Screening for Usage

Justin Hadad* and Kyle Woodward†

July 8, 2026

Abstract

In systems with significant wealth inequality, charging for a resource distorts efficient allocation across wealth types. We study a repeated allocation problem where agents vary in both their marginal value for money and their usage probability, and the principal wants to maximize resource usage. In each period agents choose whether or not to make reservations, but may randomly cancel an allocated reservation. The principal may screen on usage probabilities using two empirically-relevant instruments: monetary penalties and eligibility restrictions. We show that the optimal regulation has a simple structure: the principal never offers a refund for cancellation, and access restrictions are essentially deterministic. When agents are patient and inequality is high, eligibility-based penalties become an efficient way to screen on usage propensity without amplifying wealth-based exclusion. Conversely, when agents are impatient or inequality is mild, monetary penalties perform well in spite of the inequality they engender.

1 Introduction

When designing systems to allocate scarce resources, policymakers often rely on monetary penalties to deter undesirable behavior. In settings where users with substantially different wealth levels compete, financial penalties can create perverse distributional consequences; wealthy individuals may treat modest monetary costs as trivial, while the same penalties deter the poor. This inequality not only raises concerns with respect to fairness, but also undermines social surplus maximization.

We examine the challenge of regulation in a repeated market with inequality. To make our analysis concrete, we situate our approach in the recent regulation of campsite reservations in the state of California. Prior to 2024, reservations at state-run campsites in California could be canceled up to the evening of arrival for a nominal penalty of \$8. Frustrated with reservations at highly-desirable campsites being canceled at the last minute (which leave the campsite unfilled), in 2024 the state increased the penalty to \$8 plus the cost of one night of stay for any cancellation within a week of the arrival date Thomas ([n. d.]).[1]

Less-publicized than the financial penalties was the legislation's enabling of yearlong bans for individuals who renege on a reservation three times in a one-year span. Traditional, price-based approaches to mechanism design cannot justify this design: restrictions on future payoffs can be mapped to a single threshold in value space, the same as a posted price.[2] However, California is one of the most income-unequal states in the U.S.,

*University of Oxford; justin.hadad@balliol.ox.ac.uk

†Assign Group; kyle@assign.group

[1]Nightly rates at state-run campsites range from \$25 to \$100 per night, hence this change potentially represents a significant increase in cancellation penalty. In 2026, California further increased cancellation penalties by introducing a tiered penalty system based on time-until-arrival Parks ([n. d.]).

[2]At the core of this statement is the fact that any stationary distribution of campsite reservations in the presence of ban-based penalties can be weakly improved upon by a simple posted price. Since posted-price equilibria are time-invariant, it stands to reason that a regulator may prefer posted prices over bans in a dynamic environment.

and the utility value of a fixed financial penalty may vary substantially across the population. In such an environment, financial penalties disproportionately dissuade high cancellation-propensity poor individuals, while eligibility-based penalties shrink the pool of available consumers. It is ex ante unclear whether a designer interested in maximizing realized value in an unequal environment will prefer to impose monetary or eligibility-based penalties.

The potential scalability of equitable mechanisms based on future restrictions on consumption arises from the fact that, absent prices, agents have identical valuations for immediate consumption, and hence identical disutility from restrictions on future consumption. In the context of camping, a yearlong ban harms rich and poor alike. The common value for camping implies that camping acts like a numéraire good, unlike money which has wealth-dependent marginal value.

Ceteris paribus, screening on numéraire valuation is more socially efficient than screening on other valuations. However, screening using restrictions on future consumption comes at a cost: an agent who is removed from the market via eligibility restrictions cannot consume while under sanction, even if their consumption would improve social value. For this reason it is not obvious ex ante whether relying on financial or restriction-based punishments yields higher realized surplus.

We evaluate these tradeoffs in a model in which infinitely-lived agents have fixed wealths but random per-period usage propensities. High-wealth agents have lower marginal utility for money, and low-wealth agents have higher marginal utility for money. All else equal, an agent with a high usage propensity will attempt to use the resource while an agent with a low usage propensity will not. Our principal has two design instruments: prices (and refunds), and temporary eligibility restrictions. As discussed above, prices distort the market away from efficient allocation as high-wealth agents book the resource even when their usage propensity is relatively low, but low-wealth agents book the resource only when the usage propensity is high. Eligibility restrictions distort away from efficiency by artificially shrinking the pool of consumers.

Our analysis opens by describing essential features of optimal financial and eligibility-based penalization schemes. First, in equilibrium the principal never offers refunds for cancellation (unlike California's recent policies). In our model, full refunds turn payment into a usage-contingent transfer, and the market price will either admit all high-wealth agents or all agents. Since we show that usage, ceteris paribus, increases as the spread between high- and low-wealth thresholds decreases, refunds harm usage. Our analysis also handles the case of partial refunds. Second, in equilibrium eligibility restrictions are a randomization over at most two adjacent ban lengths. Fundamentally this is a consequence of Jensen's inequality, where restricting all cancelers' eligibility for an average amount of time keeps the eligible population larger than restricting a small fraction of cancelers' eligibilities for a long time.

Building on these results, we derive results for limiting markets. When agents are impatient, financial penalties are preferred to eligibility-based penalties. This is because when agents are impatient, eligibility-based penalties offer essentially no sanction and therefore cannot screen. The inequality generated by prices results in higher usage than the unscreened outcome generated by bans. Conversely, when agents are patient, and inequality is not too small, eligibility-based penalties outperform financial penalties. This is because when the resource is scarce, financial penalties admit only high-wealth agents. Because bans screen independent of wealth, the effect of targeting the full population dominates the loss from holding potential consumers out of the market. We show in an example that "too small" and "sufficiently scarce" are not just limiting cases.

Our results offer prescriptive suggestions for when principals interested in maximizing resource usage in the presence of wealth inequality should impose monetary penalties versus eligibility-based penalties, and vice versa.

1.1 Literature

A series of papers study regulation in markets where value is only realized after transaction. The most closely related, Ma et al. (2018), analyzes a two-period resource allocation problem where a principal wants to maximize usage, and where agents realize their value for usage only after assignment. We also study contingent payment mechanisms, but we expand the scope in three ways: our repeated allocation environment lets use affect future access; we allow both monetary and nonmonetary instruments; and we focus on the differential impact of regulation under inequality. Other work investigates the question of revenue maximization in reservation systems. Non-refundable deposits are useful screening devices and overbooking can be useful under uncertainty in demand (Georgiadis and Tang, 2014; Lyu, 2024).

Beyond money, nonmarket mechanisms can maximize surplus when the objective differs from revenue maximization. Whether monetary schemes dominate nonmarket mechanisms is subject to the distribution of the willingness to pay (Condorelli, 2013). Relatedly, money-burning and delays can make mechanisms incentive-compatible and increase surplus (Jackson and Sonnenschein, 2007; Burkett and Woodward, 2024; Patel and Urgun, 2024). Verdier and Reeling (2022) studies efficient dynamic allocation in an empirical model of rationed bear hunting permits in Michigan.

Our paper belongs to the *inequality-aware market-design* agenda which takes ethical institutional design seriously: "Policymakers, though, often attempt to maximize an objective function that evidently differs from either revenue or allocative efficiency" (Dworczak, 2024). A body of literature has emerged in this spirit and is instructive for the design of markets which exhibit wealth inequality (**?**Akbarpour et al., 2024; Mäkimattila, 2025). Pathak et al. (2021) apply these principles to the equitable distribution of COVID-19 vaccines.

2 Model

Time is discrete and indexed by $t \in \{0, 1, 2, \ldots\}$. There is a unit mass of agents who are infinitely-lived and discount future payoffs by $\delta \in [0, 1)$. In each period, agents may book at most one reservation, and at most a mass $C \in (0, 1)$ of reservations can be allocated in any period. Agents share a common value $v > 0$ from *using* an allocated reservation; without loss of generality we fix $v = 1$ throughout. If an allocated reservation is not used, the agent obtains no consumption value from that allocation.

Agents differ in their marginal value for money. There are two wealth types $w \in \{w_L, w_H\}$ with $w_H > w_L > 0$; fraction $\mu_H \in (0, 1)$ of agents has wealth type w_H, and the remaining fraction $\mu_L := 1 - \mu_H$ has wealth type w_L. Each wealth level enters as the inverse marginal value for money; throughout, a payment of size x yields utility cost x/w.

In each period, an agent has an *eligibility*: an agent is eligible if she is able to book the resource (recall that booking is necessary for allocation). In each period t, every eligible agent i draws a private usage probability $\pi_{it} \in [0, 1]$ from a cumulative distribution function F that is continuously differentiable on $(0, 1)$ with density $f(\pi) > 0$ and bounded second derivative $|f'(x)| < \bar{f}'$ for all $\pi \in (0, 1)$. Across agents and across time, the draws $\{\pi_{it}\}$ are independent. If agent i is allocated a reservation in period t, then she uses it with probability π_{it} and cancels (equivalently, does not use) with complementary probability $1 - \pi_{it}$.

In each period, each eligible agent i makes a *booking decision* $b_{it} \in \{0, 1\}$ with $b_{it} = 1$ corresponding to making a booking, and $b_{it} = 0$ corresponding to not. Let

$$A_t := \int b_{it} di$$

denote the mass of agents who book in period t.[3] Recall that an agent who is ineligible necessarily has $b_{it} = 0$.

If $A_t \le C$, then every booking is allocated a reservation. If $A_t > C$, then exactly mass C is allocated by symmetric random rationing among those who book. Let $y_{it} \in \{0, 1\}$ indicate whether agent i is allocated a reservation in period t. Conditional on booking, the allocation probability is defined

$$\beta_t := \Pr(y_{it} = 1 \mid b_{it} = 1) = \min\left\{1, \frac{C}{A_t}\right\}, \qquad \Pr(y_{it} = 1 \mid b_{it} = 0) = 0.$$

Let $c_{it} \in \{0, 1\}$ indicate cancellation, where

$$\Pr(c_{it} = 1 \mid y_{it} = 1, \pi_{it}) = 1 - \pi_{it}, \qquad \Pr(c_{it} = 1 \mid y_{it} = 0, \pi_{it}) = 0.$$

2.1 Steady state and the principal's objective

The mechanisms we study induce stationary equilibria in which, for each wealth type $w \in \{w_L, w_H\}$, the booking decision is monotone in the privately observed usage probability π. Let z_{it} be a variable such that $z_{it} = 0$ when agent i is eligible to book in period t.[4] In any stationary equilibrium in our class, there exist type-specific cutoffs

$$\underline{\pi}_w \in [0, 1] \quad \text{such that} \quad b_{it} = 1 \iff z_{it} = 0 \text{ and } \pi_{it} \ge \underline{\pi}_w \;\; \text{for wealth type } w.$$

For any cutoff $\underline{\pi} \in [0, 1]$, the following are our primary sufficient statistics:

$$\alpha(\underline{\pi}) := \Pr(\pi \ge \underline{\pi}) = 1 - F(\underline{\pi}) \qquad \text{(booking mass)}$$

$$M(\underline{\pi}) := \mathbb{E}[\pi \mathbf{1}\{\pi \ge \underline{\pi}\}] = \int_{\underline{\pi}}^{1} x \, dF(x) \qquad \text{(expected used mass by booking agents)}$$

$$m(\underline{\pi}) := \mathbb{E}[\pi \mid \pi \ge \underline{\pi}] = \frac{M(\underline{\pi})}{\alpha(\underline{\pi})} \qquad \text{(average usage rate among booking agents).}$$

The function $m(\cdot)$ is the average usage probability among those who book, and we assume for simplicity that $m(1) = 1$. Observe that m is strictly increasing on $[0, 1]$ and satisfies $m(0) = \mathbb{E}[\pi]$.

A policy may include access restrictions following cancellations, which can make some agents temporarily ineligible to book. In steady state, the mass of wealth-w agents who are eligible at the start of a period is constant over time. Denote this mass by $\mu_w \zeta_w$ where $\zeta_w \in (0, 1]$ is the steady-state fraction of wealth-w agents who are eligible.

Fix a stationary equilibrium with type-specific booking cutoffs $(\underline{\pi}_H, \underline{\pi}_L)$ and eligibility masses (ζ_H, ζ_L). The steady-state mass of agents who book is then

$$A := \mu_H \zeta_H \alpha(\underline{\pi}_H) + \mu_L \zeta_L \alpha(\underline{\pi}_L). \tag{1}$$

[3] Because draws of usage probabilities are independent across agents and across time, this definition is without loss.

[4] We define z_{it} rigorously in Section 2.2 below.

Given A, the steady-state allocation probability conditional on booking is

$$\beta := \min\left\{1, \frac{C}{A}\right\}.$$

Thus $\beta = 1$ when $A \leq C$ and $\beta = C/A$ when $A > C$.

The principal's objective is to maximize *usage*: the expected mass of allocated reservations that are used in a period. In a steady state, expected usage per period is

$$U(\underline{\pi}_H, \underline{\pi}_L; \zeta_H, \zeta_L) := \beta\left(\mu_H \zeta_H M(\underline{\pi}_H) + \mu_L \zeta_L M(\underline{\pi}_L)\right). \tag{2}$$

That is, $\mu_w \zeta_w M(\underline{\pi}_w)$ is the mass of *used* reservations that would arise from wealth type w if every booking were allocated, and β scales this by the probability a booking is allocated under capacity rationing. We assume transfers do not enter the principal's objective.[5] Prices and refunds matter only through the incentives they induce, and access restrictions matter only through their effect on steady-state eligibility and booking behavior. The principal therefore chooses a feasible policy to maximize steady-state usage (2).

2.2 Restricted class of mechanisms

At time $t = 0$, before agents receive any information, the principal commits to a policy from the class of mechanisms consisting of financial and eligibility-based penalties (defined below) and applies it uniformly over time. A policy is a triple

$$(p, r; \rho) \in \mathbb{R}_+ \times [0, p] \times \Delta(\mathbb{Z}_+),$$

where $p \geq 0$ is the posted price paid whenever a booking is allocated; $r \in [0, p]$ is the refund paid if an allocated booking is canceled; and ρ is a distribution over ban lengths $\tau \in \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$, where $\tau = 0$ corresponds to no access restriction following cancellation.

We now consider the representation of a ban. Each agent i has a ban state $z_{it} \in \mathbb{Z}_+$, interpreted as the number of periods of ineligibility remaining at the start of period t. Agent i is eligible to book in period t if and only if $z_{it} = 0$. The state evolves according to

$$z_{i,t+1} = \begin{cases} z_{it} - 1, & \text{if } z_{it} > 0, \\ \tau_{it}, & \text{if } z_{it} = 0,\ y_{it} = 1,\ \text{and } c_{it} = 1,\ \text{where } \tau_{it} \sim \rho, \\ 0, & \text{otherwise.} \end{cases}$$

Given a policy $(p, r; \rho)$ and the induced booking cutoffs $(\underline{\pi}_H, \underline{\pi}_L)$, let $\zeta_{tw}(k)$ denote the period-t mass of wealth-w agents with ban state $z_{it} = k$, and let β_t be the probability that a booking is allocated. Then

$$\begin{aligned} \zeta_{t+1,w}(\tau) &= \zeta_{t,w}(\tau + 1) + \zeta_{t,w}(0)\beta_t\big(1 - M(\underline{\pi}_w)\big)\rho(\tau), \qquad (\tau > 0), \\ \zeta_{t+1,w}(0) &= \zeta_{t,w}(1) + \zeta_{t,w}(0)\beta_t\big(1 - M(\underline{\pi}_w)\big)\rho(0) \\ &\quad + \zeta_{t,w}(0)\Big(\big(1 - \alpha(\underline{\pi}_w)\big) + \big(1 - \beta_t\big)\alpha(\underline{\pi}_w) + \beta_t M(\underline{\pi}_w)\Big). \end{aligned}$$

In steady state, $\zeta_{t+1,w}(\tau) = \zeta_{tw}(\tau)$ for all τ when $\underline{\pi}_w$ and β_t are determined by agent equilibrium behavior.

[5] In classical analyses, transfers *are* the principal's objective. In this case it is trivially true that the principal prefers prices to eligibility-based restrictions.

As $\zeta_{tw}(\cdot)$ does not depend on t, we write the steady-state eligible fraction for wealth type w as

$$\zeta_w := \zeta_w(0) \in (0, 1].$$

All together, a policy implements a steady-state outcome summarized by $(\underline{\pi}_H, \underline{\pi}_L; \zeta_H, \zeta_L)$.

2.3 Agent behavior

Fix a wealth type $w \in \{w_L, w_H\}$ and let $\pi \in [0, 1]$ denote the agent's (period-t) privately observed usage probability. Define V_w as the stationary continuation value for a type-w agent who is eligible at the start of a period, and flow utility as $\hat{V}_w := (1 - \delta)V_w$.

If allocated, the agent pays price p; if the agent cancels, she gets a refund r and faces a ban captured by an *effective* discount factor $\delta_\tau = \mathbb{E}_\rho[\delta^\tau] \in [0, \delta]$.[6] Given (π, w), her utility from booking (b) and from not booking (n) in a given period can be written

$$\hat{V}^b(\pi, w) = \beta\Big[\pi\Big((1-\delta)\Big(1 - \frac{p}{w}\Big) + \delta\hat{V}_w\Big) + (1-\pi)\Big(-(1-\delta)\frac{p-r}{w} + \delta_\tau\hat{V}_w\Big)\Big] + (1-\beta)\delta\hat{V}_w, \tag{3}$$

$$\hat{V}^n(w) = \delta\hat{V}_w. \tag{4}$$

The booking cutoff $\underline{\pi}_w$ is defined by indifference between booking and not booking:

$$\hat{V}^b(\underline{\pi}_w, w) = \hat{V}^n(w) = \delta\hat{V}_w,$$

Hence the booking cutoff for type w is

$$\underline{\pi}_w = \frac{(\delta - \delta_\tau)\,\hat{V}_w + (1-\delta)\frac{p-r}{w}}{(\delta - \delta_\tau)\,\hat{V}_w + (1-\delta)\left(1 - \frac{r}{w}\right)}. \tag{5}$$

Note that if $p - r$ is held constant, then $\underline{\pi}_w$ is increasing in p: the less surplus is retained by the agent, the less likely they are to book. Relatedly, the lower the negative consequences of a temporary ban, the more willing the agent is to run the risk of punishment.

3 Results

All elided proofs are provided in the Appendix. We begin with preliminary results which apply regardless of whether bans or prices are analyzed. Henceforth we maintain the running assumption that the available mass of bookings C is small; we discuss this assumption following some preparatory results. We first show that when available capacity is small, booking thresholds are near 1.

Lemma 1 (Small capacity implies high thresholds)**.** *Let (μ_H, μ_L, w_H, w_L) be fixed. For any $\pi < 1$, there is $\bar{C} > 0$ such that when $C < \bar{C}$, equilibrium thresholds are $\underline{\pi}^\star_H, \underline{\pi}^\star_L > \pi$.*

If the booking threshold $\underline{\pi}_w$ does not go to 1 as C becomes small, expected usage relative to capacity remains strictly under 100%. Because the principal can impose prices which bring expected usage to 100%, low booking thresholds are inconsistent with usage maximization under low available capacity.

[6] Under our mechanism restriction in steady state, the utility loss from being banned is constant.

When C is small and than C agents book, the regulator can locally increase incentives and increase bookings without tightening the allocation constraint, thereby raising expected usage.

Lemma 2 (Weak overbooking is optimal)**.** *Provided the equilibrium steady-state population mass is at least C, it is never optimal to elicit less than mass C of bookings.*

Proof. Suppose underbooking occurs. Then $(1-\underline{\pi}_H)\mu_H\zeta_H(0)+(1-\underline{\pi}_L)\mu_L\zeta_L(0)<C$. In Lemma 5 we show that by slightly reducing prices the principal can increase $(1-\underline{\pi}_w)\zeta_w(0)$; since the base mechanism is underbooking, a marginal increase in booking mass does not affect the allocation probability $\beta=1$. Thus reducing prices strictly increases expected usage. □

Intuitively, by lowering the price, the principal converts her unused capacity while affecting neither the usage probability of the units she would have allocated anyway, nor the probability that a booking is allocated. In the presence of eligibility-based punishment reducing prices lowers the steady-state mass of eligible agents $\zeta_w(0)$, but we show that this decrease is of second-order importance when capacity C is small.

Lemmas 1 and 2 justify the role of the maintained assumption that C is small. When the principal employs eligibility-based punishment, a small decrease in booking thresholds may lead to a disproportionate decrease in the mass of eligible agents. The principal may keep booking thresholds high enough that the mechanism is under-booked, in order to maintain a larger steady-state mass of eligible agents. Our results show that this is only a concern when C is not small. Because this logic applies only in the presence of eligibility-based punishments, all results that hold when C is small hold when the principal is constrained to employ only monetary punishments.

In these environments there is a natural relationship between allocational inequality, defined as the difference between $\underline{\pi}_H$ and $\underline{\pi}_L$, and the principal's objective of maximum usage.

Proposition 1 (Effect of allocational inequality on principal's outcomes)**.** *Let A, the aggregate mass that books, be small. If A is held constant, expected usage increases as allocational inequality $|\underline{\pi}_H-\underline{\pi}_L|$ decreases.*

Lemma 2 and Proposition 1 illustrate the key tensions in our analysis. When there is wealth inequality, prices exclude too many low-wealth applicants while permitting too many high-wealth applicants. This cannot be optimal for the regulator, ceteris paribus, because Proposition 1 shows that the regulator in general prefers $\underline{\pi}_L=\underline{\pi}_H$. Second, although bans treat applicants equally across wealth types, bans exclude too many potential applicants (who are in ineligibility states). Because Lemma 2 says that the regulator generally desires at least C applications, removing applicant mass is equivalent to reducing the type threshold for application.

Formally, Proposition 1 justifies our later attention to mechanisms which use only monetary punishment, or only eligibility-based punishment. When the mechanism uses only monetary penalties the expected duration of ineligibility upon cancellation is $\sum_{t=1}^{T} t\rho(t)=0$, and Lemma 5 implies that the equilibrium mass of eligible agents $\zeta_w^\star(0)$ is maximized. On the other hand, when the mechanism uses only eligibility-based penalties equations (3) and 5 imply that booking thresholds are equal across wealth types. That is, there is no allocational inequality. It follows that these mechanisms represent the key tradeoffs faced by the principal.

Because the principal's objective is to maximize expected usage, the relationship between allocational inequality and usage derived in Proposition 1 implies that, all else equal, the principal prefers mechanisms which bring $\underline{\pi}_H$ and $\underline{\pi}_L$ into alignment.

Our last preparatory result shows that optimal access-based punishment is structurally simple: if an agent cancels an allocated booking, they are punished with a ban randomized between $\tau^\star$ and $\tau^\star+1$ periods, for some $\tau^\star$. Observe that we can always hold the effective discount rate constant while moving punishment forward, which does not affect the decision to book. But moving punishment forward will increase the steady-state mass of potential applicants, which is optimal for the principal who wants to maximize steady-state eligible mass. Since δ^τ is strictly decreasing in τ, those two nonzero entries must lie on adjacent τ (or one if $\bar{\delta}/\delta = \delta^\tau$ exactly).

Proposition 2 (Optimal ineligibility duration)**.** *The optimal ineligibility-duration distribution $\rho^\star$ is supported by at most two durations. If it is supported by two durations, these durations are adjacent.*

Lemma 2 shows that, absent eligibility-based punishment, the principal can improve usage on the margin by incentivizing entry up to the value at which supply is reached. When we consider the incentives induced by eligibility-based punishment, from Proposition 2 we know the utility in the present period is higher than the same policy but delayed any number of periods.

3.1 No refunds

With slight abuse of notation, we now consider a policy (p, r) in light of the previous analysis; a ban of arbitrary length following Proposition 2 may be deployed in conjunction with the price regulation. We now show that refunds are the wrong way to encourage participation when facing wealth inequality; refunds neither ameliorate the distortion across wealth types nor screen on usage. Hence a principal optimally sets refunds to zero.

Proposition 3 (No refunds)**.** *Let A be small and fix any interior steady-state equilibrium that has a positive refund. Then there exists a local perturbation (dp, dr) with $dr < 0$ such that the adapted policy $(p+dp, r+dr)$ both (i) keeps the booking mass A fixed to first order, and (ii) strictly increases expected usage.*

The presence of refunds turns prices into a use-contingent instrument and therefore reduces the ability of prices to screen on usage probabilities. This is most clear at the extreme of full refunds, $r = p$. In this case the expected flow utility of an allocated booking is $(1 - p/w)\pi$, and agents will book as long as $p/w < 1$, independent of usage probability. Reducing refunds improves the screening ability of prices and therefore improves usage. Ironically, full refunds are perfectly fair and hence Proposition 1 suggests that this may be beneficial to expected usage. However, unlike eligibility-based penalties, which retain an ability to screen on usage, issuing full refunds in this model is equivalent to uniformly rationing allocation across all agents which is in general not usage-maximizing.

A key to the proof of Proposition 3 is the fact that refunds do not affect continuation utility beyond the incurred change in the threshold usage and allocation probability.

Lemma 3 (Refunds do not change continuation utility directly)**.** *Refunds affect continuation utility only through changes in $\underline{\pi}_w$ and A.*

So, conditional on booking thresholds $\underline{\pi}_w$ and booking mass A, the continuation utility is pinned down by the price and ban. Intuitively, if the principal were to raise r, she makes cancellation less painful in the flow payoff but there is an offsetting change in the effective continuation penalty.

3.2 Optimal regulation

Propositions 2 and 3 reduce the policy choice to a price p and a relative ban impact $(\delta - \mathbb{E}_\rho[\delta^\tau]) =: \kappa$. We first observe that there is no policy such that the threshold usage rate is higher for the high-wealth type than for the low-wealth type. Let $(p;\kappa)$ denote the policy — the price and relative ban impact — which the principal implements.

Proposition 4 (Threshold comparison)**.** *For all optimal policies $(p;\kappa)$, equilibrium usage thresholds satisfy $\underline{\pi}^\star_H \leq \underline{\pi}^\star_L$.*

We prove this by noting that there is no such price whose marginal impact on the high-wealth type is *higher* than for the low-wealth type. We thus preclude the possibility that, because the high-wealth type faces less effect of monetary punishments, their continuation utility is higher, and thus bans have a larger differential effect on them. Instead, prices have a consistent effect of screening more harshly on the low-wealth type.

We now consider a fixed economy and take the patience δ to 1. Provided inequality is not too high — that is, that μ_H is sufficiently small — access-based punishment generates strictly higher utility than the price-based mechanism.

Proposition 5 (Optimal mechanism with patient agents)**.** *Let $\mu_H < 1/2$. Holding fixed other market parameters, when δ is large, optimal eligibility-based punishment generates higher usage than optimal price-based mechanisms.*

As agents grow more patient, eligibility-based punishments become increasingly punitive in order to dissuade inefficient booking. In the patient limit agents remain ineligible with positive probability, and the eligible mass is lower under eligibility-based punishment than under monetary punishment. However, when C is small monetary punishments can only target high-wealth individuals. The tradeoff for the principal is whether to target lower-usage high-wealth individuals, or an endogenously smaller mass of all agents. As low-wealth agents become less competitive, eligibility-based punishments perform better because the unequal screening power of prices has an increasingly consequential impact on usage.

Proposition 5 brings δ to 1 holding other market parameters constant. We offer a partial counterpart describing optimal mechanisms when capacity is small.

Proposition 6 (Optimal mechanism with infinitesimal supply)**.** *Holding fixed other market parameters, when C is small, optimal price-based mechanisms generate higher usage than optimal eligibility-based mechanisms.*

When capacity is extremely small, the principal cannot use eligibility to meaningfully punish agents. The agent is unlikely to receive allocation in the foreseeable future whether or not they cancel an allocated booking and hence eligibility restrictions have limited screening power. The screening power of prices remains constant, although unfair, and price-based mechanisms generate higher usage.

We now remark on how the principal sets price p in light of there being no refunds in equilibrium. For a given level of inequality, it may be optimal for the principal to overbook her capacity.

Remark (Overbooking)**.** *The optimal policy may include a price p at which the booking mass is above C.*

If bookings are dominated by the high-wealth type, then the principal may prefer to set a policy which incentivizes the marginal low-wealth *and* high-wealth type to book, and thereby increase the fraction of low-wealth types among which the principal rations. Indeed, in an economy with sufficiently small capacity and sufficiently high inequality, there may be no such price where $\beta = 1$ *and* where the low-wealth type books. In that case, decreasing the price to encourage booking from the low-type, and then rationing among all booking agents, can increase principal surplus.

The clearest intuition for overbooking occurs in mechanisms with only monetary punishment, in which expected usage is

$$[\mu_H M(\underline{\pi}_H^\star) + \mu_L M(\underline{\pi}_L^\star)]\,\beta.$$

When the mass of bookings is below C, expected usage is strictly decreasing in $\underline{\pi}$. Hence an optimal mechanism in this class cannot underbook. Once the mass of bookings exceeds C the mechanism employs rationing and the change of usage with respect to $\underline{\pi}$ depends on the comparative relationship between the numerator and denominator, both of which decrease as $\underline{\pi}$ increases. Thus there may be an interior, overbooked optimal mechanism. Whether exact booking or overbooking is optimal depends on discrete comparison of local optima.

In Section 4 below we show that there exist economies where, absent bans, the principal will set the price low enough such that $\beta < 1$ in equilibrium. That is, overbooking is not merely a possibility. The example we develop there, where usage rates are drawn from a standard uniform distribution, confirms Proposition 5: We show that the capacity at which bans dominate prices is small, and further, we show that the interval where bans dominate prices is also the region where the overbooking price dominates the capacity-filling price.

4 A uniform economy

We now analyze an example to show how the instruments of prices and bans compare under inequality, in the "patient limit" $\delta \nearrow 1$.[7] Let the distribution of usage probabilities be uniform: in each period, each eligible agent draws a private usage probability $\pi \sim \text{Unif}[0,1]$. In the uniform distribution, our sufficient statistics from Subsection 2.1 are analytically tractable:

$$\alpha(\underline{\pi}) := \Pr(\pi \geq \underline{\pi}) = 1 - \underline{\pi}, \qquad M(\underline{\pi}) := \mathbb{E}[\pi\,\mathbf{1}\{\pi \geq \underline{\pi}\}] = \int_{\underline{\pi}}^{1} x\,dx = \frac{1-\underline{\pi}^2}{2},$$

To simplify expressions relating to wealth inequality, normalize the difference in the mass of wealth types and wealth levels as follows:

$$\mu_L = \frac{1}{1+\gamma}, \qquad \mu_H = \frac{\gamma}{1+\gamma}, \qquad \gamma > 0,$$

$$w_L = \frac{1}{1+\xi}, \qquad w_H = \frac{\xi}{1+\xi}, \qquad \xi \geq 1.$$

Under these normalizations γ captures inequality in composition, while ξ indexes wealth dispersion; jointly, the pair (γ, ξ) characterize wealth inequality in the economy. Inequality decreases as γ or ξ go to 1, and increases as γ or ξ go to ∞.

4.1 Prices

Suppose the principal may use only monetary punishments. From Proposition 3, the principal does not refund upon cancellation. When the principal has access to prices alone, she can directly influence threshold usage rates as a function of the price p she sets. The threshold condition for wealth-type w reduces to

[7] Steady-state equilibrium under monetary penalties is independent of patience. For equilibrium under eligibility restrictions we apply the tools developed in the proof of Proposition 5.

$\pi \geq \underline{\pi}_w(p) := \frac{p}{w}$, thus under the wealth normalization, becomes

$$\underline{\pi}_L(p) = \frac{p}{w_L} = \min\{1, (1+\xi)p\}, \qquad \underline{\pi}_H(p) = \frac{p}{w_H} = \frac{1+\xi}{\xi}\,p.$$

For simplicity, we typically consider $\underline{\pi}_L(p) = (1+\xi)p$ and then verify that this simplification is (or is not) consistent with $\underline{\pi}_L(\cdot) \leq 1$. Following Section 2, let $A(p)$ denote the mass of bookings and let $N(p)$ denote usage among those who book, each as a function of the price:

$$A(p) := \mu_L \alpha(\underline{\pi}_L(p)) + \mu_H \alpha(\underline{\pi}_H(p)), \qquad N(p) := \mu_L M(\underline{\pi}_L(p)) + \mu_H M(\underline{\pi}_H(p)).$$

We know from Lemma 2 that the principal allocates all of her capacity, so expected usage is given by

$$U^m(p) = \beta(p)N(p) = \frac{CN(p)}{A(p)} \tag{6}$$

where we used that $\beta(p) = \frac{C}{A(p)}$. With C fixed, maximizing $U^m(p)$ over $A(p) \geq C$ is equivalent to choosing p which satisfies $\beta(p) \leq 1$ and which maximizes (6). There are two candidates for the optimal price: the corner solution at which demand exactly meets supply, and the root which overbooks.[8] Thus we compare utility under the capacity-filling price p_C and the lower overbooking price p_{OB}.

We consider the corner solution first and derive price p_C. Setting $A(p) = C$ yelds

$$p_C(\xi, \gamma; C) = \frac{\xi(1+\gamma)}{(1+\xi)(\xi+\gamma)}(1-C). \tag{7}$$

At p_C we have $\beta = 1$, so $U^m(p_C) = N(p_C)$ and hence the utility when capacity is *exactly filled* is

$$U^{m,\mathrm{EF}}(\xi, \gamma; C) = \frac{1}{2}\Big(1 - \mathcal{K}(\xi,\gamma)(1-C)^2\Big), \qquad \mathcal{K}(\xi,\gamma) := \frac{(1+\gamma)(\gamma+\xi^2)}{(\gamma+\xi)^2}.$$

$\mathcal{K}$ is a measure of wealth inequality. Larger wealth dispersion ξ makes p/w a noisier proxy for π, which increases the inequality measure $\mathcal{K}$ and lowers welfare. We defer comparative statics on $\mathcal{K}$ to the end of this subsection.

The second candidate for the optimal price has $A(p) \geq C$ slack. Differentiating (6) delivers the root

$$p_{\mathrm{OB}}(\xi, \gamma) = \frac{\xi(1+\gamma)}{(1+\xi)(\xi+\gamma)}\left[1 - \frac{(\xi-1)\sqrt{\gamma}}{\sqrt{(1+\gamma)(\xi^2+\gamma)}}\right]. \tag{8}$$

Observe the relationship between (7) and (8): The overbooking price is a lower price, and thus overbooks, if $\frac{(\xi-1)\sqrt{\gamma}}{\sqrt{(1+\gamma)(\xi^2+\gamma)}} > C$. Define this critical capacity level as

$$C^\star(\xi,\gamma) := A\big(p_{\mathrm{OB}}(\xi,\gamma)\big) = \frac{(\xi-1)\sqrt{\gamma}}{\sqrt{(1+\gamma)(\xi^2+\gamma)}}. \tag{9}$$

Then the interior point p_{OB} is feasible if and only if $A(p_{\mathrm{OB}}) \geq C$, i.e. $C \leq C^\star(\xi,\gamma)$. Plug the overbooking

[8] By this construction, it is possible that a root sets the price p such that capacity is underbooked, but this violates Lemma 2. Note that the overbooking root may not exist, in which case the corner solution is optimal.

price p_{OB} into (6) to yield utility

$$U^{m,\text{OB}}(\xi,\gamma;C) = C \cdot \frac{N(p_{\text{OB}})}{A(p_{\text{OB}})} = C \cdot \frac{1}{1+C^\star(\xi,\gamma)} = \frac{C}{1+C^\star(\xi,\gamma)}.$$

Hence the value of $C^\star(\xi,\gamma)$ determines the optimal price and the subsequent utility level:

$$U^{m,\star}(\xi,\gamma;C) = \begin{cases} \dfrac{C}{1+C^\star(\xi,\gamma)} & \text{if } C < C^\star(\xi,\gamma), \\ \dfrac{1}{2}\Big(1-\mathcal{K}(\xi,\gamma)\,(1-C)^2\Big) & \text{if } C \geq C^\star(\xi,\gamma). \end{cases} \tag{10}$$

The planner's pricing problem therefore becomes a trade-off between (i) using a high price p_C to exactly fill capacity, which avoids rationing but worsens selection distortions as ξ rises (captured by the inequality term $\mathcal{K}(\xi,\gamma)$), and (ii) using a lower price p_{OB} that deliberately overbooks so that rationing substitutes for precise screening when the price signal is too noisy.

We now consider comparative statics on (9).

Observation 1. *As wealth inequality decreases, $\xi \to 1$, exact booking becomes preferred to overbooking, $C^* \to 0$.*

When there is no gap in wealth level the principal has no incentive to overbook, because the principal can target usage thresholds across wealth types using the same price. This argument remains valid when wealth inequality is small. Conversely, increasing levels of wealth inequality via ξ always makes the principal more likely to overbook.

Observation 2. *As the population distribution of wealth becomes imbalanced, $\gamma \to 0$ or $\gamma \to \infty$, exact booking becomes preferred to overbooking, $C^* \to 0$.*

When there are few high-wealth types the principal cannot clear the booking market without setting a price low enough to attract at least some low-wealth types; hence exact-fill pricing always wins. Similarly, when there are so many high-usage, high-wealth types that the principal can clear the booking market while never attracting any low-wealth types, exact booking always wins. For intermediate γ overbooking patterns are nonmonotone. As γ increases from 0 to ξ, more rich agents are willing to book and then cancel, and overbooking can recover the low-wealth types disincentivized from booking by a price which exactly clears the booking market. Once $\gamma > \xi$, the applicant pool becomes skewed toward the rich so the benefit of overbooking shrinks, hence C^* falls, on (ξ, ∞).

4.2 Bans

Absent money, wealth types behave symmetrically and use a common cutoff $\underline{\pi}$. The principal can implement any cutoff $\underline{\pi}$ subject to filling capacity.[9] Expected usage per period equals

$$U^b = \beta \cdot \zeta(0) \cdot M(\underline{\pi}) = \beta \cdot \zeta(0) \cdot \frac{1-\underline{\pi}^2}{2}.$$

[9] Overbooking is never optimal in a mechanism which imposes eligibility-based penalties. See the discussion in Lemma 5 for details.

Under the benchmark, the principal can choose a ban policy so that capacity is exactly filled without rationing, so $A^b = C$ and hence $\beta = 1$. Let τ be the ban length.[10] Then

$$\zeta(0)M(\underline{\pi}) = \frac{1-\underline{\pi}^2}{2+\tau(1-\underline{\pi})^2},$$

which the agent chooses $\underline{\pi}$ to maximize, hence $\tau(1-\underline{\pi})^2 = 2\underline{\pi}$. Plugging into $\zeta(0)$ gives

$$\zeta(0) = \frac{1}{1+\underline{\pi}}.$$

So for any cutoff $\underline{\pi}$ that is induced by some ban length τ, the implied steady-state eligible mass is $\zeta(0) = 1/(1+\underline{\pi})$. Equivalently, the principal can implement any $\underline{\pi} \in (0,1)$ by choosing $\tau = 2\underline{\pi}/(1-\underline{\pi})^2$.

Apply the fact that capacity is filled exactly under the ban policy, so that $C = (1-\underline{\pi})\zeta(0)$ implies

$$\frac{1-\underline{\pi}}{1+\underline{\pi}} = C \quad \Longrightarrow \quad \underline{\pi}^{b,\star}(C) = \frac{1-C}{1+C}, \quad \zeta^{b,\star}(0) = \frac{1}{1+\underline{\pi}^{b,\star}(C)} = \frac{1+C}{2}.$$

Expected usage per period equals allocated mass times expected usage probability conditional on booking:

$$U^b = C \cdot \mathbb{E}[\pi \mid \pi \geq \underline{\pi}] = C \cdot \frac{M(\underline{\pi})}{\alpha(\underline{\pi})} = C \cdot \frac{\frac{1-\underline{\pi}^2}{2}}{1-\underline{\pi}} = C \cdot \frac{1+\underline{\pi}}{2}. \tag{11}$$

and plugging $\underline{\pi}^{b,\star}(C) = \frac{1-C}{1+C}$ into (11) gives

$$U^{b,\star}(C) = C \cdot \frac{1+\underline{\pi}^{b,\star}(C)}{2} = C \cdot \frac{1+\frac{1-C}{1+C}}{2} = \frac{C}{1+C}. \tag{12}$$

Comparing (10) and (12), observe that

$$U^{b,\star}(C) > U^{m,\star}(\xi,\gamma;C) \iff C < C^\star(\xi,\gamma).$$

Observation 3. *Eligibility-based punishment is preferred to price-based allocation if and only if the price-based mechanism is not in the overbooking regime.*

Thus in our uniform example the level of inequality at which overbooking becomes optimal is the level of inequality at which bans begin to dominate the price regime. When inequality grows, a principal would prefer to either overbook and subsequently ration, or to threaten agents with a loss of access. Figure 1 illustrates this transition.

5 Conclusion

We study an empirically relevant game: a principal wants to allocate resources to a group of agents who vary in their value for money, but whose wealth levels affect the incentive properties of regulation. We motivate the use of access restrictions, especially under wealth inequality: the principal is better off threatening agents with temporary removal from the game than screening using prices. We highlight a tradeoff between a designer maintaining competition and equalizing punishment across the population.

[10] We argue in the proof of Proposition 2 that when δ is large it is sufficient to consider deterministic bans.

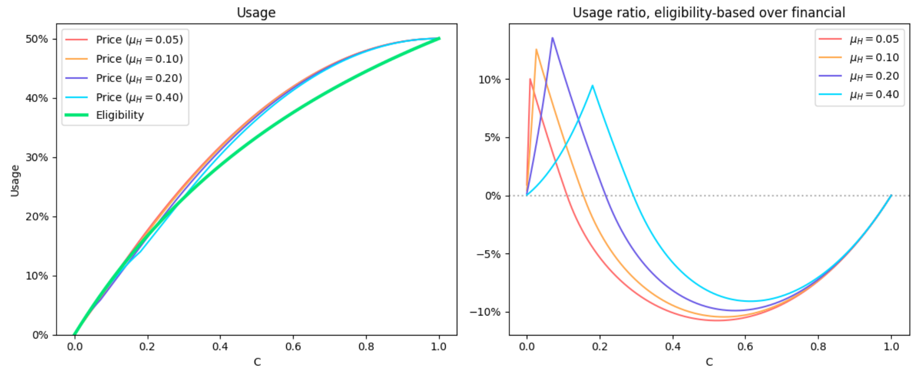

Figure 1: Performance of financial versus eligibility-based penalties. The left panel presents usage by mechanism, and the differences are marginal for low capacities C. The right panel presents relative usage $U^{b,\star}/U^{m,\star}$, and clearly illustrates Proposition 5. In both plots, $w_H/w_L = 2$.

We note that our analysis is artificially restrictive: in practice, platform designers are not constrained to select between financial and eligibility-based punishments, and may select a more exotic mechanism. We leave this possibility unaddressed for two reasons. First, we design with an eye toward practical applications, where monetary penalties and temporary bans are common features of reservation markets.[11] Second, in the general mechanism design problem the principal may be able to leverage the history of each agent's actions to determine whether they are high- or low-wealth, and discriminate appropriately in steady-state. To the best of our knowledge, this does not correspond to the mechanisms observed in practice. We note that similar difficulties arise when usage propensities persist over time, or are potentially correlated with wealth level—for instance, if wealthier agents have stronger outside options compared to the common value of usage.

References

Mohammad Akbarpour, Piotr Dworczak, and Scott Duke Kominers. 2024. Redistributive allocation mechanisms. *Journal of Political Economy* 132, 6 (2024), 1831–1875.

Justin Burkett and Kyle Woodward. 2024. Intertemporal Allocation with Unknown Discounting. (2024).

Daniele Condorelli. 2013. Market and non-market mechanisms for the optimal allocation of scarce resources. *Games and Economic Behavior* 82 (November 2013), 582–591.

Piotr Dworczak. 2024. Inequality and Market Design. *ACM SIGecom Exchanges* 22, 1 (June 2024), 83–92. https://doi.org/10.1145/3699824.3699831

George Georgiadis and Christopher S. Tang. 2014. Optimal reservation policies and market segmentation. *International Journal of Production Economics* 154 (Aug. 2014), 81–99. https://doi.org/10.1016/j.ijpe.2014.04.021

[11]Cf. our leading example of California state campsites. Additionally, the online restaurant reservation platform OpenTable states that the platform, "[...] shall have no liability for any charges made to the debit or credit card account for any failure to cancel your reservation in accordance with a Restaurant's cancellation policy ... [Additionally,] your Account will be suspended if you are a no-show for four reservations within a 12-month period." US (2026)

Matthew O. Jackson and Hugo F. Sonnenschein. 2007. Overcoming Incentive Constraints by Linking Decisions. *Econometrica* 75, 1 (2007), 241–257.

Qianjun Lyu. 2024. Optimal Refund Mechanism with Consumer Learning. https://doi.org/10.48550/arXiv.2404.14927 arXiv:2404.14927 [econ].

Hongyao Ma, Reshef Meir, David C. Parkes, and James Zou. 2018. Contingent Payment Mechanisms for Resource Utilization. https://doi.org/10.48550/arXiv.1607.06511 arXiv:1607.06511 [cs].

Mikael Mäkimattila. 2025. Inequality-Aware Regulation. (2025).

California State Parks. [n. d.]. New Reservations Policy FAQ — parks.ca.gov. https://www.parks.ca.gov/?page_id=31977. [Accessed 07-02-2026].

Rohit Patel and Can Urgun. 2024. Costly Verification and Money Burning. (2024).

Parag A Pathak, Tayfun Sönmez, M Utku Ünver, and M Bumin Yenmez. 2021. Fair allocation of vaccines, ventilators and antiviral treatments: leaving no ethical value behind in health care rationing. In *Proceedings of the 22nd ACM Conference on Economics and Computation*. 785–786.

Gregory Thomas. [n. d.]. New California law to bring big changes for campsite reservations — sfchronicle.com. https://www.sfchronicle.com/outdoors/article/california-law-parks-camping-18415520.php. [Accessed 07-02-2026].

OpenTable US. 2026. OpenTable Terms of Use - OpenTable. https://www.opentable.com/c/legal/terms-and-conditions/. [Accessed 09-02-2026].

Valentin Verdier and Carson Reeling. 2022. Welfare effects of dynamic matching: An empirical analysis. *The Review of Economic Studies* 89, 2 (2022), 1008–1037.

A Preliminary results

Proof of Lemma 1. Fix $\pi < 1$. Because $\underline{\pi}_w^\star \in [0,1]$ for all C, we may assume that $\underline{\pi}_w^\star \to P_w$ and $C \searrow 0$ for $w \in \{L, H\}$. If $P_L, P_H > 0$, then the principal can improve usage by increasing prices (this has the effect of increasing the expected usage rate of booking agents, $M(\underline{\pi}_w^\star)/\alpha(\underline{\pi}_w^\star)$, the reservation acceptance probability $\beta^\star$, and the mass of eligible agents $\zeta_w^\star(0)$). Then assume that w and w' are such that $P_{w'} = 0$ and $P_w > 0$. Expected usage along this sequence of equilibria is

$$U^\star = \mu_w M\left(\underline{\pi}_w^\star\right)\beta^\star\zeta_w^\star\left(0\right) + \mu_{w'} M\left(\underline{\pi}_{w'}^\star\right)\beta^\star\zeta_{w'}^\star\left(0\right).$$

By the assumption that $P_w > 0$, the mechanism employs rationing when C is sufficiently small. Then

$$\beta^\star = \left[\mu_w \alpha\left(\underline{\pi}_w^\star\right)\zeta_w^\star\left(0\right) + \mu_{w'}\alpha\left(\underline{\pi}_{w'}^\star\right)\zeta_{w'}^\star\left(0\right)\right]^{-1} C.$$

Consider an alternative mechanism which eliminates bans in favor of a price $p = w_H F^{-1}(1 - C/\mu_H)$. When C is small ($C < \mu_H$) this price is well defined, and such that $p > w_L$ so that only high-wealth agents will book. Expected usage in this mechanism is $U^p = \mu_H M \circ F^{-1}(1 - C/\mu_H)$.

It is then sufficient to compare $U^{\star} \gtrless U^{p}$ when C is small. That is,

$$\lim_{C\searrow 0}\frac{\mu_w M\left(\underline{\pi}_w^{\star}\right)\zeta_w^{\star}\left(0\right)+\mu_{w'}M\left(\underline{\pi}_{w'}^{\star}\right)\zeta_{w'}^{\star}\left(0\right)}{\mu_w\alpha\left(\underline{\pi}_w^{\star}\right)\zeta_w^{\star}\left(0\right)+\mu_{w'}\alpha\left(\underline{\pi}_{w'}^{\star}\right)\zeta_{w'}^{\star}\left(0\right)} \gtrless \lim_{C\searrow 0}\frac{1}{C}\mu_H M\circ F^{-1}\left(1-\frac{C}{\mu_H}\right).$$

Since $P_w > 0$ and $P_{w'} = 0$, the left-hand side is $\mathbb{E}[\pi|\pi \geq \underline{\pi}_w] < 1$. The right-hand side is $\frac{d}{dx}M(F^{-1}(1-x))|_{x=0} = 1$. It follows that $U^p > U^{\star}$ when C is small, contradicting the assumed optimality of the mechanisms inducing $(\underline{\pi}_L^{\star}, \underline{\pi}_H^{\star})$. □

Lemma 4 (Steady-State Eligibility)**.** *The steady-state equilibrium mass of agents who are eligible to book (i.e., not in an access-restriction state) is*

$$\zeta_w^{\star}\left(0\right)=\left[1+\sum_{t=1}^{T}\left(\alpha\left(\underline{\pi}_w\right)-M\left(\underline{\pi}_w\right)\right)\beta^{\star}t\rho\left(t\right)\right]^{-1}.$$

Proof. Let $\zeta_w^{\star}(z)$ be the steady-state mass of wealth-type w with state variable z, and let $\lambda_{wzz'}^{\star}$ be the steady-state probability that a wealth-w agent transitions from state variable z to state variable z'. Note that $\lambda_{wzz'} \neq 0$ only if $z = 0$ or $z = z'+1$. Then $\zeta_w^{\star}(T) = \lambda_{w0T}^{\star}\zeta_w^{\star}(0)$, and $\zeta_w^{\star}(t) = \lambda_{w0t}^{\star}\zeta_w^{\star}(0)+\lambda_{w,t+1,t}^{\star}\zeta_w^{\star}(t+1)$ for all $t < T$. Since the state variable z counts down the time until the agent may book again, $\lambda_{w,t+1,t}^{\star} = 1$ for all w, t, and hence $\zeta_w^{\star}(t) = \lambda_{w0t}^{\star}\zeta_w^{\star}(0) + \zeta_w^{\star}(t+1)$. Applying recursion gives

$$\zeta_w^{\star}\left(t\right)=\sum_{t'=t}^{T}\lambda_{w0t'}^{\star}\zeta_w^{\star}\left(0\right).$$

Since $\zeta_w^{\star}$ is a probability distribution, this implies that

$$\sum_{t=0}^{T}\sum_{t'=t}^{T}\lambda_{w0t'}^{\star}\zeta_w^{\star}\left(0\right)=1 \quad\Longrightarrow\quad \zeta_w^{\star}\left(0\right)=\left[\sum_{t=0}^{T}\sum_{t'=t}^{T}\lambda_{w0t'}^{\star}\right]^{-1}=\left[\sum_{t=0}^{T}\left(t+1\right)\lambda_{w0t}^{\star}\right]^{-1}.$$

The transition probabilities $\lambda_{w0t}^{\star}$ are defined by the probability that an agent cancels and is punished with a t-period ban. We have already argued that fraction $(\alpha(\underline{\pi}_w) - M(\underline{\pi}_w))\beta^{\star}$ of wealth-w agents will be subject to some (possibly nondegenerate) punishment arising from cancelation. Of these, fraction $\rho(t)$ will face a t-period suspension. Then $\lambda_{w0t}^{\star} = (\alpha(\underline{\pi}_w) - M(\underline{\pi}_w))\beta^{\star}\rho(t)$ for $t > 0$, and $\lambda_{w00}^{\star} = 1 - (\alpha(\underline{\pi}_w) - M(\underline{\pi}_w))\beta^{\star}(1-\rho(0))$. Then we have

$$\zeta_w^{\star}\left(0\right)=\left[1+\sum_{t=1}^{T}\left(\alpha\left(\underline{\pi}_w\right)-M\left(\underline{\pi}_w\right)\right)\beta^{\star}t\rho\left(t\right)\right]^{-1}.$$

□

Lemma 5 (Effect of Threshold on Booking Outcomes)**.** *When π is large, both $\zeta_0^{\star}(\cdot)\alpha(\cdot)$ and $\zeta_0^{\star}(\cdot)M(\cdot)$ are decreasing.*

Proof. From Lemma 4,

$$\frac{d}{d\pi}\zeta_0^{\star}\left(\pi\right)=-K\left(1-\pi\right)\alpha'\left(\pi\right)\zeta_0^{\star}\left(\pi\right)^2,\quad K=\sum_{t=1}^{T}\beta^{\star}t\rho\left(t\right).$$

Then the derivative of $\zeta_0^\star(\cdot)\alpha(\cdot)$ is

$$\begin{aligned}\frac{d}{d\pi}\left[\zeta_0^\star(\pi)\,\alpha(\pi)\right] &= -K(1-\pi)\,\alpha'(\pi)\,\zeta_0^\star(\pi)^2\,\alpha(\pi)+\alpha'(\pi)\,\zeta_0^\star(\pi)\\ &\overset{\text{sgn}}{=} K(1-\pi)\,\alpha(\pi)-\zeta_0^\star(\pi)^{-1}\\ &= K(1-\pi)\,\alpha(\pi)-\left[1+K(\alpha(\pi)-M(\pi))\right]\\ &= K(M(\pi)-\pi\alpha(\pi))-1.\end{aligned}$$

As $\pi\nearrow 1$, $M(\pi)-\pi\alpha(\pi)\searrow 0$, establishing the first result.

Similarly, the derivative of $\zeta_0^\star(\cdot)M(\cdot)$ is

$$\begin{aligned}\frac{d}{d\pi}\left[\zeta_0^\star(\pi)\,M(\pi)\right] &= -K(1-\pi)\,\alpha'(\pi)\,\zeta_0^\star(\pi)^2\,M(\pi)+M'(\pi)\,\zeta_0^\star(\pi)\\ &\overset{\text{sgn}}{=} K(1-\pi)\,M(\pi)-\pi\zeta_0^\star(\pi)^{-1}\\ &= K(1-\pi)\,M(\pi)-\left[1+K(\alpha(\pi)-M(\pi))\right]\pi\\ &= K(M(\pi)-\pi\alpha(\pi))-\pi.\end{aligned}$$

As $\pi\nearrow 1$, $M(\pi)-\pi\alpha(\pi)\searrow 0$, establishing the second result. □

A.1 Proof of Proposition 1

If the principal can directly manipulate the thresholds $\underline{\pi}_w$, the steady-state distribution of agents in state $z_i=0$ depends only on $\underline{\pi}_w$ and not directly on wealth. We can thus write $\zeta(\underline{\pi}_w)$ for $\zeta_w(0;\underline{\pi}_w)$. The total mass of bookings is

$$\mu_H\zeta(\underline{\pi}_H)\,\alpha(\underline{\pi}_H)+\mu_L\zeta(\underline{\pi}_L)\,\alpha(\underline{\pi}_L).$$

Then an infinitesimal change in $\underline{\pi}$ that leaves the total mass of bookings unchanged is such that

$$\mu_H\frac{d}{d\underline{\pi}_H}\left[\zeta(\underline{\pi}_H)\,\alpha(\underline{\pi}_H)\right]d\underline{\pi}_H+\mu_L\frac{d}{d\underline{\pi}_L}\left[\zeta(\underline{\pi}_L)\,\alpha(\underline{\pi}_L)\right]d\underline{\pi}_L=0.$$

For simplicity, write $d/d\underline{\pi}_w=\zeta_w'\alpha_w+\zeta\alpha_w'$. Then the above condition can be written as

$$\frac{\mu_H d\underline{\pi}_H}{\mu_L d\underline{\pi}_L}=-\frac{\zeta_L'\alpha_L+\zeta_L\alpha_L'}{\zeta_H'\alpha_H+\zeta_H\alpha_H'}.$$

We want to show that

$$\mu_H\frac{d}{d\underline{\pi}_H}\left[\zeta(\underline{\pi}_H)\,M(\underline{\pi}_H)\right]d\underline{\pi}_H+\mu_L\frac{d}{d\underline{\pi}_L}\left[\zeta(\underline{\pi}_L)\,\alpha(\underline{\pi}_L)\right]d\underline{\pi}_L\geq 0.$$

This is equivalent to

$$\frac{\mu_H d\underline{\pi}_H}{\mu_L d\underline{\pi}_L}\geq-\frac{\zeta_L'M_L+\zeta_LM_L'}{\zeta_H'M_H+\zeta_HM_H'}.$$

Thus it is sufficient to show

$$\frac{\zeta_L'M_L+\zeta_LM_L'}{\zeta_H'M_H+\zeta_HM_H'}\geq\frac{\zeta_L'\alpha_L+\zeta_L\alpha_L'}{\zeta_H'\alpha_H+\zeta_H\alpha_H'}. \tag{13}$$

From Lemma 5 we have that both numerators and denominators are negative. Then it is sufficient to show

$$\frac{\zeta_L' M_L + \zeta_L M_L'}{\zeta_L' \alpha_L + \zeta_L \alpha_L'} \geq \frac{\zeta_H' M_H + \zeta_H M_H'}{\zeta_H' \alpha_H + \zeta_H \alpha_H'}.$$

Since $\underline{\pi}_L > \underline{\pi}_H$, for this it is in turn sufficient to show

$$\frac{d}{d\pi}\left[\frac{\zeta' M + \zeta M'}{\zeta' \alpha + \zeta \alpha'}\right] \geq 0.$$

Following the simplifications in Lemma 4, this is

$$\begin{aligned}
\frac{d}{d\pi}\left[\frac{-K(1-\pi)\alpha' M + \zeta^{-1} M'}{-K(1-\pi)\alpha'\alpha + \zeta^{-1}\alpha'}\right] &= \frac{d}{d\pi}\left[\frac{K(1-\pi)M - \pi\zeta^{-1}}{K(1-\pi)\alpha - \zeta^{-1}}\right] \\
&= \frac{d}{d\pi}\left[\frac{K(1-\pi)M - (1+K(\alpha - M))\pi}{K(1-\pi)\alpha - (1+K(\alpha-M))}\right] \\
&= \frac{d}{d\pi}\left[\frac{K(M-\pi\alpha) - \pi}{K(M-\pi\alpha) - 1}\right] \\
&= \frac{d}{d\pi}\left[1 + \frac{1-\pi}{K(M-\pi\alpha)-1}\right] \\
&\overset{\text{sgn}}{=} -(K(M-\pi\alpha)-1) - (1-\pi)K(M' - \alpha - \pi\alpha') \\
&= 1 - K[(M-\pi\alpha) - (1-\pi)\alpha] \\
&= 1 - K(M-\alpha).
\end{aligned}$$

This is positive when π is large, establishing the desired result. □

Proof of Proposition 2. The agent's decision depends on the punishment distribution ρ only through $\mathbb{E}_{\tau\sim\rho}[\delta^\tau]$.[12] We show that for any ρ, there is a $\rho^\star$ supported on at most two ban lengths such that $\mathbb{E}_{\tau\sim\rho}[\delta^\tau] = \mathbb{E}_{\tau\sim\rho^\star}[\delta^\tau]$ and $\zeta_{\rho 0}(\pi) \leq \zeta_{\rho^\star 0}(\pi)$. That is, the principal can move from punishment distribution ρ to punishment distribution $\rho^\star$ and increase the steady-state mass of eligible agents without affecting the agents' incentives.

From Lemma 4 in Appendix A, we write the steady state mass as

$$\zeta_{\rho 0}(\pi) = \left[1 + \sum_{\tau=1}^{T} (\alpha(\pi) - M(\pi))\beta^\star \tau \rho(\tau)\right]^{-1}$$

Since $(\alpha(\pi) - M(\pi))\beta^\star$ is constant in this argument, we write

$$\zeta_{\rho 0}(\pi) = \left[1 + K\sum_{\tau=1}^{T} \tau\rho(\tau)\right]^{-1}.$$

Then we solve

$$\max_{\rho'} \left[1 + K\sum_{\tau=1}^{T} \tau\rho'(\tau)\right]^{-1}, \text{ s.t. } \rho' \in \Delta_T \text{ and } \sum_{\tau=0}^{T} \delta^\tau \rho'(\tau) = D.$$

We replace the maximization of the inverse with the minimization of the non-inverse,

$$\min_{\tilde{\rho}} \sum_{\tau=1}^{T} \tau\tilde{\rho}(\tau), \text{ s.t. } \rho' \in \Delta_T \text{ and } \sum_{\tau=0}^{T} \delta^\tau \tilde{\rho}(\tau) = D.$$

[12] The agent's utility also depends on the allocation probability β. In steady-state equilibrium this probability affects stage and continuation utilities equally, hence the effect can be ignored here.

The Lagrangian of this problem with respect to $\tilde{\rho}(\tau)$ is

$$t - \lambda_\tau + \gamma + \delta^\tau \eta = 0,$$

where λ_τ is the multiplier on $\tilde{\rho}(\tau) \geq 0$, γ is the multiplier on $\sum_{\tau=0}^{T} \rho(\tau) = 1$, and η is the multiplier on $\mathbb{E}_{\tilde{\rho}}[\delta^\tau] = D$. If $\rho^\star(\tau) > 0$ then $\lambda_\tau = 0$, and hence $\tau + \delta^\tau \eta = -\gamma$ for all τ such that $\rho^\star(\tau) = 0$. Note that $\frac{d^2}{d\tau^2}[\tau + \delta^\tau \eta] = [\ln \delta]^2 \delta^\tau > 0$, hence $\tau + \delta^\tau \eta$ is convex. It follows that there are at most two values $\{\tau_1^\star, \tau_2^\star\} \ni \tau$ such that $\rho^\star(\tau) > 0$. Additionally, since λ_τ is weakly positive convexity implies that $|\tau_2^\star - \tau_1^\star| \in \{0, 1\}$.

The result follows from the observation that with excess bookings, the principal can improve their utility by increasing the booking threshold. □

B Refunds

In this section we work with continuation values V_w; all identities translate from the main text by multiplying by $(1 - \delta)$.

B.1 Proof of Lemma 3

Introduce new definitions:

$$H(\underline{\pi}) := \mathbb{E}\left[(1 - \pi)\mathbf{1}\{\pi \geq \underline{\pi}\}\right] = \int_{\underline{\pi}}^{1} (1 - x)dF(x) = \alpha(\underline{\pi}) - M(\underline{\pi}).$$

$$G(\underline{\pi}) := M(\underline{\pi}) - \underline{\pi}\alpha(\underline{\pi}) = \mathbb{E}\left[(\pi - \underline{\pi})\mathbf{1}\{\pi \geq \underline{\pi}\}\right] = \int_{\underline{\pi}}^{1} (x - \underline{\pi})dF(x).$$

Among those who book, $H(\underline{\pi})$ is the expected unused mass; among those who book, $G(\underline{\pi})$ is the expected excess usage probability above the cutoff. Now observe the incremental gain from booking relative to not booking can be written as

$$V_b(\pi, w) - \delta V_w = \beta \left(\pi - \frac{p}{w} + \frac{r}{w}(1 - \pi) - (1 - \pi)\kappa V_w\right).$$

Agents book iff $\pi \geq \underline{\pi}_w$. Hence stationarity implies

$$\begin{aligned} V_w &= \delta V_w + \mathbb{E}\left[\left(V_b(\pi, w) - \delta V_w\right)\mathbf{1}\{\pi \geq \underline{\pi}_w\}\right] \\ &= \delta V_w + \beta \mathbb{E}\left[\left(\pi - \frac{p}{w} + \frac{r}{w}(1 - \pi) - (1 - \pi)\kappa V_w\right)\mathbf{1}\{\pi \geq \underline{\pi}_w\}\right]. \end{aligned}$$

We can write this in terms of the definitions of α, M, H:

$$(1 - \delta)V_w = vM(\underline{\pi}_w) - \frac{p}{w}\alpha(\underline{\pi}_w) + \frac{r}{w}H(\underline{\pi}_w) - \kappa V_w H(\underline{\pi}_w). \tag{14}$$

At an interior cutoff $\underline{\pi}_w \in (0, 1)$, the agent is indifferent between booking and not booking, so

$$\begin{aligned} 0 &= V_b(\underline{\pi}_w, w) - \delta V_w \\ &= \beta \left(\underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1 - \underline{\pi}_w) - (1 - \underline{\pi}_w)\kappa V_w\right), \end{aligned}$$

where for the RHS to be zero we must have

$$(1-\underline{\pi}_w)\kappa V_w = \underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1-\underline{\pi}_w). \tag{15}$$

Multiply the Bellman equation (14) by $(1-\underline{\pi}_w)$:

$$\begin{aligned}(1-\delta)(1-\underline{\pi}_w)V_w &= \beta\Big[(1-\underline{\pi}_w)\Big(M - \frac{p}{w}\alpha + \frac{r}{w}H\Big) - (1-\underline{\pi}_w)\kappa V_w H\Big] \\ &= \beta\Big[(1-\underline{\pi}_w)\Big(M - \frac{p}{w}\alpha + \frac{r}{w}H\Big) - \Big(\underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1-\underline{\pi}_w)\Big)H\Big],\end{aligned} \tag{16}$$

where in the second line we substituted $(1-\underline{\pi}_w)\kappa V_w$ from (15) (and we write $M = M(\underline{\pi}_w)$, $\alpha = \alpha(\underline{\pi}_w)$, $H = H(\underline{\pi}_w)$ for brevity).

Now expand the bracket in (16) and observe the refund terms cancel:

$$\begin{aligned}&(1-\underline{\pi}_w)\Big(M - \frac{p}{w}\alpha + \frac{r}{w}H\Big) - \Big(\underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1-\underline{\pi}_w)\Big)H \\ &\quad= (1-\underline{\pi}_w)M - (1-\underline{\pi}_w)\frac{p}{w}\alpha + \underbrace{(1-\underline{\pi}_w)\frac{r}{w}H}_{\text{refund}} - \underline{\pi}_w H + \frac{p}{w}H - \underbrace{\frac{r}{w}(1-\underline{\pi}_w)H}_{\text{refund}} \\ &\quad= \Big((1-\underline{\pi}_w)M - \underline{\pi}_w H\Big) + \frac{p}{w}\Big(H - (1-\underline{\pi}_w)\alpha\Big).\end{aligned} \tag{17}$$

Using $H = \alpha - M$, simplify both terms in parentheses:

$$(1-\underline{\pi}_w)M - \underline{\pi}_w H = (1-\underline{\pi}_w)M - \underline{\pi}_w(\alpha - M) = M - \underline{\pi}_w\alpha = G(\underline{\pi}_w), \tag{18}$$

$$H - (1-\underline{\pi}_w)\alpha = (\alpha - M) - (1-\underline{\pi}_w)\alpha = \underline{\pi}_w\alpha - M = -G(\underline{\pi}_w). \tag{19}$$

Substituting (18)–(19) into (17) yields

$$(1-\underline{\pi}_w)\Big(M - \frac{p}{w}\alpha + \frac{r}{w}H\Big) - \Big(\underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1-\underline{\pi}_w)\Big)H = \Big(v - \frac{p}{w}\Big)G(\underline{\pi}_w).$$

And finally, we return to (16) and have the simple equality:

$$(1-\delta)(1-\underline{\pi}_w)V_w = \beta\Big(1 - \frac{p}{w}\Big)G(\underline{\pi}_w). \tag{20}$$

Then observe that refunds r cancel out of (20). □

B.2 Proof of Proposition 3

The strategy is as follows: we will choose some (dp, dr) with $dr < 0$ in such a way that perturbs the cutoffs and keeps the booking mass fixed. Then we will invoke Proposition 1.

From the proof of Lemma 3, multiply (20) by κ and use (15) to replace $\kappa(1-\underline{\pi}_w)V_w$:

$$\begin{aligned}(1-\delta)\Big(\underline{\pi}_w - \frac{p}{w} + \frac{r}{w}(1-\underline{\pi}_w)\Big) &= (1-\delta)\kappa(1-\underline{\pi}_w)V_w \\ &= \beta\kappa\Big(v - \frac{p}{w}\Big)G(\underline{\pi}_w).\end{aligned}$$

This is exactly the equality which characterizes an interior cutoff; we eliminated V_w using the stationarity condition. We now define a function Φ from this equation which satisfies $\Phi_w(\underline{\pi}_w; p, r) = 0$:

$$\Phi_w(\underline{\pi}; p, r) := (1-\delta)\left(\underline{\pi} - \frac{p}{w}\right) + (1-\delta)(1-\underline{\pi})\frac{r}{w} - \beta\kappa\Big(1 - \frac{p}{w}\Big)G(\underline{\pi})$$

and which implies

$$\beta\kappa G(\underline{\pi}_w) = (1-\delta)\frac{\underline{\pi}_w - \frac{p}{w} + (1-\underline{\pi}_w)\frac{r}{w}}{1 - \frac{p}{w}}. \tag{21}$$

Now we differentiate Φ_w with respect to its three parameters.

$$\Phi_{w,\underline{\pi}}(\underline{\pi}; p, r) = (1-\delta)\left(1 - \frac{r}{w}\right) + \beta\kappa\Big(1 - \frac{p}{w}\Big)\alpha(\underline{\pi}), \tag{22}$$

$$\Phi_{w,p}(\underline{\pi}; p, r) = \frac{1}{w}\Big(-(1-\delta) + \beta\kappa G(\underline{\pi})\Big). \tag{23}$$

$$\Phi_{w,r}(\underline{\pi}; p, r) = (1-\delta)\frac{1-\underline{\pi}}{w}, \tag{24}$$

Substituting (21) into the price differential in (23) yields

$$\begin{aligned}\Phi_{w,p}(\underline{\pi}_w; p, r) &= \frac{1-\delta}{w}\left[-1 + \frac{\underline{\pi}_w - \frac{p}{w} + (1-\underline{\pi}_w)\frac{r}{w}}{1 - \frac{p}{w}}\right] \\ &= -\frac{1-\delta}{w} \cdot \frac{(1-\underline{\pi}_w)\left(1 - \frac{r}{w}\right)}{1 - \frac{p}{w}} < 0.\end{aligned} \tag{25}$$

Because $0 = \Phi_{w,\underline{\pi}} d\underline{\pi}_w + \Phi_{w,p} dp + \Phi_{w,r} dr$, the change in cutoffs satisfies

$$d\underline{\pi}_w = -\frac{\Phi_{w,p}}{\Phi_{w,\underline{\pi}}} dp - \frac{\Phi_{w,r}}{\Phi_{w,\underline{\pi}}} dr, \tag{26}$$

and now we substitute (24) and (25) into (26) and factor:

$$\begin{aligned}d\underline{\pi}_w &= \frac{(1-\delta)(1-\underline{\pi}_w)}{\Phi_{w,\underline{\pi}}(\underline{\pi}_w; p, r)}\left[\frac{1}{w}\frac{1 - \frac{r}{w}}{1 - \frac{p}{w}} dp - \frac{1}{w} dr\right] \\ &= \underbrace{\frac{(1-\delta)(1-\underline{\pi}_w)}{\Phi_{w,\underline{\pi}}(\underline{\pi}_w; p, r)} \cdot \frac{1}{w}\frac{1 - \frac{r}{w}}{1 - \frac{p}{w}}}_{=:\ \Gamma_w\ >\ 0}\Bigg(dp - \underbrace{\frac{1 - \frac{p}{w}}{1 - \frac{r}{w}}}_{=:\ \theta_w \le 1} dr\Bigg).\end{aligned} \tag{27}$$

Note two new variables: $\Gamma_w > 0$ because of (22), and $\theta_w \le 1$ because $r \le p$. Hence the local comparative static in $\underline{\pi}_w$ satisfies

$$d\underline{\pi}_w = \Gamma_w\big(dp - \theta_w dr\big). \tag{28}$$

So if the principal were to decrease the price, it would require decreasing the refund to keep $\underline{\pi}_w$ unchanged to first order. Observe that $\theta_H - \theta_L \ge 0$ so, holding fixed the change in refund, the compensating price change is larger in magnitude for the high-wealth type.

Now we will choose a joint perturbation (dp, dr) which keeps the total booking mass A fixed to first order and then show that, along that perturbation, usage increases. Define the per-type contributions under threshold $\underline{\pi}$ as

$$\mathcal{A}(\underline{\pi}) := \zeta(\underline{\pi})\alpha(\underline{\pi}), \qquad \mathcal{N}(\underline{\pi}) := \zeta(\underline{\pi})M(\underline{\pi}).$$

Then

$$A = \mu_H \mathcal{A}(\underline{\pi}_H) + \mu_L \mathcal{A}(\underline{\pi}_L), \qquad N = \mu_H \mathcal{N}(\underline{\pi}_H) + \mu_L \mathcal{N}(\underline{\pi}_L).$$

Differentiating A and substituting (28) gives

$$\begin{aligned} dA &= \mu_H \mathcal{A}'(\underline{\pi}_H) d\underline{\pi}_H + \mu_L \mathcal{A}'(\underline{\pi}_L) d\underline{\pi}_L \\ &= \mu_H \mathcal{A}'(\underline{\pi}_H) \Gamma_H \big(dp - \theta_H dr\big) + \mu_L \mathcal{A}'(\underline{\pi}_L) \Gamma_L \big(dp - \theta_L dr\big). \end{aligned} \tag{29}$$

Now we impose $dA = 0$ and define $\eta := dp/dr$. Solving (29) yields

$$\eta = \frac{\mu_H \mathcal{A}'(\underline{\pi}_H) \Gamma_H \theta_H + \mu_L \mathcal{A}'(\underline{\pi}_L) \Gamma_L \theta_L}{\mu_H \mathcal{A}'(\underline{\pi}_H) \Gamma_H + \mu_L \mathcal{A}'(\underline{\pi}_L) \Gamma_L}. \tag{30}$$

By assumption, $\underline{\pi}_w$ is large enough so that both $\mathcal{A}'(\underline{\pi}_H)$ and $\mathcal{A}'(\underline{\pi}_L)$ are negative, so every coefficient defined on the right-hand side of (30) is negative, so η is a strict convex combination of θ_H and θ_L. Hence

$$\theta_L < \eta < \theta_H. \tag{31}$$

Then for $dr < 0$, (28) and (31) together imply

$$d\underline{\pi}_H = \Gamma_H(\eta - \theta_H) dr > 0, \qquad d\underline{\pi}_L = \Gamma_L(\eta - \theta_L) dr < 0,$$

so there exists a local perturbation (dp, dr) with $dr < 0$ that reduces the gap $|\underline{\pi}_H - \underline{\pi}_L|$ to first order. The claim then follows from Proposition 1. □

B.3 Proof of Proposition 4

We will show that $\frac{d\underline{\pi}_w}{d\frac{p}{w}} > 0$ for any interior cutoff, and then use the fact that $w_H \geq w_L$ to conclude that $\underline{\pi}_H \leq \underline{\pi}_L$. Note the optimal policy has $r = 0$ and $\kappa = \delta - \delta_\tau$ for some τ. The agent books iff $\pi \geq \underline{\pi}_w$. The first steps continue the same as in Lemma 3. Stationarity implies

$$V_w = \delta V_w + \mathbb{E}\left[(V_b(\pi, w) - \delta V_w) \cdot \mathbf{1}\{\pi \geq \underline{\pi}_w\}\right]$$

Then use definitions $M(\underline{\pi}) = \mathbb{E}[\pi\, \mathbf{1}\{\pi \geq \underline{\pi}\}]$ and $\Pr(\pi \geq \underline{\pi}) = 1 - F(\underline{\pi})$ to obtain the following:

$$(1-\delta)V_w = \beta \left[vM(\underline{\pi}_w) - \frac{p}{w}(1 - F(\underline{\pi}_w)) - \kappa V_w((1 - F(\underline{\pi}_w)) - M(\underline{\pi}_w)) \right]. \tag{32}$$

At $\pi = \underline{\pi}_w$, the expected incremental gain from booking is zero, hence

$$\underline{\pi}_w - \frac{p}{w} = (1 - \underline{\pi}_w)\kappa V_w. \tag{33}$$

Multiply both sides of (32) by (33) rearranged for κ, and then simplify/substitute:

$$\begin{aligned}(1-\delta)\kappa V_w &= \beta\left[\kappa\, M(\underline{\pi}_w) - \kappa\frac{p}{w}\left(1-F(\underline{\pi}_w)\right) - \kappa^2 V_w\left(\left(1-F(\underline{\pi}_w)\right) - M(\underline{\pi}_w)\right)\right]\\ &= \beta\left[\kappa\, M(\underline{\pi}_w) - \kappa\frac{p}{w}\left(1-F(\underline{\pi}_w)\right) - \kappa\left(\frac{\underline{\pi}_w - \frac{p}{w}}{1-\underline{\pi}_w}\right)\left(\left(1-F(\underline{\pi}_w)\right) - M(\underline{\pi}_w)\right)\right]\\ &= \beta\kappa\left[\left(1-\frac{p}{w}\right)M(\underline{\pi}_w) - \underline{\pi}_w\left(1-\frac{p}{w}\right)\left(1-F(\underline{\pi}_w)\right)\right]\\ &= \beta\kappa\left(1-\frac{p}{w}\right)\left(M(\underline{\pi}_w) - \underline{\pi}_w\left(1-F(\underline{\pi}_w)\right)\right).\end{aligned}$$

Rewriting the LHS we get

$$(1-\delta)\left(\underline{\pi}_w v - \frac{p}{w}\right) = \beta\kappa\left(v-\frac{p}{w}\right)\left(M(\underline{\pi}_w) - \underline{\pi}_w\left(1-F(\underline{\pi}_w)\right)\right). \tag{34}$$

Use the function $G(\cdot)$ as defined in Appendix B.1 and note that $G(\underline{\pi}_w)\geq 0$. Now let $x := \frac{p}{w}$ where $x\in(0,v)$. Rearrange (34) and define $\Psi(\cdot)$ as follows:

$$\Psi(\underline{\pi}_w; x) := (1-\delta)(\underline{\pi}_w - x) - \beta\kappa(1-x)G(\underline{\pi}_w). \tag{35}$$

Then take its partials toward using the implicit function theorem:

$$\frac{\partial\Psi}{\partial\underline{\pi}_w}(\underline{\pi}_w; x) = (1-\delta)v - \beta\kappa(v-x)G'(\underline{\pi}_w) = (1-\delta)v + \beta\kappa(v-x)\left(1-F(\underline{\pi}_w)\right) > 0.$$

$$\frac{\partial\Psi}{\partial x}(\underline{\pi}_w; x) = -(1-\delta) + \beta\kappa G(\underline{\pi}_w) < 0$$

To see that $\frac{\partial\Psi}{\partial x}(\underline{\pi}_w; x)$ is negative, observe that at any interior solution we have $\Psi(\underline{\pi}_w; x) = 0$, so from (35) we get

$$\beta\kappa G(\underline{\pi}_w) = (1-\delta)\frac{\underline{\pi}_w v - x}{v-x} < (1-\delta).$$

Hence

$$\frac{d\underline{\pi}_w}{dx} = -\frac{\frac{\partial\Psi}{\partial x}}{\frac{\partial\Psi}{\partial\underline{\pi}_w}} > 0$$

at any interior solution, so $\underline{\pi}_w$ is strictly increasing in $x = p/w$. □

C Folk theorem

C.1 Proof of Proposition 5

For the moment, ignore the agent's wealth type w. Let $\delta^\star$ be the agent's expected discount rate conditional on punishment, $\delta^\star = \sum_{t=0}^{T}\delta^{t+1}\rho(t)$. Recall that the agent's value from booking (b) or not (n) is[13]

$$\begin{aligned}V^b(\pi) &= (1-\pi)\,\delta^\star\hat{V} + \pi\left((1-\delta)+\delta\hat{V}\right),\\ V^n(\pi) &= \delta\hat{V}.\end{aligned}$$

[13] Punishment occurs with probability $1-\pi$ because the construction of $\delta^\star$ incorporates the possibility of degenerate zero-period bans.

As derived in (5) the indifferent type $\underline{\pi}$ is such that

$$V^b(\underline{\pi}) = V^n(\underline{\pi}) \implies \underline{\pi} = \frac{(\delta-\delta^\star)\hat{V}}{(1-\delta)+(\delta-\delta^\star)\hat{V}} = \frac{\frac{\delta-\delta^\star}{1-\delta}\hat{V}}{1+\frac{\delta-\delta^\star}{1-\delta}\hat{V}}.$$

Ex ante, before the cancelation type π is realized, the agent's value is

$$\begin{aligned}\hat{V} &= \int_0^{\underline{\pi}} \delta\hat{V} dF(\pi) + \int_{\underline{\pi}}^1 (1-\pi)\delta^\star\hat{V} + \pi\left((1-\delta)+\delta\hat{V}\right) dF(\pi) \\ &= (1-\alpha(\underline{\pi}))\delta\hat{V} + \alpha(\underline{\pi})\delta^\star\hat{V} + \left((1-\delta)+(\delta-\delta^\star)\hat{V}\right) M(\underline{\pi}).\end{aligned}$$

This gives

$$((1-\delta)+(\delta-\delta^\star)(\alpha(\underline{\pi})-M(\underline{\pi})))\hat{V} = (1-\delta)M(\underline{\pi}) \implies \hat{V} = \frac{M(\underline{\pi})}{1+\frac{\delta-\delta^\star}{1-\delta}(\alpha(\underline{\pi})-M(\underline{\pi}))}.$$

It suffices now to show that when δ is large, there is some access-based punishment that improves the principal's utility over that obtained from a price-based mechanism. To see this, assume that there is $\eta \geq 0$ so that $\delta^\star = \delta^{\eta+1}$.[14] Then $\lim_{\delta\nearrow 1}(\delta-\delta^\star)/(1-\delta) = \eta$.

Because all involved terms are bounded, there is a subsequence of δ so that $M(\underline{\pi})$ and $\alpha(\underline{\pi})$ both converge. If $\lim_{\delta\nearrow 1} M(\underline{\pi}) = 0$, then $\lim_{\delta\nearrow 1}\hat{V} = 0$; this in turn implies that $\lim_{\delta\nearrow 1}\underline{\pi} = \eta$ and hence $\lim_{\delta\nearrow 1} M(\underline{\pi}) > 0$, a contradiction. Then it must be that $\lim_{\delta\nearrow 1} M(\underline{\pi}) := M^\star(\underline{\pi}) > 0$ and, since $\alpha(\cdot) \geq M(\cdot)$, $\lim_{\delta\nearrow 1}\alpha(\underline{\pi}) := \alpha^\star(\underline{\pi}) > 0$.

We thus have, in the limit,

$$\hat{V} = \frac{M^\star(\underline{\pi})}{1+\eta(\alpha^\star(\underline{\pi})-M^\star(\underline{\pi}))}, \text{ and } \underline{\pi} = \frac{\eta\hat{V}}{1+\eta\hat{V}}.$$

Together, these give

$$\underline{\pi}^\star = \frac{\eta M^\star(\underline{\pi}^\star)}{1+\eta\alpha^\star(\underline{\pi}^\star)} \implies \eta = \frac{\underline{\pi}}{M^\star(\underline{\pi}^\star)-\underline{\pi}^\star\alpha^\star(\underline{\pi})}. \tag{36}$$

It follows that for any threshold $\underline{\pi}^\star$ there is a η that achieves this threshold in the limit.

As a final step, we show that marginal increases of C from 0 result in higher realized value for the principal when they implement access-based punishments rather than prices. Note first that when C is small, price-based mechanisms serve only high-wealth agents. Then the optimal price-based wealth threshold $\underline{\pi}_H$ satisfies $1-F(\underline{\pi}_H) = C/\mu_H$, and $\frac{d}{dC}\underline{\pi}_H = -[\mu_H dF(\underline{\pi}_H)]^{-1}$.[15]

Analyzing $\frac{d}{dC}\underline{\pi}^\star$ is more involved. For sake of approximation, assume that $\delta^\star = \delta^{\eta+1}$ represents η periods of access restriction, whether or not η is an integer; Proposition 2 establishes that this approximation is reasonable. Applying Lemma 4 gives

$$\zeta_w^\star(0) = \left[1+(\alpha(\underline{\pi})-M(\underline{\pi}))\eta\right]^{-1}.$$

[14] This is a matter of letting T grow with δ, and assigning $\rho(t)$ as appropriate.

[15] Lemma 5 establishes that there is no rationing in allocation when C is small.

Then the optimal threshold under access-based punishment satisfies

$$1 - F(\underline{\pi}) = C/\zeta_w^\star(0) = [1 + (\alpha(\underline{\pi}) - M(\underline{\pi}))\eta]\, C$$
$$= \left[1 + \frac{(\alpha^\star(\underline{\pi}) - M^\star(\underline{\pi}))\,\pi}{M^\star(\underline{\pi}) - \underline{\pi}\alpha^\star(\underline{\pi})}\right] C = \underbrace{\left[\frac{(1-\underline{\pi}^\star)\, M^\star(\underline{\pi}^\star)}{M^\star(\underline{\pi}^\star) - \underline{\pi}^\star\alpha^\star(\underline{\pi}^\star)}\right]}_{h(\underline{\pi}^\star)} C.$$

We thus have

$$\frac{1 - F(\underline{\pi})}{1 - \underline{\pi}} = \frac{M(\underline{\pi})\, C}{M(\underline{\pi}) - \pi\alpha(\underline{\pi})}.$$

Note that the limit of the left-hand side, as $\underline{\pi} \nearrow 1$, is $dF(1)$. Then

$$dF(1) = \lim_{\underline{\pi}\nearrow 1} \frac{M(\underline{\pi})\, C}{M(\underline{\pi}) - \underline{\pi}\alpha(\underline{\pi})} = \lim_{\underline{\pi}\nearrow 1} \frac{M'(\underline{\pi})\, C + M(\underline{\pi})\frac{dC}{d\underline{\pi}}}{M'(\underline{\pi}) - \alpha(\underline{\pi}) - \underline{\pi}\alpha'(\underline{\pi})} = -\lim_{\underline{\pi}\nearrow 1} \frac{M'(\underline{\pi})\, C + M(\underline{\pi})\frac{dC}{d\underline{\pi}}}{\alpha(\underline{\pi})}.$$

We have already noted that $M(\underline{\pi})/\alpha(\underline{\pi}) = \mathbb{E}[\pi|\pi \geq \underline{\pi}]$. With $d^2F(\cdot)$ bounded, we also obtain

$$\lim_{\underline{\pi}\nearrow 1} \frac{M'(\underline{\pi})\, C}{\alpha(\underline{\pi})} = \lim_{\underline{\pi}\nearrow 1} \frac{M''(\underline{\pi})\, C + M'(\underline{\pi})\frac{dC}{d\underline{\pi}}}{\alpha'(\underline{\pi})} = \lim_{\underline{\pi}\nearrow 1} \frac{dC}{d\underline{\pi}}.$$

It follows that

$$dF(1) = -2\frac{dC}{d\underline{\pi}}.$$

By the inverse function theorem, $dC/d\pi = [d\underline{\pi}/dC]^{-1}$, giving

$$\left.\frac{d\underline{\pi}}{dC}\right|_{\underline{\pi}=1} = -\left[\frac{1}{2}dF(1)\right]^{-1} \gtrless -[\mu_H dF(1)]^{-1}.$$

Then $\frac{d\underline{\pi}}{dC}|_{C=0} > \frac{d\underline{\pi}_H}{dC}|_{C=0}$ when $\mu_H < \frac{1}{2}$, implying the desired result.

C.2 Proof of Proposition 6

Recall from (3) and (4) that

$$V^b(\pi) = \beta\left[\pi\left((1 - \delta(\left(1 - \frac{p}{w}\right) + \delta\hat{V}_w\right) + (1-\pi)\left(-(1-\delta)\left(\frac{p-r}{w}\right) + \delta_\tau\hat{V}_w\right)\right] + (1-\beta)\,\delta\hat{V}_w,$$
$$\hat{V}_w = \int_0^{\underline{\pi}_w} \delta\hat{V}_w dF(\pi) + \int_{\underline{\pi}_w}^1 V^b(\pi)\, dF(\pi).$$

As $C \searrow 0$, either $\underline{\pi} \nearrow 1$, in which case $\hat{V}_w \to 0$, or $\beta \searrow 0$, in which case $\hat{V}_w \to 0$. Appealing to Proposition 3, (5) implies that the limiting booking threshold is $\underline{\pi}_w = p/w$. Then a mechanism with only eligibility-based punishment offers no screening, while a mechanism with only monetary punishment offers efficient screening constrained to high-wealth agents. The latter generates higher usage, establishing the desired result.