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

[16] 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.

[17] 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

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.

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[18][19].

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

traditional judicial reasoning.

[18]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.

[19]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.

the covariance matrix[20].

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[21]. 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)[22]. For judicial effort index, I draw from verdict word count, provisions cited from existing law, and provisions written in the decision, guided by the SPC' s standards and (Liu et al., 2022; Liu, 2018)[23].

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 different case assignment systems impacts court performance, let's start from analyzing how cases are assigned under each system. The case assignment process for judges in China remains a black box. Most existing studies rely on fieldwork in individual courts to understand assignment patterns, focusing on variations across courts and potential harm from human's discretion (Gramckow

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

[21]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.

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

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

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. To address this gap, I focus on two main dimensions: (1) which types of judges are assigned more complex cases, and (2) how many cases are assigned to judges with relevant expertise.

4.1 Judge-Case Correlation

For the first dimension, I examine how judge characteristics relate to case complexity before and after the introduction of random and ML-based assignment systems. These systems are expected to influence judge-case correlations in different ways. A random assignment system is designed to eliminate such correlations by assigning cases without considering judge characteristics. This breaks existing links between traits like experience, past performance, and effort, and the types of cases judges receive. In contrast, an ML-based system uses historical data to optimize assignments. It may replicate prior patterns by learning from past assignments or introduce small adjustments to improve match quality through predictive modeling.

To assess these changes, I estimate a regression of case complexity on judge characteristics, including an interaction with the policy indicator. The analysis includes all judges who served in the same court both before and after the reform:

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

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.[24] 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 historical case complexity, historical performance, historical effort, gender and experience

[24]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.

(normalized).[25] The interaction 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[26].

Pre-policy Judge-Case Correlation Figure A6 presents the baseline correlations between judge characteristics and case complexity before the implementation of random and ML-based assignment, captured by β_2. In both random and ML court divisions, judges' current case complexity is strongly correlated with their historical complexity. A one standard deviation increase in historical complexity is associated with a 0.54 (random) and 0.56 (ML) standard deviation increase in current complexity. In random court divisions, other characteristics—historical performance, effort, experience, and gender—show no significant correlation with case complexity. In ML court divisions, historical effort and experience are positively correlated with case complexity, but the effect sizes are limited.

Change in Judge-Case Correlation Figure A7 shows the changes in judge-case correlations following policy implementation, captured by β_3. After the introduction of random or ML-based assignment, the correlation between historical and current case complexity shrinks significantly. Specifically, a one standard deviation increase in historical complexity is associated with a 0.16 (random) or 0.17 (ML) smaller standard deviation increase in current case complexity. In ML court divisions, the

[25]Historical complexity is calculated by the judge's average complexity index before the policy. Historical performance and effort indices are calculated as averages over prior quarters, following the method in Anderson (2008).

[26]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 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.

correlation between judge experience and case complexity also declines.

Post-Policy Judge-Case Correlation Figure A8 shows the post-policy correlations, $\beta_2 + \beta_3$, for both systems. While historical complexity remains positively correlated with current case complexity in both random and ML court divisions, the strength of this relationship is noticeably reduced compared to the pre-policy period. In random divisions, a one standard deviation increase in historical complexity now corresponds to a 0.42 standard deviation increase in current complexity; in ML divisions, the figure is slightly higher at 0.43. In ML courts, the correlation between experience and case complexity becomes statistically insignificant, suggesting that experience no longer affects case assignments.

Overall, both systems reduce reliance on traditional assignment patterns, particularly those based on past case complexity. By the post-policy period, neither system shows significant correlations between other judge characteristics and case complexity.

4.2 Judge Expertise Matching

Other than judge's personal characteristics such as general experience, past performance, effort and gender, the expertise of judges in handling specific types of cases is also an important dimension of judge-case assignment quality. Expertise matching refers to the alignment between a judge's specialization and the nature of the case assigned to them. When a family law specialist handles divorce proceedings or a commercial law expert oversees corporate disputes, courts can leverage accumulated knowledge and specialized procedures. This specialization potentially improves case processing efficiency and verdict quality.

I measure judge expertise using relative specialization within each court division. For each judge-case pair, I calculate the proportion of cases with the same cause of action that the judge has previously handled, averaged or median across all cases in that court division. For example, the expertise of 60% means the cases, in the given court division, during the specific quarter, are overseen by judges on average/median who have previously handled 60% of cases with the same cause of action. However, I acknowledge that this measure captures only observable specialization patterns reflected in case assignment history, not unobservable judicial skills, preferences, or informal expertise that human assigners might consider when making assignment decisions.

I apply the same generalized difference-in-differences approach using Callaway and Sant'Anna (2021), with the outcome variable now defined as the average expertise matching ratio within each court division-quarter. All other specifications, including controls and clustering, remain identical to the main analysis.

Random assignment shows no significant improvement in expertise matching compared to manual assignment (Figure A9). The point estimate suggests a small decrease of 0.02 percentage points in the expertise matching ratio, but this is not statistically distinguishable from zero. This finding aligns with the design of random assignment: by construction, it disregards judge characteristics and case features that human assigners might use to create expertise-based matches.

In contrast, ML-based assignment significantly increases expertise matching by 0.02 (Figure A9), representing approximately a 4.2% improvement over the baseline matching rate of 0.470. This improvement emerges gradually over time, consistent with the algorithm learning optimal matching patterns from historical data. The ML system appears to identify and replicate successful judge-case pairings that human assigners previously made, while potentially discovering new efficient combinations that humans might have overlooked.

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 A10 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.[27] 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.[28]

In contrast, the effort index shows an immediate decline post-policy, which stabilizes around zero in subsequent quarters.[29] 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 0.096 fewer provisions per case per quarter, also significant at the 1% level.

[27] 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.

[28] 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.

[29] 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.

Heterogeneity by Divisions Figure A12 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 A12, 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 A11), 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 A13, 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.[30]

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.

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

[30]In future versions, I intend to incorporate the alternative empirical method from Sun and Abraham (2021).

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 Trust in Court

While the previous section evaluates how algorithmic case assignment affects court performance, this section focuses on a broader institutional outcome: how these changes influence public trust in the court system. Courts derive authority not only from the law but also from citizens' belief in the fairness, impartiality, and transparency of judicial processes. Thus, understanding how algorithmic reforms shape this trust is therefore critical to evaluating their institutional impact. To answer this question, I examine how ordinary citizens' attitudes toward the Chinese court system evolve as algorithmic case assignment replaces human discretion, using a nationally representative biennial survey covering all provinces in China. I then further tested the channels behind the changes in trust.

6.1 Data and Trust Index Construction

I use data from the *China Social Survey* (CSS), a nationally representative cross-sectional survey conducted biennially between 2006 and 2021 by the Chinese Academy of Social Sciences. In addition to collecting demographic, employment, and income information, the CSS asks respondents about their values, behaviors, and institutional attitudes. This survey is dedicated to the general, representative public, where the direct experience with the court system accounts for only around 1% of the sample —their responses reflect general perceptions of trust in court.

I include data from 26 provinces that adopted either random or ML-based assignment, yielding 5,700 to 8,400 observations per wave. My outcome variable is a composite trust index, created from three survey questions: (1) how much do you trust the courts, (2) how satisfied are you with the rule of law, and (3) how fair is law enforcement. Each is rated on a 1-4 scale (the larger the better), with the option to refuse to answer. I construct the index using *Multiple Correspondence Analysis (MCA)* to reduce dimensionality and treat refusal to answer as an informative and negative response[31][32]. The MCA index is normalized such that higher values indicate greater trust. The first MCA dimension explains 12.50% of the total variation and loads positively on all three questions.

6.2 Empirical Strategy

To estimate the impact of algorithmic case assignment on public trust, I follow the same empirical strategy as in the performance evaluation section. Specifically, I use CSDiD, using the staggered adoption of algorithmic assignment across court divisions. I estimate separate regressions for random and ML-based assignment systems and use not-yet-treated provinces as the control group, since no province remains untreated in the end.

The unit of analysis is the individual-year level, and standard errors are clustered at the provincial level. I include four sets of control variables corresponding to four domains of individual characteristics that may influence trust: demographic background, socioeconomic status, ideological orientation, and exposure to or connection with the legal system. These control variables allow CSDiD to reweight untreated observations to better match the treated group, giving more weight to control units with similar characteristics.

6.3 Effects of Random Assignment

Figure A14 left panel presents the event study results for provinces that implemented random assignment. The results satisfy the parallel trends assumption. Beginning in the second year after reform, trust in the court increases relative to the not-yet-treated provinces. According to Table A5, the point estimates suggest an increase of approximately 0.1 to 0.2 standard deviations of the standardized trust index. While

[31] I use MCA instead of Principal Component Analysis (PCA) because MCA is designed for ordinal variables where the ranking of the categories matters. PCA treats variables as continuous and does not account for the rank order of categories.

[32] In most studies, nonresponse is coded as missing. However, under authoritarian regimes, silence may serve as an implicit signal of disapproval (Tannenberg, 2022; Robinson and Tannenberg, 2018). Ignoring such responses risks underestimating dissatisfaction. I also present a version where refusals are dropped in Appendix C; results remain consistent but exhibit smaller magnitudes.

small, the effect is statistically significant and persistent across different types of controls.

This increase may reflect growing perceptions of procedural fairness. Random assignment is straightforward and easily understood by the public. Even those who are not directly aware of the reform may indirectly observe improved consistency or fairness in court outcomes through word of mouth, local news, or community networks. The reform may thus enhance trust both through direct awareness and indirect inference.

6.4 Effects of ML-Based Assignment

Figure A14 right panel reports the corresponding results for ML-based assignment. The parallel trends assumption again appears satisfied. However, the estimates show no significant change in the trust index following the reform (Table A5). This suggests that while ML may improve internal efficiency or judge-case matching, these gains do not translate into higher public trust.

One likely explanation is the lack of transparency of the ML assignment system. Unlike random assignment, ML-based decision-making lacks intuitive rules. Members of the public may be unaware that a change occurred or may find the reform too complex to interpret as a fairness signal. As a result, the reform's internal benefits remain institutionally invisible to the broader public.

7 Conclusion

This paper provides new evidence on how case assignment systems affect court performance and public trust by studying China's nationwide transition from manual to algorithmic assignment systems.

Using 66 million court documents, I find that random assignment reduces court performance but increases public trust, while ML-based assignment improves case matching without affecting overall performance or trust.

This study makes three contributions to existing literature. First, it provides the first quantitative comparison of different assignment technologies in high-stakes settings, showing that administrative systems are not neutral infrastructure but important policy choices. Second, it advances the algorithmic governance literature by examining task allocation rather than binary decision-making, demonstrating that well-designed algorithms can match human performance in complex professional environments. Third, it contributes to judicial decision-making research by showing that assignment systems themselves—not just individual judge characteristics—systematically affect court outcomes.

These findings offer suggestive evidence for court reform efforts. Courts considering random assignment should weigh modest performance costs against significant gains in public trust. Those implementing ML-based systems should not expect automatic performance improvements but can achieve comparable results to human assignment while maintaining transparency. Most importantly, policymakers must recognize that seemingly neutral administrative technologies embed value trade-offs requiring explicit policy choices rather than purely technical solutions.

References

Agan, Amanda, Jennifer L Doleac, and Anna Harvey, "Misdemeanor prosecution," *The Quarterly Journal of Economics*, 2023, *138* (3), 1453–1505.

Aizer, Anna and Joseph J Doyle Jr, "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," *The Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

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 Dobbie, and Crystal S Yang, "Algorithmic Recommendations When the Stakes Are High: Evidence from Judicial Elections," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2024, pp. 633–637.

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

Arnold, David, Will Dobbie, and Peter Hull, "Building Nondiscriminatory Algorithms in Selected Data," *American Economic Review: Insights*, 2025, *7* (2), 231–249.

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L Miller, and Eleonora Patacchini, "Refining public policies with machine learning: The case of tax auditing," *Journal of Econometrics*, 2024, p. 105847.

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.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad, "Incarceration, recidivism, and employment," *Journal of political economy*, 2020, *128* (4), 1269–1324.

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.

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.

Chyn, Eric, Brigham Frandsen, and Emily C Leslie, "Examiner and Judge Designs in Economics: A Practitioner's Guide," Technical Report, National Bureau of Economic Research 2024.

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.

_ and _ , "Juvenile punishment, high school graduation, and adult crime: Evidence from idiosyncratic judge harshness," *Review of Economics and Statistics*, 2021, *103* (1), 34–47.

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.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie, "Judging judge fixed effects," *American Economic Review*, 2023, *113* (1), 253–277.

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.

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.

_ , Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction policy problems," *American Economic Review*, 2015, *105* (5), 491–495.

Kling, Jeffrey R, "Incarceration length, employment, and earnings," *American Economic Review*, 2006, *96* (3), 863–876.

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.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan, "The unreasonable effectiveness of algorithms," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2024, pp. 623–627.

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

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

Robinson, Darrel and Marcus Tannenberg, "Self-Censorship in Authoritarian States: Response bias in measures of popular support in China," 2018.

Rosenblat, Alex and Luke Stark, "Algorithmic labor and information asymmetries: A case study of Uber' s drivers," *International journal of communication*, 2016, *10*, 27.

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

Stevenson, Megan T and Jennifer L Doleac, "Algorithmic risk assessment in the hands of humans," *American Economic Journal: Economic Policy*, 2024, *16* (4), 382–414.

Stoddard, Greg, Dylan J Fitzpatrick, and Jens Ludwig, "Predicting police misconduct," Technical Report, National Bureau of Economic Research 2024.

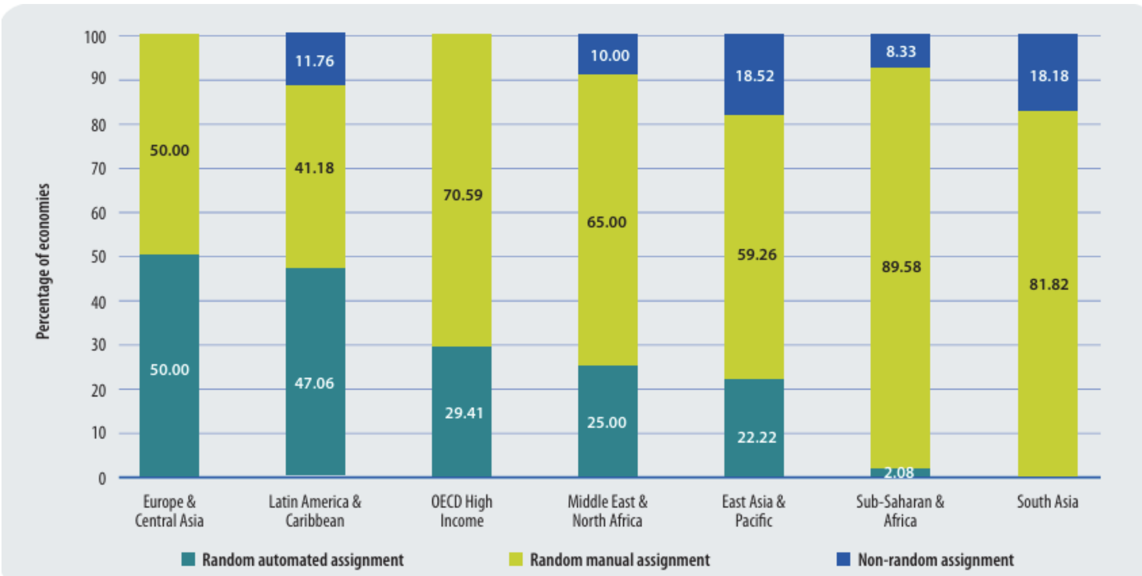
Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.

Tannenberg, Marcus, "The autocratic bias: self-censorship of regime support," *Democratization*, 2022, *29* (4), 591–610.

A Tables and Figures

A.1 Figures

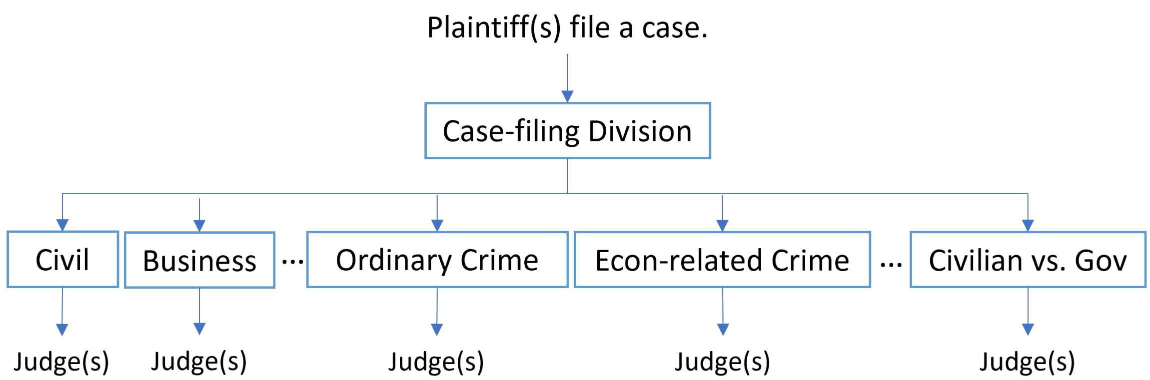
Appendix Figure A1. Case Assignment System Worldwide by 2016



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.

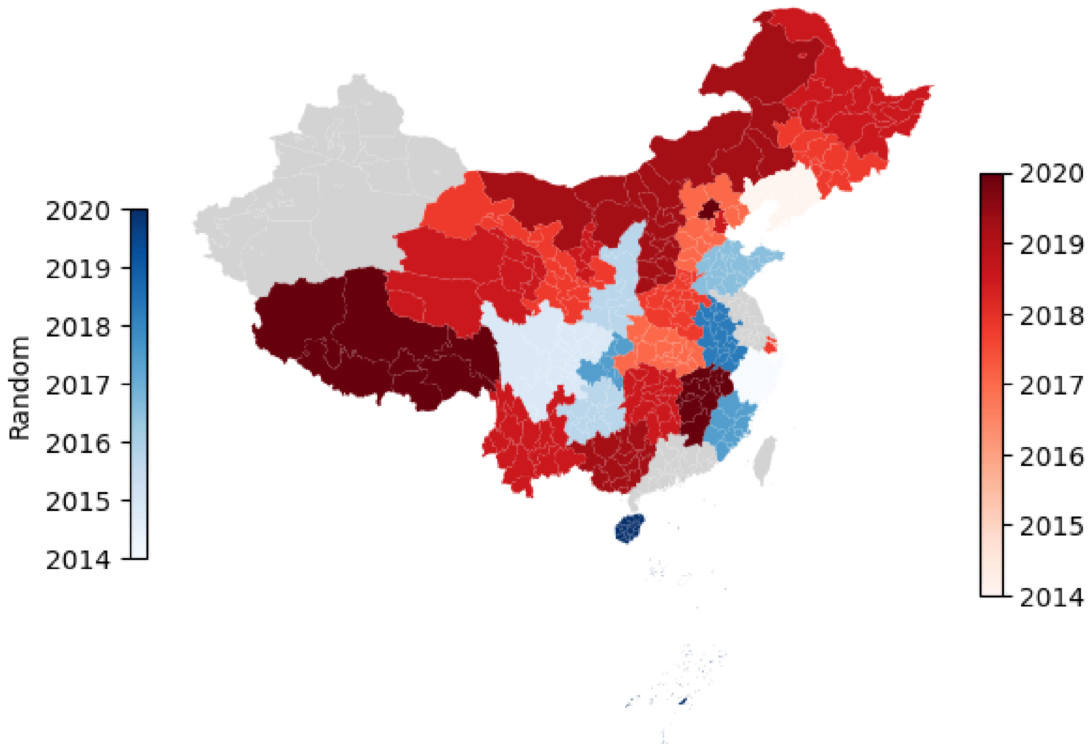
Source: Doing Business database, http://www.doingbusiness.org/data.

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.

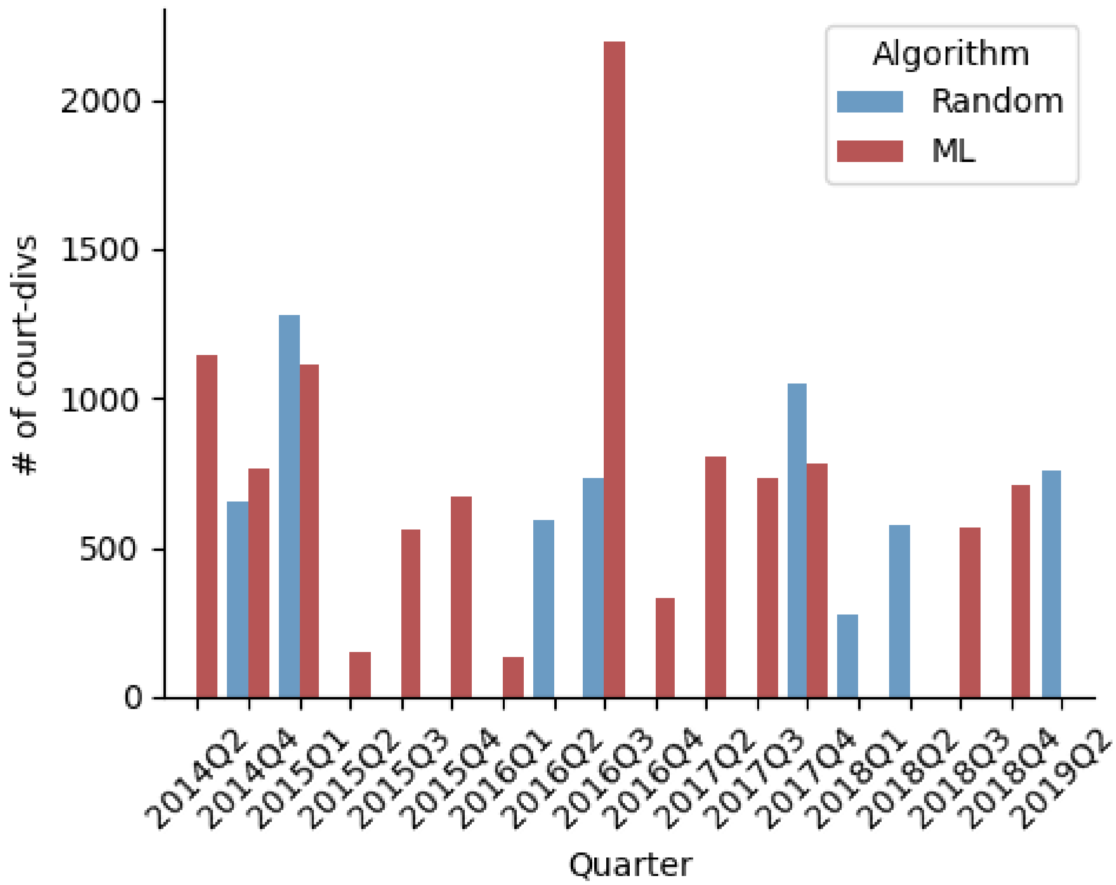
Appendix Figure A3. Algorithmic Case Assignment System Adoption by Prefecture



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.

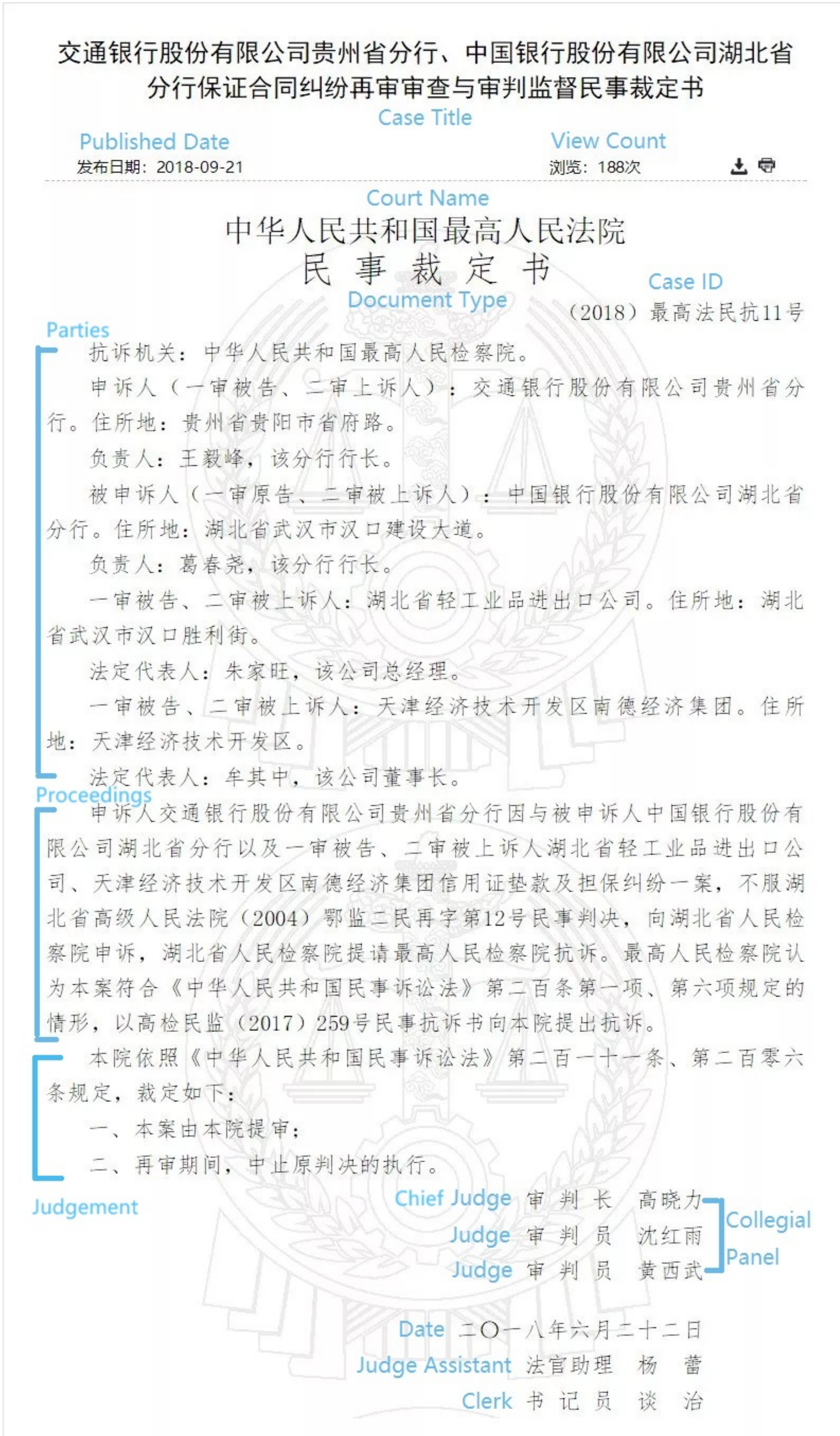
Appendix Figure A4. Algorithmic Case Assignment System Adoption by Courts



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

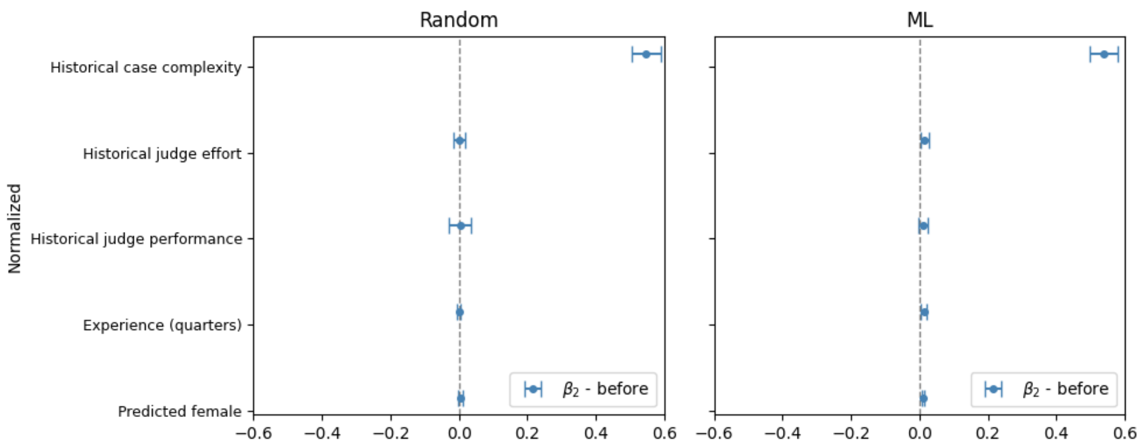
法官助理 杨 蕾

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.

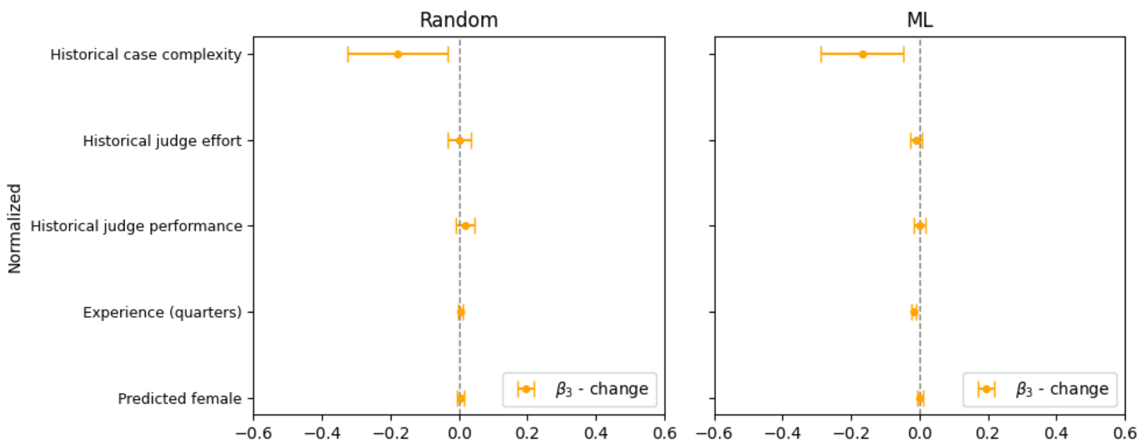
Source: China Judgment Online.

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



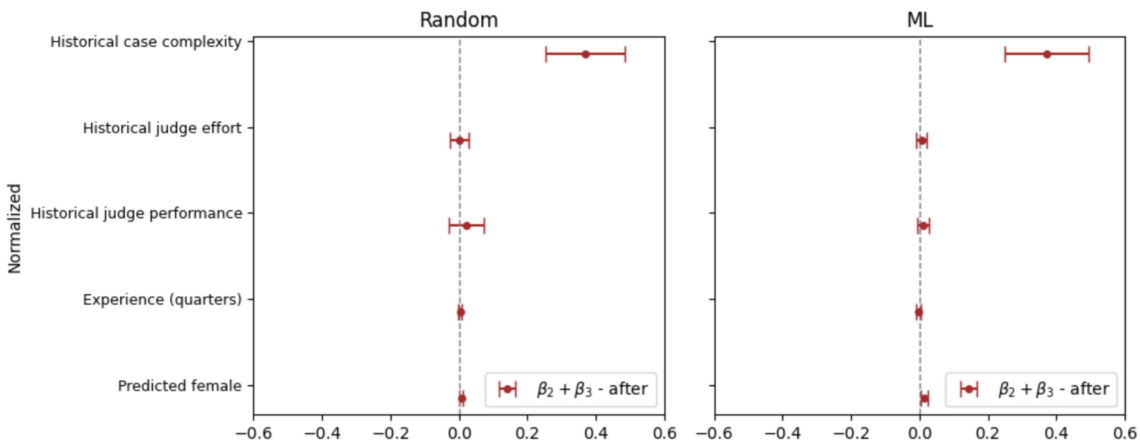
Notes: This figure illustrates case assignment patterns in random and ML-based courts during the manual era, displaying the β_2 coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how judge characteristics correlate with case complexity before the assignment system change. The analysis uses all cases involving judges observed in both the pre- and post-reform periods. Historical case complexity, judge effort and judge performance indices are standardized indices from weighted averages of multiple variables, normalized to pre-treatment observations. Experience is similarly normalized. 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



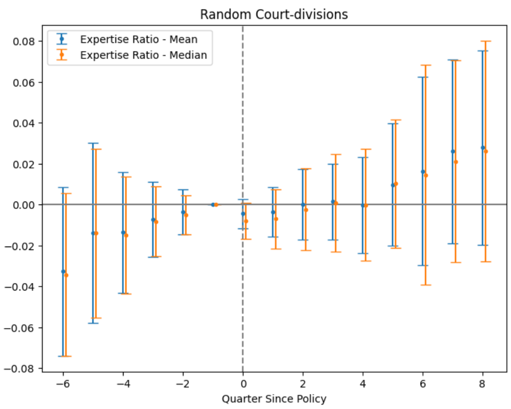
Notes: This figure illustrates the change in case assignment patterns in random and ML-based courts due to the algorithmic systems, displaying the β_3 coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how differently judge characteristics correlate with case complexity since the assignment system change. The analysis uses all cases involving judges observed in both the pre- and post-reform periods. Historical case complexity, judge effort and judge 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

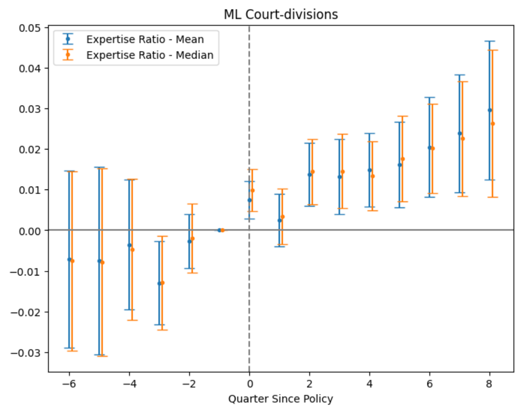


Notes: This figure illustrates case assignment patterns in random and ML-based courts after the algorithmic system adoption, displaying the $\beta_2 + \beta_3$ coefficient from the following equation: $\text{case\_complexity}_i = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{judge\_char}_{jt} + \beta_3 (\text{judge\_char} \times \text{Post})_{jt} + \epsilon_{ijt}$. This coefficient shows how judge characteristics correlate with case complexity after the assignment system change. The analysis uses all cases involving judges observed in both the pre- and post-reform periods. Historical case complexity, judge effort and judge 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 Algorithmic Assignment on Judge Expertise Ratio



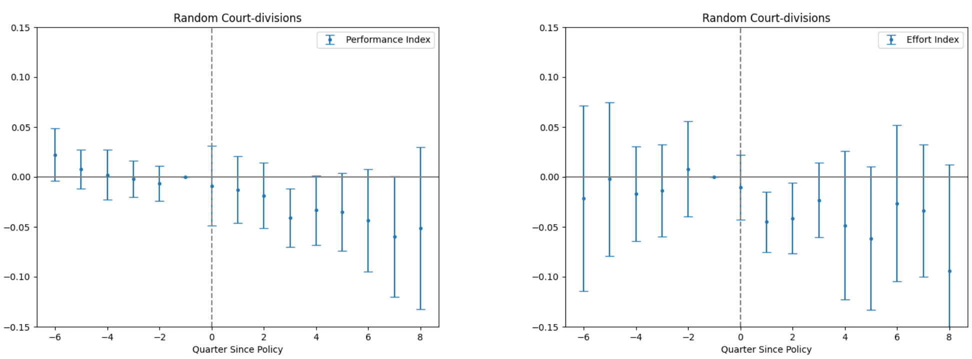
(a) Random Assignment



(b) ML Assignment

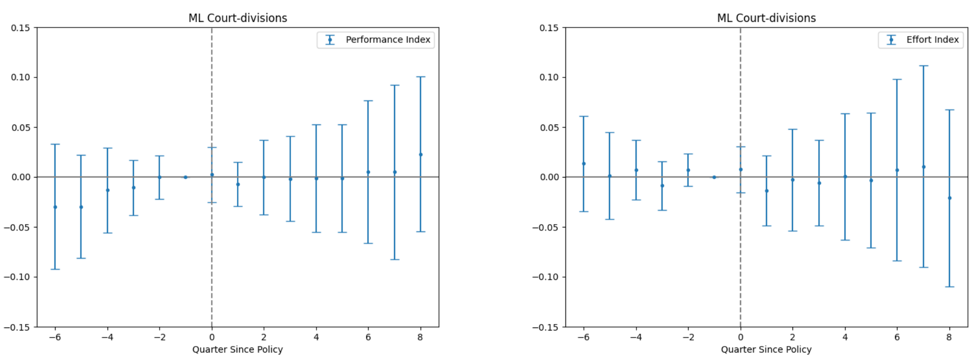
Notes: The figure presents event-study estimates of the effects of random/ML case assignment system on expertise ratio of judges. The outcome variable, judge expertise ratio, is measured as the share of cases within a judge's specialization out of the total cases assigned to the judge, I take the mean (blue) and median (orange) across all judges within a court-division in each quarter. The left panel contains the random court-divisions. The right panel contains the ML court-divisions. Only the right panel display post-treatment and anticipatory effects, with 95% confidence intervals, derived from an event-study model corresponding to Equation 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/ML assignment, while the control group comprises not-yet-implemented random/ML assignment court-divisions. Control variables include caseload inflow, and standard errors are clustered at the provincial level.

Appendix Figure A10. Event Study of Random Assignment on Performance Index and Efforts Index



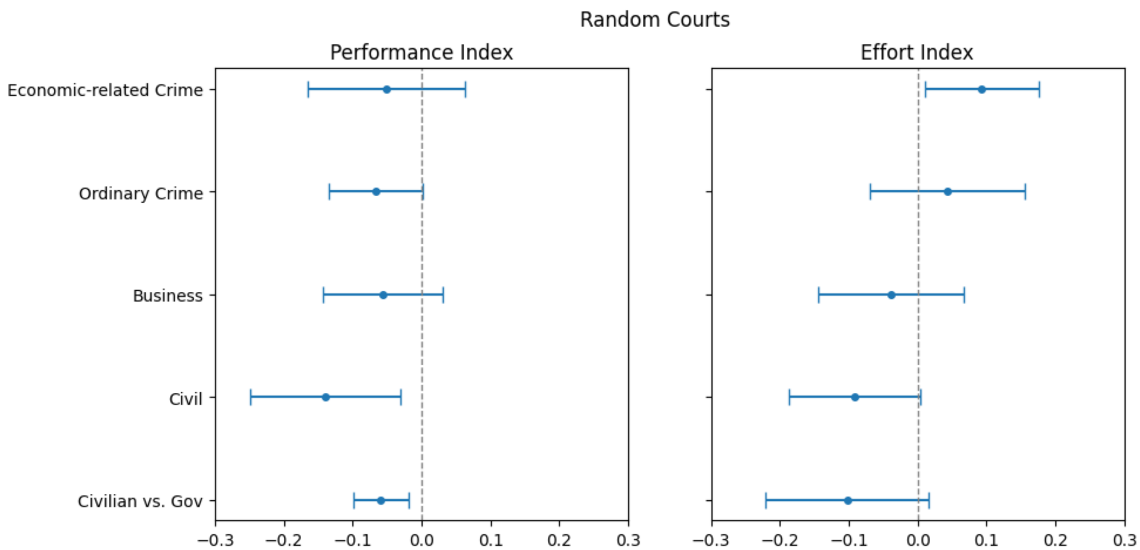
Notes: The figure presents event-study estimates of the effects of a random case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal 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 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 A11. Event Study of ML-Based Assignment on Performance Index and Efforts Index



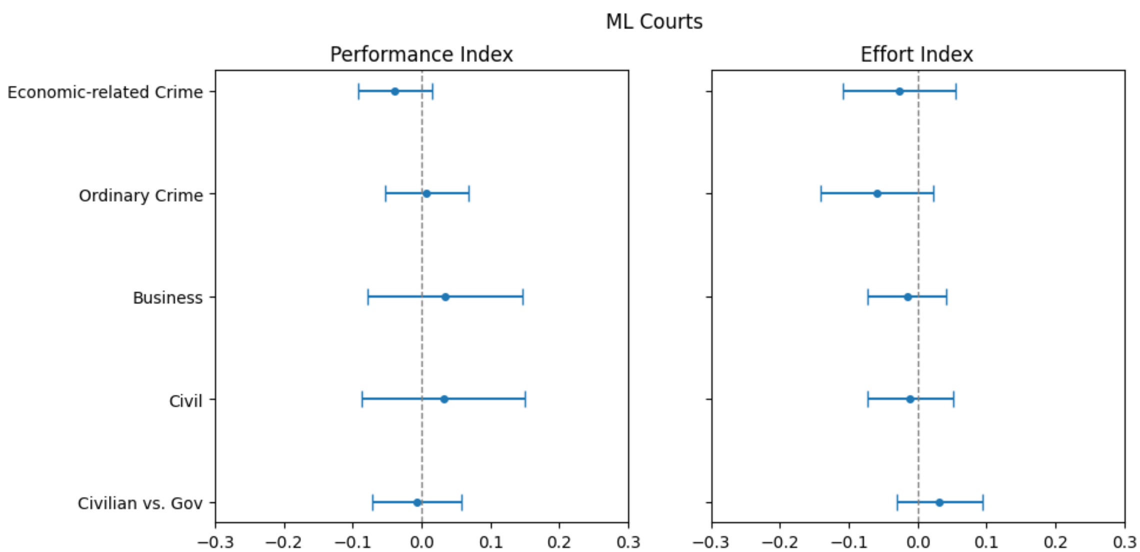
Notes: The figure presents event-study estimates of the effects of a ML-based case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, a weighted average of normalized # of case processed, not appeal rate and not reversal 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 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 A12. ATT of Random Assignment on Performance Index and Efforts Index across Divisions



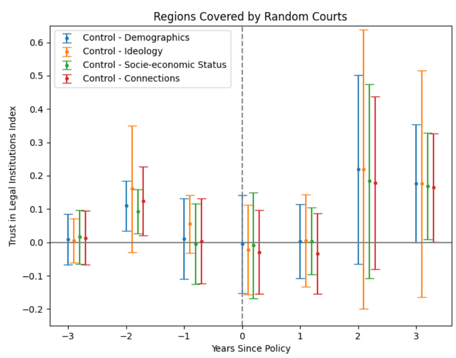
Notes: The figure shows the average treatment effects of an random case assignment system on performance (output) and effort (input) indices across five court divisions: civil vs. government, civil, business, ordinary crime, and economic-related crime. The left panel uses the performance index as the 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 **??**. 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 A13. ATT of ML-Based Assignment on Performance Index and Efforts Index across Divisions

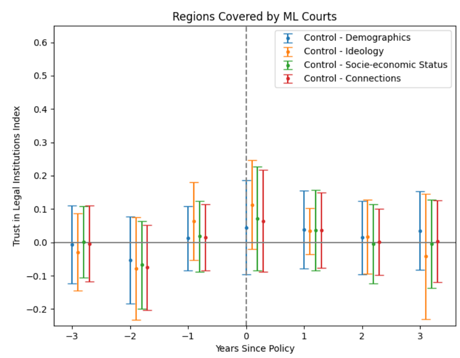


Notes: The figure shows the average treatment effects of an ML-based case assignment system on performance (output) and effort (input) indices across five court divisions: civil vs. government, civil, business, ordinary crime, and economic-related crime. The left panel uses the performance index as the 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 **??**. 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.

Appendix Figure A14. Event Study of Algorithmic Assignment on Public Trust in Court



(a) Random Assignment



(b) ML Assignment

Notes: The figure presents event-study estimates of the effects of random/ML case assignment system on public trust in court. The outcome variable, the public trust index, is measured by using multiple survey questions from the Chinese General Social Survey (CGSS) and the Chinese Social Survey (CSS), and applying the multiple correspondence analysis (MCA) method to construct a composite index. The left panel contains the random court-divisions. The right panel contains the ML court-divisions. Only the left panel display post-treatment and anticipatory effects, with the influeces manifests after two years. This effect stands 95% confidence intervals, derived from an event-study model corresponding to Equation 1. The unit of analysis is each individual surveyee and the unit of time is year. The treatment group includes surveyees located within the jurisdictions of courts already implementing random/ML assignment, while the control group comprises surveyees in not-yet-implemented random/ML assignment court-divisions jurisdiction. Each color represents a different set of control variables, capturing demographics (blue), ideology (orange), social-economic status (green) and possible court conenctions (red), and standard errors are clustered at the provincial level. The result is robust to different sets of control variables.

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.

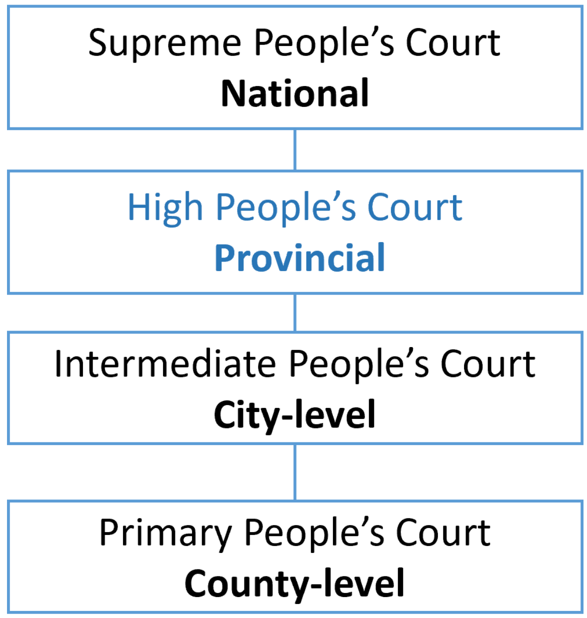
Appendix Table A5. Effect of Random/ML Assignment on Public Trust in Legal Institutions

	Outcome: Trust in Legal Institutions Index			
Random Courts	0.067*	0.071	0.064*	0.030
	(0.039)	(0.053)	(0.034)	(0.037)
N	27,490	27,490	27,490	27,490
ML Courts	0.036	0.041	0.033	0.035
	(0.044)	(0.042)	(0.050)	(0.045)
N	20,536	20,536	20,536	20,536
Control	Demographics	Ideology	Socio-economic Status	Connections
Cluster	Province	Province	Province	Province

Notes: This table presents the ATT of random/ML assignment system on the trust index in the legal institutions. The outcome variable, the public trust index, is measured by using multiple survey questions from the Chinese General Social Survey (CGSS) and the Chinese Social Survey (CSS), and applying the multiple correspondence analysis (MCA) method to construct a composite index. The first row contains the ATT of surveyees living in places implementing random assignment and the second row contains the ATT of those living in ML assignment administrative areas. Column (1)-(4) shows the ATT of random/ML after controlling for different sets of control variables: demographics, ideology, socio-economic status, and connections. 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.

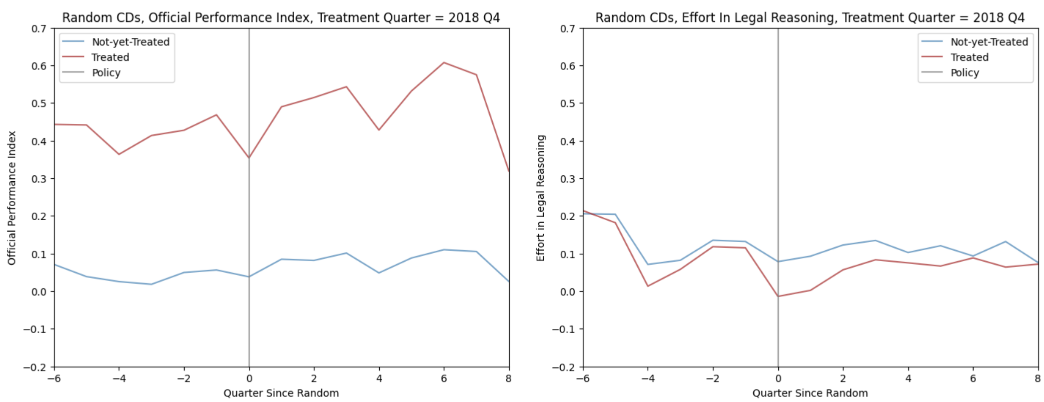
B Additional Graphs

Appendix Figure B1. Court Structure of China



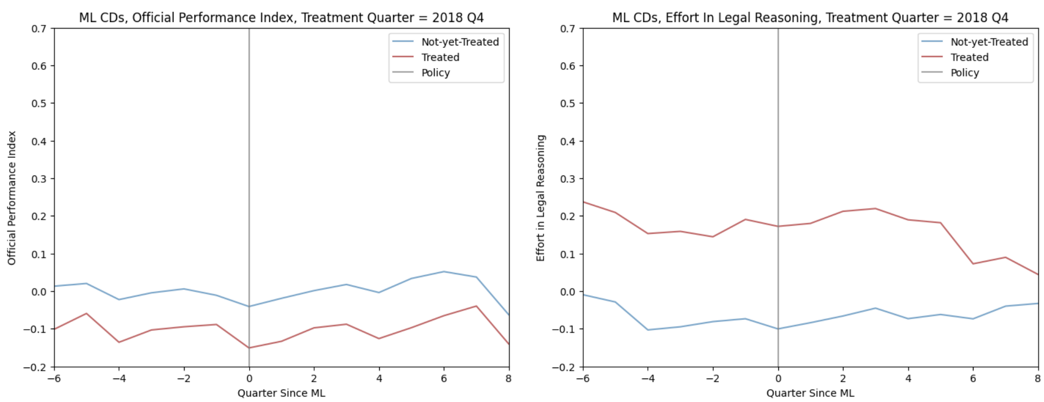
Notes: This figure illustrates the court structure in China. The Chinese court system is not a homogeneous entity. Policy documents and commands are issued top-down, with instructions cascading from the Supreme People's Court (SPC) to the primary courts. China's judicial system operates hierarchically 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



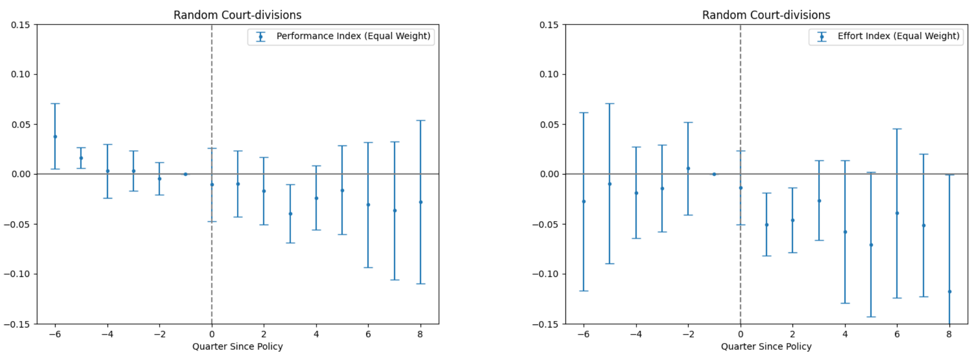
Notes: The figure illustrates descriptive evidence on the impact of random assignment on court-division outcomes by plotting performance (left panel) and effort (right panel) trends in both the control and treatment groups over time, relative to the period of random assignment adoption. This example 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



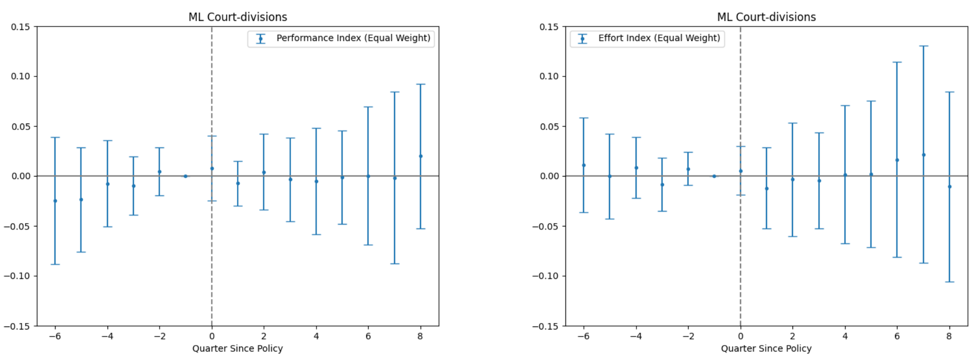
Notes: The figure illustrates descriptive evidence on the impact of ML-based assignment on court-division outcomes by plotting performance (left panel) and effort (right panel) trends in both the control and treatment groups over time, relative to the period of ML-based assignment adoption. This example 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



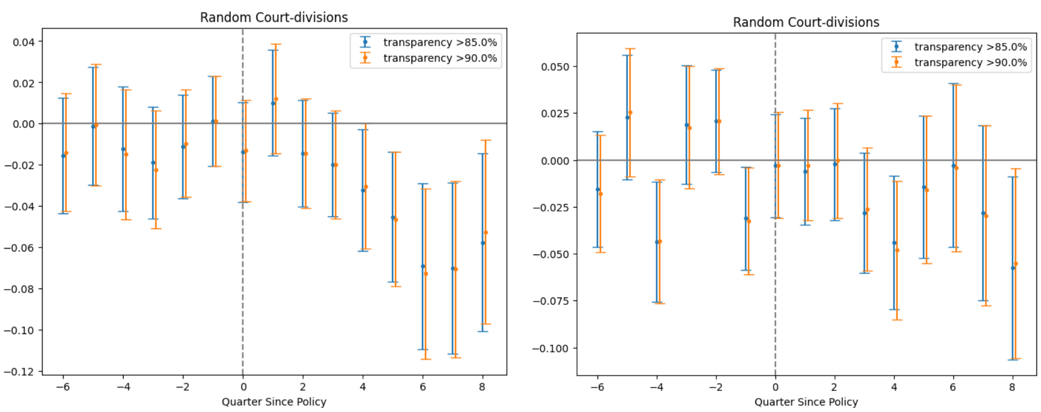
Notes: The figure presents event-study estimates of the effects of a random case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, an equal weighted average of normalized # of case processed, not appeal rate and not 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 **??**. 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



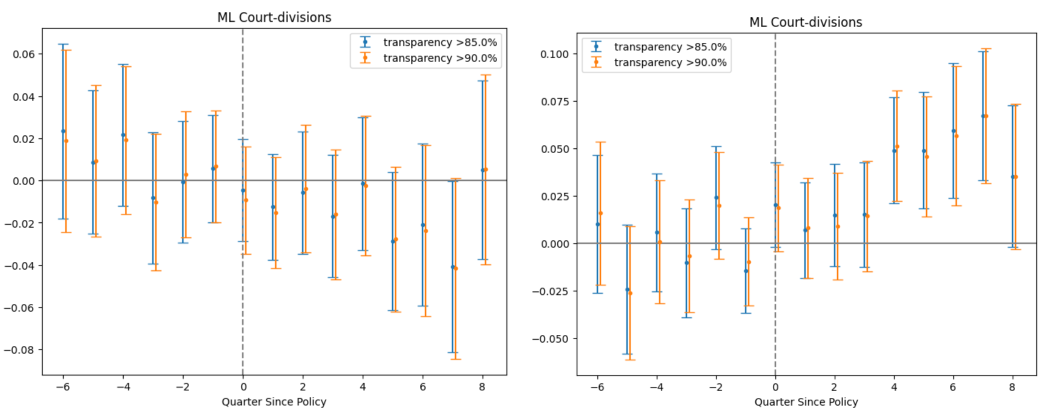
Notes: The figure presents event-study estimates of the effects of a ML-based case assignment system on performance (output) and effort (input) indices. The left panel uses the performance index as the outcome variable, an equal weighted average of normalized # of case processed, not appeal rate and 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 **??**. 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 **??**. 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 **??**. 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

	(1)	(2)	(3)	(4)
	Panel A: Outcome Variable - Performance			
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.