Judicial Assignment against Judge Shopping

Hasan Tas, Jingui Xie
TUM School of Management, Technical University of Munich, TUM Campus Heilbronn, 74076 Heilbronn, Germany
hasan.tas@tum.de, jingui.xie@tum.de

Galit B. Yom-Tov
Faculty of Data and Decision Sciences, Technion–Israel Institute of Technology, Haifa 3200003, Israel, gality@technion.ac.il

Judge shopping refers to litigants' strategic choice of filing venue to increase the chance that their case is assigned to a favorable judge. Hence, case assignment rules that determine who processes a juridical case also influence where demand is concentrated. Courts often assign cases based on specialization to improve justice and efficiency, but doing so can make particular judges targetable. Anti-shopping reforms use random assignment rules to reduce *targetability*. This weakens the *shopping incentive* and reduces filing concentration, but also creates *specialization mismatch* and reduces efficiency of case processing. We study this trade-off in a queueing game in which litigants choose between courts before cases are assigned to heterogeneous judges. We show that judge shopping increases system-wide waiting by concentrating filings in a court with targeted judges. Reducing targetability relieves this concentration but creates specialization mismatch. The optimal policy therefore balances congestion relief against processing-efficiency loss. Without a specialization advantage, eliminating targetability is optimal. When specialization improves processing, the processing-efficiency loss from specialization mismatch makes it optimal to preserve some targetability. The resulting optimum need not coincide with uniform random assignment. Greater congestion shifts the optimum toward preserving specialist matching, and at sufficiently high congestion full targetability becomes optimal.

Key words: Court operations; judge shopping; strategic routing; endogenous venue choice; judicial specialization.

1. Introduction

Courts worldwide face persistent congestion: growing backlogs delay access to justice and increase litigation costs for individuals, firms, and governments (European Commission for the Efficiency of Justice 2024, National Judicial Data Grid 2026, Randhawa et al. 2026). Scarce judicial capacity and rising caseloads are important drivers of these delays, but congestion can also arise from how demand responds to judicial assignment. Judges differ in ideology, discretion, and personal background, which can affect expected litigation outcomes (Kahan and McKenzie 2021, Ganglmair et al. 2026). Litigants anticipate those differences and exploit case assignment patterns by selecting venues that increase their chance of reaching a judge they perceive as more favorable to their case. This practice, known as *judge shopping*, can concentrate filings in particular courts, increasing delays and reducing overall system efficiency.

Two recent episodes illustrate the scale of the filing response to assignment rules. In the Western District of Texas (WDTX), a deterministic assignment rule gave patent litigants certain access to Judge Albright and led to a dramatic concentration of patent filings in the district. WDTX patent filings increased by more than

700%, with more than 20% of all U.S. patent cases concentrated on a single judge's docket (Ganglmair et al. 2026). After the court reduced assignment probability to Albright, WDTX patent filings declined by more than 70%, while the Eastern District of Texas (EDTX) received a growing share of patent filings over the same period (Castle 2024) (Appendix B). Large Chapter 11 filings similarly shifted away from the Southern District of New York after it adopted random assignment (Levitin 2023). Together, these episodes show that litigants can respond strongly to *targetability*—the probability of reaching a preferred judge—when choosing where to file, and that random assignment can substantially reduce the resulting filing concentration. Recent controversies in high-stakes litigation have prompted reforms that expand random case assignment, reduce targetability, and protect judicial impartiality and public confidence (Judicial Conference of the United States 2024).

Yet reducing targetability is not without an operational cost. Judges are scarce (European Commission for the Efficiency of Justice 2024); those with the specialization, expertise, or experience needed for a particular case type are even scarcer (Kesan and Ball 2011). Courts use dedicated dockets and specialization-based assignment to match cases with relevant judicial expertise and improve processing performance (Fabri and Langbroek 2007, Eisenberg et al. 2012, Coviello et al. 2019). Because relevant expertise may be concentrated in a narrow pool, and some courts or divisions have only one or a few eligible judges (Ganglmair et al. 2026, Gouzoules 2024), such assignment can make particular judges targetable. Lowering targetability then changes filing behavior through two channels. First, making the preferred judge harder to reach weakens the *shopping incentive* and makes the venue less attractive. Second, greater *specialization mismatch* sends more cases to nonspecialists, slowing processing and raising waiting, which further discourages filing there. These forces reduce filing concentration, but at a cost: the resulting congestion relief comes with lower processing efficiency. A judicial-administration review captures this tension: random assignment may curb judge shopping but reduce specialization and efficiency; "the challenge is to achieve a proper balance" (Jayawickrama 2011, p. 43).

This trade-off raises three questions: How can judge shopping under high targetability influence system-wide waiting? How do targetability, judicial specialization, and congestion shape equilibrium filing and waiting across courts? How should courts choose the targetability level as litigants place greater value on reaching a preferred judge and courts become more congested?

We study these questions using a queueing-game framework in which litigants strategically choose where to file before judicial assignment, while the assignment probability determines both the targetability of preferred judges and the efficiency of case processing. We characterize equilibrium filing behavior and the assignment probability that minimizes system-wide waiting. Our analysis shows that judge shopping can concentrate filings in the targeted court and increase system-wide waiting; reducing targetability can mitigate this concentration but creates specialization mismatch and lowers processing efficiency. Congestion moderates judge shopping by making litigants more responsive to waiting differences across courts. Consequently,

the optimal assignment probability depends on the interaction among the shopping benefit, specialization advantage, and congestion: greater shopping benefits favor lower targetability, whereas greater congestion can favor preserving targetability. Without a specialization advantage, minimum targetability is optimal; with specialization, partial or full targetability can be optimal.

This paper contributes to the emerging literature on court operations and, more broadly, to the design of service systems with strategic demand in three ways.

- First, we introduce judge shopping and judicial assignment as an operational problem in court systems. We show that the assignment probability can shape congestion by influencing where litigants choose to file. This creates a distinctive feature of court operations: the probability with which cases are assigned to targeted judges also affects the demand arriving at the court. Our analysis therefore identifies targetability as a demand-management lever for mitigating congestion without changing judicial capacity, while also quantifying the processing-efficiency cost of lower targetability when judges are specialized.
- Second, we develop a queueing-game analysis that captures the joint determination of strategic filing, judicial assignment, specialization, and endogenous congestion. We provide a complete characterization of the symmetric filing equilibrium and identify conditions under which asymmetric equilibria arise despite ex ante symmetric case types. The analysis reveals that endogenous case mix can decouple case volume, utilization, and waiting: a judge receiving more cases need not have both higher utilization and longer waiting. More broadly, these results extend strategic queueing analysis to a setting in which the assignment probability affects not only congestion after demand enters the system but also strategic choices over where that demand enters.
- Third, we characterize the optimal assignment probability by balancing the congestion relief from limiting judge shopping against the processing-efficiency loss from specialization mismatch. Without a specialization advantage, minimum targetability is optimal, whereas specialization can make partial or full targetability optimal. A key result is that congestion shifts this trade-off toward greater targetability: greater congestion disciplines judge shopping by making litigants more responsive to waiting differences, while simultaneously increasing the cost of inefficient case–judge matching as the system approaches capacity. Consequently, the optimal assignment probability increases with congestion, and dedicated specialist assignment can become optimal at sufficiently high congestion. These results identify when courts should favor minimum, partial, or full targetability.

The remainder of the paper is organized as follows. Section 2 reviews the related literature, Section 3 presents the model, Section 4 characterizes the symmetric filing equilibrium and its comparative statics, Section 5 quantifies the operational cost of judge shopping, and Section 6 analyzes the assignment policy. Section 7 examines equilibrium multiplicity, its policy implications, and the extension to larger specialist pools and capacity expansion. Section 8 discusses the managerial and institutional implications, and Section 9 concludes. All technical proofs are provided in Appendix A.

2. Literature Review

Our work lies at the intersection of court operations, queueing games with strategic customers, and skill-based routing and flexible-service systems. These streams study how courts manage judicial capacity, how customers respond to congestion and service differences, and how specialization and server capability affect processing performance. We bring these elements together by allowing assignment to affect both where litigants file and how efficiently their cases are processed.

2.1. Court Operations and Judge Shopping

Courts are public service systems that process cases with limited heterogeneous judicial capacity under uncertain arrivals. Their performance affects delays, public expenditure, timely access to justice, litigation costs, and confidence in legal institutions. Yet research on court operations has received limited attention in the Operations Management (OM) literature compared with other service systems.

Early analytical work viewed courts as interconnected flow systems. Blumstein and Larson (1969) modeled workloads, costs, resource requirements, and interactions across the criminal-justice process, while Buchanan and Fennell (1981) studied information, capacity, and case-flow control for managing judicial delay. Early management-science and OR applications studied court delay through sequencing, network simulation, and simulation-based scheduling (Nagel et al. 1978, Taylor et al. 1982, Kwak et al. 1984). More recent OM work studies congestion, judicial workload, capacity allocation, and litigation flow using queueing and service-system models (Bakshi et al. 2025, Freund and Weng 2026, Feigin et al. 2026, Randhawa et al. 2026, Kadzelashvily et al. 2026).

Most of this literature emphasizes the supply side of court operations: given a realized caseload, how should courts schedule cases, assign judges, allocate capacity, and redesign workflows? This includes judge and session scheduling (Brooks 2012, Bray et al. 2016), staffing and resource reallocation (Gupta and Bolia 2024, Peyrache and Zago 2024), and efficiency and workflow redesign (Falavigna et al. 2015, Azaria et al. 2023, 2024). Demand-side responses before cases enter the processing system have received much less attention. There are exceptions: Shamir and Shamir (2012) endogenize prosecutors' charging decisions, and Peyrache and Zago (2024) allow demand to respond to disposition time. Neither, however, studies strategic venue choice induced by judicial assignment. Recent work identifies endogenous demand, strategic litigant behavior, and the interaction between assignment and specialization as important open questions (Randhawa et al. 2026).

A separate legal and empirical literature documents the demand-side mechanism of judge shopping, whereby litigants select a court or division to increase the probability of reaching a preferred judge or judge pool (Gouzoules 2024, Ganglmair et al. 2026). Studies of patent and large Chapter 11 bankruptcy filings show that high targetability can make particular venues more attractive and generate substantial filing concentration (Ganglmair et al. 2026, Levitin 2023). This evidence establishes that assignment rules

affect where demand enters the court system and has motivated institutional reforms intended to limit strategic targeting (Judicial Conference of the United States 2024). Taken together, this literature suggests that reducing targetability limits judge shopping by making a preferred judge harder to reach. Our model reproduces this demand response, but shows that it does not necessarily reduce system-wide waiting.

Legal and institutional sources document specialization-based judicial assignment. The Supreme Court of India notes that judicial rosters may account for judges' subject-matter specialization (Supreme Court of India 2018, para. 19, pp. 23–24). In the Israel Supreme Court, judges receive a greater share of cases in their areas of expertise than would arise under random assignment (Eisenberg et al. 2012). Assignment rules can therefore affect both the likelihood of reaching a particular judge and how well cases are matched to judicial expertise. Because specialization can improve processing performance (Zimmer 2009, Coviello et al. 2019, Castelliano et al. 2021, Baum 2009), reducing targetability may deter judge shopping while weakening valuable specialist matching. These findings suggest that sending more cases to specialist judges should improve system performance. In our setting, however, specialist assignment can also attract strategic filings to the targeted court, so this system-level implication need not hold.

Random assignment is a longstanding institutional benchmark in courts and a prominent response to judge shopping (Kahan and McKenzie 2021, Ganglmair et al. 2026, Levitin 2023, Judicial Conference of the United States 2024). At the same time, actual assignment can depart from uniform random assignment because of specialization, expertise, experience, and other institutional considerations (Eisenberg et al. 2012, Thorley 2020). In our two-judge setting, uniform random assignment, $p = 1/2$, therefore provides a natural anti-shopping benchmark. We show that this benchmark has no special operational status: it can be infeasible when it removes too much specialist processing capacity, and even when it is feasible, the waiting-minimizing policy can lie on either side of $p = 1/2$.

Within court operations, Ata et al. (2023) are the closest: they study when defendants accept plea offers under judicial rotation and capacity constraints. We instead study where litigants file before assignment and how the assignment probability affects the use of specialist judges.

2.2. Queueing Games and Strategic Customer Behavior

Our paper also relates to queueing games, where customers anticipate congestion when deciding whether, when, and where to seek service. Since Naor (1969), this literature has shown that strategic choices can produce demand patterns that differ from the system optimum and motivate interventions such as admission control, pricing, information disclosure, priority, and routing. A classical alternative-server benchmark is Bell and Stidham (1983), who compare individually chosen and socially optimal allocations across alternative servers. Their inefficiency arises because customers choose where to obtain service without accounting for the congestion they impose on others. Our setting adds a distinct mechanism: the assignment probability changes both the private attractiveness of a venue and the processing efficiency of the cases routed there.

A related economics literature interprets waiting as a nonprice allocation mechanism. Barzel (1974) treats waiting as a time price, while Holt and Sherman (1982) model arrival time as a bid in a waiting-line auction and show that changes in the allocation procedure can be offset by strategic behavior. These mechanisms are conceptually related but distinct from ours: litigants do not choose queue position or waiting time; they choose a venue, and waiting arises endogenously from the resulting filing allocation.

Queueing games show that service attractiveness and congestion jointly shape customer routing. Strategic routing can create inefficient demand patterns and even destabilize queueing networks (Parlaktürk and Kumar 2004), while reducing search frictions can redirect demand toward higher-quality providers and increase congestion (Yang et al. 2019). Similar feedback appears in healthcare. Hospital choice and waiting can be jointly determined (van de Klundert et al. 2024), while Dong et al. (2019) show that patients respond to delay information when choosing among emergency service providers, thereby changing coordination and waiting across the hospital network.

A more closely related stream studies shopping for preferred providers or evaluators. In healthcare, patients engage in doctor shopping to continue unresolved care or obtain preferred treatment (Biernikiewicz et al. 2019). Huang et al. (2022) directly study how doctor shopping interacts with congestion. Similar strategic selection arises when firms seek favorable auditors, certifiers, or rating agencies (DeFond et al. 2025, Deng et al. 2026, Sangiorgi and Spatt 2017). In these settings, customers or firms directly select the provider or evaluator they prefer. Judge shopping instead has a two-stage structure: litigants choose a court rather than a judge, and the assignment probability determines their chance of reaching the preferred judge.

Other work studies strategic choice among service options rather than preferred providers. Nageswaran (2023) studies whether customers should be allowed to select their preferred vaccine, Xu et al. (2016) consider choice between deterministic and probabilistic service, and Kerner et al. (2026) allows customers to choose between dedicated and flexible service. More broadly, service-design models show how service rules can change customer choices and the demand entering a system (Afèche and Pavlin 2016, Yang 2021, Yang et al. 2022). The service-menu model of Caldentey et al. (2024) is closest within this stream: customers choose service classes that determine their server pools, but customer–server fit affects the matching reward rather than the processing rate.

Our analysis also relates to skill-based routing and flexible-service systems, which study how heterogeneous demand is matched to servers with different capabilities (Chen et al. 2020, Wallace and Whitt 2005, Tsitsiklis and Xu 2017). In the canonical parallel-server formulation reviewed by Chen et al. (2020), customer-class arrival rates are exogenous, customers wait in class-specific queues, and routing matches them to compatible server pools with class–server-dependent service rates. Relatedly, Chen et al. (2026) show that selective flexibility can improve performance under exogenous demand asymmetry when non-dedicated servers process customers more slowly. Our architecture differs in two respects. Litigants choose between courts before assignment, making venue-specific demand endogenous, and cases are assigned upon arrival

to judge-specific queues, so each judge queue contains a mixture of specialist and nonspecialist cases. The same assignment probability therefore changes both where demand enters and how efficiently that demand is processed. As a result, an assignment rule that performs well for fixed demand need not remain optimal when the same rule also changes demand.

Rodriguez et al. (2025) show that probabilistic scheduling can discourage customers from misreporting urgency to obtain priority. Their scheduling policy changes service priority and reporting incentives without creating specialization mismatch. In our setting, litigants instead choose where to file, and changing the assignment probability also changes whether cases reach specialist judges.

Our contribution relative to these streams. We make filing demand endogenous to judicial assignment while allowing the same assignment rule to change processing efficiency. This joint response creates the central trade-off between congestion relief from filing reallocation and the processing-efficiency loss from specialization mismatch.

3. Model Formulation

3.1. Court System and Assignment Rule

Consider two courts, indexed by $m \in \{1, 2\}$, both serving two case types, indexed by $i \in \{1, 2\}$. Each court has a narrow assignment pool of two judges, indexed by $j \in \{1, 2\}$, where judge $j = i$ is the dedicated specialist judge for case type i. In Court 1, these specialist judges are also the judges that litigants seek to target. The assignment rule in Court 1 may need to depart from dedicated specialist assignment to limit judge shopping. The two courts represent a targeted venue and a relevant alternative, so changes in one court reallocate demand rather than eliminate it. Such reallocation arises in bankruptcy and patent venue choice (Antill and Bellon 2026, Castle 2024) and, beyond courts, in hospital choice when patients respond to delay information (Dong et al. 2019).

Cases of each type arrive according to independent Poisson processes with rate $\lambda > 0$. We focus on litigants' choice between courts based on long-run congestion expectations. The total volume of filed cases is taken as given, so we focus on where cases are filed rather than whether they are filed at all. This corresponds to the venue-choice effect of judge shopping, which Ganglmair et al. (2026) distinguish from its effect on whether a case is filed. Let $q_i \in [0, 1]$ denote the filing share of type-i cases choosing Court 1; the remaining share $1 - q_i$ files in Court 2. Each litigant is too small to affect waiting times, so the filing shares are interpreted as population fractions. Each case generates a common value V, assumed large enough that every litigant files in one of the two courts. Judges process specialist cases at rate $\theta\mu$ when $j = i$ and nonspecialist cases at rate $\mu > 0$ when $j \neq i$, where $\theta \geq 1$ (see, e.g., Chen et al. 2026). We refer to λ/μ as the *baseline load* and to θ as the *specialization advantage*. Throughout the comparative statics, we hold the nonspecialist service rate μ fixed, so variation in λ/μ comes from λ, while variation in θ changes the specialist service rate $\theta\mu$.

Conditional on whether a case is assigned to its specialist, service times are independent of arrivals and of one another and are exponentially distributed with rate $\theta\mu$ for specialist processing and μ for nonspecialist

processing. Each judge processes cases on a first-come, first-served basis. This specification is consistent with queueing models of court operations: Clarke and Metzler (2025) use the same exponential-service and first-come, first-served assumptions for criminal trials.

Litigants incur waiting cost at rate $c > 0$ while in queue. A type-i litigant receives an additional private benefit $\delta > 0$ when it files in Court 1 and reaches its dedicated specialist judge $j = i$. We refer to δ as the *shopping benefit*; it captures a judge-specific private advantage to the litigant, distinct from the processing benefit of specialization, and is not itself treated as a system benefit. For example, historical judge-specific conversion rates are observable from public bankruptcy dockets, and Chapter 11 litigants may prefer judges with lower propensities to convert reorganization cases to Chapter 7 liquidation (Hüther and Kleiner 2026). Both courts have specialist judges, but judge-shopping incentives are tied only to Court 1's specialists. Court 2 is the alternative venue without a corresponding preferred judge, consistent with settings in which litigants can choose among multiple eligible venues but target a particular judge in only one of them (Kahan and McKenzie 2021, Ganglmair et al. 2026). We maintain $\delta > 0$ in the judge-shopping model and use $\delta = 0$ only as the no-shopping benchmark.

The administrator chooses the assignment probability applied in Court 1, $p \in [0, 1]$: a type-i case is assigned to its dedicated specialist judge with probability p and to the nonspecialist judge with probability $1 - p$. We use *assignment probability* when discussing the administrator's choice and *targetability* when discussing the litigant's probability of reaching the preferred judge.

Judicial scholarship characterizes judge shopping through the probability of reaching a preferred or favorable judge, and empirical evidence shows that demand responds to changes in this probability (Kahan and McKenzie 2021, Gouzoules 2024, Ganglmair et al. 2026). This terminology is important because p does not measure the amount of randomization. Both $p = 0$ and $p = 1$ are deterministic, whereas $p = 1/2$ is uniform random assignment across the two judges. Uniform random assignment is one possible policy, not the only policy available to the administrator (Thorley 2020). The choices $p = 0$, $0 < p < 1$, and $p = 1$ are called *minimum*, *partial*, and *full targetability*, respectively.

The *shopping incentive* from filing in Court 1 is $p\delta$, whereas Court 2 offers no shopping benefit. The effectiveness of random assignment as an anti-shopping intervention cannot be evaluated independently of the shopping benefit. When the shopping benefit is large, even a substantially reduced probability of reaching the preferred judge can leave judge shopping attractive. Botoman provides a direct legal example: a fifty-percent assignment probability was regarded as "a high probability," making the preferred venue "a gamble worth taking" (Botoman 2018). An assignment policy restricted to $p \geq 1/2$ may therefore fail to weaken judge shopping enough to change filing behavior.

For this reason, p is a routing probability over the full interval $[0, 1]$, rather than a measure of randomization over $[1/2, 1]$. Uniform random assignment, $p = 1/2$, is one routing benchmark. Values $p < 1/2$ represent

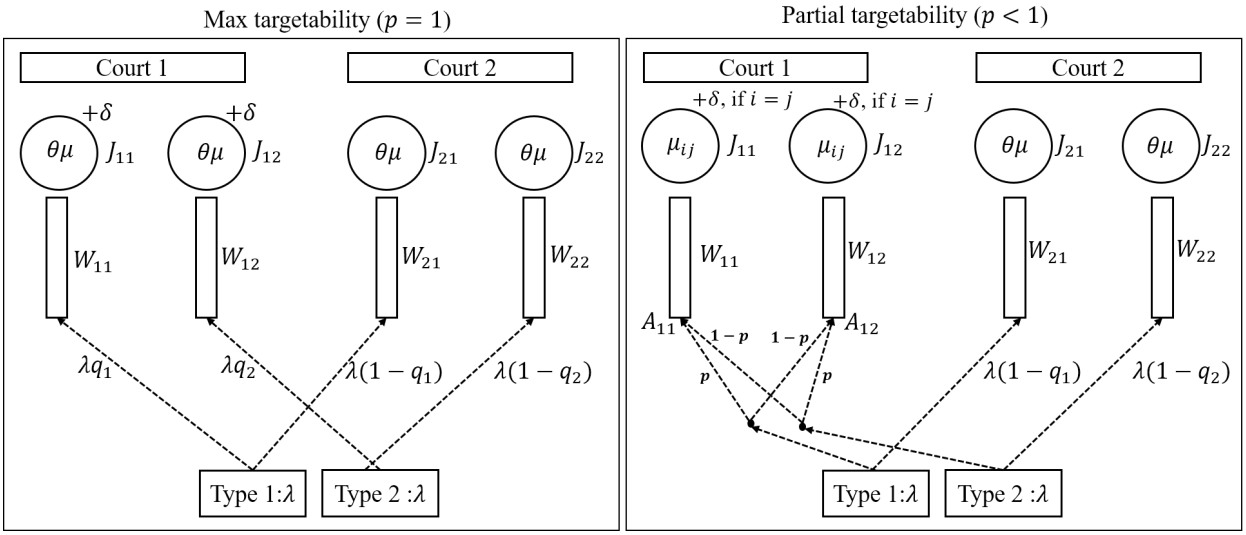


Figure 1 Specialist and nonspecialist flows under Court 1 assignment probability p.

Notes. Judges J_{m1} and J_{m2} denote judges 1 and 2 in court m. Under probabilistic assignment, the effective arrival rates to the two Court 1 queues are $A_{11} = \lambda q_1 p + \lambda q_2(1-p)$ and $A_{12} = \lambda q_1(1-p) + \lambda q_2 p$. $\mu_{ji} = \theta\mu$ if $i = j$, and $\mu_{ji} = \mu$ otherwise.

policies that route cases to the nonspecialist more often than to the dedicated specialist. When $\theta > 1$, this routing creates *specialization mismatch*. Figure 1 illustrates the assignment mechanism.

For a case filed in Court 1,

$$\text{expected service time} = \frac{p}{\theta\mu} + \frac{1-p}{\mu}, \qquad \text{second moment of service time} = \frac{2p}{\theta^2\mu^2} + \frac{2(1-p)}{\mu^2}.$$

For $\theta > 1$, the expected service time and its second moment decrease with p. Thus, lower targetability worsens Court 1 processing. The first moment governs utilization, while the second moment also enters the $M/G/1$ waiting-time formula derived in the next Subsection.

3.2. Judge-Level Waiting Times

Each judge's docket is modeled as a stochastic queue, and litigants choose where to file based on steady-state expected waiting times, waiting cost, and shopping incentive. This focus on long-run congestion is consistent with Feigin et al. (2026), who use flow-based measures to study court delay. Throughout the paper, waiting time means expected time in queue before service begins, not total case duration. We denote it by $W_{mj}(\mathbf{q}, p)$, where $\mathbf{q} = (q_1, q_2)$ is the filing profile. p is a standing assignment rule rather than a dynamic rule that responds to current queue lengths. If assignment probabilities instead responded to current queue lengths, litigants could also engage in judge shopping through strategic filing timing (Hüther and Kleiner 2026). We leave this distinct dimension of judge shopping outside our scope and focus on venue choice under a standing assignment rule. For fixed filing shares and assignment probability, filing and judicial assignment are independent across cases.

Court 2 remains fully specialized, meaning that each judge serves a single case type. Thus, both dockets are standard $M/M/1$ queues with expected waiting times

$$W_{21}(q_1) = \frac{(\lambda/\mu)(1-q_1)}{\theta\mu[\theta-(\lambda/\mu)(1-q_1)]}, \tag{1}$$

$$W_{22}(q_2) = \frac{(\lambda/\mu)(1-q_2)}{\theta\mu[\theta-(\lambda/\mu)(1-q_2)]}. \tag{2}$$

In Court 1, probabilistic assignment mixes specialist and nonspecialist cases within each judge's docket. Judge 1 receives specialist and nonspecialist arrivals at rates $\lambda p q_1$ and $\lambda(1-p)q_2$, while Judge 2 receives them at rates $\lambda p q_2$ and $\lambda(1-p)q_1$, respectively. We denote the corresponding effective arrival rates by $A_{11} = \lambda q_1 p + \lambda q_2(1-p)$ and $A_{12} = \lambda q_1(1-p) + \lambda q_2 p$. For $0 < p < 1$, each Court 1 docket mixes cases processed at specialist and nonspecialist rates. Thus, each docket is an $M/G/1$ queue; at $p \in \{0,1\}$, it reduces to an $M/M/1$ queue. Applying the Pollaczek–Khinchine formula yields

$$W_{11}(\mathbf{q},p) = \frac{(\lambda/\mu)[pq_1 + (1-p)\theta^2 q_2]}{\theta\mu[\theta-(\lambda/\mu)(pq_1+(1-p)\theta q_2)]}, \tag{3}$$

$$W_{12}(\mathbf{q},p) = \frac{(\lambda/\mu)[(1-p)\theta^2 q_1 + pq_2]}{\theta\mu[\theta-(\lambda/\mu)((1-p)\theta q_1+pq_2)]}. \tag{4}$$

Let ρ_{mj} denote the utilization of judge j in Court m. The four utilization rates are

$$\begin{aligned} \rho_{11}(\mathbf{q},p) &= \frac{\lambda}{\theta\mu}[pq_1+(1-p)\theta q_2], \quad & \rho_{12}(\mathbf{q},p) &= \frac{\lambda}{\theta\mu}[(1-p)\theta q_1 + pq_2], \\ \rho_{21}(q_1) &= \frac{\lambda}{\theta\mu}(1-q_1), & \rho_{22}(q_2) &= \frac{\lambda}{\theta\mu}(1-q_2). \end{aligned} \tag{5}$$

These utilization rates determine which filing profiles the system can sustain. A filing profile $\mathbf{q} = (q_1, q_2)$ is admissible under assignment probability p if every judge is stable. Any equilibrium filing profile must therefore be admissible:

$$\rho_{mj} < 1, \qquad m,j \in \{1,2\}. \tag{6}$$

For a fixed filing profile and assignment probability, a court's *processing capacity* is the baseline-load threshold below which both of its judge queues remain stable. The total arrival rate is 2λ. Even if every judge processed only specialist cases, the four judges together could process at rate at most $4\theta\mu$. Therefore, $0 < \lambda/\mu < 2\theta$ is necessary for any admissible filing profile—and hence any equilibrium—under any assignment policy. We refer to this inequality as the *global stability requirement* and maintain it throughout.

3.3. Litigant Utilities

Litigants choose between the two filing venues by comparing the expected shopping benefit with expected waiting costs. Let U_{mi} denote the expected utility of a type-i litigant filing in Court m.

For a type-i litigant, filing in Court 1 yields

$$U_{1i}(\mathbf{q}; p) = V + p\delta - c\Big[pW_{1i}(\mathbf{q}, p) + (1-p)W_{1,3-i}(\mathbf{q}, p)\Big], \tag{7}$$

Here, $3-i$ denotes the judge other than i. The expected waiting cost averages over the two possible judicial assignments. Filing in Court 2 yields

$$U_{2i}(q_i) = V - cW_{2i}(q_i). \tag{8}$$

Because V is common to both venues, it cancels from the filing decision. The relevant difference is therefore the shopping benefit from Court 1 versus the expected waiting-cost difference between the two courts.

For a given assignment probability p, each case type chooses the court with greater expected utility. In equilibrium, a case type splits its filings across both courts only when the two courts offer equal expected utility; if it files exclusively in one court, that court must offer at least as much expected utility as the other.

4. Equilibrium Analysis

Litigants choose between courts by trading off waiting times against the shopping incentive, while their filing choices in turn determine waiting times. Two types of equilibria arise: symmetric and asymmetric. We use the symmetric equilibrium for the main comparative statics and policy analysis. In asymmetric equilibria, the two case types file at different rates. At empirically plausible specialization levels, however, these asymmetric equilibria are rare and do not appear at the optimal policy in our main analysis. We therefore study them in Section 7 as a robustness check.

4.1. Wardrop Equilibrium

Definition 1 (Wardrop equilibrium). For fixed $p \in [0,1]$, an admissible filing profile $\mathbf{q} = (q_1, q_2)$ is a Wardrop equilibrium if, for each case type $i \in \{1,2\}$,

$$\begin{aligned} q_i = 0 \quad &\Rightarrow U_{1i} \le U_{2i}, \\ 0 < q_i < 1 \ &\Rightarrow U_{1i} = U_{2i}, \\ q_i = 1 \quad &\Rightarrow U_{1i} \ge U_{2i}. \end{aligned} \tag{9}$$

All utilities are evaluated at $(\mathbf{q}, p)$.

For a given assignment probability p, let

$$Q(p) = \big\{\mathbf{q} \in [0,1]^2 : \rho_{mj}(\mathbf{q}, p) < 1 \text{ for all } m, j, \text{ and } \mathbf{q} \text{ satisfies the Wardrop conditions in (9)}\big\}$$

denote the set of Wardrop equilibrium filing profiles. We call p *feasible* if $Q(p) \neq \emptyset$. Define $\Delta_i(\mathbf{q}; p) := U_{1i}(\mathbf{q}; p) - U_{2i}(q_i)$.

Definition 2 (Equilibrium classes). For fixed p, an equilibrium $\mathbf{q} \in Q(p)$ is *symmetric* if $q_1 = q_2$, and *asymmetric* otherwise. It is *interior* if $0 < q_i < 1$ for both case types, and a *boundary equilibrium* otherwise.

Table 1 classifies these equilibria by symmetry and whether filing shares are interior or on the boundary.

Table 1 Classification of Wardrop equilibria

	Interior	Boundary
Symmetric	$0 < q_1 = q_2 < 1$ $\Delta_1 = \Delta_2 = 0$	$q_1 = q_2 = 0: \quad \Delta_1 \le 0,\ \Delta_2 \le 0$ $q_1 = q_2 = 1: \quad \Delta_1 \ge 0,\ \Delta_2 \ge 0$
Asymmetric	$0 < q_1, q_2 < 1, \quad q_1 \ne q_2$ $\Delta_1 = \Delta_2 = 0$	$q_i = 0,\ 0 < q_{3-i} < 1: \quad \Delta_i \le 0,\ \Delta_{3-i} = 0$ $q_i = 1,\ 0 < q_{3-i} < 1: \quad \Delta_i \ge 0,\ \Delta_{3-i} = 0$ $(q_i, q_{3-i}) = (1,0): \quad \Delta_i \ge 0,\ \Delta_{3-i} \le 0$

Note. $\Delta_i = U_{1i} - U_{2i}$ denotes the utility difference between filing in Court 1 and Court 2 for case type i. Asymmetric boundary equilibria occur in type-swapped pairs.

4.2. Symmetric Equilibrium

Under a symmetric filing profile $q_s := q_1 = q_2$, the two case types face identical waiting times in each court:

$$W_1(q_s, p) := W_{11}((q_s, q_s), p) = W_{12}((q_s, q_s), p), \qquad W_2(q_s) := W_{21}(q_s) = W_{22}(q_s).$$

An interior symmetric filing profile is a Wardrop equilibrium if and only if

$$W_1(q_s, p) - W_2(q_s) = p\frac{\delta}{c}. \tag{10}$$

The *relative shopping benefit*, δ/c, measures the additional waiting time a litigant is willing to accept to reach the preferred judge. Equation (10) therefore implies $W_1 > W_2$ whenever $p > 0$: litigants are willing to accept a longer wait in Court 1 for a greater chance of reaching the preferred judge. A similar trade-off arises in differentiated service queues, where customers may accept longer waits for more attractive service options (Yang et al. 2019). The waiting times for both types are

$$W_1(q_s, p) = \frac{(\lambda/\mu) q_s \left[p + (1-p)\theta^2\right]}{\theta\mu \left[\theta - (\lambda/\mu) q_s \left[p + (1-p)\theta\right]\right]}, \tag{11}$$

$$W_2(q_s) = \frac{(\lambda/\mu)(1-q_s)}{\theta\mu \left[\theta - (\lambda/\mu)(1-q_s)\right]}. \tag{12}$$

Lemma 1 **(Symmetric admissibility)**. *For fixed p, an interior symmetric filing profile is admissible if and only if*

$$\max\left\{0,\ 1 - \frac{\theta}{\lambda/\mu}\right\} < q_s < \min\left\{1,\ \frac{\theta}{(\lambda/\mu)[p + (1-p)\theta]}\right\}. \tag{13}$$

This interval is nonempty if and only if

$$\left(\frac{\lambda}{\mu} - \theta\right)[p + (1-p)\theta] < \theta. \tag{14}$$

The proof is provided in Appendix A.1. The lower and upper bounds are the Court 2 and Court 1 stability limits, respectively. Condition (14) is the capacity-feasibility condition for a symmetric filing profile at assignment probability p. At $p = 1$, the nonemptiness condition reduces to $\lambda/\mu < 2\theta$, exactly the global stability requirement. Hence, under global stability, a symmetric equilibrium always exists somewhere in

the policy space. At a symmetric filing share q_s, Court 1's processing capacity is $\theta/\{q_s[p + (1-p)\theta]\}$. For $\theta > 1$, any reduction in targetability lowers processing capacity by sending more cases to nonspecialists. Relative to full targetability, processing capacity at p is $1/[p + (1-p)\theta]$ of its full-targetability level. The capacity loss therefore increases as targetability p falls and as the specialization advantage θ grows. For example, under uniform random assignment, Court 1 retains $2/(1+\theta)$ of its full-targetability processing capacity. Uniform random assignment therefore reduces Court 1's processing capacity by sending more cases to nonspecialists. As a result, Court 1 can become unstable at loads that dedicated specialist assignment can serve.

Within this admissible region, the Wardrop condition determines how filings are divided between the two courts. Substituting (11)–(12) into (10) reduces the equilibrium condition to

$$a_s(p)q_s^2 + b_s(p)q_s + c_s(p) = 0, \tag{15}$$

where the coefficients $a_s(p)$, $b_s(p)$, and $c_s(p)$ are given in Appendix A.2.

Proposition 1 **(Symmetric equilibrium)**. *Fix $p \in [0,1]$ satisfying the capacity-feasibility condition* (14).

(i) Lower boundary: $q_s = 0$. The profile $q_s = 0$ is never a symmetric Wardrop equilibrium.

(ii) Upper boundary: $q_s = 1$. The profile $q_s = 1$ is a symmetric equilibrium if and only if

$$\frac{\lambda}{\mu}\,[p + (1-p)\theta] < \theta \tag{16}$$

and

$$p\frac{\delta}{c} \geq \frac{(\lambda/\mu)\left[p + (1-p)\theta^2\right]}{\theta\mu\left[\theta - (\lambda/\mu)\left[p + (1-p)\theta\right]\right]}. \tag{17}$$

(iii) Interior: $0 < q_s < 1$. Otherwise, there is a unique interior symmetric equilibrium. If $(\theta, p) \neq (1,0)$, it is

$$q_s(p) = \frac{-b_s(p) + \sqrt{b_s(p)^2 - 4a_s(p)c_s(p)}}{2a_s(p)}. \tag{18}$$

At $(\theta, p) = (1,0)$, the unique symmetric equilibrium is $q_s = 1/2$.

Proposition 1 shows that litigants never abandon Court 1 entirely. All litigants choose Court 1 only when it can accommodate all filings and the shopping benefit compensates for its longer waiting time; otherwise, filings are split across the two courts.

4.3. Comparative Statics

We examine how the relative shopping benefit, assignment probability, specialization advantage, and baseline load affect equilibrium filing and waiting times. Throughout this subsection, total derivatives are evaluated along the equilibrium, while partial derivatives hold q_s fixed.

Proposition 2 **(Comparative statics)**. *Fix an interior symmetric equilibrium. Waiting-time derivatives are evaluated along the equilibrium; partial derivatives hold q_s fixed.*

(i) Relative shopping benefit. For $p > 0$, $dq_s/d(\delta/c) > 0$, $dW_1/d(\delta/c) > 0$, and $dW_2/d(\delta/c) < 0$. At $p = 0$, all three derivatives equal zero.

(ii) Assignment probability. For $p \in (0, 1)$,

$$\frac{dq_s}{dp} = \underbrace{\frac{\delta/c}{\partial W_1/\partial q_s - \partial W_2/\partial q_s}}_{\text{shopping-incentive channel}} + \underbrace{\frac{-\partial W_1/\partial p}{\partial W_1/\partial q_s - \partial W_2/\partial q_s}}_{\text{specialization-mismatch channel}} > 0,$$

$$\frac{dW_2}{dp} = \frac{\partial W_2}{\partial q_s}\frac{dq_s}{dp} < 0, \qquad \frac{dW_1}{dp} = \frac{dW_2}{dp} + \frac{\delta}{c}.$$

Consequently,

$$\frac{dW_1}{dp} > 0 \iff -\frac{\partial W_2}{\partial q_s}\frac{dq_s}{dp} < \frac{\delta}{c}.$$

Equality and the reverse inequality yield zero and negative responses, respectively.

(iii) Specialization advantage. For $\theta > 1$ and $0 < p < 1$,

$$\frac{dW_1}{d\theta} = \frac{dW_2}{d\theta} < 0, \qquad \frac{dq_s}{d\theta} = \frac{\partial W_2/\partial\theta - \partial W_1/\partial\theta}{\partial W_1/\partial q_s - \partial W_2/\partial q_s}. \quad \text{Hence, } \frac{dq_s}{d\theta} > 0 \iff \frac{\partial W_1}{\partial\theta} < \frac{\partial W_2}{\partial\theta}.$$

At $p = 0$, $\partial W_1/\partial\theta = 0$ and $dq_s/d\theta < 0$. At $p = 1$, $q_s > 1/2$ and $dq_s/d\theta > 0$. For $0 < p < 1$, there exists a unique threshold $\eta_\theta(\lambda/\mu, p, \theta; \mu) > 0$ such that $dq_s/d\theta < 0$ when $\delta/c < \eta_\theta$, $dq_s/d\theta = 0$ when $\delta/c = \eta_\theta$, and $dq_s/d\theta > 0$ when $\delta/c > \eta_\theta$. Moreover, $\partial\eta_\theta/\partial(\lambda/\mu) > 0$.

(iv) Baseline load.

$$\frac{dq_s}{d(\lambda/\mu)} < 0 \iff q_s\frac{\partial W_1}{\partial q_s} > -(1 - q_s)\frac{\partial W_2}{\partial q_s}.$$

Equality and the reverse inequality yield zero and positive responses, respectively. At $p = 1$ and $\delta > 0$, $q_s > 1/2$ and $dq_s/d(\lambda/\mu) < 0$.

The proof is provided in Appendix A.3. We discuss Proposition 2 in the following.

Relative shopping benefit. This result confirms the basic shopping prediction: when reaching a preferred judge becomes more valuable, more litigants file in Court 1. They are willing to accept a longer wait there, so waiting time rises in Court 1 and falls in Court 2. Related service-choice models similarly show that more attractive service can redirect demand toward preferred providers (Yang et al. 2019). At $p = 0$, the shopping benefit does not affect filing because the preferred judge cannot be targeted.

Assignment probability. Lower targetability reduces Court 1's filing share, but it need not reduce Court 1's waiting time because specialization mismatch can outweigh the resulting congestion relief. The filing share falls through two channels. The *shopping-incentive channel* operates directly: lowering targetability reduces the shopping incentive, $p\delta$, so some litigants shift from Court 1 toward Court 2. The *specialization-mismatch*

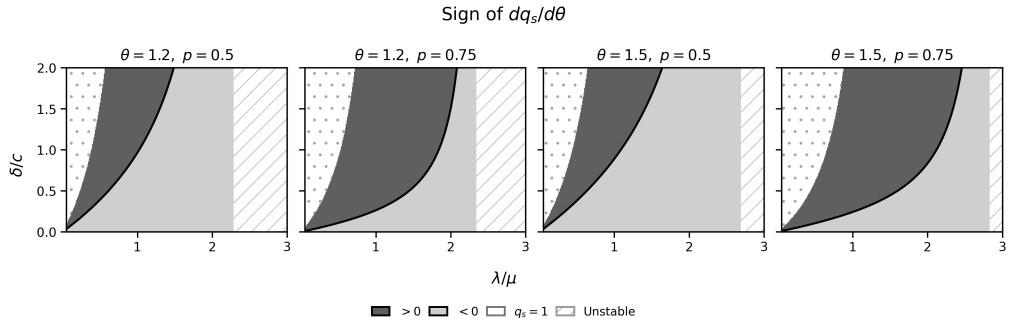


Figure 2 Effect of a greater specialization advantage on Court 1's symmetric-equilibrium filing share. Dark regions show $dq_s/d\theta > 0$, light regions show $dq_s/d\theta < 0$, and the solid curve marks $dq_s/d\theta = 0$. Dotted hatching identifies the boundary $q_s = 1$, and diagonal hatching identifies parameter values that violate queue stability. Each panel fixes (θ, p) and varies $(\lambda/\mu, \delta/c)$.

channel operates through waiting: lowering targetability sends more cases to nonspecialists, raising Court 1 waiting for a given filing share and shifting additional litigants toward Court 2.

As filings move from Court 1 to Court 2, congestion falls in Court 1 and rises in Court 2. This congestion feedback limits the filing shift generated by both channels. In particular, the denominator $\partial W_1/\partial q_s - \partial W_2/\partial q_s$ in dq_s/dp captures how strongly waiting in the two courts responds to a change in the filing share, q_s.

This filing reallocation relieves congestion in Court 1, whereas specialization mismatch creates a processing-efficiency loss. Court 1 waiting falls when the congestion relief outweighs the processing-efficiency loss, remains unchanged when the two effects exactly offset each other, and rises when the processing-efficiency loss dominates. Court 2 waiting rises as filings shift toward Court 2; hence, when Court 1 waiting also rises, waiting increases in both courts.

Specialization advantage. A greater specialization advantage lowers equilibrium waiting in both courts. The filing response depends on the direct waiting reductions at a fixed filing share: Court 1 attracts more filings when specialization reduces its waiting more than Court 2's.

Under minimum targetability, a greater specialization advantage does not directly reduce waiting time at Court 1 but reduces waiting time at Court 2, thereby shifting filings toward Court 2. Under full targetability, specialization produces a greater direct waiting reduction in the more heavily loaded Court 1, so its filing share rises. With partial targetability, a greater specialization advantage shifts filings toward Court 2 when $\delta/c < \eta_\theta$ and toward Court 1 when $\delta/c > \eta_\theta$. Moreover, η_θ increases with baseline load. Thus, under high load, a greater specialization advantage is less likely to draw additional filings precisely when faster processing is most valuable. Figure 2 illustrates this threshold. We normalize time by setting $\mu = 1$, so δ/c is measured in units of the nonspecialist mean service time $1/\mu$. As baseline load rises, a greater relative shopping benefit is required for a greater specialization advantage to increase Court 1's filing share.

Baseline load. Baseline load operates through the congestion feedback. Greater baseline load amplifies congestion in both courts. Under partial targetability, the direction of the filing response is parameter dependent, so higher baseline load can shift filings either toward or away from Court 1. Under full targetability and a positive shopping benefit, however, judge shopping concentrates more than half of filings in Court 1. Because Court 1 has the larger filing share, its waiting time increases more with additional load, which redirects some filings toward Court 2. Higher baseline load therefore limits the filing concentration induced by judge shopping.

Together, these results show how the shopping incentive and specialization mismatch interact with congestion to determine equilibrium filing and waiting.

5. Operational Cost of Judge Shopping

To isolate the operational cost created by shopping, we first hold full targetability fixed and compare positive shopping benefit with the no-shopping benefit benchmark. We then allow p to vary when analyzing assignment design.

5.1. System Performance Objective Function

We evaluate assignment policy using system-wide waiting. For each case type, q_i weights expected waiting in Court 1, $pW_{1i}(\mathbf{q}, p) + (1-p)W_{1,3-i}(\mathbf{q}, p)$, while $1 - q_i$ weights waiting in Court 2, $W_{2i}(q_i)$. Summing these terms across the two case types gives the System Performance Objective (SPO) function. The shopping incentive $p\delta$ enters Court 1 utility in Equation (7) and therefore affects SPO through the equilibrium filing shares and resulting waiting times. The SPO objective places no separate penalty on successful targeting.

For every admissible filing profile $\mathbf{q}$ under policy p, define

$$SPO(\mathbf{q}; p) := \sum_{i=1}^{2} q_i \left[pW_{1i}(\mathbf{q}, p) + (1-p)W_{1,3-i}(\mathbf{q}, p) \right] + \sum_{i=1}^{2} (1 - q_i) W_{2i}(q_i). \tag{19}$$

Total arrival volume and the waiting-cost rate c are fixed, so minimizing SPO is equivalent to minimizing total expected waiting cost. Our main policy analysis evaluates SPO along the symmetric equilibrium characterized in Proposition 1 while Section 7 examines equilibrium multiplicity and policy robustness.

Along the symmetric equilibrium, $SPO\big((q_s(p), q_s(p)); p\big) = 2SPO_s(p)$, where

$$SPO_s(p) = q_s(p)W_1(q_s(p), p) + \big(1 - q_s(p)\big)W_2(q_s(p)). \tag{20}$$

Thus, minimizing SPO_s yields the same assignment probability as minimizing SPO along the symmetric-equilibrium path. The algebraic reduction is shown in Appendix A.4.

At an interior symmetric equilibrium, the Wardrop condition (10) allows Equation (20) to be written as

$$SPO_s(p) = W_2(q_s(p)) + q_s(p)p\frac{\delta}{c}. \tag{21}$$

Equation (21) rewrites system-wide waiting as the sum of two components. The term $q_s(p)p$ is the equilibrium share of litigants who successfully reach their preferred judge, so $q_s(p)p\delta/c$ is the time-equivalent benefit from successful targeting. By the Wardrop condition, $p\delta/c$ is also the additional waiting that Court 1 filers accept relative to Court 2. This has the same time-price interpretation as rationing by waiting (Barzel 1974): at an interior equilibrium, the expected shopping benefit is exactly offset at the margin by the additional waiting cost of filing in Court 1. The other term, $W_2(q_s(p))$, captures congestion in the alternative court. These two components move in opposite directions with targetability: Proposition 2(ii) gives $dW_2/dp < 0$, whereas $d[q_s(p)p\delta/c]/dp = (\delta/c)[q_s(p) + p\, dq_s/dp] > 0$. Thus, decreasing targetability reduces successful targeting in Court 1 but shifts more filings to Court 2, increasing congestion there. A similar mechanism appears in Rodriguez et al. (2025): an operational policy changes users' strategic incentives, and the resulting behavioral response enters the manager's waiting-cost objective. Their policy operates through service priority and reporting incentives, whereas here the assignment probability operates through venue choice and specialist matching.

5.2. Operational Cost under Full Targetability

Proposition 3 **(Operational cost of judge shopping under full targetability)**. *Fix $p = 1$ and suppose that an admissible symmetric equilibrium exists.*

(i) No-shopping benchmark. If $\delta = 0$,

$$q_s(1) = \frac{1}{2}, \qquad W_1 = W_2 = SPO_s(1) = \frac{\lambda/\mu}{\theta\mu(2\theta - \lambda/\mu)}, \qquad W_1 + W_2 = \frac{2(\lambda/\mu)}{\theta\mu(2\theta - \lambda/\mu)}.$$

(ii) Positive shopping benefit. For every $\delta > 0$, $q_s(1) > 1/2$. Relative to the no-shopping benchmark, judge shopping strictly increases W_1 and strictly decreases W_2, but strictly increases both the unweighted aggregate $W_1 + W_2$ and $SPO_s(1)$. Under full targetability, judge shopping increases system-wide waiting by concentrating filings in Court 1.

The proof is provided in Appendix A.5. When $\delta = 0$, dedicated specialist assignment combines full specialist matching with balanced filings across the two courts. When $\delta > 0$, the same assignment rule also attracts more filings to Court 1 and strictly increases system-wide waiting relative to the no-shopping benchmark. Dedicated specialist assignment therefore cannot be evaluated as a pure processing benefit once litigants choose where to file.

6. Optimal Assignment Policy

The administrator chooses p before litigants select a court and therefore anticipates the equilibrium filing response. The waiting-minimizing assignment probability along the symmetric equilibrium is

$$p^* \in \arg\min_{p\in[0,1]} SPO_s(p) \qquad \text{s.t.} \qquad (q_s(p), q_s(p)) \in Q(p). \tag{22}$$

Throughout the policy analysis, *optimal* refers to minimizing $SPO_s(p)$ over this feasible policy set. Because the environment is symmetric, we use the symmetric equilibrium as the analytical benchmark. We do not assume, however, that it is uniquely selected when asymmetric equilibria coexist. Section 7 therefore examines whether the optimal policy changes when asymmetric equilibria are taken into account under different parameter specifications.

Before presenting the optimal assignment policy, we use two benchmarks to isolate the forces behind the policy trade-off. Holding filing shares fixed isolates the processing effect of assignment, while removing the specialization advantage isolates the venue-choice effect.

Proposition 4 **(Assignment under fixed symmetric filing shares)**. *Fix a symmetric filing profile* $q_1 = q_2 = q_s \in (0, 1]$ *that is admissible for at least one assignment probability, and suppose that its filing shares do not respond to* p*. If* $\theta > 1$*, then* $SPO(\mathbf{q}; p)$ *is strictly decreasing in* p *over the assignment probabilities for which the fixed filing profile is admissible. Hence dedicated specialist assignment,* $p = 1$*, uniquely minimizes system-wide waiting.*

The proof is provided in Appendix A.6. This prescription depends on the symmetry of the fixed filing profile. With asymmetric demand, cross-serving can help balance workloads even without a filing response, as in the flexibility configurations studied by Chen et al. (2026). We focus on symmetric case types to isolate how endogenous filing changes the assignment prescription. The policy reverses when filing shares respond to p but assignment has no processing consequence.

Proposition 5 **(No-specialization benchmark)**. *When* $\theta = 1$*, assignment does not create specialization mismatch, and minimum targetability minimizes system-wide waiting. Therefore,* $p^* = 0$*.*

The proof is provided in Appendix A.7. Together, the two benchmarks show that the policy trade-off arises because assignment affects both filing and processing. In the full model, lower targetability can create congestion relief through filing reallocation but also a processing-efficiency loss through specialization mismatch. We next characterize the resulting optimum for $\theta > 1$. Differentiating $SPO_s(p)$ with respect to p along the symmetric equilibrium path and using the symmetric Wardrop condition (10) yields the first-order condition

$$0 = p\left(\frac{\delta}{c}\right)^2 + \left[q_s(p)\frac{\partial W_1}{\partial q_s} + \big(1 - q_s(p)\big)\frac{\partial W_2}{\partial q_s} - p\frac{\partial W_1}{\partial p}\right]\frac{\delta}{c} - \frac{\partial W_1}{\partial p}\frac{\partial W_2}{\partial q_s}, \tag{23}$$

where all derivatives are evaluated at $(q_s(p), p)$. For $p > 0$, Equation (23) has a unique positive root in the relative shopping benefit. Let $\tau(q, p) > 0$ denote this threshold when the filing share is q. Its explicit expression is given in Appendix A.8. Thus, along an interior symmetric equilibrium,

$$SPO_s'(p) = 0 \quad \Longleftrightarrow \quad \frac{\delta}{c} = \tau(q_s(p), p)\,. \tag{24}$$

Together with the symmetric Wardrop equation (10), (24) characterizes the partial-targetability optimum and its equilibrium filing share.

LEMMA 2 **(Shape of the objective function along the symmetric-equilibrium path)**. *Fix $\theta > 1$ and suppose the global stability requirement $0 < \lambda/\mu < 2\theta$ holds. The assignment probabilities supporting a symmetric equilibrium form a connected interval containing $p = 1$. If the feasible policy range ends because a queue becomes unstable, system-wide waiting grows without bound as p approaches that boundary. Hence such a boundary cannot be optimal.*

For assignment probabilities with $p > 0$ and $0 < q_s(p) < 1$, $\text{sign}\{SPO_s'(p)\} = \text{sign}[\delta/c - \tau(q_s(p), p)]$. On the interior-equilibrium range, $SPO_s(p)$ has at most one stationary point. If it exists, it is a strict minimum and the unique minimizer over that range; otherwise, $SPO_s(p)$ is monotone on that range.

Increasing targetability strengthens the shopping incentive but reduces specialization mismatch. At an interior symmetric equilibrium, the threshold $\tau(q_s(p), p)$ determines which effect dominates. If the relative shopping benefit δ/c is below this threshold, a small increase in targetability reduces system-wide waiting; if it is above the threshold, the same increase raises system-wide waiting. This shape result applies to the interior-equilibrium range. The theorem below also accounts for the $q_s = 1$ range and its transition from interior filing, identifies the global optimum, and characterizes its response to the relative shopping benefit and baseline load.

THEOREM 1 **(Optimal assignment policy)**. *Fix $\theta > 1$, suppose the global stability requirement holds, and consider the symmetric equilibrium.*

(i) ***Minimum targetability.*** *If $p = 0$ is feasible, then $SPO_s'(0^+) < 0$. Hence $p^* = 0$ is not optimal for any finite δ/c when $\theta > 1$.*

(ii) ***Full versus partial targetability.*** *If $q_s(1) < 1$, then $p^* = 1 \iff SPO_s'(1^-) \leq 0 \iff \delta/c \leq \tau(q_s(1), 1)$. If these conditions fail, the unique optimum satisfies $0 < p^* < 1$ and (24). If $q_s(1) = 1$, full targetability is not globally optimal and the unique optimum satisfies $0 < p^* < 1$.*

(iii) ***Relative shopping benefit.*** *For fixed $\theta > 1$ and λ/μ, full targetability, $p^* = 1$, is optimal for all sufficiently small positive values of δ/c. When $p = 0$ is feasible, $p^* \to 0$ as $\delta/c \to \infty$.*

(iv) ***Baseline load.*** *For every fixed $\theta > 1$ and finite δ/c, full targetability, $p^* = 1$, is optimal for all λ/μ sufficiently close to 2θ. At $\lambda/\mu < 2\theta$, some assignment probabilities below one remain feasible.*

The proof is provided in Appendix A.9. Part (i) rules out minimum targetability when $\theta > 1$ because raising p restores specialist matching. Part (ii) shows that filing concentration can instead make dedicated specialist assignment suboptimal. The full model can therefore have a unique partial-targetability optimum.

Having characterized the full-, partial-, and minimum-targetability conditions, we next examine how the partial-targetability optimum responds to the relative shopping benefit and baseline load.

PROPOSITION 6 **(Comparative statics of the optimal assignment probability)**. *Fix $\theta > 1$ and consider the partial-targetability optimum characterized in Theorem 1, with $0 < p^* < 1$ and $0 < q_s(p^*) < 1$. Then:*

(i) Relative shopping benefit. $\partial p^*/\partial(\delta/c) < 0$.

(ii) Baseline load. $\partial p^*/\partial(\lambda/\mu) > 0$.

The proof is provided in Appendix A.10. A greater relative shopping benefit strengthens the shopping incentive. Court 1 attracts more filings, increasing the filing concentration that lower targetability can reduce. The optimal assignment probability consequently falls as δ/c increases.

Higher baseline load increases congestion but raises, rather than lowers, the optimal assignment probability. Related flexibility models with exogenous demand also find that high workload can favor dedicated service when cross-service is slower (Chen et al. 2026). Our setting reveals an additional demand-side mechanism: congestion itself disciplines strategic venue choice. Under full targetability, Court 1 receives more than half of filings, so rising congestion pushes some litigants back toward Court 2 and reduces the filing concentration created by judge shopping. Higher baseline load therefore both increases the processing cost of reducing targetability and reduces the filing concentration that lower targetability can correct. The optimal assignment probability consequently rises with baseline load, and at sufficiently high baseline load full targetability becomes optimal.

Figure 3 visualizes the policy regions characterized in Theorem 1. The top row reports the filing concentration generated by full targetability: a greater relative shopping benefit draws more filings to Court 1, whereas higher baseline load limits this concentration. The middle row shows the policy response. When $\theta > 1$, lower targetability weakens the shopping incentive but creates more specialization mismatch. Reducing targetability is optimal when the reduction in filing concentration outweighs the processing-efficiency loss from specialization mismatch. When δ/c is small or $q_s(1)$ is close to $1/2$, the congestion-relief gain from reducing targetability is smaller, so preserving specialist matching can make full targetability optimal. The bottom row reports the resulting filing allocation at p^*. For example, when $\theta = 1.5$, $\lambda/\mu = 2.5$, and $\delta/c = 2$, full targetability is optimal with $q_s(1) \approx 0.524$, so filings are nearly balanced. The corresponding utilizations are $\rho_{11} = \rho_{12} \approx 0.873$ in Court 1 and $\rho_{21} = \rho_{22} \approx 0.794$ in Court 2. Full targetability can be optimal when shopping creates little filing concentration even while both courts remain below their stability boundaries.

We next turn to implications of this policy characterization: how p^* compares with litigants' preferred rule and common assignment benchmarks, and how it changes when successful targeting receives a direct institutional penalty.

Corollary 1 **(Litigants' preferred targetability)**. *Define aggregate litigant utility as* $\lambda \sum_{i=1}^{2} [q_i U_{1i}(\mathbf{q}; p) + (1 - q_i) U_{2i}(q_i)]$. *Along the symmetric equilibrium, this quantity is maximized under full targetability,* $p = 1$. *Moreover, full targetability supports no asymmetric equilibrium, so the assignment rule preferred by litigants has a unique symmetric equilibrium.*

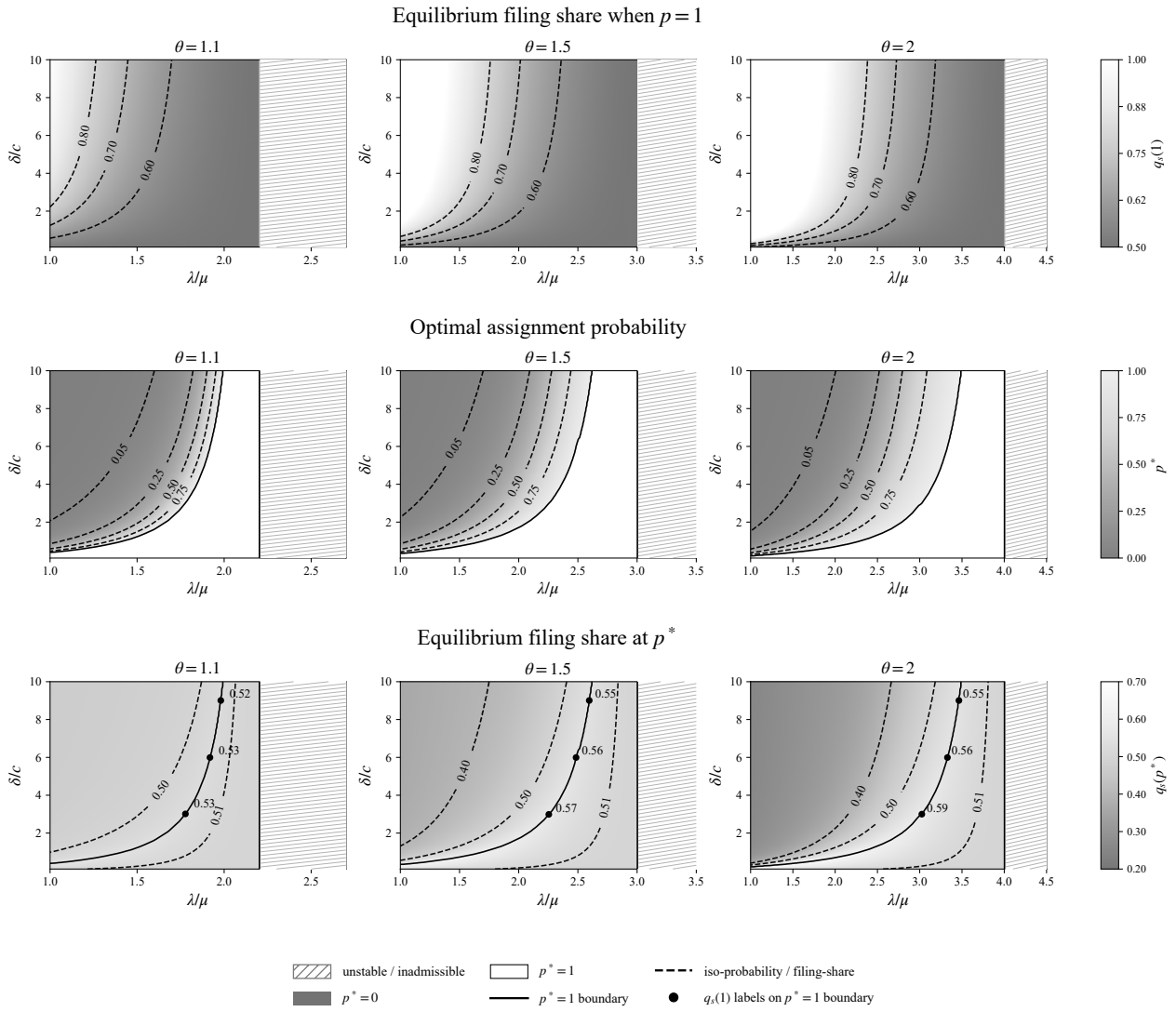


Figure 3 Symmetric-equilibrium policy maps for $\theta \in \{1.1, 1.5, 2\}$. Rows report the filing share under full targetability $q_s(1)$, the optimal assignment probability p^*, and the filing share under the optimal policy $q_s(p^*)$, respectively. In the second row, darker shades indicate lower p^*, while white denotes $p^* = 1$. Solid curves delimit the $p^* = 1$ region, and dashed curves connect parameter combinations with the same indicated value. Black dots in the bottom row report $q_s(1) = q_s(p^*)$ at selected points on the $p^* = 1$ boundary. Hatching denotes instability or inadmissibility. The $p^* = 1/2$ contour marks uniform random assignment, a particular partial-targetability policy.

The proof is provided in Appendix A.11. Accordingly, we evaluate p^* against two benchmarks: full targetability, $p = 1$, and uniform random assignment, $p = 1/2$.

Figure 4 compares p^* with uniform random assignment and full targetability across the evaluated parameter regions. When the shopping benefit is high and load is low or moderate, the optimum can lie below both benchmarks because reducing targetability shifts filings away from Court 1 and relieves congestion. As load increases, the optimum shifts toward preserving specialist matching, and uniform random assignment can perform substantially worse because more cases are processed by nonspecialists. Full targetability performs

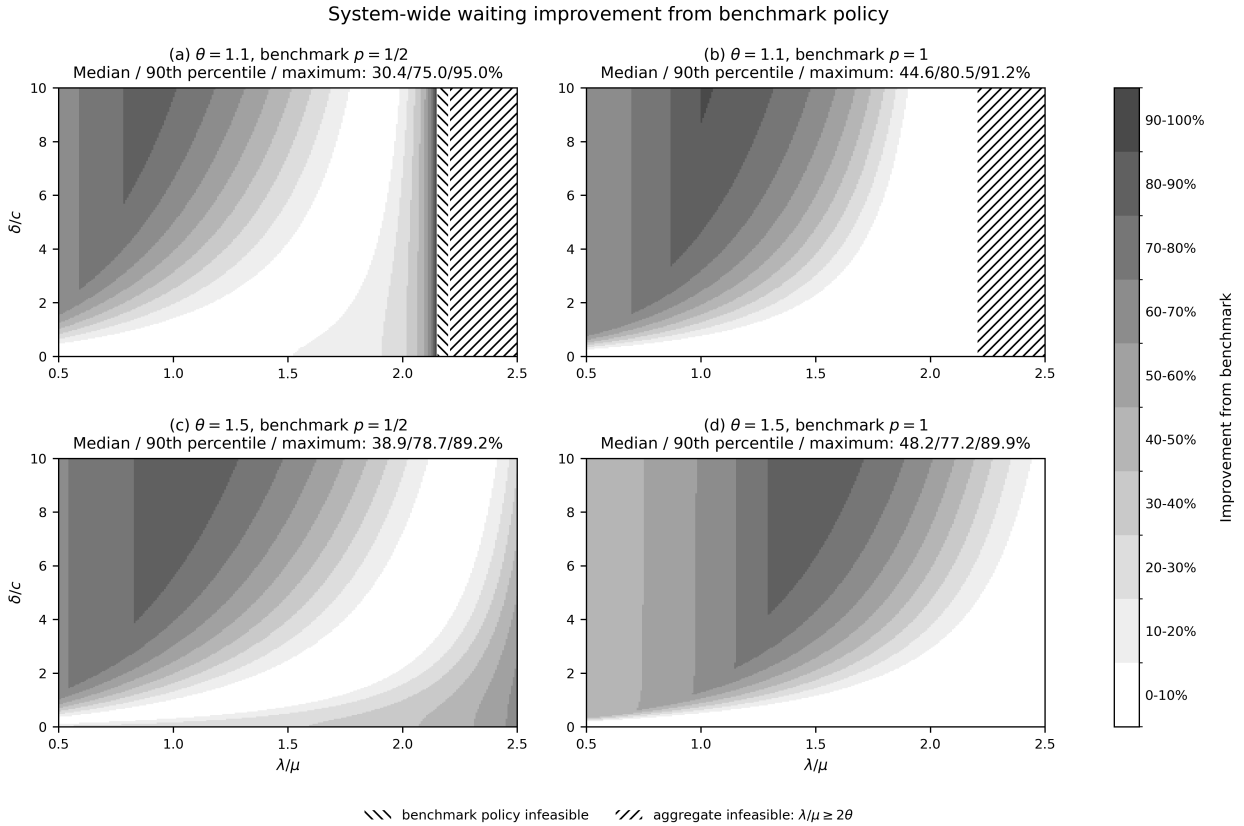


Figure 4 System-wide waiting reduction under optimal targetability. Panels report the percentage reduction in system-wide waiting obtained by replacing the indicated benchmark, $p = 1/2$ or $p = 1$, with p^*. Hatching identifies infeasibility of the benchmark policy or violation of the global stability requirement, as indicated in the legend. Median, 90th percentile, and maximum values summarize the improvements separately in each panel.

poorly in a different region, where the shopping incentive creates strong filing concentration. Thus, the two benchmarks can depart from the optimum for different reasons: full targetability can preserve too much filing concentration, whereas uniform random assignment can create too much specialization mismatch. Across the evaluated parameter regions, the median reduction in system-wide waiting under p^* ranges from 30.4% to 48.2%, with larger differences in the upper tail.

The comparison above evaluates assignment policies only through system-wide waiting. An administrator may also care directly about successful targeting, in addition to its effect on congestion. We next consider an extension in which the administrator assigns an institutional penalty $\gamma \geq 0$ to successful targeting. Equation (19) is augmented by $\gamma p(\delta/c) \sum_{i=1}^{2} q_i$. The term $p \sum_{i=1}^{2} q_i$ measures successful targeting, while δ/c scales the litigants' private benefit from successful targeting in waiting-time units. The parameter γ measures the strength of the institutional concern about successful targeting. The extension changes the administrator's objective function and optimal policy but not litigants' equilibrium filing behavior for a given p.

COROLLARY 2 **(Institutional concern)**. *Fix $\theta > 1$. For $\gamma \geq 0$, let p^*_γ minimize*

$$SPO^\gamma_s(p) := SPO_s(p) + \gamma q_s(p) p \frac{\delta}{c}$$

over the same policy domain as p^*. Then:

(i) For any two institutional weights $\gamma_1 < \gamma_2$, $p^*_{\gamma_2} \leq p^*_{\gamma_1}$. Thus, the optimal assignment probability is weakly decreasing in γ.

(ii) If $p = 0$ is feasible, there exists a finite threshold $\bar{\gamma} > 0$ such that $p = 0$ is optimal for every $\gamma \geq \bar{\gamma}$, and is the unique optimum for every $\gamma > \bar{\gamma}$.

The proof is provided in Appendix A.12. At an interior symmetric equilibrium, Equation (21) gives $SPO^\gamma_s(p) = W_2(q_s(p)) + (1+\gamma) q_s(p) p\delta/c$. Moreover, Equation (10) implies that the additional term equals $\gamma q_s(p)[W_1(q_s(p), p) - W_2(q_s(p))]$. Thus, the institutional penalty can be viewed as penalizing the equilibrium waiting difference between the two courts, weighted by Court 1's filing share. From the administrator's point of view, a stronger institutional concern therefore favors lower targetability and, from an operational point of view, a smaller waiting difference between the courts.

7. Equilibrium Multiplicity and Robustness

The preceding policy analysis follows the symmetric equilibrium. We now study whether additional asymmetric equilibria can coexist with it, what structure they have, and whether equilibrium multiplicity changes the policy conclusion. We also examine whether the results depend on the baseline narrow-pool structure by expanding each specialist pool to n specialist judges, k of whom are targeted in Court 1. We first derive the condition that separates the symmetric and asymmetric branches for interior equilibria.

7.1. Asymmetric Equilibria

For an interior equilibrium, each case type must be indifferent between the two filing venues. Substituting (7)–(8) into the Wardrop conditions (9) yields

$$p\frac{\delta}{c} = pW_{11}(\mathbf{q}, p) + (1-p)W_{12}(\mathbf{q}, p) - W_{21}(q_1), \tag{25}$$

$$p\frac{\delta}{c} = (1-p)W_{11}(\mathbf{q}, p) + pW_{12}(\mathbf{q}, p) - W_{22}(q_2). \tag{26}$$

Adding the two equilibrium conditions gives the sum equation

$$W_{11}(\mathbf{q}, p) + W_{12}(\mathbf{q}, p) - W_{21}(q_1) - W_{22}(q_2) = 2p\frac{\delta}{c}, \tag{27}$$

whereas subtracting them gives the difference equation

$$(2p-1)\left[W_{11}(\mathbf{q}, p) - W_{12}(\mathbf{q}, p)\right] = W_{21}(q_1) - W_{22}(q_2). \tag{28}$$

The sum equation equates the two case types' total Court 1 waiting disadvantage to their total shopping benefit, while the difference equation determines whether their Court 1 filing shares are equal. Substituting (1)–(4) into the difference equation yields the following result.

Lemma 3 **(Interior-equilibrium branches)**. *For every interior Wardrop equilibrium,*

$$(q_1 - q_2)B(q_1, q_2, p) = 0, \tag{29}$$

where the explicit expression for $B(q_1, q_2, p)$ is given in Appendix A.13. Hence every interior equilibrium lies on one of two branches: a symmetric branch with $q_1 = q_2$, or a possible asymmetric branch with $q_1 \neq q_2$ and $B(q_1, q_2, p) = 0$.

This factorization separates the symmetric and asymmetric interior branches. We now characterize the asymmetric branch. The proof and the explicit form of B are provided in Appendix A.13.

Corollary 3 **(Coexistence with the symmetric equilibrium)**. *If an asymmetric Wardrop equilibrium exists for a given assignment probability p, then the same p also supports a unique symmetric Wardrop equilibrium. Hence asymmetric equilibria can arise only through equilibrium multiplicity; they never exist in isolation.*

The proof is provided in Appendix A.14. Multiplicity adds possible filing patterns without removing the symmetric benchmark used in the policy analysis.

For any asymmetric equilibrium, let q_h and q_l denote the higher and lower Court 1 filing shares. We first consider the interior case, $0 < q_l < q_h < 1$. Its main operational feature is that filing volume, utilization, and waiting need not move together when judges receive different case mixes.

Proposition 7 **(Necessary structure of interior asymmetric equilibria)**. *If $\theta = 1$, no interior asymmetric Wardrop equilibrium exists. Suppose $\theta > 1$ and $\mathbf{q} = (q_h, q_l)$, with $0 < q_l < q_h < 1$, is an interior asymmetric Wardrop equilibrium. Then:*

(i) Type symmetry. It has a type-swapped counterpart $\mathbf{q} = (q_l, q_h)$ with the same total Court 1 filing share and the same system-wide waiting.

(ii) Targetability window. It can exist only if $1/2 < p < \theta^2/(1 + \theta^2)$. When $\lambda/\mu > \theta$, the upper bound becomes tighter; its exact load-dependent expression is given in Appendix A.15.

(iii) Waiting-time ordering. The equilibrium waiting times satisfy $W_{21}(q_h) < W_{22}(q_l)$ and $W_{11}((q_h, q_l), p) < W_{12}((q_h, q_l), p)$.

(iv) Filing-share ordering. Let q_s denote the unique symmetric equilibrium supported at the same assignment probability. Then

$$q_h > q_s > q_l. \tag{30}$$

(v) Case-volume and utilization ordering. Under $(q_1, q_2) = (q_h, q_l)$, the effective arrival rates satisfy $A_{11} = \lambda[pq_h + (1 - p)q_l] > A_{12} = \lambda[(1 - p)q_h + pq_l]$. Moreover, $\rho_{11}((q_h, q_l), p) < \rho_{12}((q_h, q_l), p)$ if and only if $p < \theta/(1 + \theta)$. The two judges have equal utilization at $p = \theta/(1 + \theta)$, and the ordering reverses above this threshold. Together with part (iii), the Court 1 judge receiving more cases therefore has shorter waiting on both sides of the threshold.

The case-mix mechanism explains the volume–utilization–waiting ordering in parts (iii) and (v). Relative to a specialist case, a nonspecialist case contributes θ times as much to utilization and θ^2 times as much to the service-time second moment entering waiting (3)–(4), (5). This case-mix difference allows the judge receiving more cases to have shorter waiting. The resulting Court 1 waiting advantage offsets the high-filing type's lower waiting in Court 2, allowing both types' Wardrop conditions to hold with $q_h > q_l$. The targetability window is necessary, not sufficient; the proof and tighter load-dependent necessary conditions are provided in Appendix A.15.

The boundary cases introduce the exact same mechanism.

Proposition 8 **(Necessary structure of boundary asymmetric equilibria)**. *Every boundary asymmetric Wardrop equilibrium has a type-swapped counterpart with the same system-wide waiting and can exist only if*

$$\frac{1}{2} < p < \frac{\theta^2}{1+\theta^2}. \tag{31}$$

Moreover:

(i) If one case type files only in Court 2 while the other splits its filings across the two courts, $q_l = 0$ *and* $0 < q_h < 1$*, then* $\lambda/\mu < \theta$*.*

(ii) If one case type files only in Court 1 while the other splits its filings across the two courts, $0 < q_l < 1$ *and* $q_h = 1$*, and* $\lambda/\mu > \theta$*, then* $\theta < \lambda/\mu < \theta + 1/2$*.*

(iii) Complete separation, $q_l = 0$ *and* $q_h = 1$*, can occur only if* $\theta > 2$*,*

$$\frac{\lambda}{\mu} < \theta, \qquad \left(\frac{\lambda}{\mu}\right)(1-p) < 1,$$

and the corner-feasibility condition (A.41) *holds.*

The proof and regime-specific Wardrop conditions are provided in Appendix A.16. Boundary asymmetry requires both specialization and partial targetability. Complete separation additionally requires $\theta > 2$ and stable queues at both boundary filing shares. The upper endpoint $\theta^2/(1+\theta^2)$ increases with θ, so the necessary targetability window widens with specialization and collapses to an empty interval at $\theta = 1$.

Corollary 4 **(Equilibrium symmetry without specialization)**. *Suppose* $\theta = 1$ *and* $0 < \lambda/\mu < 2$*. For every* $p \in [0,1]$*, the Wardrop equilibrium is unique and symmetric.*

The proof is provided in Appendix A.17. Specialization is necessary for equilibrium multiplicity. In the absence of specialization advantage (i.e., $\theta = 1$), targetability can change where litigants file but cannot generate asymmetric filing.

The analytical results characterize the forms and necessary conditions for asymmetric equilibria but do not determine where they exist. Figure 5 illustrates their occurrence under $\theta = 3$, a deliberately strong specialization benchmark relative to the empirical ranges in Appendix C. The symmetric equilibrium persists

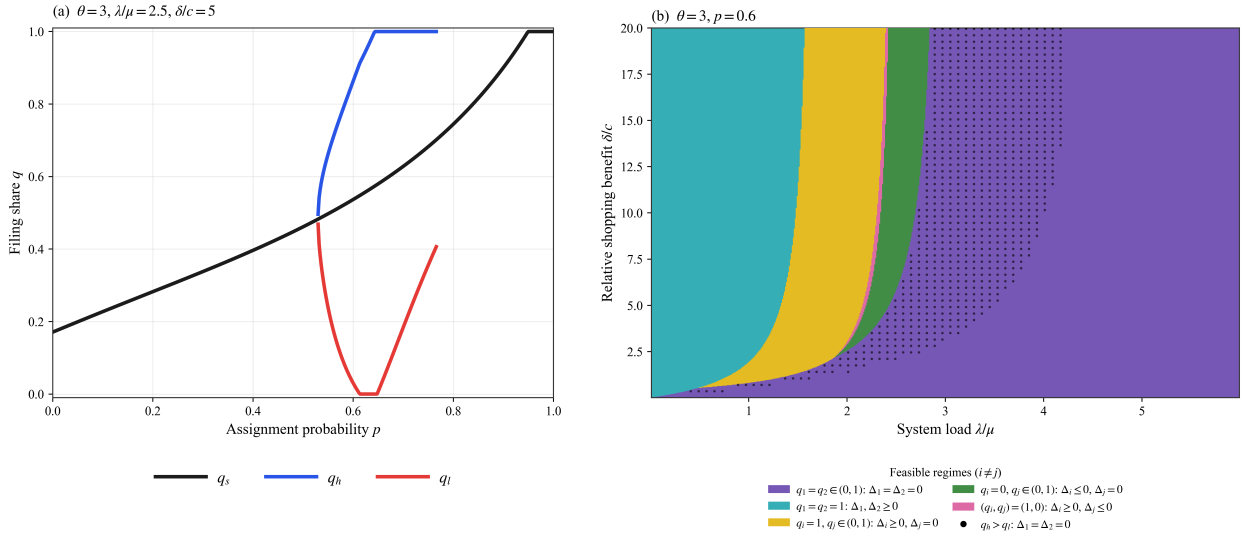


Figure 5 Symmetric and asymmetric Wardrop equilibrium structure under strong specialization. Panel (a) reports the symmetric filing share q_s and the high and low asymmetric filing shares q_h and q_l as the assignment probability varies for $\theta=3$, $\lambda/\mu=2.5$, and $\delta/c=5$. Panel (b) reports the Wardrop equilibrium regimes supported over baseline load λ/μ and relative shopping benefit δ/c for $\theta=3$ and $p=0.6$. Regimes are classified by the filing-share and Δ_i conditions in Table 1.

throughout, while asymmetric equilibria appear only over an intermediate range of targetability and move between interior and boundary forms. For example, at $\theta=3$, $\lambda/\mu=2.5$, $\delta/c=5$, and $p=0.6$, the asymmetric equilibria are approximately $(q_1,q_2)=(0.863,0.030)$ and $(0.030,0.863)$, while the symmetric equilibrium is $q_s=0.537$.

7.2. Robustness to Equilibrium Multiplicity

The symmetric optimum p^* minimizes system-wide waiting conditional on the symmetric equilibrium. To determine whether equilibrium multiplicity changes this policy conclusion, we use a conservative criterion that evaluates each policy at its highest-waiting supported equilibrium.

For each feasible p, define $\overline{SPO}(p)=\sup_{\mathbf{q}\in Q(p)} SPO(\mathbf{q};p)$. The equilibrium-robust assignment probability is

$$p^*_{\text{asym}}\in\arg\min_{p\in[0,1]} \quad \overline{SPO}(p) \quad \text{s.t.} \quad Q(p)\neq\varnothing.$$

The subscript indicates that the policy also accounts for asymmetric equilibria; it does not require an asymmetric equilibrium to remain supported at p^*_{asym}. In the numerical analysis, no asymmetric equilibrium is supported at p^* for $\theta\in\{1.1,1.5,2\}$, which include an empirically grounded value and larger specialization values used in the main analysis. Multiplicity at p^* appears only in the $\theta=3$ extreme-specialization stress test. Thus, at the evaluated values $\theta\in\{1.1,1.5,2\}$, accounting for asymmetric equilibria leaves the optimal assignment policy unchanged. This shows that the main policy conclusion does not depend on selecting the symmetric equilibrium over other supported equilibria in these parameter regions. Appendix D reports the equilibrium computations, numerical implementation, and robustness results.

Table 2 Assignment optimization versus capacity expansion

	Baseline	Untargeted expansion	Targeted expansion	Optimize p
(n, k, p)	$(12, 5, 1)$	$(13, 5, 1)$	$(13, 6, 1)$	$(12, 5, 0.58)$
Targetability pk/n	0.42	0.38	0.46	0.24
$(\widetilde{W}_1, \widetilde{W}_2, SPO_s)$	$(2.94, 0.23, 2.21)$	$(2.65, 0.15, 2.11)$	$(3.14, 0.14, 2.54)$	$(2.44, 0.88, 1.59)$
Change in SPO_s	—	−4.67%	+14.87%	**−28.21**%

7.3. Larger Specialist Pools and Capacity Expansion

We extend the model to specialist pools with multiple judges. Suppose that each case type has n specialist judges in each court. In Court 1, $1 \le k \le n$ of the n specialists for each case type are targeted, while the remaining $n - k$ are not; no judge in Court 2 is targeted. Cases assigned to a specialist pool are equally shared among its n judges.

The specialist assignment probability remains p. A type-i case is assigned to its own specialist pool with probability p and to the other specialist pool with probability $1 - p$. At $p = 1$, each judge in type i's specialist pool receives cases at rate $\lambda q_i/n$, and targetability is k/n. For $0 < p < 1$, each judge receives specialist cases at rate $\lambda p q_i/n$ and nonspecialist cases at rate $\lambda(1-p)q_{3-i}/n$, while targetability is pk/n.

Thus, targetability depends on the targeted share k/n, rather than on pool size n alone. Increasing n reduces targetability when k remains fixed, but not when k grows proportionally with n. Judge shopping can therefore remain operationally important in a large pool when litigants prefer a sufficiently large subset of its judges, particularly when k/n is close to one.

The queueing structure is unchanged. At $p = 1$, each judge serves only specialist cases, while at $p = 0$, each judge serves only nonspecialist cases; each docket is therefore an $M/M/1$ queue at either endpoint. For $0 < p < 1$, each docket mixes specialist and nonspecialist cases and remains an $M/G/1$ queue. Appendix E derives the resulting waiting times and shows that the equilibrium and policy analysis continues to apply after accounting for per-judge arrival rates and targetability.

The extension also separates the effects of capacity expansion from those of assignment design. Consider $\theta = 2$, $\lambda/\mu = 28$, $\delta/c = 6.5$, and $\mu = 1$, starting from a system with $n = 12$ and $k = 5$. We consider two forms of capacity expansion, each adding one judge to every specialist pool in both courts. An untargeted capacity expansion changes (n, k) to $(n+1, k)$, increasing specialist-pool capacity without increasing the number of targeted judges. A targeted capacity expansion changes (n, k) to $(n+1, k+1)$, increasing both specialist-pool capacity and the number of targeted judges. Table 2 compares these two expansions with optimizing p at the original capacity.

At the original capacity, optimizing the assignment probability yields $p^* \approx 0.58$ and reduces system-wide waiting by 28.21%. This gain is substantially larger than that from a one-unit capacity expansion while keeping $p = 1$: the untargeted expansion reduces system-wide waiting by only 4.67%, whereas the targeted expansion increases it by 14.87%. Thus, assignment optimization at existing capacity can be considerably more effective than expanding capacity while leaving the assignment probability unchanged.

Keeping $p = 1$ and $k = 5$, achieving at least the 28.21% reduction from assignment optimization requires expanding each specialist pool from $n = 12$ to $n = 19$, or seven additional untargeted specialists per pool. If capacity is nevertheless expanded by one unit, re-optimizing p reduces system-wide waiting by 40.70% under the untargeted expansion and by 38.16% under the targeted expansion. Thus, additional capacity can complement assignment design, but it does not replace it.

The adverse effect of capacity expansion does not require targetability itself to increase. With $k = n$, $p = 1$, $\theta = 2$, $\delta/c = 2$, $\lambda/\mu = 24$, and $\mu = 1$, expanding the specialist pools from $n = 12$ to $n = 13$ increases SPO_s from 1.74 to 1.81, or 4.34%, even though targetability remains unchanged. Appendix E shows this effect analytically near the boundary between the $q_s = 1$ and interior regimes.

8. Discussion and Managerial Implications

Recent reforms seek to curb judge shopping by expanding random assignment (Judicial Conference of the United States 2024). Their operational effects depend on both the filing response to lower targetability and the processing-efficiency loss from specialization mismatch. More broadly, assignment can simultaneously regulate where demand enters a service system and determine how well that demand is matched to heterogeneous servers. The analysis has three direct implications for assignment design. First, administrators should evaluate anti-shopping reforms by their operational consequences, not by deterrence alone. Under full targetability, judge shopping is operationally costly because the resulting filing concentration raises system-wide waiting even though waiting falls in an alternative court. The relevant operational question is how a reform changes filing concentration and system-wide waiting. Conceptually, this extends the classical alternative-server setting of Bell and Stidham (1983): the administrator does not directly allocate filings across courts, but changes the attractiveness of Court 1 through the assignment rule, and litigants reallocate themselves in response.

Second, neither uniform random assignment nor dedicated specialist assignment should be treated as a universal operational default. Lower targetability weakens the shopping incentive but increases specialization mismatch. The filing shift creates congestion relief, whereas specialization mismatch creates a processing-efficiency loss. Assignment design should balance these opposing operational consequences rather than evaluate deterrence or specialist matching in isolation. Partial targetability can be a deliberate operational choice rather than an administrative compromise. In particular, the optimal policy does not call for deliberately assigning cases to nonspecialists simply to eliminate targetability. This is consistent with institutional practice: randomized systems can preserve specialization, expertise, and experience (Eisenberg et al. 2012, Thorley 2020).

Two related consequences sharpen this second implication. First, litigants and the administrator need not prefer the same assignment rule. Litigants prefer full targetability, whereas minimizing system-wide waiting may require lower targetability. Second, the waiting objective already captures the operational effect of judge

shopping through equilibrium filing and waiting. If the administrator additionally places a direct institutional penalty on successful targeting, the optimal policy shifts further toward lower targetability; when minimum targetability is operationally feasible, a sufficiently strong institutional concern can make it optimal.

Third, administrators should account for congestion when choosing how much targetability to preserve. Greater congestion limits the filing concentration created by judge shopping while making the processing loss from specialization mismatch more costly. Both effects shift the optimal policy toward higher targetability, and sufficiently high congestion makes full targetability optimal even when lower-targetability policies remain feasible.

Larger pools and capacity expansion. Section 7.3 shows that optimizing assignment at existing capacity can outperform substantial capacity expansion while p is held fixed, and that capacity expansion can even increase system-wide waiting through the induced filing response. More broadly, this illustrates why operational changes cannot be evaluated holding user behavior fixed: strategic responses can offset their direct performance gains, as in classical models of waiting-line allocation (Holt and Sherman 1982). This complements court-operations evidence that additional judicial capacity need not translate proportionally into better performance, while scheduling, process redesign, and resource reallocation can generate substantial gains with existing capacity (Beenstock and Haitovsky 2004, Bakshi et al. 2025, Azaria et al. 2023, Gupta and Bolia 2024, Peyrache and Zago 2024). Administrators should therefore evaluate capacity expansion jointly with assignment design.

Empirical grounding. Empirical evidence informs the model's three main mechanisms. Reported specialization effects are smaller than those generated by $\theta = 2$, so we use $\theta \leq 2$ in the main analysis (Appendix C). Judge-level outcome differences provide an observable basis for the shopping benefit (Hüther and Kleiner 2026), while immigration-court data document substantial within-court outcome dispersion (Table 4). Finally, the WDTX–EDTX filing reallocation after the 2022 assignment reform is consistent with the filing response in Proposition 2 (Appendix B).

Scope and limitations. The model uses two courts and two case types and represents each court's effective assignment pool with two judges. The two courts represent a targeted venue and a relevant alternative available to the same litigants, as in the venue-choice episodes motivating the model (Ganglmair et al. 2026, Levitin 2023) and the WDTX–EDTX filing pattern in Appendix B. The baseline two-judge formulation is most directly applicable to narrow-pool settings in which specialization-based assignment is operationally valuable and makes preferred judges sufficiently targetable to support judge shopping (Judicial Conference of the United States 2024, Eisenberg et al. 2012, Thorley 2020). Section 7.3 shows, however, that the equilibrium and policy structure extends to larger specialist pools. The mechanism therefore does not require a literally two-judge pool; it requires that a sufficiently large subset of eligible specialists remain targetable. We hold total filing volume fixed, so q_i describes venue allocation rather than participation. Allowing endogenous participation or balking would make total filing volume responsive to assignment. We also assume a common

shopping benefit and waiting cost across litigants; allowing heterogeneity in these preferences would make venue choice vary across litigants within a case type.

We use a standing assignment rule rather than one that changes with current docket conditions. Load-dependent or time-varying assignment can create opportunities for strategic filing timing because the probability of reaching a preferred judge may depend on when a case is filed (Hüther and Kleiner 2026). Such a setting would add a timing decision to the venue-choice problem studied here.

We also treat the specialist and nonspecialist processing rates, $\theta\mu$ and μ, as fixed, as in related server-capability models (Chen et al. 2026). In practice, processing performance may evolve with the composition of cases assigned to a judge. Repeated exposure to a case type can improve processing performance (Coviello et al. 2019), so assignment may affect the development of expertise. This issue is especially relevant when $p^* < 1$. Future work could allow each judge's processing rate to depend on the composition of assigned cases (Pinker and Shumsky 2000, Delasay et al. 2016).

The model focuses on long-run expected waiting rather than day-to-day changes in queue length. Service times are exponential conditional on specialist matching, using available court evidence as a benchmark (Clarke and Metzler 2025). Alternative service-time distributions may change quantitative waiting levels and policy thresholds.

The filing-reallocation pattern in Appendix B and the judge-level outcome variation in Appendix C are descriptive. The specialization values in Appendix C provide empirical benchmarks rather than direct estimates of θ. The numerical results should be interpreted as model-based comparisons over the evaluated parameter regions rather than estimates of realized effects in a particular court.

9. Conclusion

This paper shows that judicial assignment shapes venue choice as well as case processing. Under full targetability, judge shopping concentrates filings in the targeted court and raises system-wide waiting relative to the no-shopping benchmark. Reducing targetability shifts filings away from the targeted court, but it also creates specialization mismatch by sending more cases to slower nonspecialists. The optimal assignment rule therefore balances congestion relief from filing reallocation against the processing-efficiency loss from specialization mismatch. Two benchmarks identify the extremes of this trade-off. Without a specialization advantage, minimum targetability is optimal because reducing targetability carries no processing-efficiency cost. When symmetric filing shares are held fixed instead, dedicated specialist assignment minimizes system-wide waiting. With endogenous filing and specialist processing, neither extreme is generally optimal. Stronger shopping incentives shift the optimum toward lower targetability, whereas greater congestion shifts it toward higher targetability and can make dedicated specialist assignment optimal. The larger-pool extension further shows that assignment optimization at existing capacity can outperform substantial capacity expansion, while strategic filing can make added capacity increase rather than decrease system-wide waiting. Overall,

assignment and capacity decisions should be evaluated by their joint effects on where litigants file and how efficiently their cases are processed.

References

Afèche P, Pavlin JM (2016) Optimal price/lead-time menus for queues with customer choice: Segmentation, pooling, and strategic delay. *Management Science* 62(8):2412–2436.

Antill S, Bellon A (2026) The real effects of bankruptcy forum shopping. Technical Report CES-26-29, U.S. Census Bureau, Center for Economic Studies.

Ata B, Hester R, Wein LM, Zhou Y (2023) Analyzing the effects of judicial rotation on criminal sentencing: An operations management perspective, working paper.

Azaria S, Ronen B, Shamir N (2023) Justice in time: A theory of constraints approach. *Journal of Operations Management* 69(7):1202–1208.

Azaria S, Ronen B, Shamir N (2024) Alleviating court congestion: The case of the jerusalem district court. *INFORMS Journal on Applied Analytics* 54(3):267–281.

Bakshi N, Kim J, Randhawa RS (2025) Service operations for justice-on-time: A data-driven queueing approach. *Manufacturing & Service Operations Management* 27(1):305–321.

Barzel Y (1974) A theory of rationing by waiting. *Journal of Law and Economics* 17(1):73–95.

Baum L (2009) Probing the effects of judicial specialization. *Duke Law Journal* 58(7):1667–1684.

Beenstock M, Haitovsky Y (2004) Does the appointment of judges increase the output of the judiciary? *International Review of Law and Economics* 24(3):351–369.

Bell CE, Stidham S Jr (1983) Individual versus social optimization in the allocation of customers to alternative servers. *Management Science* 29(7):831–839.

Biernikiewicz M, Taieb V, Toumi M (2019) Characteristics of doctor-shoppers: A systematic literature review. *Journal of Market Access & Health Policy* 7(1):1595953.

Blumstein A, Larson R (1969) Models of a total criminal justice system. *Operations Research* 17(2):199–232.

Botoman A (2018) Divisional judge-shopping. *Columbia Human Rights Law Review* 49(2):297–344.

Bray RL, Coviello D, Ichino A, Persico N (2016) Multitasking, multiarmed bandits, and the italian judiciary. *Manufacturing & Service Operations Management* 18(4):545–558.

Brooks JP (2012) The court of appeals of virginia uses integer programming and cloud computing to schedule sessions. *Interfaces* 42(6):544–553.

Buchanan JR, Fennell RD (1981) Model-directed information systems for management of the federal courts. *Management Science* 27(8):887–903.

Caldentey R, Hillas LA, Gupta V (2024) Designing service menus for bipartite queueing systems. *Operations Research* 73(3):1496–1534.

Castelliano C, Grajzl P, Alves A, Watanabe E (2021) Adjudication forums, specialization, and case duration: Evidence from brazilian federal courts. *Justice System Journal* 42(1):50–77.

Castle L (2024) Texas patent pendulum swings back from west to eastern district. Bloomberg Law, URL `https://news.bloomberglaw.com/ip-law/texas-patent-pendulum-swings-back-from-west-to-eastern-district`, accessed: 2026-07-05.

Chen J, Dong J, Shi P (2020) A survey on skill-based routing with applications to service operations management. *Queueing Systems* 96(1):53–82.

Chen Y, Xie J, Yang N, Zhang ZG, Zhu T (2026) Impact of server capability on pooling configuration in stochastic service systems. *Production and Operations Management* 35(7):2830–2850.

Cissner AB, Labriola M, Rempel M (2013) Testing the effects of New York's domestic violence courts: A statewide impact evaluation. Technical Report NCJ 242583, Center for Court Innovation, New York, NY.

Clarke DR, Metzler A (2025) Estimating the delay of criminal trials: Evidence from canada. *Journal of Law and Empirical Analysis* 2(1):33–59.

Coviello D, Ichino A, Persico N (2019) Measuring the gains from labor specialization. *Journal of Law and Economics* 62(3):403–426.

DeFond M, Zhang J, Zhao Y (2025) Do managers successfully shop for auditors who allow them to opportunistically report positive news? evidence from accounting estimates. *Management Science* 71(11):9256–9289.

Delasay M, Ingolfsson A, Kolfal B (2016) Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research* 64(4):867–885.

Deng M, Li J, Zhou N (2026) An information-based theory of auditor switching. *Management Science* Articles in Advance.

Dong J, Yom-Tov E, Yom-Tov GB (2019) The impact of delay announcements on hospital network coordination and waiting times. *Management Science* 65(5):1969–1994.

Eisenberg T, Fisher T, Rosen-Zvi I (2012) Does the judge matter? exploiting random assignment on a court of last resort to assess judge and case selection effects. *Journal of Empirical Legal Studies* 9(2):246–290.

European Commission for the Efficiency of Justice (2024) European Judicial Systems: CEPEJ Evaluation Report, 2024 Evaluation Cycle (2022 Data), Part 1: General Analyses. Technical report, Council of Europe.

Fabri M, Langbroek PM (2007) Is there a right judge for each case? a comparative study of case assignment in six european countries. *European Journal of Legal Studies* 1(2):292–315.

Falavigna G, Ippoliti R, Manello A, Ramello GB (2015) Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research* 240(2):592–601.

Feigin PD, Gjonça E, Mandelbaum A, Shapkova Kocevska K, Yom-Tov GB (2026) Operational performance of first instance courts: Augmenting disposition time via little's law (in albania, israel and north macedonia). Bos J, Silva M, Zago A, eds., *Handbook on the Efficiency of Courts of Justice: A Comparative Analysis Based on Court Data*, 1–17 (Cambridge University Press).

Freund D, Weng W (2026) The dedicated docket in U.S. immigration courts: An analysis of fairness and efficiency properties. *Manufacturing & Service Operations Management* Forthcoming.

Ganglmair B, Helmers C, Love BJ (2026) Randomly albright: The end of judge shopping in the western district of texas? *Journal of Empirical Legal Studies* Early view.

Gouzoules A (2024) Choosing your judge. *SMU Law Review* 77(4):699–749.

Gupta M, Bolia NB (2024) Redistribution of judicial resources for improved performance. *Annals of Operations Research* 342(3):2147–2168.

Holt J Charles A, Sherman R (1982) Waiting-line auctions. *Journal of Political Economy* 90(2):280–294.

Huang F, Guo P, Wang Y (2022) Modeling patients' illness perception and equilibrium analysis of their doctor shopping behavior. *Production and Operations Management* 31(3):1216–1234.

Hüther N, Kleiner K (2026) Justice good as random? *Journal of Finance* .

Jayawickrama N (2011) The kenyan judicial service code of conduct and ethics: A review of its compliance with international standards, an outline of implementation modalities, and a performance evaluation mechanism. Kenya Law.

Judicial Conference of the United States (2024) Conference acts to promote random case assignment. United States Courts, published March 12, 2024.

Kadzelashvily G, Daw A, Yom-Tov GB (2026) A two-time-scale hawkes process for modeling multi-stakeholder court case dynamics. *Proceedings of the 2026 Winter Simulation Conference*, forthcoming.

Kahan M, McKenzie TA (2021) Judge shopping. *Journal of Legal Analysis* 13(1):341–379.

Kerner Y, Oz B, Shneer S (2026) Opaque service: The case of server selection by strategic customers. *Queueing Systems* 110, article 23.

Kesan JP, Ball GG (2011) Judicial experience and the efficiency and accuracy of patent adjudication: An empirical analysis of the case for a specialized patent trial court. *Harvard Journal of Law & Technology* 24(2):393–467.

Kwak NK, Kuzdrall PJ, Schniederjans MJ (1984) Felony case scheduling policies and continuances: A simulation study. *Socio-Economic Planning Sciences* 18(1):37–43.

Levitin AJ (2023) Judge shopping in chapter 11 bankruptcy. *University of Illinois Law Review* 2023(2):351–418.

Lex Machina Data Team (2023) Patent litigation report 2023. Technical report, Lex Machina.

Lex Machina Data Team (2024) Patent litigation report 2024. Technical report, Lex Machina.

Lex Machina Data Team (2025) Patent litigation report 2025. Technical report, Lex Machina.

Nagel S, Neef M, Munshaw N (1978) Bringing management science to the courts to reduce delay. *Judicature* 62(3):128–143.

Nageswaran L (2023) Implications of vaccine shopping during pandemic. *Production and Operations Management* 32(4):1133–1149.

Naor P (1969) The regulation of queue size by levying tolls. *Econometrica* 37(1):15–24.

National Judicial Data Grid (2026) National Judicial Data Grid: District Court of India Pending Dashboard. https://njdg.ecourts.gov.in/, age-wise pending cases; accessed: 24 June 2026.

Parlaktürk AK, Kumar S (2004) Self-interested routing in queueing networks. *Management Science* 50(7):949–966.

Peyrache A, Zago A (2024) The inefficiency of courts of justice: Industry structure, capacity and misallocation. *Journal of Productivity Analysis* 62(2):217–238.

Pinker EJ, Shumsky RA (2000) The efficiency-quality trade-off of cross-trained workers. *Manufacturing & Service Operations Management* 2(1):32–48.

Randhawa R, Feigin PD, Kim J, Mandelbaum A, Mok HW, Tas H, Weinshall K, Yom-Tov GB, Bakshi N (2026) Accessible, timely, and fair justice systems: Grand challenges for msom, presented at the 2026 INFORMS MSOM Conference SIG Day.

Rodriguez AE, Ibrahim R, Zhan D (2025) On customer (Dis-)honesty in unobservable queues: The role of lying aversion. *Management Science* 71(1):844–860.

Sangiorgi F, Spatt C (2017) Opacity, credit rating shopping, and bias. *Management Science* 63(12):4016–4036.

Shamir J, Shamir N (2012) The role of prosecutor's incentives in creating congestion in criminal courts. *Review of Law & Economics* 8(3):579–618.

Supreme Court of India (2018) Shanti bhushan v. supreme court of india through its registrar and another. (2018) 8 SCC 396.

Taylor BW, Keown AJ, Barrett RT (1982) Analyzing court system congestion with Q-GERT network modeling and simulation. *Computers & Operations Research* 9(3):163–171.

Thorley D (2020) Randomness pre-considered: Recognizing and accounting for "de-randomizing" events when utilizing random judicial assignment. *Journal of Empirical Legal Studies* 17(2):342–382.

TRAC (2026) Judge-by-judge asylum decisions in immigration courts. TRAC Reports, fY 2021–2026; accessed September 5, 2026.

Tsitsiklis JN, Xu K (2017) Flexible queueing architectures. *Operations Research* 65(5):1398–1413.

Unified Patents (2026) Patent dispute report: 2025 in review. Unified Patents.

van de Klundert J, Cominetti R, Liu Y, Kong Q (2024) The interdependence between hospital choice and waiting time—with a case study in urban china. *Journal of Choice Modelling* 52:100509.

Wallace RB, Whitt W (2005) A staffing algorithm for call centers with skill-based routing. *Manufacturing & Service Operations Management* 7(4):276–294.

Williams MS, Eyre R (2021) Patent pilot program: Final report. Technical report, Federal Judicial Center, Washington, DC.

Xu X, Lian Z, Li X, Guo P (2016) A hotelling queue model with probabilistic service. *Operations Research Letters* 44(5):592–597.

Yang L (2021) Invite your friend and you'll move up in line: Optimal design of referral priority programs. *Manufacturing & Service Operations Management* 23(5):1139–1156.

Yang L, Cui S, Wang Z (2022) Design of Covid-19 testing queues. *Production and Operations Management* 31:2204–2221.

Yang L, Debo LG, Gupta V (2019) Search among queues under quality differentiation. *Management Science* 65(8):3605–3623.

Zimmer MB (2009) Overview of specialized courts. *International Journal for Court Administration* 2(1):46–60.

Appendix A: Proofs of Analytical Results

A.1. Proof of Lemma 1

At a symmetric filing profile $q_1 = q_2 = q_s$, the two Court 1 judges have the same utilization, and the two Court 2 judges have the same utilization. Using (5), queue stability therefore reduces to $(\lambda/\mu)q_s[p + (1-p)\theta] < \theta$ in Court 1 and $(\lambda/\mu)(1-q_s) < \theta$ in Court 2. The first condition gives an upper bound on the Court 1 filing share, while the second gives a lower bound. Combining these bounds with $0 < q_s < 1$ gives

$$\max\left\{0,\ 1-\frac{\theta}{\lambda/\mu}\right\} < q_s < \min\left\{1,\ \frac{\theta}{(\lambda/\mu)[p+(1-p)\theta]}\right\},$$

which is (13).

It remains to determine when this interval is nonempty. If $\lambda/\mu \le \theta$, its lower endpoint is zero, so the interval is nonempty. If $\lambda/\mu > \theta$, the Court 2 lower bound must lie below the Court 1 upper bound: $1 - \theta/(\lambda/\mu) < \theta/\{(\lambda/\mu)[p + (1-p)\theta]\}$. Rearranging gives $(\lambda/\mu - \theta)[p + (1-p)\theta] < \theta$, which is (14).

A.2. Proof of Proposition 1

The proof follows the three cases in Proposition 1. We first rule out the lower boundary $q_s = 0$, then characterize the upper boundary $q_s = 1$, and finally establish the unique interior equilibrium.

At a symmetric filing profile $q_1 = q_2 = q_s$, waiting times are given by (11)–(12). By Lemma 1, the admissible interior filing shares are those satisfying (13).

(i) Lower boundary: $q_s = 0$. At $q_s = 0$, Court 1 has no demand, so $W_1(0, p) = 0$. If $\lambda/\mu \ge \theta$, Court 2 is unstable and the profile is not admissible. If $\lambda/\mu < \theta$, then $W_2(0) > 0$, so $\Delta_1 = \Delta_2 = p\delta + cW_2(0) > 0$. The Wardrop conditions at $q_s = 0$ instead require both utility differences to be nonpositive. Therefore, $q_s = 0$ is never a symmetric Wardrop equilibrium.

(ii) Upper boundary: $q_s = 1$. At $q_s = 1$, Court 2 has no demand, so $W_2(1) = 0$. Its stability condition is (16). Conditional on stability, $\Delta_1 = \Delta_2 = p\delta - cW_1(1, p)$. Thus the Wardrop condition is (17). Hence $q_s = 1$ is an equilibrium if and only if both conditions in Proposition 1(ii) hold.

(iii) Interior: $0 < q_s < 1$. If $q_s = 1$ is not an admissible Wardrop equilibrium, the remaining symmetric equilibrium must be interior. We first establish existence and uniqueness using the equilibrium waiting difference $W_1(q_s, p) - W_2(q_s) - p\delta/c$.

If $\lambda/\mu < \theta$, the lower endpoint is $q_s = 0$, where this difference is negative. If $\lambda/\mu \ge \theta$, the lower endpoint is $1 - \theta/(\lambda/\mu)$. As q_s approaches this endpoint from above, the Court 2 stability denominator approaches zero, so $W_2(q_s) \to \infty$ and the equilibrium waiting difference tends to $-\infty$.

At the upper endpoint, there are again two cases. If the Court 1 stability bound is the active endpoint, then $W_1(q_s, p) \to \infty$, so the equilibrium waiting difference becomes positive. If the active endpoint is $q_s = 1$, failure of (17) gives $W_1(1, p) - p\delta/c > 0$.

On the admissible interval, both waiting times are continuous, and

$$\frac{\partial}{\partial q_s}\left[W_1(q_s, p) - W_2(q_s)\right] = \frac{\lambda/\mu}{\mu}\left[\frac{p + (1-p)\theta^2}{[\theta - (\lambda/\mu)q_s[p+(1-p)\theta]]^2} + \frac{1}{[\theta - (\lambda/\mu)(1-q_s)]^2}\right] > 0. \tag{A.1}$$

Thus the equilibrium waiting difference is continuous and strictly increasing. Since it is negative at the lower endpoint and positive at the upper endpoint, (10) has a unique interior solution.

It remains to derive its closed form and verify that this solution lies in the admissible interval. Substituting (11) and (12) into (10) and multiplying by the positive stability denominators gives

$$\begin{aligned}\frac{\lambda}{\mu}q_s\left[p+(1-p)\theta^2\right]\left[\theta-\frac{\lambda}{\mu}(1-q_s)\right]-\frac{\lambda}{\mu}(1-q_s)\left[\theta-\frac{\lambda}{\mu}q_s\left[p+(1-p)\theta\right]\right]\\ -p\frac{\delta}{c}\theta\mu\left[\theta-\frac{\lambda}{\mu}q_s\left[p+(1-p)\theta\right]\right]\times\left[\theta-\frac{\lambda}{\mu}(1-q_s)\right]=0.\end{aligned} \tag{A.2}$$

Collecting powers of q_s yields (15), with coefficients

$$\begin{aligned}a_s(p)&=\theta\left(\frac{\lambda}{\mu}\right)^2\left\{(1-p)(\theta-1)+p\mu\frac{\delta}{c}\left[p+(1-p)\theta\right]\right\},\\ b_s(p)&=\frac{\lambda}{\mu}\left\{\theta\left[1+p+(1-p)\theta^2\right]-\frac{\lambda}{\mu}(1-p)\theta(\theta-1)\right.\\ &\qquad\left.+p\theta\mu\frac{\delta}{c}\left(\left[p+(1-p)\theta\right]\left[\theta-\frac{\lambda}{\mu}\right]-\theta\right)\right\},\\ c_s(p)&=-\theta\left\{\frac{\lambda}{\mu}+p\theta\mu\frac{\delta}{c}\left[\theta-\frac{\lambda}{\mu}\right]\right\}.\end{aligned}$$

If $(\theta,p)\neq(1,0)$, then $a_s(p)>0$. At $(\theta,p)=(1,0)$, the equilibrium equation is linear and gives $q_s=1/2$.

The two algebraic roots are $q_s^{\pm}(p)=[-b_s(p)\pm\sqrt{b_s(p)^2-4a_s(p)c_s(p)}]/[2a_s(p)]$.

We first show that the negative-sign root is not admissible. If $\lambda/\mu\leq\theta$, then $c_s(p)<0$. Since $a_s(p)>0$, the two roots have opposite signs: $q_s^-(p)q_s^+(p)=c_s(p)/a_s(p)<0$. The unique admissible root is positive, so $q_s^-(p)<0$.

Now suppose $\lambda/\mu>\theta$. Evaluating the left-hand side of (15) at the Court 2 lower stability bound gives

$$\begin{aligned}&a_s(p)\left(1-\frac{\theta}{\lambda/\mu}\right)^2+b_s(p)\left(1-\frac{\theta}{\lambda/\mu}\right)+c_s(p)\\ &\quad=-\theta\left[\theta-\left(\frac{\lambda}{\mu}-\theta\right)\left[p+(1-p)\theta\right]\right]<0,\end{aligned} \tag{A.3}$$

where the final inequality follows from the capacity-feasibility condition (14). Since $a_s(p)>0$, the Court 2 lower stability bound lies strictly between the two roots: $q_s^-(p)<1-\theta/(\lambda/\mu)<q_s^+(p)$. Thus $q_s^-(p)$ violates the Court 2 lower stability bound in (13).

We next verify that the positive-sign root lies below the upper endpoint of the admissible interval. Suppose first that the Court 1 stability bound is the active upper endpoint, so $\theta/\{(\lambda/\mu)[p+(1-p)\theta]\}\leq 1$. Evaluating the left-hand side of (15) at this boundary and using (A.2) gives

$$\begin{aligned}&a_s(p)\left(\frac{\theta}{(\lambda/\mu)\left[p+(1-p)\theta\right]}\right)^2+b_s(p)\left(\frac{\theta}{(\lambda/\mu)\left[p+(1-p)\theta\right]}\right)+c_s(p)\\ &\quad=\frac{\lambda}{\mu}\frac{\theta}{(\lambda/\mu)\left[p+(1-p)\theta\right]}\left[p+(1-p)\theta^2\right]\left[\theta-\frac{\lambda}{\mu}\left(1-\frac{\theta}{(\lambda/\mu)\left[p+(1-p)\theta\right]}\right)\right]>0.\end{aligned} \tag{A.4}$$

The Court 2 lower stability bound lies between the two roots, as shown above. Because the quadratic opens upward, its positive value at the Court 1 upper stability bound implies $q_s^+(p)<\theta/\{(\lambda/\mu)[p+(1-p)\theta]\}$.

If instead the admissible interval extends to $q_s=1$, an interior equilibrium means that the upper-boundary Wardrop condition (17) fails. Hence $W_1(1,p)-p\delta/c>0$. By (A.2), the left-hand side of (15) is therefore positive at $q_s=1$, so $q_s^+(p)<1$.

Combining the lower- and upper-bound arguments shows that $q_s^+(p)$ satisfies (13). It is therefore the unique admissible interior root and coincides with (18).

The comparative statics below differentiate the interior equilibrium, so we record the required regularity here. On every admissible interior region, the waiting-time denominators in (11)–(12) are strictly positive. Hence the waiting times are smooth in the model parameters. Because (A.1) is strictly positive, the implicit-function theorem implies that the unique interior filing share q_s, and therefore $W_1(q_s, p)$, $W_2(q_s)$, and $SPO_s(p)$, vary smoothly with these parameters while μ is held fixed. At a transition from $0 < q_s < 1$ to $q_s = 1$, any sequence of interior equilibria approaching the transition has a limit satisfying the limiting Wardrop conditions. Uniqueness implies that this limit is $q_s = 1$. Therefore $q_s(p)$ and $SPO_s(p)$ are continuous across the transition.

A.3. Proof of Proposition 2

We prove the four comparative-static results in the same order as Proposition 2. Throughout, q_s denotes the admissible interior symmetric equilibrium. Total derivatives follow the equilibrium filing share, while partial derivatives hold q_s fixed. Queue stability guarantees that the denominators in (11)–(12) are strictly positive. The symmetric Wardrop condition is (10), and (A.1) shows that its left-hand side is strictly increasing in q_s.

(i) Relative shopping benefit Differentiating (10) along the equilibrium with respect to δ/c gives

$$\frac{dq_s}{d(\delta/c)} = \frac{\mu p}{(\lambda/\mu)\left[\frac{p+(1-p)\theta^2}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2} + \frac{1}{[\theta-(\lambda/\mu)(1-q_s)]^2}\right]}. \tag{A.5}$$

The denominator is strictly positive. Hence $dq_s/d(\delta/c) > 0$ for $p > 0$, while $dq_s/d(\delta/c) = 0$ at $p = 0$. Thus a greater relative shopping benefit increases Court 1's equilibrium filing share whenever the preferred judge is targetable.

Using (A.5) in the total derivatives of the two waiting times gives

$$\frac{dW_2(q_s)}{d(\delta/c)} = -\frac{p/[\theta-(\lambda/\mu)(1-q_s)]^2}{\frac{p+(1-p)\theta^2}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2} + \frac{1}{[\theta-(\lambda/\mu)(1-q_s)]^2}} \le 0,$$

$$\frac{dW_1(q_s,p)}{d(\delta/c)} = p\frac{\left[p+(1-p)\theta^2\right]/[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2}{\frac{p+(1-p)\theta^2}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2} + \frac{1}{[\theta-(\lambda/\mu)(1-q_s)]^2}} \ge 0.$$

Both inequalities are strict when $p > 0$. Therefore, a greater relative shopping benefit raises equilibrium waiting in Court 1 and lowers it in Court 2. At $p = 0$, neither filing nor waiting responds to δ/c.

(ii) Assignment probability Differentiating (10) with respect to p gives

$$\frac{dq_s}{dp} = \frac{\frac{\delta}{c} - \frac{\partial W_1(q_s,p)}{\partial p}}{\frac{\partial W_1(q_s,p)}{\partial q_s} - \frac{\partial W_2(q_s)}{\partial q_s}} = \frac{\frac{\delta}{c}}{\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}} + \frac{-\frac{\partial W_1}{\partial p}}{\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}}. \tag{A.6}$$

The first term is the shopping-incentive channel in Proposition 2(ii), while the second is the specialization-mismatch channel.

Holding the filing share fixed,

$$\frac{\partial W_1(q_s,p)}{\partial p} = -\frac{(\lambda/\mu)q_s(\theta-1)\left[\theta+1-(\lambda/\mu)q_s\right]}{\mu\left[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]\right]^2} \le 0. \tag{A.7}$$

The derivative is strictly negative when $\theta > 1$ and equals zero when $\theta = 1$. Because $\partial W_1/\partial q_s - \partial W_2/\partial q_s > 0$ by (A.1) and $\delta/c > 0$, (A.6) implies $dq_s/dp > 0$. Thus higher targetability increases Court 1's equilibrium filing share.

Moreover, $\partial W_2/\partial q_s = -(\lambda/\mu)/\{\mu[\theta-(\lambda/\mu)(1-q_s)]^2\} < 0$. Since $dq_s/dp > 0$, it follows that $dW_2/dp < 0$. Differentiating (10) along the equilibrium gives $dW_1/dp = dW_2/dp + \delta/c$. Hence Court 1 waiting increases with targetability if and only if $-dW_2/dp < \delta/c$, as stated in Proposition 2(ii).

Using (A.6) and (A.7), this condition can also be written as

$$\frac{dW_1(q_s,p)}{dp} > 0 \quad \Longleftrightarrow \quad \frac{\delta}{c} > \frac{(\lambda/\mu)q_s(\theta-1)\left[\theta+1-(\lambda/\mu)q_s\right]}{\mu\left[p+(1-p)\theta^2\right]\left[\theta-(\lambda/\mu)(1-q_s)\right]^2}.$$

Equality gives $dW_1/dp = 0$, while reversing the inequality gives $dW_1/dp < 0$.

(iii) Specialization advantage Because the right-hand side of the symmetric Wardrop condition (10) does not depend on θ, the two equilibrium waiting times move together: $dW_1(q_s,p)/d\theta = dW_2(q_s)/d\theta$. We first determine how the filing share responds.

Implicit differentiation of (10) gives

$$\frac{dq_s}{d\theta} = \frac{\dfrac{q_s p\left[2\theta-(\lambda/\mu)q_s\{p+(1-p)\theta(2-\theta)\}\right]}{\left[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]\right]^2} - \dfrac{(1-q_s)\left[2\theta-(\lambda/\mu)(1-q_s)\right]}{\left[\theta-(\lambda/\mu)(1-q_s)\right]^2}}{\theta^2\left[\dfrac{p+(1-p)\theta^2}{\left[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]\right]^2} + \dfrac{1}{\left[\theta-(\lambda/\mu)(1-q_s)\right]^2}\right]}. \tag{A.8}$$

The denominator in (A.8) is strictly positive. Therefore,

$$\frac{dq_s}{d\theta} > 0 \quad \Longleftrightarrow \quad \frac{q_s p\left[2\theta-(\lambda/\mu)q_s\{p+(1-p)\theta(2-\theta)\}\right]}{\left[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]\right]^2} > \frac{(1-q_s)\left[2\theta-(\lambda/\mu)(1-q_s)\right]}{\left[\theta-(\lambda/\mu)(1-q_s)\right]^2}. \tag{A.9}$$

This is the Appendix form of the comparison $\partial W_1/\partial\theta < \partial W_2/\partial\theta$ reported in Proposition 2(iii).

We next show why the sign changes once as the relative shopping benefit increases. For $0 < p < 1$ and $\theta > 1$, differentiating the difference between the two sides of (A.9) with respect to q_s gives

$$2\theta^2\left[\frac{p\left[1+(\lambda/\mu)q_s(1-p)(\theta-1)\right]}{\left[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]\right]^3} + \frac{1}{\left[\theta-(\lambda/\mu)(1-q_s)\right]^3}\right] > 0. \tag{A.10}$$

Thus the expression determining the sign of $dq_s/d\theta$ is strictly increasing in the Court 1 filing share.

At the no-shopping benchmark $\delta/c = 0$, the Wardrop condition gives $W_1(q_s,p) = W_2(q_s)$. If $q_s = 1/2$ is admissible, then $W_1(1/2,p) > W_2(1/2)$, because $p+(1-p)\theta^2 > 1$ and $p+(1-p)\theta > 1$. The strictly increasing waiting difference therefore gives $q_s < 1/2$. If $q_s = 1/2$ is not admissible, Court 2 remains stable there under global stability, so the Court 1 upper stability bound is at most $1/2$. The admissible equilibrium again satisfies $q_s < 1/2$.

Using the no-shopping Wardrop equality to simplify the numerator of (A.8) gives

$$-\frac{q_s\theta^2(1-p)^2(1-q_s)(\theta-1)\left[1-q_s\{1+p+(1-p)\theta(2-\theta)\}\right]}{\left[p+(1-p)\theta^2\right]\left[1-q_s\{1+p+(1-p)\theta\}\right]^2} < 0.$$

Indeed, $1+p+(1-p)\theta(2-\theta) < 2$, while $q_s < 1/2$. Hence $dq_s/d\theta < 0$ at $\delta/c = 0$.

Now consider the upper endpoint of the admissible interior region. If the Court 1 stability bound is active, then $\theta-(\lambda/\mu)q_s[p+(1-p)\theta] \downarrow 0$, while $2\theta-(\lambda/\mu)q_s\{p+(1-p)\theta(2-\theta)\} \to \theta[p+(1-p)\theta^2]/[p+(1-p)\theta] > 0$. The first term in the numerator of (A.8) therefore tends to positive infinity.

If instead the upper endpoint is $q_s = 1$, the numerator of (A.8) becomes

$$\frac{p\left[2\theta-(\lambda/\mu)\{p+(1-p)\theta(2-\theta)\}\right]}{\left[\theta-(\lambda/\mu)[p+(1-p)\theta]\right]^2} > 0.$$

The inequality follows from Court 1 stability, $(\lambda/\mu)[p + (1-p)\theta] < \theta$, together with $2[p + (1-p)\theta] - \{p + (1-p)\theta(2-\theta)\} = p + (1-p)\theta^2 > 0$.

Thus the expression determining the sign of $dq_s/d\theta$ is negative at the no-shopping benchmark, positive at the upper endpoint, and strictly increasing in q_s. Since (A.5) gives $dq_s/d(\delta/c) > 0$ for $p > 0$, there is a unique positive value of the relative shopping benefit at which the specialization response changes sign. Denote this value by $\eta_\theta(\lambda/\mu, p, \theta; \mu) > 0$. Then $dq_s/d\theta < 0$ when $\delta/c < \eta_\theta$, equals zero when $\delta/c = \eta_\theta$, and is positive when $\delta/c > \eta_\theta$.

At the threshold, the expression determining the sign of $dq_s/d\theta$ is continuously differentiable in δ/c and λ/μ. Its derivative with respect to δ/c is strictly positive because (A.10) is positive and (A.5) gives $dq_s/d(\delta/c) > 0$. The implicit-function theorem therefore implies that $\eta_\theta(\lambda/\mu, p, \theta; \mu)$ is continuously differentiable in λ/μ.

We now show that this threshold increases with baseline load. At $\delta/c = \eta_\theta(\lambda/\mu, p, \theta; \mu)$, the condition $dq_s/d\theta = 0$ is equivalent to $\partial W_1/\partial\theta = \partial W_2/\partial\theta$. Differentiating this equality and the symmetric Wardrop condition with respect to λ/μ, while holding p, θ, and μ fixed, gives

$$\left(\frac{\partial^2 W_1}{\partial q_s \partial\theta} - \frac{\partial^2 W_2}{\partial q_s \partial\theta}\right)\frac{dq_s}{d(\lambda/\mu)} + \frac{\partial^2 W_1}{\partial(\lambda/\mu)\partial\theta} - \frac{\partial^2 W_2}{\partial(\lambda/\mu)\partial\theta} = 0,$$
$$\left(\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}\right)\frac{dq_s}{d(\lambda/\mu)} + \frac{\partial W_1}{\partial(\lambda/\mu)} - \frac{\partial W_2}{\partial(\lambda/\mu)} = p\frac{\partial\eta_\theta}{\partial(\lambda/\mu)}.$$

Solving the first equality for $dq_s/d(\lambda/\mu)$, substituting into the second, and using (A.9) gives

$$\operatorname{sign}\left[\frac{\partial\eta_\theta}{\partial(\lambda/\mu)}\right] = \operatorname{sign}\left[\left[p + (1-p)\theta^2\right] - \frac{p\left[1 + (1-p)(\theta-1)(\lambda/\mu)q_s\right]}{\sqrt{1 - 2(1-p)(\lambda/\mu)q_s + (1-p)(\lambda/\mu)^2 q_s^2}}\right]. \tag{A.11}$$

The square root is strictly positive because $1 - 2(1-p)(\lambda/\mu)q_s + (1-p)(\lambda/\mu)^2 q_s^2 = p + (1-p)[1 - (\lambda/\mu)q_s]^2 > 0$. Both quantities being compared in (A.11) are therefore positive, so their ordering can be checked after squaring. It is enough to show

$$\left[p + (1-p)\theta^2\right]^2\left[1 - 2(1-p)(\lambda/\mu)q_s + (1-p)(\lambda/\mu)^2 q_s^2\right]$$
$$- p^2\left[1 + (1-p)(\theta-1)(\lambda/\mu)q_s\right]^2 > 0. \tag{A.12}$$

The left-hand side of (A.12) is quadratic in $(\lambda/\mu)q_s$. Its discriminant is $-4p\theta^2(1-p)^2[p + (1-p)\theta^2]^3 < 0$, while its value at $(\lambda/\mu)q_s = 0$ is $[p + (1-p)\theta^2]^2 - p^2 > 0$. Hence it is strictly positive, so $\partial\eta_\theta(\lambda/\mu, p, \theta; \mu)/\partial(\lambda/\mu) > 0$. Thus higher baseline load raises the relative shopping benefit required for a greater specialization advantage to increase Court 1's filing share, as stated in Proposition 2(iii).

The two endpoint assignment policies give the remaining signs directly. At $p = 0$, the first term in the numerator of (A.8) vanishes, so $dq_s/d\theta < 0$. At $p = 1$, (A.9) becomes

$$\frac{q_s[2\theta - (\lambda/\mu)q_s]}{[\theta - (\lambda/\mu)q_s]^2} > \frac{(1-q_s)[2\theta - (\lambda/\mu)(1-q_s)]}{[\theta - (\lambda/\mu)(1-q_s)]^2}.$$

The difference between the left- and right-hand sides equals

$$\frac{\theta^2[2\theta - (\lambda/\mu)](2q_s - 1)}{[\theta - (\lambda/\mu)q_s]^2[\theta - (\lambda/\mu)(1-q_s)]^2} > 0,$$

because $q_s > 1/2$, the queues are stable, and $\lambda/\mu < 2\theta$. Therefore $dq_s/d\theta|_{p=1} > 0$.

Finally, substituting (A.8) into the total derivative of $W_2(q_s)$ gives

$$\frac{dW_2(q_s)}{d\theta} = -\frac{\lambda/\mu}{\mu\theta^2\left[[p+(1-p)\theta^2][\theta-(\lambda/\mu)(1-q_s)]^2+[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2\right]} \times\Bigg\{[p+(1-p)\theta^2](1-q_s)[2\theta-(\lambda/\mu)(1-q_s)] + q_s p\,[2\theta-(\lambda/\mu)q_s\{p+(1-p)\theta(2-\theta)\}]\Bigg\} < 0.$$

The first term in braces is strictly positive by Court 2 stability. If $p+(1-p)\theta(2-\theta)\le 0$, the second term is nonnegative. Otherwise, Court 1 stability gives $2\theta-(\lambda/\mu)q_s\{p+(1-p)\theta(2-\theta)\} > \theta[p+(1-p)\theta^2]/[p+(1-p)\theta] > 0$. At $p=0$, the second term vanishes. Thus the expression in braces is strictly positive, giving $dW_1(q_s,p)/d\theta = dW_2(q_s)/d\theta < 0$.

(iv) Baseline load Differentiating (10) with respect to λ/μ, while holding p, θ, and μ fixed, gives

$$\frac{dq_s}{d(\lambda/\mu)} = -\frac{\dfrac{q_s[p+(1-p)\theta^2]}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2}-\dfrac{1-q_s}{[\theta-(\lambda/\mu)(1-q_s)]^2}}{(\lambda/\mu)\left[\dfrac{p+(1-p)\theta^2}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^2}+\dfrac{1}{[\theta-(\lambda/\mu)(1-q_s)]^2}\right]}. \tag{A.13}$$

The denominator is strictly positive. Therefore,

$$\frac{dq_s}{d(\lambda/\mu)} < 0 \quad\Longleftrightarrow\quad q_s\frac{\partial W_1}{\partial q_s} > -(1-q_s)\frac{\partial W_2}{\partial q_s}, \tag{A.14}$$

which is the condition stated in Proposition 2(iv).

For $0<p<1$ and $\theta>1$, the numerator of (A.13) is strictly increasing in q_s. Its derivative is

$$\frac{[p+(1-p)\theta^2]\,[\theta+(\lambda/\mu)q_s[p+(1-p)\theta]]}{[\theta-(\lambda/\mu)q_s[p+(1-p)\theta]]^3}+\frac{\theta+(\lambda/\mu)(1-q_s)}{[\theta-(\lambda/\mu)(1-q_s)]^3}>0.$$

At the no-shopping benchmark $\delta/c=0$, the equilibrium satisfies $q_s < 1/[1+p+(1-p)\theta]$. To see this, evaluate the two waiting times at equality: their denominators are the same, while $p+(1-p)\theta^2 > p+(1-p)\theta$, so Court 1 waiting would be larger than Court 2 waiting, contradicting the no-shopping Wardrop equality $W_1=W_2$. Using this equality in the numerator of (A.13) shows that its sign is the sign of $q_s[1+p+(1-p)\theta]-1$, which is strictly negative.

At the upper endpoint of the admissible interior region, the numerator is strictly positive. If the Court 1 stability boundary is active, its first term tends to positive infinity. If instead the upper endpoint is $q_s=1$, the second term vanishes while the first remains positive. Since the numerator is strictly increasing in q_s, it crosses zero exactly once.

Because (A.5) gives $dq_s/d(\delta/c)>0$ for $p>0$, this crossing corresponds to a unique positive value $\eta_\lambda(\lambda/\mu,p,\theta;\mu)$ of the relative shopping benefit. For $\delta/c<\eta_\lambda$, $dq_s/d(\lambda/\mu)>0$; at $\delta/c=\eta_\lambda$, the derivative is zero; and for $\delta/c>\eta_\lambda$, it is negative.

Finally, consider full targetability. At $p=1$ and $\delta>0$, Proposition 1 and (10) imply $q_s>1/2$. Condition (A.14) then becomes

$$\frac{q_s}{[\theta-(\lambda/\mu)q_s]^2} > \frac{1-q_s}{[\theta-(\lambda/\mu)(1-q_s)]^2}.$$

The difference between the left- and right-hand sides is

$$\frac{(2q_s-1)\left[\theta^2-(\lambda/\mu)^2q_s(1-q_s)\right]}{[\theta-(\lambda/\mu)q_s]^2[\theta-(\lambda/\mu)(1-q_s)]^2}>0.$$

The numerator is positive because $q_s>1/2$ and queue stability implies $(\lambda/\mu)^2q_s(1-q_s)<\theta^2$. Hence $dq_s/d(\lambda/\mu)|_{p=1}<0$.

A.4. Derivation of Equation (20)

Under $q_1 = q_2 = q_s$, both case types have the same filing share, with $W_{11} = W_{12} = W_1$ and $W_{21} = W_{22} = W_2$. Substituting these equalities into (19) gives $SPO((q_s, q_s); p) = 2[q_s W_1(q_s, p) + (1 - q_s)W_2(q_s)] = 2SPO_s(p)$. The factor two is positive and independent of p, so minimizing SPO along the symmetric equilibrium is equivalent to minimizing SPO_s.

A.5. Proof of Proposition 3

At full targetability, $W_1(q_s, 1) = (\lambda/\mu)q_s/\{\theta\mu[\theta - (\lambda/\mu)q_s]\}$ and $W_2(q_s) = (\lambda/\mu)(1 - q_s)/\{\theta\mu[\theta - (\lambda/\mu)(1 - q_s)]\}$.

The waiting difference $W_1(q_s, 1) - W_2(q_s)$ is strictly increasing in q_s and equals zero at $q_s = 1/2$. At the no-shopping benchmark $\delta = 0$, the Wardrop condition (10) therefore gives $q_s = 1/2$, which yields the benchmark waiting times reported in the proposition.

When $\delta > 0$, an interior equilibrium satisfies $W_1(q_s, 1) - W_2(q_s) = \delta/c > 0$, so $q_s > 1/2$; the same filing-share inequality holds at the upper boundary $q_s = 1$. Since $W_1(q_s, 1)$ increases with q_s while $W_2(q_s)$ decreases, judge shopping raises Court 1 waiting and lowers Court 2 waiting relative to the no-shopping benchmark.

It remains to show that both system-wide waiting measures increase. The functions $W_1(q_s, 1) + W_2(q_s)$ and $q_s W_1(q_s, 1) + (1 - q_s)W_2(q_s)$ both have zero first derivative at $q_s = 1/2$. Their second derivatives are

$$\frac{d^2}{dq_s^2}[W_1(q_s, 1) + W_2(q_s)] = \frac{2(\lambda/\mu)^2}{\mu}\left\{\frac{1}{[\theta - (\lambda/\mu)q_s]^3} + \frac{1}{[\theta - (\lambda/\mu)(1 - q_s)]^3}\right\} > 0,$$

$$\frac{d^2}{dq_s^2}[q_s W_1(q_s, 1) + (1 - q_s)W_2(q_s)] = \frac{2(\lambda/\mu)\theta}{\mu}\left\{\frac{1}{[\theta - (\lambda/\mu)q_s]^3} + \frac{1}{[\theta - (\lambda/\mu)(1 - q_s)]^3}\right\} > 0.$$

Thus both expressions are uniquely minimized at $q_s = 1/2$. Since every symmetric equilibrium with $\delta > 0$ under full targetability has $q_s > 1/2$, judge shopping strictly increases both $W_1 + W_2$ and SPO_s relative to the no-shopping benchmark.

A.6. Proof of Proposition 4

Holding $q_1 = q_2 = q_s$ fixed gives

$$SPO(\mathbf{q}; p) = 2\left[q_s W_1(q_s, p) + (1 - q_s)W_2(q_s)\right].$$

Because $W_2(q_s)$ does not depend on p, differentiating Equation (11) gives

$$\frac{\partial SPO(\mathbf{q}; p)}{\partial p} = -\frac{2(\lambda/\mu)q_s^2(\theta - 1)[\theta + 1 - (\lambda/\mu)q_s]}{\mu\left[\theta - (\lambda/\mu)q_s[p + (1 - p)\theta]\right]^2} < 0.$$

The inequality follows from $\theta > 1$, $q_s > 0$, and queue stability. Moreover, increasing p reduces the processing load in Court 1 and does not affect Court 2, so $p = 1$ is admissible whenever the fixed filing profile is admissible for some p. Hence $p = 1$ uniquely minimizes system-wide waiting over the admissible assignment probabilities.

A.7. Proof of Proposition 5

When $\theta = 1$, assignment does not affect processing: for a fixed filing share, the waiting times are independent of p. The symmetric waiting functions reduce to

$$W_1(q_s) = \frac{(\lambda/\mu)q_s}{\mu[1 - (\lambda/\mu)q_s]}, \qquad W_2(q_s) = \frac{(\lambda/\mu)(1 - q_s)}{\mu[1 - (\lambda/\mu)(1 - q_s)]}.$$

Thus targetability affects equilibrium only through the shopping incentive $p\delta/c$. Under the maintained global stability requirement $\lambda/\mu < 2$, every $p \in [0, 1]$ supports a symmetric equilibrium.

At $p = 0$, (10) gives $W_1 = W_2$, so symmetry of the two waiting functions implies $q_s(0) = 1/2$. For every $p > 0$, an interior equilibrium satisfies $W_1 - W_2 = p\delta/c > 0$, so $q_s(p) > 1/2$; if the equilibrium reaches the boundary $q_s = 1$, lowering p still creates no processing-efficiency loss and can only weaken the shopping incentive.

System-wide waiting as a function of the filing share is

$$SPO_{s,\theta=1}(q_s) = \frac{(\lambda/\mu)q_s^2}{\mu[1-(\lambda/\mu)q_s]} + \frac{(\lambda/\mu)(1-q_s)^2}{\mu[1-(\lambda/\mu)(1-q_s)]}.$$

The first term is strictly convex because $d^2\{(\lambda/\mu)q_s^2/[\mu(1-(\lambda/\mu)q_s)]\}/dq_s^2 = 2(\lambda/\mu)/\{\mu[1-(\lambda/\mu)q_s]^3\} > 0$; the second term is strictly convex by symmetry. Hence $SPO_{s,\theta=1}(q_s)$ is strictly convex. It is also symmetric around $q_s = 1/2$, so its derivative is zero there. Therefore system-wide waiting is uniquely minimized at $q_s = 1/2$ and strictly increases for $q_s > 1/2$.

Since $p = 0$ gives $q_s = 1/2$, while every $p > 0$ gives $q_s > 1/2$ or $q_s = 1$, minimum targetability uniquely minimizes system-wide waiting. Therefore $p^* = 0$.

A.8. Proof of Lemma 2

Having characterized the symmetric equilibrium, we now determine how system-wide waiting changes with targetability.

Connected policy range. By Lemma 1, the condition for the symmetric admissibility interval to be nonempty is (14). If $\lambda/\mu \le \theta$, it holds for every $p \in [0,1]$. If $\lambda/\mu > \theta$, its left-hand side decreases with p because $\theta > 1$. At $p = 1$, the condition becomes $\lambda/\mu - \theta < \theta$, which follows from the global stability requirement $\lambda/\mu < 2\theta$. Hence the assignment probabilities supporting a symmetric equilibrium form a connected interval containing $p = 1$.

We next consider the upper boundary $q_s = 1$. Proposition 1(ii) requires (16) and (17). The stability condition continues to hold as p increases because $p + (1-p)\theta$ decreases with p. The Wardrop condition also becomes easier to satisfy because $d[p\delta/c - W_1(1,p)]/dp = \delta/c - \partial W_1(1,p)/\partial p > 0$. Thus, once $q_s = 1$ is an equilibrium, it remains an equilibrium at every higher assignment probability. Because (16) implies (14), this boundary does not create a separate policy interval. On the $q_s = 1$ range, $SPO_s(p) = W_1(1,p)$ and $SPO_s'(p) = \partial W_1(1,p)/\partial p < 0$, so system-wide waiting decreases with targetability.

Queue-stability boundaries. Along an interior symmetric equilibrium, system-wide waiting is given by (20). If the Court 1 stability denominator approaches zero, then $q_s(p) \to \theta/\{(\lambda/\mu)[p+(1-p)\theta]\} > 0$ and $W_1(q_s(p),p) \to \infty$. Hence $q_s(p)W_1(q_s(p),p) \to \infty$.

If instead the Court 2 stability denominator approaches zero, then $1 - q_s(p) \to \theta/(\lambda/\mu) > 0$ and $W_2(q_s(p)) \to \infty$. Hence $(1-q_s(p))W_2(q_s(p)) \to \infty$.

Therefore $SPO_s(p) \to \infty$ whenever an interior symmetric equilibrium approaches a queue-stability boundary. A stability-induced open endpoint cannot minimize system-wide waiting.

Interior policy range. Fix $\delta/c > 0$ and consider $p > 0$ with $0 < q_s(p) < 1$. Differentiating the symmetric Wardrop condition (10) gives

$$\frac{dq_s}{dp} = \frac{\delta/c - \partial W_1/\partial p}{\partial W_1/\partial q_s - \partial W_2/\partial q_s} > 0. \tag{A.15}$$

Using $SPO_s(p) = W_2(q_s(p)) + pq_s(p)\delta/c$ and substituting (A.15) gives

$$SPO_s'(p) = \frac{1}{\partial W_1/\partial q_s - \partial W_2/\partial q_s}\left\{p\left(\frac{\delta}{c}\right)^2 + \left[q_s(p)\frac{\partial W_1}{\partial q_s} + (1-q_s(p))\frac{\partial W_2}{\partial q_s} - p\frac{\partial W_1}{\partial p}\right]\frac{\delta}{c} - \frac{\partial W_1}{\partial p}\frac{\partial W_2}{\partial q_s}\right\}. \tag{A.16}$$

All derivatives are evaluated at $(q_s(p), p)$. The denominator is strictly positive, so $SPO_s'(p) = 0$ is equivalent to (23).

For $p > 0$, define

$$\begin{aligned} A(q,p) &:= q\frac{\partial W_1(q,p)}{\partial q} + (1-q)\frac{\partial W_2(q)}{\partial q} - p\frac{\partial W_1(q,p)}{\partial p}, \\ \tau(q,p) &:= \frac{-A(q,p) + \sqrt{A(q,p)^2 + 4p\frac{\partial W_1(q,p)}{\partial p}\frac{\partial W_2(q)}{\partial q}}}{2p}. \end{aligned} \tag{A.17}$$

Because $\partial W_1/\partial p < 0$ and $\partial W_2/\partial q < 0$, $\tau(q,p) > 0$ is the unique positive root of (23). The other root is

$$\tau_-(q,p) = \frac{-A(q,p) - \sqrt{A(q,p)^2 + 4p(\partial W_1(q,p)/\partial p)(\partial W_2(q)/\partial q)}}{2p} < 0.$$

The positive root is $\tau(q,p)$, defined in (A.17). Hence (A.16) can be written as

$$SPO_s'(p) = \frac{p}{\partial W_1/\partial q_s - \partial W_2/\partial q_s}\left[\frac{\delta}{c} - \tau_-(q_s(p),p)\right]\left[\frac{\delta}{c} - \tau(q_s(p),p)\right].$$

Because the first two factors are positive,

$$\operatorname{sign}\{SPO_s'(p)\} = \operatorname{sign}\left[\delta/c - \tau(q_s(p),p)\right]. \tag{A.18}$$

This proves the derivative-sign characterization in the lemma.

It remains to show that every interior stationary point is a strict minimum. Let $\widehat{p} \in (0,1)$ satisfy $SPO_s'(\widehat{p}) = 0$, with $0 < q_s(\widehat{p}) < 1$. All filing shares and waiting-time derivatives below are evaluated at $(q_s(\widehat{p}), \widehat{p})$.

Differentiating $SPO_s(p) = W_2(q_s(p)) + pq_s(p)\delta/c$ gives $-\partial W_2/\partial q_s - \widehat{p}(\delta/c) = q_s(\delta/c)/q_s'(\widehat{p}) > 0$ at the stationary point. The stationary condition also yields

$$\begin{aligned} &q_s\left[\frac{\partial W_1}{\partial q_s}\frac{\delta}{c} - \left(-\frac{\partial W_2}{\partial q_s}\right)\left(-\frac{\partial W_1}{\partial p}\right)\right] \\ &\quad = \left[\frac{\delta}{c} - \frac{\partial W_1}{\partial p}\right]\left[(1-q_s)\left(-\frac{\partial W_2}{\partial q_s}\right) - \widehat{p}\frac{\delta}{c}\right]. \end{aligned}$$

Suppose first that the last bracket is nonnegative. Then

$$\frac{1}{q_s'(\widehat{p})} = \frac{\partial W_1/\partial q_s - \partial W_2/\partial q_s}{\delta/c - \partial W_1/\partial p} \le \frac{\partial W_1/\partial q_s}{-\partial W_1/\partial p} < \frac{\theta}{(\lambda/\mu)q_s^2(\theta-1)}.$$

The final inequality follows from the exact identity

$$\begin{aligned} &\frac{\theta}{(\lambda/\mu)q_s^2(\theta-1)} - \frac{\partial W_1/\partial q_s}{-\partial W_1/\partial p} \\ &\quad = \frac{(\theta+1)\{\theta - (\lambda/\mu)q_s[\widehat{p} + (1-\widehat{p})\theta]\}}{(\lambda/\mu)q_s^2(\theta-1)[\theta + 1 - (\lambda/\mu)q_s]} > 0. \end{aligned}$$

Twice differentiating the Wardrop equation and the objective, and imposing the stationary condition, gives

$$\begin{aligned} SPO_s''(\widehat{p}) = \frac{[q_s'(\widehat{p})]^2}{\delta/c - \partial W_1/\partial p}\Bigg\{ & q_s\frac{\delta}{c}\frac{\partial^2 W_1}{\partial p^2}\left[\frac{1}{q_s'(\widehat{p})} - \frac{\partial W_1/\partial q_s}{-\partial W_1/\partial p}\right] \\ & \times\left[\frac{1}{q_s'(\widehat{p})} - \frac{\theta}{(\lambda/\mu)q_s^2(\theta-1)}\right] \\ & + 2\frac{\delta}{c}\left(-\frac{\partial W_2}{\partial q_s}\right) + \frac{\partial^2 W_2}{\partial q_s^2}\left[(1-q_s)\frac{\delta}{c} - \frac{\partial W_1}{\partial p}\right]\Bigg\}. \end{aligned} \tag{A.19}$$

Here

$$\frac{\partial^2 W_1}{\partial p^2} = \frac{2(\lambda/\mu)q_s(\theta-1)(-\partial W_1/\partial p)}{\theta - (\lambda/\mu)q_s[\widehat{p} + (1-\widehat{p})\theta]} > 0,$$

and

$$\frac{\partial^2 W_2}{\partial q_s^2} = \frac{2(\lambda/\mu)(-\partial W_2/\partial q_s)}{\theta - (\lambda/\mu)(1-q_s)} > 0.$$

Both factors in the product in (A.19) are nonpositive, and the remaining terms are strictly positive. Thus $SPO_s''(\widehat{p}) > 0$.

Now suppose $(1-q_s)(-\partial W_2/\partial q_s) - \widehat{p}(\delta/c) < 0$. The stationary identity and $[\widehat{p} + (1-\widehat{p})\theta]^2 \le \widehat{p} + (1-\widehat{p})\theta^2$ imply $\frac{[\widehat{p}+(1-\widehat{p})\theta][\theta-(\lambda/\mu)(1-q_s)]}{\theta-(\lambda/\mu)q_s[\widehat{p}+(1-\widehat{p})\theta]} < \sqrt{(-\partial W_1/\partial p)/(\delta/c)}$. Combining this bound with the stationary identity and $\delta/c - \partial W_1/\partial p \ge 2\sqrt{(\delta/c)(-\partial W_1/\partial p)}$ gives

$$\begin{aligned} &- \frac{[\widehat{p} + (1-\widehat{p})\theta](\delta/c)}{\theta - (\lambda/\mu)q_s[\widehat{p} + (1-\widehat{p})\theta]} \left[(1-q_s)\left(-\frac{\partial W_2}{\partial q_s}\right) - \widehat{p}\frac{\delta}{c}\right] \\ &\quad < \frac{q_s(-\partial W_2/\partial q_s)(-\partial W_1/\partial p)}{2[\theta - (\lambda/\mu)(1-q_s)]} \\ &\quad < \frac{(-\partial W_2/\partial q_s)[(1-q_s)(\delta/c) - \partial W_1/\partial p]}{\theta - (\lambda/\mu)(1-q_s)}. \end{aligned}$$

An equivalent expression for the second derivative is

$$\begin{aligned} SPO_s''(\widehat{p}) = &\frac{2(\delta/c)^2}{\partial W_1/\partial q_s - \partial W_2/\partial q_s} \\ &+ \frac{2(\lambda/\mu)q_s(\delta/c)}{(\partial W_1/\partial q_s - \partial W_2/\partial q_s)[-\partial W_2/\partial q_s - \widehat{p}(\delta/c)]} \\ &\times \left\{ \frac{[(1-q_s)(-\partial W_2/\partial q_s) - \widehat{p}(\delta/c)][\theta(\delta/c) + (\theta-1)\partial W_2/\partial q_s]}{\theta - (\lambda/\mu)q_s[\widehat{p} + (1-\widehat{p})\theta]} \right. \\ &\left. + \frac{(-\partial W_2/\partial q_s)[(1-q_s)(\delta/c) - \partial W_1/\partial p]}{\theta - (\lambda/\mu)(1-q_s)} \right\}. \end{aligned} \tag{A.20}$$

Moreover, $-\partial W_2/\partial q_s - \widehat{p}(\delta/c) > 0$ implies $\theta(\delta/c) + (\theta-1)\partial W_2/\partial q_s < [\widehat{p} + (1-\widehat{p})\theta](\delta/c)$. Since the bracket multiplying this expression in (A.20) is negative, the preceding bounds show that the term in braces is strictly positive. Therefore $SPO_s''(\widehat{p}) > 0$ in this case as well. Every interior stationary point is therefore a strict local minimum.

There can be at most one such point. If two stationary points $p_1 < p_2$ existed, the derivative would be positive immediately to the right of p_1 and negative immediately to the left of p_2, requiring a local maximum between them. This contradicts the positive second derivative at every stationary point. Hence $SPO_s(p)$ has at most one interior stationary point. If one exists, the objective decreases before it and increases after it; otherwise it is monotone. Together with the connected policy range, upper-boundary behavior, and queue-stability argument, this proves Lemma 2.

A.9. Proof of Theorem 1

A.9.1. (i) Minimum targetability At $p = 0$, $W_1(q_s, 0) = (\lambda/\mu)q_s/\{\mu[1 - (\lambda/\mu)q_s]\}$. The symmetric Wardrop condition (10) gives $W_1(q_s(0), 0) = W_2(q_s(0))$. Substituting the waiting-time expressions and multiplying by the positive stability denominators yields $(1+\theta^2)q_s(0) - 1 = (\theta-1)(\lambda/\mu)q_s(0)[1 - q_s(0)] > 0$. Hence $q_s(0) > 1/(1+\theta^2)$.

Using the same Wardrop equality gives $q_s(0)(\partial W_1/\partial q_s) + (1 - q_s(0))(\partial W_2/\partial q_s) = [q_s(0)/(1 - q_s(0))](\partial W_1/\partial q_s)[1 - (1+\theta^2)q_s(0)] < 0$, where all derivatives are evaluated at $(q_s(0), 0)$. The sign follows from $q_s(0) > 1/(1+\theta^2)$ and $\partial W_1/\partial q_s > 0$.

Equation (A.15) also gives $q_s'(0) > 0$, because $\partial W_1/\partial p < 0$. When differentiating $SPO_s(p) = q_s(p)W_1(q_s(p), p) + (1 - q_s(p))W_2(q_s(p))$, the term $q_s'(0)[W_1 - W_2]$ vanishes by the Wardrop equality. The remaining terms are $q_s(0)(\partial W_1/\partial p) < 0$ and $q_s'(0)[q_s(0)(\partial W_1/\partial q_s) + (1 - q_s(0))(\partial W_2/\partial q_s)] < 0$. Therefore $SPO_s'(0^+) < 0$. A small increase in targetability from $p = 0$ reduces system-wide waiting, so minimum targetability cannot be optimal when $\theta > 1$.

A.9.2. (ii) Full versus partial targetability Suppose first that $q_s(1) < 1$. Since policies supporting $q_s = 1$ form an upper interval, the symmetric equilibrium is interior over the entire feasible policy range. By Lemma 2, the objective has at most one stationary point, and any stationary point is a strict minimum.

Full targetability is optimal if and only if

$$SPO_s'(1^-) \le 0. \tag{A.21}$$

By (A.18), this is equivalent to $\delta/c \le \tau(q_s(1), 1)$.

If the inequality is strict, the shape result implies that the objective cannot have turned upward before $p = 1$. If equality holds, the stationary-point proof in Appendix A.8 also applies at $p = 1$: $0 < q_s(1) < 1$, the stability denominators are positive, and no step divides by $1 - p$. Thus $SPO_s''(1^-) > 0$, so $SPO_s'(p) < 0$ immediately to the left of one. An earlier stationary point would make the derivative positive thereafter, contradicting this sign. Hence $p = 1$ is again the global minimizer.

If $SPO_s'(1^-) > 0$, full targetability is not optimal. Part (i) rules out $p = 0$ when feasible, while Appendix A.8 rules out a stability-induced open endpoint. The unique optimum is therefore interior: $0 < p^* < 1$, with (24) determining p^*.

Now suppose $q_s(1) = 1$. Stability at $p = 1$, (16), implies $\lambda/\mu < \theta$, so $p = 0$ is feasible. Direct substitution gives $W_1(1/[\theta(\theta - \lambda/\mu) + \lambda/\mu], 0) = W_1(1, 1)$ and $W_2(1/[\theta(\theta - \lambda/\mu) + \lambda/\mu]) < W_1(1, 1)$. Since $W_1(q_s, 0) - W_2(q_s)$ is strictly increasing, (10) implies $q_s(0) < 1/[\theta(\theta - \lambda/\mu) + \lambda/\mu]$. Hence $SPO_s(0) < SPO_s(1)$.

Moreover, Lemma 2 shows that $SPO_s(p)$ decreases over the range supporting $q_s = 1$, so every policy on that range has system-wide waiting at least $SPO_s(1)$. None can be globally optimal. Part (i) excludes $p = 0$, and a stability-induced open endpoint cannot minimize the objective. The unique optimum therefore lies on the interior-equilibrium range and satisfies $0 < p^* < 1$, $0 < q_s(p^*) < 1$, and (24).

A.9.3. (iii) Relative shopping benefit

Small relative shopping benefit. At $\delta/c = 0$, full targetability assigns every case to its specialist. At $p = 1$, (10) gives $W_1(q_s, 1) = W_2(q_s)$. Since the two fully specialized courts then have the same waiting-time function, $q_s(1) = 1/2$. Evaluating (A.16) at this point gives $SPO_s'(1^-) = -(\partial W_1/\partial p)(\partial W_2/\partial q_s)/(\partial W_1/\partial q_s - \partial W_2/\partial q_s) < 0$. By continuity, the same inequality holds for all sufficiently small positive values of δ/c. Hence (A.21) gives $p^* = 1$, and therefore $p^* \to 1$ as $\delta/c \downarrow 0$.

Large relative shopping benefit. Fix $\theta > 1$ and suppose $p = 0$ is feasible. Because the shopping incentive vanishes at $p = 0$, $q_s(0) > 0$ and $SPO_s(0)$ do not depend on δ/c.

Fix $\varepsilon \in (0, 1)$ and consider any feasible $p \ge \varepsilon$. If the equilibrium is interior, the Wardrop condition gives $SPO_s(p) = W_2(q_s(p)) + pq_s(p)(\delta/c)$. Since $q_s(p) \ge q_s(0)$, $SPO_s(p) \ge \varepsilon q_s(0)(\delta/c)$. This lower bound exceeds $SPO_s(0)$ for all sufficiently large δ/c, uniformly over these interior-equilibrium policies.

If instead $q_s(p) = 1$, queue stability requires $\lambda/\mu < \theta$. Under this condition, the comparison in part (ii) gives $SPO_s(0) < W_1(1, 1)$. Since $\partial W_1(1, p)/\partial p < 0$, $SPO_s(p) = W_1(1, p) \ge W_1(1, 1) > SPO_s(0)$. If $\lambda/\mu \ge \theta$, no feasible policy supports $q_s(p) = 1$, so only the interior case applies.

Thus every feasible $p \ge \varepsilon$ eventually produces more system-wide waiting than $p = 0$. Any minimizing assignment probability therefore satisfies $p^* < \varepsilon$. Since ε is arbitrary, $p^* \to 0$ as $\delta/c \to \infty$.

A.9.4. (iv) Baseline load Fix $\theta > 1$ and finite δ/c. For $\theta < \lambda/\mu < 2\theta$, full filing in Court 1 violates queue stability, so the equilibrium at $p = 1$ is interior. The Wardrop condition (10) becomes $\delta/c = (\lambda/\mu)(2q_s(1) - 1)/\{\mu[\theta - (\lambda/\mu)q_s(1)][\theta - (\lambda/\mu)(1 - q_s(1))]\}$. The stability denominators are positive, so $q_s(1) > 1/2$. The Court 1 upper bound in (13) gives $q_s(1) < \theta/(\lambda/\mu)$. Therefore, as $\lambda/\mu \uparrow 2\theta$,

$$q_s(1) \downarrow 1/2. \tag{A.22}$$

Using (10), $SPO_s(p) = W_2(q_s(p)) + pq_s(p)\delta/c$. Hence

$$SPO_s'(1^-) = q_s'(1)\left[\frac{\delta}{c} - \frac{\lambda/\mu}{\mu[\theta - (\lambda/\mu)(1 - q_s(1))]^2}\right] + \frac{q_s(1)\delta}{c}. \tag{A.23}$$

Equation (A.6) evaluated at $p = 1$ gives

$$q_s'(1) = \frac{\dfrac{\delta}{c} + \dfrac{(\lambda/\mu)q_s(1)(\theta - 1)[\theta + 1 - (\lambda/\mu)q_s(1)]}{\mu[\theta - (\lambda/\mu)q_s(1)]^2}}{\dfrac{\lambda/\mu}{\mu[\theta - (\lambda/\mu)q_s(1)]^2} + \dfrac{\lambda/\mu}{\mu[\theta - (\lambda/\mu)(1 - q_s(1))]^2}}. \tag{A.24}$$

The Wardrop condition also implies $[\theta - (\lambda/\mu)q_s(1)]/[\theta - (\lambda/\mu)(1 - q_s(1))] = 1 - \mu(\delta/c)[\theta - (\lambda/\mu)q_s(1)]$. Both stability denominators approach zero, so their ratio approaches one. Using this ratio and (A.22) in (A.24) gives $q_s'(1) \to (\theta - 1)/4 > 0$. At the same time, $(\lambda/\mu)/\{\mu[\theta - (\lambda/\mu)(1 - q_s(1))]^2\} \to \infty$. Hence (A.23) gives $SPO_s'(1^-) \to -\infty$ as $\lambda/\mu \uparrow 2\theta$.

Therefore $SPO_s'(1^-) \leq 0$ for all sufficiently high loads, and (A.21) implies $p^* = 1$. This result is not caused by lower-targetability policies becoming infeasible. For every finite $\lambda/\mu < 2\theta$, the capacity condition is strict at $p = 1$, so policies strictly below one remain feasible. For sufficiently high finite loads, $SPO_s'(1^-) < 0$, so full targetability strictly outperforms nearby feasible lower-targetability policies.

A.10. Proof of Proposition 6

Fix $\theta > 1$, and consider the partial-targetability optimum p^* characterized in Theorem 1. Along the symmetric equilibrium, the left-hand side of (23) equals

$$\left(\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}\right) SPO_s'(p).$$

At $p = p^*$, $SPO_s'(p^*) = 0$. Differentiating this identity along the symmetric equilibrium therefore shows that the derivative of the left-hand side of (23) with respect to p equals

$$\left(\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}\right) SPO_s''(p^*) > 0.$$

We use this positive derivative to establish the two comparative statics.

A.10.1. (i) Relative shopping benefit Throughout this part, all waiting-time derivatives are evaluated at $(q_s(p^*), p^*)$. Holding q_s and p^* fixed, the derivative of the left-hand side of (23) with respect to δ/c is $2p^*(\delta/c) + q_s\partial W_1/\partial q_s + (1 - q_s)\partial W_2/\partial q_s - p^*\partial W_1/\partial p$. Using the stationary condition to replace the last three terms reduces this derivative to

$$p^*\frac{\delta}{c} + \frac{(\partial W_1/\partial p)(\partial W_2/\partial q_s)}{\delta/c} > 0. \tag{A.25}$$

The expression is positive because $p^* > 0$, $\delta/c > 0$, $\partial W_1/\partial p < 0$, and $\partial W_2/\partial q_s < 0$.

We next show that the left-hand side of (23) has a strictly positive partial derivative with respect to q_s at the optimum. All filing shares and waiting-time derivatives below are evaluated at $(q_s(p^*), p^*)$. The stationary condition gives $-\partial W_2/\partial q_s - p^*(\delta/c) = q_s(\delta/c)/q_s'(p^*) > 0$.

Differentiating the left-hand side of (23) with respect to q_s, and using the stationary identities in Appendix A.8, gives

$$\begin{aligned}
&\frac{(\delta/c)[-\partial W_2/\partial q_s - p^*(\delta/c)]}{q_s} \\
&\quad + \frac{(\lambda/\mu)[-\partial W_2/\partial q_s - p^*(\delta/c)](-\partial W_1/\partial p)}{\theta + 1 - (\lambda/\mu)q_s} \\
&\quad + \frac{2(\lambda/\mu)[p^* + (1-p^*)\theta](\delta/c)}{\theta - (\lambda/\mu)q_s[p^* + (1-p^*)\theta]}\left[(1-q_s)\left(-\frac{\partial W_2}{\partial q_s}\right) - p^*\frac{\delta}{c}\right] \\
&\quad + \frac{\partial^2 W_2}{\partial q_s^2}\left[(1-q_s)\frac{\delta}{c} - \frac{\partial W_1}{\partial p}\right].
\end{aligned}$$

The first, second, and fourth terms are strictly positive. If the bracket in the third term is nonnegative, the result follows.

If that bracket is negative, the bound established in Appendix A.8, with $\widehat{p} = p^*$, makes the magnitude of the third term strictly smaller than $(\lambda/\mu)q_s(-\partial W_2/\partial q_s)(-\partial W_1/\partial p)/[\theta - (\lambda/\mu)(1-q_s)]$. Meanwhile, $\partial^2 W_2/\partial q_s^2 = 2(\lambda/\mu)(-\partial W_2/\partial q_s)/[\theta - (\lambda/\mu)(1-q_s)]$. Since $0 < q_s < 1$ and $(1-q_s)(\delta/c) - \partial W_1/\partial p > -\partial W_1/\partial p > 0$, the fourth term strictly exceeds this bound. Thus the partial derivative with respect to q_s is strictly positive in both cases.

Now hold $p = p^*$ fixed and increase δ/c along the symmetric equilibrium. Equation (A.5) shows that q_s increases. The left-hand side of (23) therefore increases both directly, by (A.25), and through its strictly positive derivative with respect to q_s. Because its derivative with respect to p along the symmetric equilibrium is strictly positive, restoring the stationary condition requires p^* to decrease. Hence $\partial p^*/\partial(\delta/c) < 0$.

A.10.2. (ii) Baseline load Consider an interior optimum. Hold $p = p^*$, δ/c, θ, and μ fixed while varying λ/μ. All filing shares and waiting-time derivatives below are evaluated at $(q_s(p^*), p^*)$. The filing response is given by (A.13). The stationary condition implies $-\partial W_2/\partial q_s - p^*(\delta/c) > 0$ and $q_s(\partial W_1/\partial q_s)/(-\partial W_2/\partial q_s) < (1-q_s) + (-\partial W_1/\partial p)/(\delta/c)$.

Consider the ratio of the Court 2 stability denominator to the Court 1 stability denominator, multiplied by $p^* + (1-p^*)\theta$. If this ratio is at least one, it is no greater than its square. The inequality $[p^* + (1-p^*)\theta]^2 \le p^* + (1-p^*)\theta^2$ bounds that square above by $(\partial W_1/\partial q_s)/(-\partial W_2/\partial q_s)$. Multiplying by q_s and using the stationary bound therefore makes the product strictly smaller than $(1-q_s) + (-\partial W_1/\partial p)/(\delta/c)$.

If the ratio is less than one, the Court 1 stability denominator exceeds the Court 2 stability denominator, because $p^* + (1-p^*)\theta \ge 1$. Comparing their definitions gives $[p^* + (1-p^*)\theta]q_s < 1 - q_s$, and hence $q_s < 1/2$. The ratio multiplied by q_s is then smaller than $q_s < 1 - q_s$. Thus the final bracket in the following expression is strictly positive in either case.

Differentiating the left-hand side of (23) with respect to λ/μ along the symmetric equilibrium gives

$$\begin{aligned}
&\frac{\partial W_2/\partial q_s}{(\lambda/\mu)(\partial W_1/\partial q_s - \partial W_2/\partial q_s)}\left\{\left[\frac{\partial W_1}{\partial q_s} - \frac{\partial W_2}{\partial q_s}\right]\left(-\frac{\partial W_1}{\partial p}\right)\right. \\
&\quad + \left[-\frac{\partial W_2}{\partial q_s} - p^*\frac{\delta}{c}\right]\left(-\frac{\partial^2 W_1}{\partial q_s \partial p}\right) \\
&\quad \left. + 2\frac{\lambda}{\mu}\frac{\partial W_1}{\partial q_s}\left[\frac{(1-q_s)(\delta/c) - \partial W_1/\partial p}{\theta - (\lambda/\mu)(1-q_s)} - \frac{[p^* + (1-p^*)\theta]q_s(\delta/c)}{\theta - (\lambda/\mu)q_s[p^* + (1-p^*)\theta]}\right]\right\}.
\end{aligned} \tag{A.26}$$

The explicit waiting-time expression gives $-\partial^2 W_1/(\partial q_s \partial p) > 0$. Its denominator is μ times the cube of the positive Court 1 stability denominator. Its numerator is $(\lambda/\mu)(\theta - 1)$ times the sum of $\theta + 1$ multiplied by that stability denominator and $2(\lambda/\mu)q_s[p^* + (1 - p^*)\theta^2]$. Every term in braces in (A.26) is therefore strictly positive, whereas the leading factor is negative. Thus the total derivative with respect to λ/μ is strictly negative.

The total derivative of the same stationary expression with respect to p, along the symmetric equilibrium at the optimum, is $(\partial W_1/\partial q_s - \partial W_2/\partial q_s)SPO_s''(p^*) > 0$. Implicit differentiation therefore gives $\partial p^*/\partial(\lambda/\mu) > 0$, proving Proposition 6.

A.11. Proof of Corollary 1

Aggregate litigant utility is $\lambda \sum_{i=1}^{2}[q_i U_{1i}(\mathbf{q}; p) + (1 - q_i)U_{2i}(q_i)]$.

Consider first an interior symmetric equilibrium $q_1 = q_2 = q_s(p)$. By (9), both courts give each case type the same equilibrium utility: $U_{1i}(\mathbf{q}; p) = U_{2i}(q_s(p)) = V - cW_2(q_s(p))$. Aggregate litigant utility is therefore $2\lambda[V - cW_2(q_s(p))]$. Proposition 2 gives $dW_2(q_s(p))/dp < 0$, so litigant utility strictly increases with targetability along every interior symmetric-equilibrium segment.

If instead $q_s = 1$, aggregate litigant utility is $2\lambda[V + p\delta - cW_1(1, p)]$. Its derivative is $2\lambda[\delta - c\,\partial W_1(1, p)/\partial p] > 0$, because $\delta > 0$ and $\partial W_1(1, p)/\partial p \le 0$.

The two expressions agree at any transition between the interior and $q_s = 1$ regimes. At such a point, $W_2(1) = 0$, and the Wardrop condition holds with equality, so $p\delta = cW_1(1, p)$. Hence aggregate litigant utility equals $2\lambda V$ from either side. It is therefore continuous across the transition. Since $q_s = 0$ is never a symmetric equilibrium, these two cases cover the entire symmetric-equilibrium path. Thus aggregate litigant utility is maximized at full targetability, $p = 1$.

Finally, Propositions 7 and 8 show that every asymmetric equilibrium requires $p < 1$. Hence full targetability supports no asymmetric equilibrium. Proposition 1 gives the unique symmetric equilibrium at $p = 1$, so the litigant-preferred assignment rule has a unique Wardrop equilibrium.

A.12. Proof of Corollary 2

Let $T(p) = q_s(p)p\delta/c$, so $SPO_s^{\gamma}(p) = SPO_s(p) + \gamma T(p)$. Changing γ does not change the feasible policy set or equilibrium filing for a given p.

We first show that successful targeting increases with targetability. Along an interior symmetric equilibrium, Proposition 2 gives $dq_s/dp > 0$, so $T'(p) = (\delta/c)[q_s(p) + p\,dq_s/dp] > 0$. If $q_s(p) = 1$, then $T(p) = p\delta/c$, which is also strictly increasing. Thus $T(p)$ is strictly increasing over the entire feasible policy range.

Now take $\gamma_2 > \gamma_1$. Optimality of $p^*_{\gamma_1}$ under γ_1 and $p^*_{\gamma_2}$ under γ_2 gives

$$SPO_s(p^*_{\gamma_1}) + \gamma_1 T(p^*_{\gamma_1}) \le SPO_s(p^*_{\gamma_2}) + \gamma_1 T(p^*_{\gamma_2}),$$
$$SPO_s(p^*_{\gamma_2}) + \gamma_2 T(p^*_{\gamma_2}) \le SPO_s(p^*_{\gamma_1}) + \gamma_2 T(p^*_{\gamma_1}).$$

Adding the two inequalities gives $(\gamma_2 - \gamma_1)[T(p^*_{\gamma_2}) - T(p^*_{\gamma_1})] \le 0$. Since T is strictly increasing, $p^*_{\gamma_2} \le p^*_{\gamma_1}$. A stronger institutional concern therefore weakly lowers optimal targetability.

Suppose now that $p = 0$ is feasible. By Lemma 2, the feasible policy range is connected and contains $p = 1$; since it also contains $p = 0$, it is $[0, 1]$. Moreover, $T(0) = 0 < T(p)$ for every $p > 0$.

For a given $p > 0$, policy $p = 0$ weakly outperforms p under the augmented objective exactly when $\gamma \geq [SPO_s(0) - SPO_s(p)]/T(p)$. The smallest weight that satisfies this condition for every $p > 0$ is

$$\bar{\gamma} = \sup_{0<p\leq 1} \frac{SPO_s(0) - SPO_s(p)}{T(p)}.$$

This threshold is finite. For every $\varepsilon > 0$, the ratio is continuous and bounded on $[\varepsilon, 1]$. As $p \downarrow 0$, it converges to $-SPO_s'(0^+)/[q_s(0)\delta/c]$, which is finite.

Hence every $\gamma \geq \bar{\gamma}$ makes $p = 0$ optimal, and every $\gamma > \bar{\gamma}$ makes it uniquely optimal. Moreover, Theorem 1 gives $SPO_s'(0^+) < 0$, so the limiting ratio above is positive and therefore $\bar{\gamma} > 0$.

If $p = 0$ is infeasible, changing γ does not change the feasible policy set, so minimum targetability cannot be optimal for any γ.

A.13. Proof of Lemma 3

The difference equation (28) separates the symmetric and asymmetric branches of an interior equilibrium. Define

$$B(q_1, q_2, p) = \frac{(2p-1)K(q_1, q_2, p)}{\mu^2\theta^2(1-\rho_{11})(1-\rho_{12})} + \frac{1}{(\theta\mu - \lambda + \lambda q_1)(\theta\mu - \lambda + \lambda q_2)},$$

where $K(q_1, q_2, p) = p(\theta^2 + 1) - \theta^2 + (\lambda/\mu)p(1-p)(\theta - 1)(q_1 + q_2)$.

Because every Wardrop equilibrium is admissible, all queue-stability denominators below are strictly positive. Substituting (1)–(4) gives

$$W_{11}(\mathbf{q}, p) - W_{12}(\mathbf{q}, p) = \frac{\lambda(q_1 - q_2)K(q_1, q_2, p)}{\mu^2\theta^2(1-\rho_{11})(1-\rho_{12})},$$
$$W_{21}(q_1) - W_{22}(q_2) = -\frac{\lambda(q_1 - q_2)}{(\theta\mu - \lambda + \lambda q_1)(\theta\mu - \lambda + \lambda q_2)}.$$

Thus both waiting-time differences are proportional to $q_1 - q_2$. Substituting these expressions into (28) and dividing by $\lambda > 0$ gives $(q_1 - q_2)B(q_1, q_2, p) = 0$. Hence every interior equilibrium lies on one of two branches: $q_1 = q_2$, or $q_1 \neq q_2$ with $B(q_1, q_2, p) = 0$, as stated in Lemma 3.

A.14. Proof of Corollary 3

Let (q_h, q_l), with $0 \leq q_l < q_h \leq 1$, be an asymmetric Wardrop equilibrium. Because the equilibrium is admissible, all four judge-level stability conditions in (6) hold.

Adding the two Court 1 stability conditions and using (5) gives $(q_h + q_l)/2 < \theta/\{(\lambda/\mu)[p + (1-p)\theta]\}$. Adding the two Court 2 stability conditions gives $(q_h + q_l)/2 > 1 - \theta/(\lambda/\mu)$. Since $0 \leq q_l < q_h \leq 1$, we also have $0 < (q_h + q_l)/2 < 1$. Therefore, $(q_h + q_l)/2$ satisfies the symmetric admissibility interval (13).

The midpoint need not equal the symmetric equilibrium filing share q_s; its role is only to show that the symmetric admissibility interval is nonempty at the same assignment probability p. Proposition 1 then implies that this p supports a unique symmetric Wardrop equilibrium, either with $0 < q_s < 1$ or with $q_s = 1$. Hence every asymmetric Wardrop equilibrium coexists with a unique symmetric Wardrop equilibrium.

A.15. Proof of Proposition 7

An asymmetric interior equilibrium can be written, without loss of generality, as $(q_1, q_2) = (q_h, q_l)$, where $0 < q_l < q_h < 1$.

We first rule out asymmetry when there is no specialization advantage. If $\theta = 1$, the function in Lemma 3 satisfies $K(q_h, q_l, p) = 2p - 1$, so

$$B(q_h, q_l, p) = \frac{(2p-1)^2}{\mu^2(1-\rho_{11})(1-\rho_{12})} + \frac{1}{(\mu - \lambda + \lambda q_h)(\mu - \lambda + \lambda q_l)} > 0.$$

All denominators are positive by admissibility. Thus $B(q_h, q_l, p) = 0$ is impossible, and Lemma 3 rules out every interior asymmetric equilibrium when $\theta = 1$. We therefore consider $\theta > 1$ for the remainder of the proof.

(i) Type symmetry. The waiting-time expressions (1)–(4) are symmetric across the two case types: $W_{11}((q_l, q_h), p) = W_{12}((q_h, q_l), p)$, $W_{12}((q_l, q_h), p) = W_{11}((q_h, q_l), p)$, $W_{21}(q_l) = W_{22}(q_l)$, and $W_{22}(q_h) = W_{21}(q_h)$. Substituting these identities into the utility differences gives $\Delta_1((q_l, q_h); p) = \Delta_2((q_h, q_l); p)$ and $\Delta_2((q_l, q_h); p) = \Delta_1((q_h, q_l); p)$. Therefore, if (q_h, q_l) satisfies the Wardrop conditions, then (q_l, q_h) satisfies the same conditions with the case-type labels reversed. The two profiles also have the same total Court 1 filing share and, by (19), the same system-wide waiting.

(ii) Targetability window. By Lemma 3, an asymmetric interior equilibrium requires $B(q_h, q_l, p) = 0$. All denominators in B are strictly positive, and its second term is strictly positive. Therefore, $(2p-1)K(q_h, q_l, p) < 0$.

We first show that $p \le 1/2$ is impossible. Adding the two Court 1 stability conditions in (6) gives $(\lambda/\mu)(q_h + q_l) < 2\theta/[p + (1-p)\theta]$. At $p = 0$, $K(q_h, q_l, 0) = -\theta^2 < 0$. For $0 < p \le 1/2$, substituting the stability bound into the definition of K gives $K(q_h, q_l, p) < p(\theta^2+1) - \theta^2 + 2\theta p(1-p)(\theta-1)/[p + (1-p)\theta]$. Since $p + (1-p)\theta > (1-p)\theta$, this is smaller than $p(\theta^2+1) - \theta^2 + 2p(\theta-1) = p(\theta^2 + 2\theta - 1) - \theta^2 \le -(\theta-1)^2/2 < 0$, where the final inequality uses $p \le 1/2$. Hence $K(q_h, q_l, p) < 0$ throughout $p \le 1/2$. But then $(2p-1)K(q_h, q_l, p) \ge 0$, contradicting the necessary condition above. Therefore every interior asymmetric equilibrium requires $p > 1/2$.

When $p > 1/2$, the condition $B(q_h, q_l, p) = 0$ requires $K(q_h, q_l, p) < 0$. Since $q_h + q_l > 0$, $K(q_h, q_l, p) > p(\theta^2 + 1) - \theta^2$. Therefore $K(q_h, q_l, p) < 0$ requires $p < \theta^2/(1+\theta^2)$. This gives the global targetability window reported in Proposition 7.

When baseline load exceeds the specialist processing rate, $\lambda/\mu > \theta$, Court 2 stability gives the tighter bounds $q_h > 1 - \theta/(\lambda/\mu)$ and $q_l > 1 - \theta/(\lambda/\mu)$. Hence $q_h + q_l > 2[1 - \theta/(\lambda/\mu)]$. Substituting this lower bound into K gives

$$K(q_h, q_l, p) > p(\theta^2+1) - \theta^2 + 2p(1-p)(\theta-1)\left(\frac{\lambda}{\mu} - \theta\right). \tag{A.27}$$

For $K(q_h, q_l, p) < 0$, the right-hand side of (A.27) must also be negative. Solving this quadratic inequality and combining it with $p > 1/2$ gives the necessary window

$$\frac{1}{2} < p < \begin{cases} \dfrac{\theta^2}{1+\theta^2}, & \dfrac{\lambda}{\mu} \le \theta, \\[2ex] \dfrac{\theta^2 + 1 + 2(\theta-1)(\lambda/\mu - \theta) - \sqrt{\left[\theta^2 + 1 + 2(\theta-1)(\lambda/\mu - \theta)\right]^2 - 8\theta^2(\theta-1)(\lambda/\mu - \theta)}}{4(\theta-1)(\lambda/\mu - \theta)}, & \dfrac{\lambda}{\mu} > \theta. \end{cases} \tag{A.28}$$

The load-dependent upper bound is stricter when $\lambda/\mu > \theta$, because above this bound the lower bound on K in (A.27) becomes nonnegative, which is incompatible with $K(q_h, q_l, p) < 0$. As $\lambda/\mu \downarrow \theta$, this upper bound converges to $\theta^2/(1+\theta^2)$. This proves the targetability window in (A.28).

(iii) Waiting-time ordering. Because $q_h > q_l$, Court 2 waiting is lower for the type with the higher Court 1 filing share: $W_{21}(q_h) < W_{22}(q_l)$. Part (ii) gives $p > 1/2$. Hence the difference equation (28) gives $(2p-1)[W_{11}((q_h,q_l),p) - W_{12}((q_h,q_l),p)] = W_{21}(q_h) - W_{22}(q_l) < 0$, so $W_{11}((q_h,q_l),p) < W_{12}((q_h,q_l),p)$.

(iv) Filing-share ordering. Let q_s denote the unique symmetric equilibrium supported at the same assignment probability p, whose existence follows from Corollary 3.

First consider the type-swapped asymmetric equilibrium (q_l, q_h). At the symmetric profile (q_l, q_l), both Court 1 waiting times are strictly lower than at (q_l, q_h), while the relevant Court 2 waiting time is unchanged. The Wardrop equalities (25)–(26) therefore imply $W_1(q_l, p) - W_2(q_l) < p\delta/c$.

Next consider (q_h, q_l). If the symmetric profile (q_h, q_h) is stable, both Court 1 waiting times are strictly higher than at (q_h, q_l), while the relevant Court 2 waiting time is unchanged. Therefore $W_1(q_h, p) - W_2(q_h) > p\delta/c$. Since $W_1(q_s, p) - W_2(q_s)$ is strictly increasing by (A.1), the symmetric Wardrop equality (10) gives $q_h > q_s > q_l$.

If the symmetric profile (q_h, q_h) is not stable, the symmetric waiting-time difference diverges before q_s reaches q_h. By (11)–(12) and (A.1), it is continuous and strictly increasing over the admissible interval. The unique symmetric equilibrium again satisfies $q_h > q_s > q_l$, which proves (30).

(v) Case-volume and utilization ordering. Under $(q_1, q_2) = (q_h, q_l)$, the effective arrival rates in Figure 1 satisfy $A_{11} - A_{12} = \lambda(2p-1)(q_h - q_l) > 0$, because $p > 1/2$ and $q_h > q_l$. Thus Court 1 Judge 1 receives more cases than Judge 2.

Using (5), $\rho_{11}((q_h,q_l),p) - \rho_{12}((q_h,q_l),p) = \lambda(q_h - q_l)[p - (1-p)\theta]/(\theta\mu)$. Because $\lambda > 0$, $\mu > 0$, $\theta > 1$, and $q_h - q_l > 0$, the utilization difference has the sign of $p - (1-p)\theta$. Hence $\rho_{11}((q_h,q_l),p) < \rho_{12}((q_h,q_l),p)$ if and only if $p < \theta/(1+\theta)$, with equal utilization at $p = \theta/(1+\theta)$ and the ordering reversed above this threshold. Part (iii) shows that $W_{11}((q_h,q_l),p) < W_{12}((q_h,q_l),p)$ throughout. Thus the Court 1 judge receiving more cases has shorter waiting both below and above the utilization threshold, as stated in Proposition 7.

A.16. Proof of Proposition 8

Swapping the two filing shares interchanges the two utility differences: $\Delta_1((q_2,q_1);p) = \Delta_2((q_1,q_2);p)$ and $\Delta_2((q_2,q_1);p) = \Delta_1((q_1,q_2);p)$. Thus every boundary asymmetric equilibrium has a type-swapped counterpart satisfying the corresponding Wardrop inequalities. By (19), the two profiles also generate the same system-wide waiting. It is therefore sufficient to analyze one profile from each pair.

A.16.1. Part (i): $q_i = 0$, $0 < q_{3-i} < 1$, $\Delta_i \le 0$, and $\Delta_{3-i} = 0$ Consider $(q_1, q_2) = (0, q_h)$, where $0 < q_h < 1$, $\Delta_1 \le 0$, and $\Delta_2 = 0$. By type symmetry, the profile $(q_h, 0)$ satisfies the same conditions with the case types reversed.

The two Court 1 waiting times are

$$W_{11}((0,q_h),p) = \frac{(\lambda/\mu)(1-p)\theta^2 q_h}{\theta\mu[\theta - (\lambda/\mu)(1-p)\theta q_h]}, \tag{A.29}$$

$$W_{12}((0,q_h),p) = \frac{(\lambda/\mu)pq_h}{\theta\mu[\theta - (\lambda/\mu)pq_h]}. \tag{A.30}$$

In Court 2, $W_{21}(0) = (\lambda/\mu)/\{\theta\mu[\theta - \lambda/\mu]\}$ and $W_{22}(q_h) = (\lambda/\mu)(1-q_h)/\{\theta\mu[\theta - (\lambda/\mu)(1-q_h)]\}$.

Substituting $(0, q_h)$ into (6) gives $(\lambda/\mu)(1-p)q_h < 1$ together with $\lambda/\mu < \theta$; under the latter condition, the other two stability conditions hold automatically. For $p < 1$, the admissible filing shares therefore satisfy $0 < q_h < \min\{1, 1/[(\lambda/\mu)(1-p)]\}$, with $\lambda/\mu < \theta$. The case $p = 1$ will be ruled out directly by the Wardrop conditions below.

Because type 1 files entirely in Court 2 while type 2 splits its filings, $p(\delta/c) \leq pW_{11}((0,q_h),p) + (1-p)W_{12}((0,q_h),p) - W_{21}(0)$. For the type that splits,

$$p\delta/c = (1-p)W_{11}((0,q_h),p) + pW_{12}((0,q_h),p) - W_{22}(q_h). \tag{A.31}$$

Using the equality in the inequality gives the equivalent boundary condition

$$(2p-1)[W_{11}((0,q_h),p) - W_{12}((0,q_h),p)] \geq W_{21}(0) - W_{22}(q_h). \tag{A.32}$$

Since $q_h > 0$ and Court 2 waiting decreases with the Court 1 filing share, $W_{21}(0) - W_{22}(q_h) > 0$. Hence the left-hand side of (A.32) must be strictly positive.

Using (A.29)–(A.30) gives

$$W_{12}((0,q_h),p) - W_{11}((0,q_h),p) = \frac{(\lambda/\mu)q_h[p-(1-p)\theta^2 + (\lambda/\mu)p(1-p)(\theta-1)q_h]}{\mu[\theta-(\lambda/\mu)pq_h][\theta-(\lambda/\mu)(1-p)\theta q_h]}. \tag{A.33}$$

We now obtain the targetability restriction. At $p = 1/2$, the left-hand side of (A.32) is zero, so the condition cannot hold. At $p = 0$, the bracket in (A.33) is $-\theta^2 < 0$. For $0 < p < 1/2$, stability gives $(\lambda/\mu)p(1-p)(\theta-1)q_h \leq p(\theta-1)$, so the bracket is no greater than $p-(1-p)\theta^2 + p(\theta-1) = \theta[p-(1-p)\theta] < 0$. Thus $W_{12}((0,q_h),p) < W_{11}((0,q_h),p)$ for $p < 1/2$. The left-hand side of (A.32) is then negative. Hence $p > 1/2$.

For $p > 1/2$, the same boundary condition requires $W_{11}((0,q_h),p) > W_{12}((0,q_h),p)$. By (A.33), this is equivalent to

$$p-(1-p)\theta^2 + (\lambda/\mu)p(1-p)(\theta-1)q_h < 0. \tag{A.34}$$

Because the final term is nonnegative, (A.34) requires $p < \theta^2/(1+\theta^2)$. In particular, $p = 1$ is impossible. Thus these profiles satisfy the global necessary window (31).

It remains to determine whether the splitting type has an admissible filing share. Over the stability interval, $\partial W_{11}/\partial q_h > 0$, $\partial W_{12}/\partial q_h > 0$, and $\partial W_{22}/\partial q_h < 0$. Hence $(1-p)W_{11}((0,q_h),p) + pW_{12}((0,q_h),p) - W_{22}(q_h)$ is continuous and strictly increasing in q_h. Therefore (A.31) has at most one admissible solution.

At $q_h \downarrow 0$, this expression converges to $-W_{21}(0) < 0$. If $(\lambda/\mu)(1-p) \geq 1$, the upper endpoint is the Court 1 stability boundary and $W_{11}((0,q_h),p) \to \infty$. Hence (A.31) has one admissible solution for every finite $\delta/c > 0$.

If $(\lambda/\mu)(1-p) < 1$, the upper endpoint is $q_h = 1$. An interior solution exists if and only if

$$p\delta/c < (1-p)W_{11}((0,1),p) + pW_{12}((0,1),p). \tag{A.35}$$

Equality in (A.35) gives the corner $(0,1)$, not a profile with $0 < q_h < 1$. Finally, the unique solution of (A.31) is a boundary Wardrop equilibrium if and only if it also satisfies $\Delta_1 \leq 0$, equivalently (A.32). By type symmetry, $(q_h, 0)$ has the same stability, existence, uniqueness, and targetability conditions.

A.16.2. Part (ii): $q_i = 1$, $0 < q_{3-i} < 1$, $\Delta_i \geq 0$, and $\Delta_{3-i} = 0$ Consider $(q_1,q_2) = (1,q_l)$, where $0 < q_l < 1$, $\Delta_1 \geq 0$, and $\Delta_2 = 0$. Type symmetry gives the counterpart $(q_l, 1)$.

Applying (6) gives $(\lambda/\mu)[p + (1-p)\theta q_l] < \theta$, $(\lambda/\mu)[(1-p)\theta + pq_l] < \theta$, and $(\lambda/\mu)(1-q_l) < \theta$. Thus

$$\max\{0, 1-\theta/(\lambda/\mu)\} < q_l < \min\left\{1, \frac{\theta/(\lambda/\mu)-p}{(1-p)\theta}, \frac{\theta/(\lambda/\mu)-(1-p)\theta}{p}\right\}. \tag{A.36}$$

When $\lambda/\mu > \theta$, the lower bound in (A.36) is positive. Court 1 stability implies $p < \theta/(\lambda/\mu)$ and $1 - p < 1/(\lambda/\mu)$. Adding these inequalities gives $\lambda/\mu < \theta + 1$, so $\theta^2 - (\lambda/\mu)(\theta - 1) \geq 1 > 0$. Requiring the Court 2 lower bound to lie below both Court 1 upper bounds therefore gives

$$\frac{\theta(\lambda/\mu - 1)}{(\lambda/\mu)(\theta - 1) + \theta} < p < \frac{\theta(\theta + 1 - \lambda/\mu)}{\theta^2 - (\lambda/\mu)(\theta - 1)}. \tag{A.37}$$

This interval is nonempty if and only if $\theta < \lambda/\mu < \theta + 1/2$. Hence these boundary profiles can remain stable above $\lambda/\mu = \theta$, but only over a load interval of width $1/2$. As $\lambda/\mu \downarrow \theta$, the probability bounds in (A.37) approach $1 - 1/\theta$ and 1. As $\lambda/\mu \uparrow \theta + 1/2$, both approach $p = \theta/(\theta + 1)$.

The high-share type files entirely in Court 1, so $W_{21}(1) = 0$, while the low-share type splits its filings. Their Wardrop conditions are

$$p\delta/c \geq pW_{11}((1, q_l), p) + (1 - p)W_{12}((1, q_l), p).$$

and

$$p\delta/c = (1 - p)W_{11}((1, q_l), p) + pW_{12}((1, q_l), p) - W_{22}(q_l). \tag{A.38}$$

Combining them gives

$$(2p - 1)[W_{12}((1, q_l), p) - W_{11}((1, q_l), p)] \geq W_{22}(q_l). \tag{A.39}$$

At $p = 1/2$, the left-hand side of (A.39) is zero. For $p < 1/2$, $(1 - p)\theta^2 + pq_l > p + (1 - p)\theta^2 q_l$ and $(1 - p)\theta + pq_l > p + (1 - p)\theta q_l$, so $W_{12}((1, q_l), p) > W_{11}((1, q_l), p)$. The left-hand side of (A.39) is then nonpositive, which is impossible. Hence $p > 1/2$.

For $p > 1/2$, the same condition requires $W_{12}((1, q_l), p) > W_{11}((1, q_l), p)$. Clearing the positive stability denominators gives

$$(1 - p)\theta^2 - p - (\lambda/\mu)p(1 - p)(\theta - 1)(1 + q_l) > 0. \tag{A.40}$$

In particular, $(1 - p)\theta^2 - p > 0$, so $p < \theta^2/(1 + \theta^2)$. Thus these profiles also satisfy the global window (31); (A.40) is the narrower load- and filing-dependent condition.

Finally, over the stability interval $\partial W_{11}/\partial q_l > 0$, $\partial W_{12}/\partial q_l > 0$, and $\partial W_{22}/\partial q_l < 0$. Therefore $(1 - p)W_{11}((1, q_l), p) + pW_{12}((1, q_l), p) - W_{22}(q_l)$ is continuous and strictly increasing in q_l. The Wardrop equality (A.38) therefore has at most one admissible solution, and it has one when this increasing expression crosses $p\delta/c$ over (A.36). That solution is a boundary Wardrop equilibrium if and only if it also satisfies (A.39). Hence, for fixed parameters, there is at most one equilibrium of the form $(1, q_l)$, together with its type-swapped counterpart $(q_l, 1)$.

A.16.3. Part (iii): Complete separation, $(q_i, q_{3-i}) = (1, 0)$, $\Delta_i \geq 0$, and $\Delta_{3-i} \leq 0$ Consider $(q_1, q_2) = (1, 0)$, with $\Delta_1 \geq 0$ and $\Delta_2 \leq 0$; type symmetry gives the counterpart $(0, 1)$. At this profile, $W_{11}(p) = (\lambda/\mu)p/\{\theta\mu[\theta - (\lambda/\mu)p]\}$, $W_{12}(p) = (\lambda/\mu)(1 - p)/\{\mu[1 - (\lambda/\mu)(1 - p)]\}$, and $W_{22} = (\lambda/\mu)/\{\theta\mu[\theta - \lambda/\mu]\}$. The Court 2 waiting time of the high-share type is zero.

Queue stability requires $\lambda/\mu < \theta$ and $(\lambda/\mu)(1 - p) < 1$. Hence complete separation is impossible when $\lambda/\mu \geq \theta$.

The two Wardrop inequalities are $p(\delta/c) \geq pW_{11}(p) + (1 - p)W_{12}(p)$ and $p(\delta/c) \leq (1 - p)W_{11}(p) + pW_{12}(p) - W_{22}$. For both to hold,

$$(2p - 1)[W_{12}(p) - W_{11}(p)] \geq W_{22}. \tag{A.41}$$

For $p \le 1/2$, stability gives $\theta^2(1-p) - p - (\lambda/\mu)p(1-p)(\theta-1) \ge \theta[\theta(1-p) - p] \ge 0$. The waiting-time expressions therefore imply $W_{12}(p) \ge W_{11}(p)$. The left-hand side of (A.41) is then nonpositive, whereas $W_{22} > 0$. Thus $p > 1/2$, and the same condition requires $W_{12}(p) > W_{11}(p)$. Direct subtraction gives

$$W_{12}(p) - W_{11}(p) = \frac{(\lambda/\mu)\left[\theta^2(1-p) - p - (\lambda/\mu)p(1-p)(\theta-1)\right]}{\theta\mu\left[1-(\lambda/\mu)(1-p)\right]\left[\theta-(\lambda/\mu)p\right]}.$$

Hence complete separation satisfies the global necessary window $1/2 < p < \theta^2/(1+\theta^2)$.

Complete separation also requires $\theta > 2$. To see this, suppose instead that $1 < \theta \le 2$. Over the stable range,

$$W_{22} - (2p-1)W_{12}(p) = \frac{(\lambda/\mu)\left\{1-(\lambda/\mu)(1-p) - \theta[\theta-\lambda/\mu](2p-1)(1-p)\right\}}{\theta\mu[\theta-\lambda/\mu][1-(\lambda/\mu)(1-p)]}.$$

The expression in braces is affine in λ/μ. At $\lambda/\mu = 0$, it is at least $1/2$, because $(2p-1)(1-p) \le 1/8$. At $\lambda/\mu = \theta$, it equals $1-\theta(1-p) > 0$. It is therefore positive throughout the stable range. Consequently, $(2p-1)[W_{12}(p) - W_{11}(p)] < (2p-1)W_{12}(p) < W_{22}$, contradicting (A.41).

Clearing the positive denominators in (A.41) yields the exact condition

$$(2p-1)\left(\theta - \frac{\lambda}{\mu}\right)\left[\theta^2(1-p) - p - \frac{\lambda}{\mu}p(1-p)(\theta-1)\right] \ge \left[1 - \frac{\lambda}{\mu}(1-p)\right]\left[\theta - \frac{\lambda}{\mu}p\right].$$

Together with queue stability, the two Wardrop inequalities are equivalent to

$$W_{11}(p) + \frac{1-p}{p}W_{12}(p) \le \frac{\delta}{c} \le \frac{1-p}{p}W_{11}(p) + W_{12}(p) - \frac{W_{22}}{p}.$$

There is no additional required ordering between $1/(1-p)$ and θ.

A.17. Proof of Corollary 4

When $\theta = 1$, Proposition 7 rules out interior asymmetric equilibria. For boundary asymmetric equilibria, the necessary window in Proposition 8 becomes $1/2 < p < 1/2$, which is empty. Hence every Wardrop equilibrium must be symmetric.

Under $\theta = 1$, the capacity condition (14) reduces to $\lambda/\mu < 2$, the global stability requirement. Proposition 1 then gives a unique symmetric Wardrop equilibrium for every $p \in [0, 1]$.

Appendix B: Cross-Court Filing Reallocation

Following the July 2022 assignment reform, patent filings shifted from WDTX to EDTX (Castle 2024). Table 3 shows this reallocation: WDTX's share of reported Texas patent filings fell while EDTX's share rose.

Appendix C: Specialization Benchmarks and Judge-Level Outcome Variation

Magnitude of the specialization advantage. Empirical studies report faster case processing in specialized settings, but they do not directly estimate the service-rate ratio θ. A simple $M/M/1$ calculation helps put the reported magnitudes in perspective (Clarke and Metzler 2025). Suppose a nonspecialist judge processes cases at rate μ, a specialist judge processes the same case type at rate $\theta\mu$, and both face the same arrival rate $\lambda < \mu$. Their mean disposition times are $T_N = 1/(\mu-\lambda)$ and $T_S = 1/(\theta\mu-\lambda)$, respectively. Thus $T_S/T_N = (1-\rho)/(\theta-\rho)$, where $\rho = \lambda/\mu < 1$. For $\theta = 2$, the specialist disposition time would be 50% of the nonspecialist time at $\rho = 0$, about 33% at $\rho = 0.5$, and about 17% at $\rho = 0.8$. The empirical differences reported in the literature are much smaller. The Federal Judicial Center reports disposition-time ratios of approximately $266/316 = 0.84$, $526/566 = 0.93$, and $271/301 = 0.90$ across termination modes (Williams and Eyre 2021), while Cissner et al. (2013) report $197/260 = 0.76$ for specialized relative to conventional domestic-violence courts. These comparisons are only rough benchmarks rather than estimates of θ, but they suggest that $\theta = 2$ is a deliberately generous operational value; we therefore use $\theta = 3$ only as an extreme-specialization stress test.

Table 3 Reported Patent Filings Across Major Texas Districts

Year	WDTX share of reported Texas total (filings)	EDTX share of reported Texas total (filings)	Reported Texas total filings
2020	64.2% (863)	29.4% (395)	1,344
2021	66.0% (984)	30.0% (448)	1,492
2022*	61.2% (878)	32.8% (470)	1,435
2023	42.9% (519)	51.8% (627)	1,211
2024	24.7% (375)	70.4% (1,069)	1,519
2025	27.4% (519)	66.2% (1,256)	1,897

* WDTX randomized patent-case assignment across judges in July 2022, reducing targetability.

Notes. Shares are calculated relative to the reported Texas total, defined as WDTX + EDTX + NDTX. NDTX is omitted from the main columns because it accounts for approximately 5.5% of reported Texas patent filings on average over 2020–2025. WDTX and EDTX together therefore account for approximately 94.5% of the reported total. SDTX is not included because it is not reported in the public top-district tables used here.

Sources: (Lex Machina Data Team 2023, 2024, 2025, Unified Patents 2026).

Judge-level outcome differences. Observed judicial outcomes provide a concrete basis for the judge-specific private advantage represented by δ. In Chapter 11 bankruptcy, historical conversion rates differ across judges and across courts and predict future conversion decisions, so litigants may prefer judges with lower propensities to convert a reorganization case to Chapter 7 liquidation (Hüther and Kleiner 2026). Immigration courts provide another example: asylum grant rates can differ substantially across judges within the same court, and the magnitude of these differences also varies across courts. Table 4 compares four judges in each of five illustrative immigration courts. The difference between the highest and lowest grant rates ranges from 1.4 percentage points in Baton Rouge to 53.5 percentage points in Chaparral. These examples illustrate both judge-level outcome differences and substantial cross-court variation in their dispersion, providing an observable counterpart to the judge-specific private advantage represented by δ.

Table 4 Judge-Level Asylum Grant-Rate Variation in Selected Immigration Courts

Immigration court	Reported judges	Asylum grant-rate range	Within-court spread (percentage points)
Baton Rouge	4	0.0%–1.4%	1.4
Omaha	4	0.9%–4.7%	3.8
Harlingen	4	14.1%–27.3%	13.2
Tucson	4	9.6%–31.5%	21.9
Chaparral	4	0.4%–53.9%	53.5

Notes. The table reports judges with at least 100 decided asylum cases during the covered period; these judges need not constitute the full set of judges in each court. Grant-rate ranges and spreads are calculated across the four reported judges. The data cover fiscal years 2021–2026, with fiscal year 2026 limited to its first nine months.

Source: TRAC (2026).

Appendix D: Numerical Analysis of Asymmetric Equilibria

The analytical results show that asymmetric equilibria are possible. But two questions remain. How often do they appear? And do they change the assignment policy?

D.1. Numerical Setup

We use four numerical analyses. The first two show where asymmetric equilibria appear. The third checks whether they appear at p^*. The fourth asks how the assignment probability changes when policy is evaluated at the highest-waiting supported equilibrium. Table 5 shows the parameter values used in each analysis.

Table 5 Numerical Designs for the Asymmetric-Equilibrium Analysis

Analysis	Specialization values	Step size	Purpose
Broad numerical analysis	$\theta = 1.25, 1.50, \ldots, 4.00$	$\Delta(\lambda/\mu) = 0.1$, $\Delta(\delta/c) = 0.25$, and $\Delta p = 0.01$	Where asymmetric equilibria appear and which types appear
Lower-specialization numerical analysis	$\theta \in \{1.25, 1.5, 1.75, 2\}$	$\delta/c \in \{2.5, 5, 7.5, 10\}$, $\Delta(\lambda/\mu) = 0.005$, and $\Delta p = 0.005$	How often asymmetric equilibria appear for $\theta \leq 2$
Analysis at p^*	$\theta \in \{1.1, 1.5, 2, 3\}$	Fine load–relative-benefit policy grid; p^* computed over 1,501 candidate assignment probabilities	Whether asymmetric equilibria appear at p^*
Equilibrium-robust probability analysis	$\theta = 3$	Complete overlap region; p^*, both endpoints, and assignment probabilities spaced 0.01 apart over $[0, 1]$	Assignment probability that minimizes the maximum system-wide waiting across all supported equilibria

Notes. Every load grid satisfies the global stability requirement $0 < \lambda/\mu < 2\theta$. Values above $\theta = 3$ in the numerical analysis are extreme sensitivity analyses and do not determine the main conclusions.

D.2. Equilibrium Computation

We first use the analytical results to rule out cases where an asymmetric equilibrium cannot exist. In particular, asymmetry is possible only when $1/2 < p < \theta^2/(1 + \theta^2)$, together with the tighter conditions for interior and one-boundary equilibria in Appendices A.15 and A.16. Complete-separation profiles $(1, 0)$ and $(0, 1)$ are checked directly whenever their queues are stable.

For an interior asymmetric equilibrium, we solve (27) together with $B(q_h, q_l, p) = 0$. We use many starting values so that the search does not depend on one initial guess. We write the solution as (q_h, q_l), with $q_h > q_l$; the type-swapped solution follows by symmetry. We keep a solution only if both filing shares are between zero and one, all queues are stable, and the Wardrop conditions hold. We count numerically identical solutions only once.

D.3. Where Do Asymmetric Equilibria Appear?

We first ask whether asymmetric equilibria are common when specialization is not very strong. They are not. We find none at $\theta = 1.25$, very few at $\theta = 1.5$, and more as specialization advantage becomes greater. Even at $\theta = 2$, only 2.3% of the $(\lambda/\mu, \delta/c, p)$ combinations we check have an asymmetric equilibrium.

To study this, we use $\theta \in \{1.25, 1.5, 1.75, 2\}$ and $\delta/c \in \{2.5, 5, 7.5, 10\}$. Baseline load changes in steps of 0.005, and p changes in steps of 0.005.

Table 6 gives two counts. The first asks whether asymmetry appears for at least one value of p at each $(\theta, \lambda/\mu, \delta/c)$. The second also fixes p. The first count shows how many parameter settings can produce asymmetry. The second shows how often asymmetry appears across all values of p that we check.

Table 6 Incidence of Asymmetric Equilibria in the Lower-Specialization Region

θ	Parameter triples with asymmetry	Share of parameter triples	Full grid cells with asymmetry	Share of full grid cells	Median observed p-span	Maximum observed p-span
1.25	0/1,996	0.0%	0/401,196	0.0%	–	–
1.50	27/2,396	1.1%	133/481,596	< 0.1%	2.0 pp	3.5 pp
1.75	325/2,796	11.6%	4,811/561,996	0.9%	7.5 pp	11.0 pp
2.00	681/3,196	21.3%	15,005/642,396	2.3%	11.5 pp	16.0 pp

Notes. pp denotes percentage points. A parameter triple is a fixed $(\theta, \lambda/\mu, \delta/c)$ combination and is counted as asymmetric if at least one value of p supports an asymmetric equilibrium. A full grid cell is a fixed $(\theta, \lambda/\mu, \delta/c, p)$ combination. The observed p-span is the difference between the largest and smallest grid values of p supporting asymmetry at a parameter triple. These shares are descriptive summaries of the specified grid, not probabilities over a parameter distribution.

Figure 6 shows where these equilibria appear. A greater specialization advantage makes asymmetric equilibria more common and allows them over a wider range of p. But they still appear only at intermediate values of p.

D.4. What Happens When Specialization Is Strong?

We next ask what happens when specialization advantage is much greater. At $\theta = 3$, asymmetric equilibria become much more common, but they still appear only over a limited range of p.

Asymmetry appears for 40.9% of the parameter settings we check. When p is also fixed, only 7.6% of the cases have an asymmetric equilibrium. All of these equilibria appear between $p = 0.51$ and $p = 0.78$. The median range of p is 0.20.

Table 7 shows which types of asymmetric equilibria appear. Almost all are interior or upper-boundary equilibria. Together, these two types account for 95.6% of the cases. Lower-boundary and complete-separation equilibria are rare.

Table 7 Asymmetric-Equilibrium Composition in the $\theta = 3$ Numerical Analysis

Asymmetric equilibrium class	Grid-cell observations	Share
Interior	8,101	44.7%
Upper boundary	9,204	50.8%
Lower boundary	692	3.8%
Corner	109	0.6%
Total	18,106	100.0%

Notes. The table classifies grid cells supporting an asymmetric equilibrium after type-swapped counterparts are collapsed. Interior equilibria satisfy $0 < q_l < q_h < 1$. Upper-boundary equilibria have $q_h = 1$, lower-boundary equilibria have $q_l = 0$, and corner equilibria have $(q_h, q_l) = (1, 0)$.

D.5. Do Asymmetric Equilibria Appear at p^*?

The next question is directly relevant for the assignment policy: do asymmetric equilibria appear at p^*? For the specialization levels used in our main analysis, the answer is no. We find no asymmetric equilibrium at p^* for $\theta = 1.1$, 1.5, or 2. At $\theta = 3$, asymmetry appears at p^* in only 4.0% of the cases we check.

For each parameter setting, we first find p^* from the symmetric equilibrium. We then check whether an asymmetric equilibrium also exists at that same value of p.

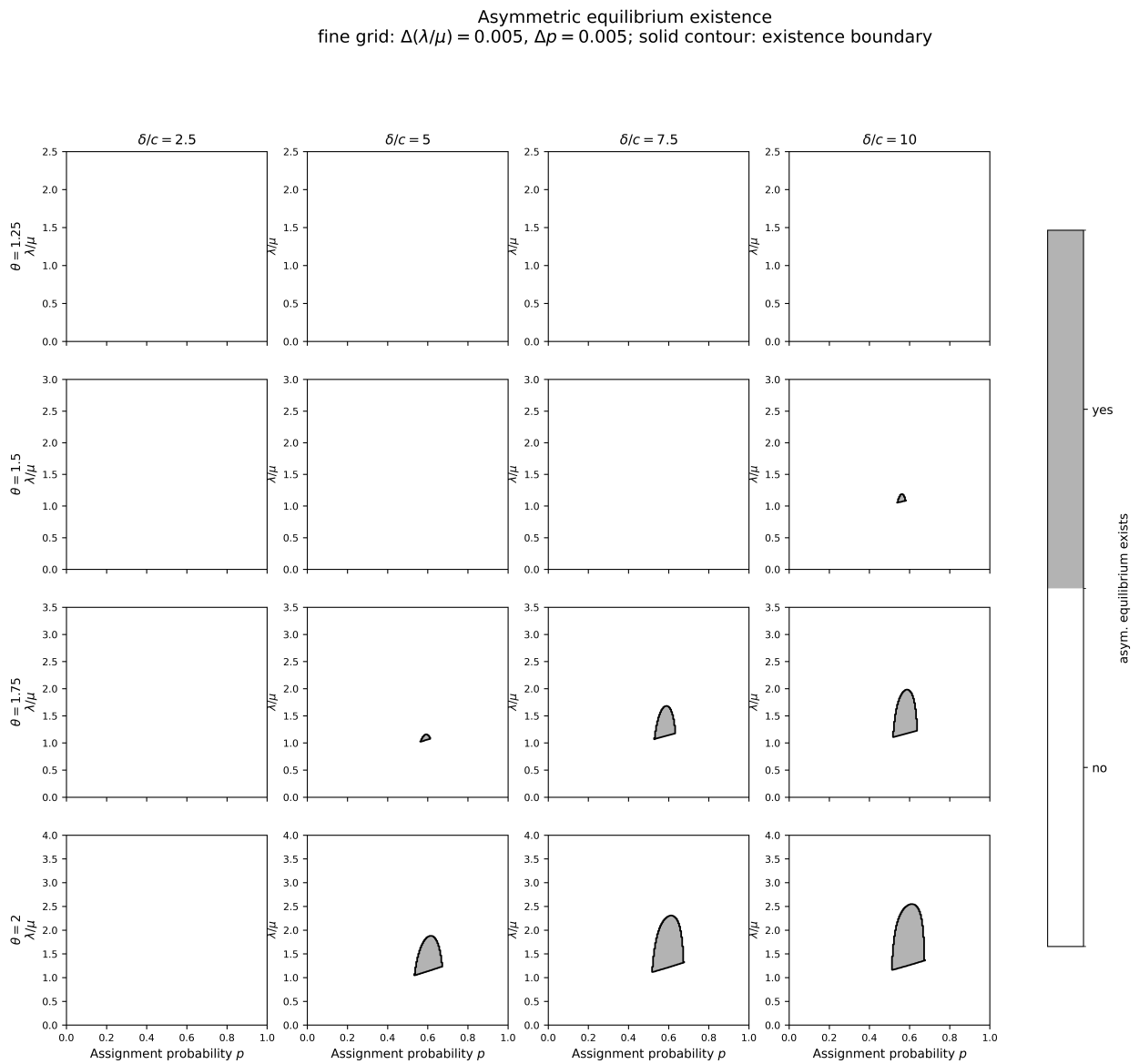


Figure 6 Asymmetric-equilibrium existence in the lower-specialization fine grid. Rows correspond to $\theta \in \{1.25, 1.5, 1.75, 2\}$, and columns correspond to $\delta/c \in \{2.5, 5, 7.5, 10\}$. Within each panel, the horizontal axis is the assignment probability p, and the vertical axis is λ/μ. Gray cells support at least one asymmetric equilibrium; white cells do not. The solid contour marks the observed existence boundary. The grid increments are $\Delta(\lambda/\mu) = 0.005$ and $\Delta p = 0.005$.

At $\theta = 3$, asymmetry at p^* appears only when both load and the relative shopping benefit are high: $3.84 \le \lambda/\mu \le 4.49$ and $5.71 \le \delta/c \le 20$. In these cases, $0.5513 \le p^* \le 0.734$. Strong specialization alone is therefore not enough.

When an asymmetric equilibrium does appear at p^*, it always has more system-wide waiting than the symmetric equilibrium. The increase is 12.8% at the median and reaches 34.1% at the maximum. We next ask whether a small change in p can reduce this extra waiting.

Table 8 Asymmetric-Equilibrium Incidence at the Symmetric-Optimal Assignment Probability

θ	Policy-grid points	Asymmetry at p^*	Share
1.10	331,670	0	0.0%
1.50	487,750	0	0.0%
2.00	682,850	0	0.0%
3.00	1,073,050	42,645	4.0%

Notes. Each policy-grid point is a fixed $(\theta, \lambda/\mu, \delta/c)$ combination evaluated at the assignment probability p^* minimizing system-wide waiting along the symmetric equilibrium path. The reported shares are grid summaries rather than probabilities over a parameter distribution.

D.6. How Much Does the Equilibrium-Robust Assignment Probability Differ from p^*?

We ask how far the equilibrium-robust assignment probability moves from p^* when policy is evaluated by the highest-waiting supported equilibrium.

For each parameter setting, we evaluate p^* and the feasible assignment probabilities on the grid $0, 0.01, \ldots, 1$. At each candidate, we use the highest system-wide waiting among the supported equilibria. We approximate p^*_{asym} by the candidate that minimizes this worst-case waiting, breaking ties in favor of the candidate closest to p^*.

Table 9 Equilibrium-Robust Assignment and Waiting Outcomes in the $\theta = 3$ Overlap Region

Statistic	Result
Overlap-region cells	42,645
Median signed assignment-probability difference	−2.0 percentage points
Median absolute assignment-probability difference	3.0 percentage points
90th-percentile absolute assignment-probability difference	6.0 percentage points
Maximum absolute assignment-probability difference	8.0 percentage points
Cells with $p^*_{\text{asym}} < p^*$	29,090 (68.2%)
Cells with $p^*_{\text{asym}} = p^*$	139 (0.3%)
Cells with $p^*_{\text{asym}} > p^*$	13,416 (31.5%)
Median reduction in $\overline{SPO}$	10.3%
90th-percentile reduction in $\overline{SPO}$	18.5%
Maximum reduction in $\overline{SPO}$	22.6%
Median symmetric-equilibrium cost	1.0%
90th-percentile symmetric-equilibrium cost	3.6%
Maximum symmetric-equilibrium cost	5.9%
Cells where the robust policy eliminates asymmetry	37,061 (86.9%)
Cells where asymmetry remains under the robust policy	5,584 (13.1%)
Median remaining asymmetric-equilibrium penalty	0.3%
90th-percentile remaining asymmetric-equilibrium penalty	0.8%
Maximum remaining asymmetric-equilibrium penalty	1.4%

Notes. The sample contains the 42,645 parameter cells at which asymmetry is supported at p^*. Assignment-probability differences are reported in percentage points. The reduction in $\overline{SPO}$ compares $\overline{SPO}(p^*_{\text{asym}})$ with $\overline{SPO}(p^*)$. The symmetric-equilibrium cost is the percentage increase in symmetric-equilibrium waiting from p^* to p^*_{asym}. When the robust policy still supports asymmetric equilibria, the remaining penalty is the percentage by which the maximum system-wide waiting among those equilibria exceeds symmetric-equilibrium waiting under the same policy.

Table 9 shows that the median absolute difference between p^*_{asym} and p^* is 3 percentage points, and the largest difference is 8 percentage points. The robust assignment probability is lower than p^* in 68.2% of the cases and higher in 31.5%.

These small changes can reduce waiting under the worst equilibrium. At the median, worst-case equilibrium system-wide waiting falls by 10.3%, while waiting under the symmetric equilibrium rises by only 1.0%. In 86.9% of the cases, moving from p^* to p^*_{asym} eliminates the asymmetric equilibria that coexist at p^*. In the remaining cases, the additional system-wide waiting associated with asymmetry is small: 0.3% at the median and 1.4% at the maximum.

Figure 7 shows these results across the parameter space. Panel (a) shows how much worst-case waiting falls when we replace p^* with p^*_{asym}. Panel (b) shows how much the assignment probability changes.

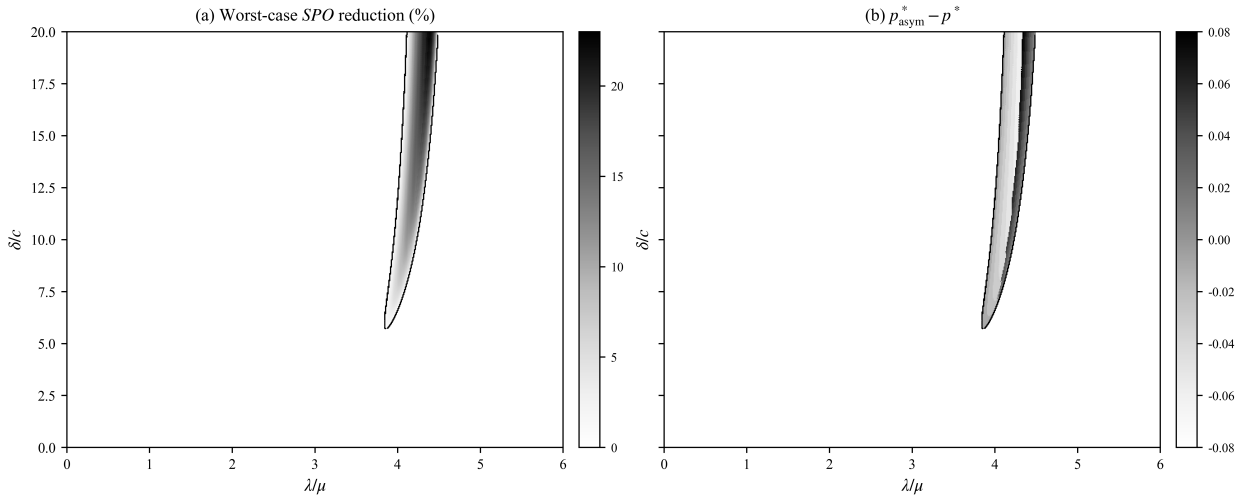


Figure 7 Equilibrium-robust assignment in the $\theta = 3$ stress test. Panel (a) reports the percentage reduction in $\overline{SPO}$ when p^* is replaced by p^*_{asym}. Panel (b) reports the signed difference $p^*_{\text{asym}} - p^*$. Negative values correspond to lower targetability and positive values to higher targetability. Black outlines delimit the $\theta = 3$ overlap region; cells outside this region are left blank.

Appendix E: Robustness to Larger Specialist Pools

Consider the extension in Section 7.3. Each case type has n specialist judges in each court, where $n \geq 1$. In Court 1, $k \in \{1, \ldots, n\}$ of the n specialists for each case type are targeted judges. Targeted and nontargeted specialists have the same processing rates; they differ only in whether assignment to them generates the shopping benefit. A type-i case is assigned to its specialist pool with probability p and to the other specialist pool with probability $1 - p$. A natural assumption is that cases assigned to either pool are equally shared among its n judges.

Hence, each judge in type i's Court 1 specialist pool receives specialist cases at rate $\lambda p q_i / n$ and nonspecialist cases at rate $\lambda(1-p)q_{3-i}/n$.

Applying the Pollaczek–Khinchine formula, the waiting time at any judge in type i's Court 1 specialist pool is

$$\widetilde{W}_{1i}(\mathbf{q}, p) = \frac{(\lambda/\mu)\left[p\frac{q_i}{n} + (1-p)\theta^2\frac{q_{3-i}}{n}\right]}{\theta\mu\left[\theta - (\lambda/\mu)\left(p\frac{q_i}{n} + (1-p)\theta\frac{q_{3-i}}{n}\right)\right]}.$$

Court 2 remains fully specialized. Type-i cases are equally shared among their n specialist judges, so the waiting time at any of these judges is

$$\widetilde{W}_{2i}(q_i) = \frac{(\lambda/\mu)\frac{1-q_i}{n}}{\theta\mu\left[\theta - (\lambda/\mu)\frac{1-q_i}{n}\right]}.$$

The change in targetability is separate from this division of case flow. The specialist assignment probability remains p. Conditional on assignment to the specialist pool, however, only k of its n judges are targeted. Because cases are equally shared among the n specialists, the probability of reaching a targeted judge conditional on specialist assignment is k/n. Targetability is therefore pk/n. A type-i litigant's Court 1 utility becomes $\widetilde{U}_{1i} = V + (pk/n)\delta - c[p\widetilde{W}_{1i} + (1-p)\widetilde{W}_{1,3-i}]$, whereas $\widetilde{U}_{2i} = V - c\widetilde{W}_{2i}$.

Under a symmetric filing profile $q_1 = q_2 = q_s$, all judges in the two Court 1 specialist pools have the same waiting time $\widetilde{W}_1(q_s, p)$, and all judges in the two Court 2 specialist pools have the same waiting time $\widetilde{W}_2(q_s)$. In particular,

$$\widetilde{W}_1(q_s, p) = \frac{(\lambda/\mu)\frac{q_s}{n}[p + (1-p)\theta^2]}{\theta\mu\left[\theta - (\lambda/\mu)\frac{q_s}{n}[p + (1-p)\theta]\right]},$$

and

$$\widetilde{W}_2(q_s) = \frac{(\lambda/\mu)\frac{1-q_s}{n}}{\theta\mu\left[\theta - (\lambda/\mu)\frac{1-q_s}{n}\right]}.$$

The resulting equilibrium has the same algebraic form as the baseline model, $p(k/n)(\delta/c) = \widetilde{W}_1(q_s, p) - \widetilde{W}_2(q_s)$. In the waiting-time expressions, dividing each flow among n judges is algebraically equivalent to evaluating the baseline model at $\lambda/\mu \mapsto (\lambda/\mu)/n$. In the Wardrop condition, replacing one targeted specialist by k targeted specialists among n specialists is algebraically equivalent to $\delta/c \mapsto (k/n)(\delta/c)$. These substitutions are only an algebraic mapping: neither λ nor total arrival volume changes, and the total arrival rate remains 2λ.

Because the general utilities have the same form under these substitutions, the same mapping applies to the full Wardrop system, including asymmetric equilibria. It follows directly that $\widetilde{q}_s(p; \lambda/\mu, \delta/c) = q_s(p; (\lambda/\mu)/n, (k/n)(\delta/c))$. The global stability requirement also scales with the number of specialist judges and becomes $0 < \lambda/\mu < 2n\theta$. Because system-wide waiting retains the same form, the optimal assignment probability satisfies $\widetilde{p}^*(\lambda/\mu, \delta/c) = p^*((\lambda/\mu)/n, (k/n)(\delta/c))$. Therefore, for fixed n and $k \geq 1$, the equilibrium characterization, comparative statics, and optimal-policy results of the baseline model continue to hold under the corresponding scaled parameter values.

Capacity expansion. Fix a common feasible p. We first consider interior symmetric equilibria. Because each waiting time is strictly increasing in the corresponding per-judge filing rate, it is sufficient to compare per-judge filings.

First consider an untargeted capacity expansion, $(n, k) \to (n+1, k)$. Suppose $q_s^{n+1,k}(p)/(n+1) \geq q_s^{n,k}(p)/n$. Since total per-judge filing falls from $1/n$ to $1/(n+1)$, Court 2 filing per judge must fall. Hence $\widetilde{W}_1 - \widetilde{W}_2$ increases, whereas the Wardrop right-hand side weakly falls from $p(k/n)(\delta/c)$ to $p[k/(n+1)](\delta/c)$, a contradiction. Thus $q_s^{n+1,k}(p)/(n+1) < q_s^{n,k}(p)/n$.

Suppose instead that Court 2 filing per judge does not fall. Conservation then implies $q_s^{n,k}(p)/n - q_s^{n+1,k}(p)/(n+1) \geq 1/[n(n+1)]$. Since $q_s^{n,k}(p) \leq 1$, it follows that $q_s^{n+1,k}(p)/(n+1) \leq [n/(n+1)]q_s^{n,k}(p)/n$. Substituting this bound into the Court 1 waiting expression gives $\widetilde{W}_1^{n+1} < [n/(n+1)]\widetilde{W}_1^n$. Therefore, $\widetilde{W}_1^n - \widetilde{W}_1^{n+1} > \widetilde{W}_1^n/(n+1) > pk(\delta/c)/[n(n+1)]$, where the second inequality follows from the old Wardrop condition. Subtracting the two Wardrop conditions

gives the opposite weak inequality, a contradiction. Hence Court 2 filing per judge also falls, so both $\widetilde{W}_1$ and $\widetilde{W}_2$ strictly decrease.

The Wardrop condition gives $SPO_s^{n,k}(p) = \widetilde{W}_2^n + pk(\delta/c)q_s^{n,k}(p)/n$. The first term strictly decreases and the second weakly decreases. Thus system-wide waiting strictly decreases when both equilibria are interior.

The comparison also holds at boundary equilibria. If the old equilibrium has $q_s^{n,k}(p) = 1$, direct substitution gives $\widetilde{W}_1^{n+1}(1,p) < [n/(n+1)]\widetilde{W}_1^n(1,p) \le pk(\delta/c)/(n+1)$. Thus the new equilibrium also has full filing in Court 1 and strictly lower system-wide waiting.

If the old equilibrium is interior but the new one has $q_s^{n+1,k}(p) = 1$, then $1/(n+1) < q_s^{n,k}(p)/n$. Otherwise Court 1 waiting would not fall, contradicting the new Wardrop inequality $\widetilde{W}_1^{n+1} \le pk(\delta/c)/(n+1)$ and the old Wardrop equality $\widetilde{W}_1^n = \widetilde{W}_2^n + pk(\delta/c)/n > pk(\delta/c)/(n+1)$. Consequently, $SPO_s^{n+1,k}(p) \le pk(\delta/c)/(n+1) < pk(\delta/c)q_s^{n,k}(p)/n < SPO_s^{n,k}(p)$.

The strict reduction therefore holds for every common feasible p. An old feasible policy remains feasible after the untargeted expansion. Optimality in the new system gives $SPO_s^{n+1,k}(p^*_{n+1,k}) \le SPO_s^{n+1,k}(p^*_{n,k})$, while the fixed-policy comparison gives $SPO_s^{n+1,k}(p^*_{n,k}) < SPO_s^{n,k}(p^*_{n,k})$. Hence untargeted capacity expansion also strictly reduces optimized system-wide waiting.

An untargeted capacity expansion lowers both $(\lambda/\mu)/n$ and $(k/n)(\delta/c)$. The separate comparative statics do not establish the net change in p^*, because these parameters change together.

Now consider a targeted capacity expansion, $(n,k) \to (n+1,k+1)$, with both equilibria interior. If Court 2 filing per judge did not fall, conservation would imply that Court 1 filing per judge falls. Then $\widetilde{W}_1 - \widetilde{W}_2$ would fall, whereas the Wardrop right-hand side weakly increases because $(k+1)/(n+1) \ge k/n$, a contradiction. Hence Court 2 filing per judge and $\widetilde{W}_2$ strictly decrease.

For Court 1, if $q_s^{n,k}(p) > n/(n+1)$, then $q_s^{n+1,k+1}(p)/(n+1) < q_s^{n,k}(p)/n$ mechanically. Otherwise, evaluating the new Wardrop condition at the old Court 1 per-judge filing rate shows that Court 1 filing per judge falls if and only if $\lambda/\mu > \mu p(n-k)(\delta/c)[\theta - (\lambda/\mu)(1-q_s^{n,k}(p))/n]\{\theta - (\lambda/\mu)[1/(n+1) - q_s^{n,k}(p)/n]\}$. Thus Court 1 filing and waiting may either increase or decrease. When $k = n$, the right-hand side is zero, so Court 1 filing per judge necessarily falls.

Finally,

$$SPO_s^{n+1,k+1}(p) - SPO_s^{n,k}(p) = \widetilde{W}_2^{n+1} - \widetilde{W}_2^n + p\frac{\delta}{c}\left[(k+1)\frac{q_s^{n+1,k+1}(p)}{n+1} - k\frac{q_s^{n,k}(p)}{n}\right].$$

The first term is negative and the second has no fixed sign, so the effect on $SPO_s(p)$ is ambiguous. A targeted capacity expansion also lowers $(\lambda/\mu)/n$ and weakly raises $(k/n)(\delta/c)$; both established comparative statics therefore imply $p^*_{n+1,k+1} \le p^*_{n,k}$, with strict inequality for interior optima.

The adverse effect of capacity expansion does not require targetability to increase. Consider $k = n$ and $p = 1$, so targetability remains equal to one as n changes, and let $x = (\lambda/\mu)/n$ denote per-judge baseline load. The $q_s = 1$ equilibrium gives way to an interior equilibrium at $\bar{x} = \theta^2\mu(\delta/c)/[1+\theta\mu(\delta/c)]$. As $x \downarrow \bar{x}$ through interior equilibria, $\lim_{x\downarrow\bar{x}} dSPO_s/dx = [\delta/c - \bar{x}/(\mu\theta^2)]\lim_{x\downarrow\bar{x}} dq_s/dx$. Equation (A.13) gives $\lim_{x\downarrow\bar{x}} dq_s/dx = -\theta^2/\{\bar{x}[\theta^2 + (\theta-\bar{x})^2]\} < 0$, while the boundary condition $\delta/c = \bar{x}/[\theta\mu(\theta-\bar{x})]$ implies $\delta/c > \bar{x}/(\mu\theta^2)$. Hence, locally, reducing per-judge load through capacity expansion can increase system-wide waiting even though targetability is unchanged.