Online Appendix

How Markets End: Insurance Exhaustion and the Limits of Priority

A Proofs

The proofs use the *marginal rationing convention* of Section 2: when surviving supply exactly exhausts the demand of the class containing the marginal reach applicant, the zero-mass boundary type may be served with the probability that makes her indifferent; this selection does not alter aggregate flows. The equilibrium, welfare, priority, and commitment results use the nondegenerate configuration of Assumption 3. The payoff and regularity assumptions remain in force throughout; Proposition 1 replaces only the nondegenerate configuration with its stated conditions. Phase-specific clauses apply on nonempty phases. The hysteresis result combines an analytic root construction with a numerical sign calculation; the remaining numbered claims are analytic.

A.1 Preliminaries

The two lemmas below reduce the strategic problem to the threshold s^* and the admission technology to the marginal admission rate p_m; later proofs build on this pair.

Proof of Lemma 1. Fix the aggregates, and let $c_2 = (s^* - q_2)^+$ be the safety cutoff. For $s < c_2$, Assumption 2 gives $p_1(s) = 0$. Under score priority or a global lottery, applying to either tier then leads to the same stage-2 eligibility after rejection; under buffered priority, a safety rejection has weakly higher priority than a failed reach application. With the stated tie-breaking convention, every such score applies to safety.

For $s \geq c_2$, a safety application is admitted and yields v_2. A reach application yields $U_1(s) = p_1(s)v_1 + (1-p_1(s))W_2(s)$, where $W_2(s)$ is the continuation value after reach rejection. Under score order W_2 is nondecreasing in s; under either lottery variant it is constant within the relevant class. Write $D(s) = U_1(s) - v_2$. For $s' > s$,

$$D(s') - D(s) \;=\; \big[p_1(s') - p_1(s)\big]\big(v_1 - W_2(s')\big) \;+\; \big(1 - p_1(s)\big)\big[W_2(s') - W_2(s)\big] \;\geq\; 0,$$

because p_1 and W_2 are nondecreasing and $W_2 \leq v_1 - d < v_1$. Thus strict safety choices lie below strict reach choices, and the zero set of D between them may be an interval. Selecting safety below one boundary point of that interval and reach above it gives a threshold best response. If the zero set has no positive length, every pure selection is threshold up to the zero-mass indifferent type. Together with the behavior below c_2, this proves the lemma. $\square$

With behavior reduced to a threshold, the reach tier's clearing condition determines the admission chances of the marginal applicant.

Proof of Lemma 2. Suppose the reach tier is oversubscribed at s^*: $1-s^* > q_1$. The admitted mass at cutoff κ is $A(\kappa) = \int_{s^*}^{1} p_1(s;\kappa)\,ds$, continuous in κ because p_1 is, and strictly decreasing in κ wherever $0 < A(\kappa) < 1 - s^*$. Since $A(\kappa) = 1 - s^* > q_1$ for κ low and $A(\kappa) = 0$ for $\kappa > 1 + \varepsilon$, the intermediate value theorem delivers $\kappa(s^*)$ with $A(\kappa(s^*)) = q_1$, and strict monotonicity on the relevant range delivers uniqueness. Because A is jointly continuous and strictly decreasing in its second argument, $\kappa(\cdot)$ is continuous, and $p_m(s^*) = p_1(s^*;\kappa(s^*))$ is continuous as a composition of continuous maps. The cutoff representation follows Azevedo and Leshno (2016). Weak monotonicity of p_m, and strict monotonicity once its value is interior, are imposed by Assumption 2. Under uniform noise both hold throughout the computed range (Online Appendix B); other noise distributions are covered by Assumption 2 directly. $\square$

A.2 Proofs for Section 3

The proof proceeds in three steps: the bang-bang structure of the marginal insurance value, the shape of the incentive function g, and the accounting identities locating the phase boundaries.

Proof of Theorem 1. Step 1: the marginal insurance value is ordered and bang-bang. Under score priority, surviving seats are assigned to stage-2 claimants in score order. Failed reach applicants occupy $[s^*, 1]$; safety-tier rejects, when the safety tier is oversubscribed, occupy $[0, s^* - q_2)$. The marginal reach applicant, score s^*, is therefore the last served among failed reach applicants and ahead of every reject. Write $S_1 = \rho\mu q_1$ for surviving reach-tier supply on the relevant oversubscribed-reach range. Away from the two allocation boundaries, her continuation value is

$$W_2^m = \begin{cases} 0, & S < D_f, \\ v_2 - d, & S \geq D_f > S_1, \\ v_1 - d, & S_1 \geq D_f. \end{cases}$$

At equality, the marginal-rationing convention selects the appropriate convex combination for the zero-mass boundary type. The third line can occur at off-equilibrium, sufficiently cautious thresholds. It cannot occur at or below s_a^*: the last clause of Assumption 3 gives

$$S_1 \le \mu q_1 < (1-\mu)(1-s_a^*-q_1) \le D_f(s) \qquad \text{for every } s \le s_a^*.$$

Along the equilibrium path, Step 3 below verifies the stronger fact $S_1 < D_f$, so every equilibrium rescue of the marginal applicant is indeed a safety seat.

Step 2: g increases within branches and jumps up across them. On the branch $W_2^m = 0$, $g_0(s) = p_m(s)v_1 - v_2$; on the branch $W_2^m = v_2 - d$, $g_1(s) = p_m(s)(v_1 - v_2 + d) - d$. On the reach-seat branch,

$$g_R(s) = d\, p_m(s) + (v_1 - v_2 - d).$$

All three are weakly increasing. The first two are strictly increasing on $[s_a^*, s_c^*]$, because their two roots require admission probabilities strictly between zero and one. The reach-seat branch is unavailable for $s \le s_a^*$ by Step 1; whenever it is available for $s > s_a^*$, it lies weakly above the already positive safety-seat branch because $\partial g/\partial W_2^m = 1 - p_m \ge 0$ (strictly when $p_m < 1$), and hence is positive. Moreover, $g_0(s) < 0$ and $g_1(s) < 0$ for $s < s_a^*$, while any continuation value weakly raises g above $g_0(s) > 0$ for $s > s_c^*$. Thus any equilibrium lies in $[s_a^*, s_c^*]$.

Let $h(s) = S(s) - D_f(s)$. When the safety tier is oversubscribed, $h(s) = \rho\mu(q_1 + q_2) - (1-\mu)(1-s-q_1)$ and $h'(s) = 1 - \mu > 0$. When it is undersubscribed,

$$h(s) = \rho[q_2 + \mu q_1 - (1-\mu)s] - (1-\mu)(1-s-q_1),$$

so $h'(s) = (1-\mu)(1-\rho) \ge 0$. Thus the unserved branch lies below the unique boundary $h = 0$ and the served branch above it whenever $\rho < 1$ and the boundary lies in the relevant score interval. At $\rho = 1$, h is constant on the undersubscribed region; its sign selects one branch unless $q_1 + q_2 = 1 - \mu$, the nongeneric knife edge. At the switch the incentive jumps by

$$\big[p_m(v_1 - v_2 + d) - d\big] - \big[p_m v_1 - v_2\big] \;=\; (1-p_m)\,(v_2 - d) \;>\; 0.$$

If $S_1 - D_f$ also crosses zero, it does so from below because $d(S_1 - D_f)/ds = 1 - \mu > 0$ on this range; continuation rises again from $v_2 - d$ to $v_1 - d$, producing another weakly positive jump and no new downcrossing. The ordered branch values therefore admit a unique stable upcrossing: an interior root s_c^* or s_a^*, or a pinned boundary at which marginal rationing makes the zero-mass boundary type indifferent.

Step 3: phase boundaries. At the cautious candidate, $S \le D_f$ evaluated at s_c^* reads $\rho\,\mu(q_1+q_2) \le (1-\mu)(1-q_1-s_c^*)$, i.e. $\rho \le \rho_0$. In that case s_c^* lies weakly on the unserved branch, $g(s_c^*) = p_m(s_c^*)\,v_1 - v_2 = 0$ by the definition of s_c^*, and Step 2 gives uniqueness: the cautious phase. For $\rho_0 < \rho < \rho^*$ we have $s^*_{\text{pin}} < s^*_c$, so $g < 0$ everywhere on the unserved branch ($p_m(s^*) < p_m(s^*_c) = v_2/v_1$ there), while on the served branch $g > 0$ because $s^*_{\text{pin}} > q_2 > s^*_a$ in this phase and hence $p_m(s^*_{\text{pin}}) > p_m(s^*_a) = d/(v_1 - v_2 + d)$. The equilibrium is therefore the pinned point: solving $S = D_f$, i.e. $\rho\mu(q_1 + q_2) = (1 - \mu)(1 - s^* - q_1)$, gives the closed form $s^*(\rho) = 1 - q_1 - \rho\mu(q_1 + q_2)/(1 - \mu)$. Differentiating, an additional unit of surviving supply raises failed-reach demand by exactly one unit: $(1 - \mu)\,|ds^*/d\rho| = \mu(q_1 + q_2)$, the exhaustion property. The pin reaches the safety-tier subscription boundary when $s^*(\rho) = q_2$, which solves to $\rho^* = (1 - \mu)(1 - q_1 - q_2)/(\mu(q_1 + q_2))$. Beyond it, any candidate with $s^* < q_2$ leaves the safety tier undersubscribed at stage 1 and supply becomes $S = \rho[(q_2 - s^*) + \mu(q_1 + s^*)]$. Define $h(s^*) = S - D_f$ on this region; $\partial h/\partial s^* = (1 - \mu)(1 - \rho) \geq 0$ and $\partial h/\partial \rho > 0$, so the root of $h = 0$ is strictly decreasing in ρ for $\rho < 1$. The same branch logic applies: while the root exceeds s^*_a the equilibrium rides it downward; once supply is slack at s^*_a the marginal applicant is served with certainty and the equilibrium is the full-insurance root, $p_m(s^*_a) = d/(v_1 - v_2 + d)$, obtained from the indifference $p\,v_1 + (1 - p)(v_2 - d) = v_2$. Hence $s^*(\rho)$ decreases toward s^*_a in the undersubscribed exhaustion phase. Finally, $\rho^* < 1$ if and only if $(1 - \mu)(1 - q_1 - q_2) < \mu(q_1 + q_2)$, the capacity-ample condition.

Solving the undersubscribed equation $h = 0$ explicitly gives

$$s^* = \frac{(1 - \mu)(1 - q_1) - \rho(q_2 + \mu q_1)}{(1 - \mu)(1 - \rho)}.$$

Setting this expression equal to s^*_a yields the stated ρ_a. The ordering $\rho_0 < \rho^*$ follows from $s^*_c > q_2$; when $\rho^* < 1$, the inequalities $\rho^* < \rho_a < 1$ follow respectively from $s^*_a < q_2$ and $q_1 + q_2 > 1 - \mu$. The primitive last clause of Assumption 3 is strongest at $(\rho, s^*) = (1, s^*_a)$ and implies $\rho\mu q_1 < D_f$ on every earlier branch: on either pinned branch $D_f = S > \rho\mu q_1$, and on the cautious branch $D_f \geq S$. Thus the marginal rescue used above is indeed a safety seat throughout the path. At the knife edge $\rho^* = 1$, h is flat on the undersubscribed region at the endpoint; the selected equilibrium depends on the treatment of the zero-mass boundary type, hence the theorem's restriction to generic uniqueness. □

The welfare corollaries evaluate this path branch by branch.

Proof of Corollary 1. Let $V = q_1 v_1 + q_2 v_2$. For $\rho \leq \rho^*$, both tiers fill in stage 1 and every surviving vacancy is assigned, so

$$W(\rho) = (1 - \mu)V + \rho\mu[V - d(q_1 + q_2)].$$

This gives the first derivative in the statement.

On the undersubscribed pinned branch, write $s = s^*(\rho)$. Stage 1 contributes $(1-\mu)(q_1 v_1 + s v_2)$; stage 2 contributes $\rho\mu q_1(v_1 - d) + \rho[q_2 - (1-\mu)s](v_2 - d)$. From Theorem 1,

$$\frac{ds}{d\rho} = -\frac{C}{(1-\mu)(1-\rho)^2}.$$

Differentiating and simplifying yields

$$\frac{dW}{d\rho} = \mu q_1(v_1 - v_2) - \frac{Cd}{(1-\rho)^2}.$$

Its derivative is $-2Cd/(1-\rho)^3 < 0$, so it has at most one zero, the candidate displayed in the corollary.

On the full-insurance branch, $s = s_a^*$ and failed-reach demand $D_f = (1-\mu)(1 - s_a^* - q_1)$ is fixed. Assumption 3 implies that a mass $\rho\mu q_1$ receives reach seats and the remaining $D_f - \rho\mu q_1$ receives safety seats. Thus $dW/d\rho = \mu q_1(v_1 - v_2) > 0$. Since welfare increases before the threshold, is strictly concave on the pinned undersubscribed branch, and increases on the final branch, the global maximum must lie in the stated finite set. The two strict inequalities eliminate the other candidates. If $\rho^* > 1$, the first formula applies throughout $[0, 1]$. □

Proof of Corollary 2. The proof organizes the shape established in the proof of Corollary 1.

Step 1: candidate reduction. On $[0, \rho^*)$ the derivative is the constant $\mu[q_1(v_1 - d) + q_2(v_2 - d)] > 0$, so no point below ρ^* is optimal. On $(\rho_a, 1]$ the derivative is $\mu q_1(v_1 - v_2) > 0$, so no point of $[\rho_a, 1)$ is optimal. On (ρ^*, ρ_a) the derivative $\varphi(\rho) = \mu q_1(v_1 - v_2) - Cd/(1-\rho)^2$ is strictly decreasing; its unique zero, when it exists in the interval, is $\hat{\rho}$ as displayed. Hence every global optimum lies in $\{\rho^*, \hat{\rho}, 1\}$, proving (i).

Step 2: exact welfare gaps. The definitions of ρ^* and ρ_a imply

$$x = 1 - \rho^* = \frac{C}{\mu q}, \qquad u = 1 - \rho_a = \frac{C}{q_2 + \mu q_1 - (1-\mu)s_a^*}.$$

Integrating the middle derivative from ρ^* to ρ_a and the final derivative from ρ_a to one gives

$$W(1) - W(\rho^*) = Ax - Cd\left(\frac{1}{u} - \frac{1}{x}\right).$$

Because

$$\frac{1}{u} - \frac{1}{x} = \frac{q_2 + \mu q_1 - (1-\mu)s_a^* - \mu q}{C} = \frac{(1-\mu)(q_2 - s_a^*)}{C},$$

the first identity in part (ii) follows. If $u < y < x$, integrate instead from $\hat{\rho} = 1 - y$:

$$
\begin{aligned}
W(1) - W(\hat{\rho}) &= A(1-\hat{\rho}) - Cd\left(\frac{1}{u} - \frac{1}{1-\hat{\rho}}\right) \\
&= Ay - Ay^2\left(\frac{1}{u} - \frac{1}{y}\right) = Ay\left(2 - \frac{y}{u}\right),
\end{aligned}
$$

where $Cd = Ay^2$. This proves (ii).

Step 3: selection. The candidate $\hat{\rho}$ lies in (ρ^*, ρ_a) exactly when $u < y < x$. There it strictly beats ρ^*, and the second identity says that it beats 1 exactly when $y > 2u$. If $y \geq x$, the middle derivative is nonpositive at the threshold and strictly decreases thereafter, so the comparison reduces to ρ^* against 1; the first identity gives the condition in part (iii). If $y < x$ but $y \leq u$, welfare rises throughout the middle branch; if $u < y \leq 2u$, its interior peak does not beat completion. Complete reliability is optimal in both cases. The equality cases follow from the same two identities.

Step 4: the merger boundary. Since $y = x$ if and only if

$$
d = \frac{Ax^2}{C} = d_2,
$$

d_2 is the algebraic value where $\hat{\rho}$ reaches ρ^*. This equality describes a candidate merger inside the present model only if $d_2 < d_4$ and the remaining nondegeneracy clauses continue to hold there. Under those domain conditions, at the merger

$$
W(1) - W(\rho^*) = Ax\left(2 - \frac{x}{u(d_2)}\right).
$$

Thus the capped point strictly beats completion exactly when $x > 2u(d_2)$. Under the strict inequality, continuity makes $\hat{\rho}$ uniquely optimal just below d_2 and ρ^* uniquely optimal just above it. Under the reverse strict inequality, 1 is uniquely optimal at d_2. At equality the two are tied. If $d_2 \geq d_4$, then $s_a^*(d_2) \geq q_2$ and the claimant composition used to derive u and the middle welfare derivative has already changed; likewise, failure of another maintained clause invalidates the same substitution. No policy boundary follows from an out-of-domain algebraic merger. This proves (iv).

Step 5: boundary completion. As $d \downarrow 0$, $y \downarrow 0$, whereas $u = C/[q_2 + \mu q_1 - (1-\mu)s_a^*]$ stays bounded away from zero. Hence $y < 2u$ for all sufficiently small d, and completion is optimal. Equivalently, the first welfare gap converges to $Ax > 0$; this argument does not require s_a^* to converge to zero. As $d \uparrow d_4$, $s_a^* \uparrow q_2$ and $u \uparrow x$, while again $W(1) - W(\rho^*) \to Ax > 0$. If $y < x$ near the boundary, then $y < 2u$; if $y \geq x$, the positive first gap selects completion.

This proves (v). $\square$

A.3 Proofs for Section 4

The design theorem compares protocols by a single statistic: the probability that the marginal reach applicant is served at the cautious candidate. The proof shows that this statistic decides whether caution survives, then computes it protocol by protocol.

Proof of Theorem 2. Preliminary: any insurance breaks caution. At the cautious candidate, $p_m(s_c^*) = v_2/v_1$. Any positive continuation value $W_2^m > 0$ gives

$$g(s_c^*) \;=\; p_m(s_c^*)\big(v_1 - W_2^m\big) - \big(v_2 - W_2^m\big) \;=\; W_2^m\Big(1 - \frac{v_2}{v_1}\Big) \;>\; 0,$$

so the cautious outcome is an equilibrium if and only if the marginal applicant is unserved.

(a) Under score priority the marginal applicant has strictly positive service probability when $S > D_f$ and zero probability when $S < D_f$; at equality the marginal-rationing convention can select zero service to sustain the cautious root. Evaluated at s_c^*, caution therefore survives through $\rho = \rho_0$, so the largest feasible reliability sustaining it is $\min\{\rho_0, 1\}$. Under buffered priority the buffer is served first, failed reach applicants last, and within that class by score, so the marginal applicant is the last claimant in the entire market. The analogous boundary is $S = D_f + D_r$, again with zero service selected at equality. Evaluated at any threshold at which both classes are nonempty, demand telescopes:

$$D_f + D_r \;=\; (1-\mu)\big[(1 - s^* - q_1) + (s^* - q_2)\big] \;=\; (1-\mu)(1 - q_1 - q_2),$$

independent of s^*. Before truncation at one, buffered priority's erosion boundary is ρ^*, hence $\rho_{\text{buf}} = \min\{\rho^*, 1\}$. Under the lottery the marginal applicant receives a positive share of the value of the best $\min\{S, D\}$ seats whenever $S > 0$, so erosion begins at once: $\rho_{\text{lot}} = 0$. Before truncation, $\rho_0 < \rho^*$ because $s_c^* > q_2$. Maximality follows from part (b) whenever $\rho^* < 1$ and is immediate from feasibility otherwise.

(b) For $\rho > \rho^*$, surviving supply at the cautious candidate satisfies $S = \rho\mu(q_1 + q_2) > (1-\mu)(1 - q_1 - q_2) = D_f + D_r$, total stage-2 demand. Every claimant is willing, since $v_2 - d > 0$. Non-wastefulness forces every claimant to be served under any priority order. The marginal applicant's continuation value is at least $v_2 - d$, and g is increasing in this value because $\partial g/\partial W_2^m = 1 - p_m > 0$. Therefore

$$g(s_c^*) \;\geq\; \frac{v_2}{v_1}(v_1 - v_2 + d) - d \;=\; \frac{(v_1 - v_2)(v_2 - d)}{v_1} \;>\; 0$$

by Assumption 1. The cautious outcome is not an equilibrium.

(c) The full-insurance target $\phi(d) = d/(v_1 - v_2 + d)$ has $\phi'(d) = (v_1 - v_2)/(v_1 - v_2 + d)^2 > 0$, so $s_a^* = p_m^{-1}(\phi(d))$ is strictly increasing in d by Assumption 2. For the full-insurance root to reach the cautious root would require $\phi(d) \geq v_2/v_1$, equivalent to $d(v_1 - v_2) \geq v_2(v_1 - v_2)$, i.e. $d \geq v_2$. This is exactly the boundary of, and lies outside the interior of, Assumption 1: the rescued safety seat has no positive continuation value there. □

Proof of Corollary 3.

At the cautious root, any positive continuation value makes the marginal reach applicant deviate (preliminary step of Theorem 2). Under a fixed claimant order, her service probability is weakly increasing in surviving supply, so the supplies that sustain caution form the interval $[0, B(\pi)]$. Substituting

$$S = \rho\mu(q_1 + q_2)$$

gives the displayed formula for ρ_π. A global lottery gives the marginal applicant positive service probability at every $S > 0$, hence $B(\text{lot}) = 0$. Score priority puts the other failed reach applicants ahead of her, giving the stated $B(\text{score})$. Buffered priority also interposes every safety reject, so the telescoping identity gives the stated $B(\text{buf})$. No non-wasteful claimant order can interpose more than the entire positive-mass claimant pool before the zero-mass marginal type, proving the upper bound. □

The impossibility in part (b) treats safety-tier seats as homogeneous; the dispersion proposition asks how far residual-seat quality relaxes that benchmark.

Proof of Proposition 1. The proof constructs the safety-vacancy ladder, establishes a unique stable crossing, and signs its response to dispersion. Because a deliberate safety application yields the constant v_H, while failed-reach continuation values are nondecreasing in score under the stated ordering, the threshold argument of Lemma 1 continues to apply. Since $W \geq 0$, $g(s_c^H) = W(s_c^H)(1 - v_H/v_1) \geq 0$, so the relevant stable crossing lies weakly below $s_c^H < q_2$. With $\rho = 1$, stage-1 safety choices therefore occupy depths $[0, s^*]$. Safety vacancies weakly above depth x total

$$V(x) \;=\; \mu \min\{x, s^*\} \;+\; (x - s^*)^+,$$

melt vacancies among filled depths plus never-filled depths. Before any safety vacancy is used, failed reach applicants take the μq_1 reach-tier melt vacancies. Hence the marginal applicant's safety depth solves

$$\mu q_1 + V(x) = D_f = (1 - \mu)(1 - s^* - q_1).$$

On the deep branch $x > s^*$,

$$\mu q_1 + \mu s^* + x - s^* = (1-\mu)(1-s^*-q_1) \quad \Longrightarrow \quad x = 1-\mu-q_1 = x_0.$$

On the shallow branch $x \le s^*$,

$$\mu q_1 + \mu x = (1-\mu)(1-s^*-q_1) \quad \Longrightarrow \quad x = \frac{(1-\mu)(1-s^*)-q_1}{\mu}.$$

The two formulas meet at $s^* = x_0$. The shallow depth reaches zero at $s_R = 1 - q_1/(1-\mu)$, where reach-tier vacancies alone equal D_f.

If the marginal student applies to safety she takes the top seat and obtains v_H. If she reaches and fails at $s^* < s_R$, she obtains $W(s^*) = \max\{0, v(x(s^*)) - d\}$. Indifference gives the stated fixed point. The incentive

$$g(s) = p_m(s)(v_1 - W(s)) - (v_H - W(s))$$

is nondecreasing. On either safety-ladder branch, $W \le v_H - d < v_H$, so any interval on which $p_m = 0$ has $g < 0$ and cannot contain a root. Once p_m is strictly between zero and one, it is strictly increasing by Assumption 2; together with nondecreasing W, this makes

$$g'(s) = p_m'(s)(v_1 - W) + (1-p_m)W'(s) > 0$$

wherever differentiable and a crossing is possible. At s_R, continuation value jumps weakly upward to $v_1 - d$, so g cannot acquire a downcrossing. Above s_R,

$$g(s) = d\, p_m(s) + (v_1 - v_H - d).$$

If $d < v_1 - v_H$ this expression is already positive when $p_m = 0$; if $d > v_1 - v_H$, its root has an interior admission probability and is strict. In the remaining equality case, the stated condition $p_m(s_R) > 0$ again makes the post-jump incentive strictly positive. The stable equilibrium is therefore unique, allowing a pinned crossing at s_R.

Holding $s < s_R$ fixed, greater dispersion lowers $v(x)$ whenever $x > 0$, hence weakly lowers W and g because $\partial g/\partial W = 1 - p_m > 0$. At any root below s_R with $W > 0$ the decrease is strict, so the unique crossing moves strictly right. At and above s_R, by contrast, reach-tier melt vacancies alone insure the marginal applicant and the continuation value $v_1 - d$ is independent of v_L; this yields only the global weak comparison. Finally, if $s_c^H < s_R$ and $v(x(s_c^H)) \le d$, then $W = 0$ at and below s_c^H, so the unique root is s_c^H. If $s_c^H > s_R$, the continuation value switches to a reach seat before that root is attained. If $s_c^H = s_R$, reach supply exactly exhausts failed-reach demand and the remaining safety supply gives the zero-mass boundary type positive continuation value. Either case precludes full restoration by safety-seat dispersion alone. $\square$

The next proof turns to the self-confirming character of the aggressive regime.

Proof of Theorem 3. At these parameters $S = \rho\mu(q_1 + q_2) < D_r(s_c^*)$, so the buffer absorbs all surviving supply and the cautious outcome is an equilibrium. In an aggressive candidate $s^* < q_2$, so $D_r = 0$ and the buffer disappears. If residual reach and safety seat masses are R_1 and R_2, random service within the failed-reach class gives the marginal applicant

$$W_2^m = \frac{R_1(v_1 - d) + R_2(v_2 - d)}{D_f}$$

when $R_1 + R_2 < D_f$; if supply is larger, the numerator uses the best mass D_f of residual seats. This expression counts reach vacancies as well as safety vacancies.

For the parameter vector

$$(q_1, q_2, v_1, v_2, d, \mu, \varepsilon, \rho) = (0.2, 0.75, 1, 0.8, 0.05, 0.062, 0.25, 0.60),$$

direct evaluation gives

s	.7433	.7435	.7788	.7791	.7899	.7901
$g(s)$	$-.000252$	.000320	.000236	$-.000155$	$-.000998$	.001003.

Continuity on each bracket gives stable upcrossings at 0.743388 and 0.790000, with a downcrossing at 0.778982. The first root is aggressive and the last cautious. All six inequalities are strict, so they persist under a sufficiently small perturbation of the primitives; coexistence therefore holds on an open parameter set. Under threshold best-response dynamics the downcrossing separates the two basins, giving the stated hysteresis. $\square$

The commitment result formalizes the time-inconsistency reading of the design problem.

Proof of Proposition 2. (i) Given any technologically surviving seat and any eligible, willing claimant, serving the claimant raises realized welfare by $v_j - d > 0$. Hence every subgame-perfect equilibrium uses all such seats, the announcement is nonbinding, and effective reliability is $\bar{\rho}$. Students anticipate the score-priority equilibrium at $\bar{\rho}$. (ii) Commitment replaces this continuation by the constrained choice $\arg\max_{\rho \le \bar{\rho}} W(\rho)$. Under the two strict conditions in Corollary 1, welfare increases up to ρ^* and remains strictly below $W(\rho^*)$ thereafter. The constrained optimum is therefore $\min\{\rho^*, \bar{\rho}\}$, and its value relative to discretion is positive exactly when $\bar{\rho} > \rho^*$. $\square$

A.4 Proofs for Section 5

The ladder results are exercises in geometric summation and telescoping. In the common-value benchmark acceptance requires $\Delta_k > c$ and $v_R > d$; under multiplicative match values the corresponding conditions are $s_{k+1}\Delta_k > c$ and $s_u v_R > d$.

Proof of Proposition 3. The proposition conditions on a surviving top-rung vacancy. Incumbent upgrade k, for $k = 1, \ldots, R-1$, requires the preceding $k-1$ subsequent vacancies to survive and therefore has probability ρ^{k-1}. Terminal absorption is an additional placement and requires all $R-1$ subsequent vacancies to survive, probability ρ^{R-1}. Expected incumbent upgrades are $\sum_{k=0}^{R-2} \rho^k = (1-\rho^{R-1})/(1-\rho)$; adding terminal absorption gives $\sum_{k=0}^{R-1} \rho^k = (1-\rho^R)/(1-\rho)$. The absorption and death probabilities follow. $\square$

Proof of Proposition 4. Each realized move k upgrades an incumbent from v_{k+1} to v_k, a gain of Δ_k at switching cost c; terminal absorption gives the unassigned agent $v_R - d$. Hence

$$W_{\text{open}} \;=\; \sum_{k=1}^{R-1} \rho^{k-1}(\Delta_k - c) \;+\; \rho^{R-1}(v_R - d), \qquad W_{\text{closed}} \;=\; v_1 - d,$$

the closed protocol assigning the top vacancy directly to an unassigned agent. At $\rho = 1$ the sum telescopes: $W_{\text{open}} = (v_1 - v_R) - (R-1)c + v_R - d = v_1 - d - (R-1)c$, so the difference is exactly $(R-1)c$. For $\rho < 1$, every term of W_{open} is increasing in ρ under the participation conditions ($\Delta_k > c$, $v_R > d$), strictly so for $k \geq 2$, hence $W_{\text{open}}(\rho) < W_{\text{open}}(1)$ and the difference strictly exceeds $(R-1)c$: with common values, every move re-houses an already-housed agent at a real resource cost, and chain mortality compounds the waste. $\square$

Proof of Proposition 5. Incumbent $k+1$ upgrades from rung $k+1$ to k with probability ρ^{k-1}, producing $s_{k+1}\Delta_k - c$. Terminal absorption produces $s_u v_R - d$ with probability ρ^{R-1}. Closed eligibility produces $s_u v_1 - d$. Summing gives the first display in the proposition. At $\rho = 1$, the gross allocation difference is

$$\sum_{k=2}^{R} s_k(v_{k-1} - v_k) \;-\; s_u(v_1 - v_R) \;=\; \sum_{k=2}^{R} s_k \Delta_{k-1} \;-\; s_u \sum_{k=2}^{R} \Delta_{k-1} \;=\; \sum_{k=2}^{R} (s_k - s_u)\,\Delta_{k-1},$$

and subtracting $(R-1)c$ gives the second display. The sum is the ladder's aggregate sorting gain. A positive sign gives strict preference for open eligibility; a zero sign gives indifference.

□

A.5 Proofs for Section 6

Both empirical propositions follow from the accounting identities.

Proof of Proposition 6. Write $c_2 = (s^* - q_2)^+$ for the main-round safety cutoff and S_1 for surviving reach-tier supply. A positive mass of failed reach applicants receives safety seats only if $S_1 < D_f$. An ACE satisfying this premise must have $s^* > 0$. Indeed, at $s^* = 0$ the separation condition gives $p_m(0) = 0$; because $S_1 < D_f$, the marginal failed applicant receives at most a safety seat, so $W_2^m \leq v_2 - d < v_2$ and $g(0) < 0$, contradicting marginal indifference. If $s^* \leq q_2$, there are no safety rejects. If instead $S \leq D_f$, score priority exhausts available supply within the failed-reach class before reaching any safety reject. In either case every safety-seat assignee has score at least s^*, so the aftermarket safety cutoff is at least $s^* > c_2$.

Now suppose $s^* > q_2$ and $S > D_f$. All failed reach applicants are served. Because $S_1 < D_f$, they exhaust the reach-tier vacancies and use $D_f - S_1$ safety seats, leaving the positive mass $\min\{S - D_f, D_r\}$ of safety seats assigned to safety rejects. Those rejects have scores strictly below $c_2 = s^* - q_2$; serving a positive mass therefore makes the aftermarket safety cutoff strictly smaller than c_2. Finally, on the path of Theorem 1, every oversubscribed outcome has $S \leq D_f$, while every outcome with $S > D_f$ has $s^* \leq q_2$. □

Proof of Proposition 7. Rescale the accounting of Theorem 1 by cohort mass N, holding capacities fixed: demands become $D_f = (1 - \mu)(N(1 - s^*) - q_1)$ and supply is unchanged, so the pinned locus and its boundary at the safety subscription margin $Ns^* = q_2$ deliver

$$\rho^*(N) \;=\; \frac{(1 - \mu)(N - q_1 - q_2)}{\mu(q_1 + q_2)},$$

strictly increasing in N. Let $z = Ns^*$ be the mass applying to safety. Before the crossing, pinned aftermarket volume is $\hat{\rho}\mu(q_1 + q_2)$, locally constant in N. After the crossing, the undersubscribed pin satisfies

$$\hat{\rho}[q_2 + \mu q_1 - (1 - \mu)z] = (1 - \mu)(N - z - q_1),$$

so $dz/dN = 1/(1 - \hat{\rho})$ and placement volume $M = (1 - \mu)(N - z - q_1)$ has

$$\frac{dM}{dN} = -\frac{(1 - \mu)\hat{\rho}}{1 - \hat{\rho}} < 0.$$

At the critical mass $N^* = q_1 + q_2 + \hat{\rho}\mu(q_1 + q_2)/(1 - \mu)$, both formulas give $z = q_2$ and the same volume. The derivative calculation applies on the undersubscribed pinned branch adjacent to the crossing; later branches require their own comparative statics. The path is therefore continuous with the stated local kink. □

A.6 Additional institutional extension: supply-side leakage

The main analysis takes aftermarket supply as the vacancies surviving from the centralized round. A platform also competes for schools, and its capacity to pursue distributional goals depends on the quality of schools' off-platform option. The following calculation isolates that participation margin.

Supply-side participation bound. Applicants have uniform scores, and an off-platform channel independently reaches a fraction $1 - \varphi$ at every score, with $q_1 < 1 - \varphi$. A reach-tier school selects the top q_1 students it reaches, so its off-platform mean score is

$$\bar{A}_{\text{off}}(\varphi) = 1 - \frac{q_1}{2(1 - \varphi)}.$$

On the platform, a distributional constraint $r \in [0, \bar{r}]$ gives the school a continuous, strictly decreasing mean-score index $\bar{A}_{\text{on}}(r)$ with $\bar{A}_{\text{on}}(0) = 1 - q_1/2$. The largest sustainable constraint,

$$r^*(\varphi) = \max\{r \in [0, \bar{r}] : \bar{A}_{\text{on}}(r) \geq \bar{A}_{\text{off}}(\varphi)\},$$

is weakly increasing in φ, strictly wherever the participation constraint binds in the interior, and satisfies $r^*(0) = 0$.

Derivation. Independent thinning leaves off-platform score density $1 - \varphi$ on $[0, 1]$. Filling q_1 seats therefore sets cutoff $1 - q_1/(1 - \varphi)$, and the mean of the admitted uniform interval is $1 - q_1/[2(1 - \varphi)]$. This outside option is strictly decreasing in φ and equals $1 - q_1/2 = \bar{A}_{\text{on}}(0)$ at zero friction. Since $\bar{A}_{\text{on}}$ is continuous and strictly decreasing, a lower outside option weakly expands the set of r satisfying participation, strictly at an interior binding solution. At $\varphi = 0$, every $r > 0$ gives $\bar{A}_{\text{on}}(r) < \bar{A}_{\text{on}}(0) = \bar{A}_{\text{off}}(0)$, so $r^*(0) = 0$.

The bound complements the demand-side mechanism in the main text. Kapor et al. (2024) document a sizeable off-platform sector in Chile and show that platform expansion increased on-platform capacity, with gains concentrated among lower-SES students. The participation bound is the participation counterpart of Avery and Pathak (2021): the constraint on redistribution is spatial in their setting and off-platform here.

B Numerical Verification

This appendix checks the closed-form predictions against equilibria computed directly from the fixed-point equation. The baseline parameters are $q_1 = 0.2$, $q_2 = 0.75$, $v_1 = 1$, $v_2 = 0.8$, $d = 0.05$, $\varepsilon = 0.25$; Table B4 modifies $q_2 = 0.80$ and sets $v_H = 0.8$ so the homogeneous case nests the baseline safety value. Tables B1 and B2 compare the pinned branches, the interior threshold, and the welfare candidates; Table B3 locates erosion under buffered priority; Tables B4 and B5 report the dispersion fixed point and equilibrium multiplicity; Table B6 traces the welfare selection rule, including a configuration where complete reliability is optimal throughout. Calculations following Table B6 locate the domain boundaries of the welfare and cutoff comparisons. Two final checks certify regularity. On the full oversubscribed range the marginal admission rate is weakly increasing: at the benchmark it is zero through $s^* = 0.552786$, then strictly increases when interior. On a 101-point grid spanning the equilibrium-relevant interval $[s_a^*, s_c^*]$, its smallest adjacent increment is 0.00380, and the separation inequality in Assumption 2 has slack at least 0.350.

Table B1: Score priority: pinned closed form and the threshold ($\mu = 0.05$, $\rho^* = 1.00$)

ρ	computed s^*	theory $s^*_{\text{pin}}(\rho)$	phase	p_m	W
0.30	0.7850	0.7850	pinned	0.755	0.7713
0.70	0.7650	0.7650	pinned	0.626	0.7863
0.95	0.7525	0.7525	pinned	0.564	0.7957
1.00	0.6417	—	endpoint selection	0.200	0.7925

At $\mu = 0.05$, $\rho^* = 1$ lies at the endpoint: the equilibrium threshold tracks the pinned closed form below one, while full service of the zero-mass boundary type selects the full-insurance root. Table B1 certifies the pinned branch and the endpoint selection; when ρ^* is interior, $s^*(\rho)$ is continuous there. Raising melt to $\mu = 0.062$ moves ρ^* into the interior.

Table B2: Score priority: interior threshold and welfare candidates ($\mu = 0.062$, $\rho^* = 0.7963$, $\rho_a = 0.9252$)

ρ	0.70	0.79	0.7963	0.80	0.85	0.90	0.9252	1.00
s^*	0.7560	0.7504	0.7500	0.7488	0.7275	0.6849	0.6417	0.6417
W	0.78306	0.78726	0.78755	0.78750	0.78663	0.78475	0.78279	0.78298

Table B2 confirms the interior case: the equilibrium threshold is continuous at ρ^* and follows the undersubscribed pin until ρ_a. Welfare peaks globally at ρ^*, falls to ρ_a, then recovers slightly through the option value of surviving reach-tier vacancies; $W(1)$ remains below $W(\rho^*)$. Table B3 locates the discontinuous erosion of the cautious equilibrium under buffered priority (score within class).

Table B3: Buffered priority (score within class): erosion at the predicted threshold ($\mu = 0.062$, theory $\rho^* = 0.7963$)

ρ	computed s^*	regime
0.30–0.79	0.7900	cautious ($s^* = s_c^*$ sustained)
0.80	0.7488	aggressive (collapse)

The cautious threshold $s_c^* = 0.7900$ survives at every grid point through $\rho = 0.79$ and collapses at $\rho = 0.80$, so the grid brackets the predicted threshold $\rho^* = 0.7963$. Table B4 verifies the dispersion fixed point of Proposition 1.

Table B4: Dispersion disciplines reaching ($\rho = 1$, $\mu = 0.062$, $q_2 = 0.80$, $v_H = 0.8$)

v_L	0.80	0.65	0.50	0.35	0.20	0.05
s^*	0.6417	0.7337	0.7511	0.7585	0.7630	0.7661
$x(s^*)$	0.7380	0.7380	0.5395	0.4281	0.3598	0.3127

The homogeneous column reproduces the full-insurance root; the first two columns lie on the deep branch $x_0 = 1 - \mu - q_1 = 0.738$, and the remaining columns use the endogenous shallow depth. Every step of dispersion raises the threshold, but reach-tier melt imposes the ceiling $s_R = 0.7868$. Table B5 exhibits multiplicity under the scramble variant.

Table B5: Multiplicity under buffered priority, scramble variant (within-class lottery) ($\mu = 0.062$, $\rho = 0.60$)

root s^*	type	regime
0.7434	stable	aggressive
0.7790	unstable	boundary
0.7900	stable	cautious

The two stable roots straddle an unstable threshold: the market can rest in either regime at identical primitives, and the unstable root marks the tipping point of the hysteresis argument. Table B6 verifies the exact selection rule of Corollary 2.

Table B6: Optimal reliability: delay-cost transitions and completion throughout

configuration	d_1	d_2	d_3	d_4
benchmark, $\varepsilon = .25$ (ρ^*=.796)	.002083	.008578	.237399	.247214
q_1=.25, q_2=.72, v_2=.70, μ=.08, ε=.25	.018613	.049819	.457821	.566025
q_1=.15, q_2=.82, v_2=.85, μ=.05, ε=.25	.001555	.009565	.270546	.283013
completion: q_1=.30, q_2=.60, v_2=.80, μ=.20, ε=.15	—	.037037	—	.044949

Here d_1 solves $y(d) = 2u(d)$, $d_2 = Ax^2/C$ is the candidate-merger point, d_3 solves $(1 - \mu)d(q_2 - s_a^*(d)) = Ax$, and d_4 solves $s_a^*(d) = q_2$. The first three rows have one root of each comparison equation and therefore exhibit complete–throttle–absorb–complete on $(0, d_4)$. In the last row neither comparison equation has a root: direct welfare maximization selects complete reliability throughout, including at d_2. There, the uniform-noise inverse gives $s_a^* \geq 0.55$, hence $u \geq 5/11$; below $d_2 = 0.037037$, $y < x = 5/9 < 2u$, while above d_2, $(1 - \mu)d(q_2 - s_a^*) < 0.8(0.045)(0.05) < Ax = 1/150$, so completion dominates both capped candidates throughout $(0, d_4)$, with $W(1) - W(\rho^*) = 0.006466$ at d_2. The replication code evaluates the displayed roots, compares both closed-form welfare gaps with directly computed equilibria, and matches the finite candidate maximizer to a reliability grid built from those direct roots.

The merger comparison in Corollary 2(iv) uses $d_2 < d_4$ to place the algebraic merger before claimant composition changes. Consider

$$(q_1, q_2, \mu, \varepsilon, v_1, v_2, d) = (.20, .72, .30, .10, 1, .50, .02),$$

where the threshold ordering at the displayed delay is $s_a^* = .707544 < q_2 < s_c^* = .775000 < 1 - q_1$, and the primitive safety-rescue inequality is strict: $0.064719 - 0.060000 > 0$. The corresponding boundary values are

$$d_2 = .086641 > d_4 = .059017, \qquad s_a^*(d_2) = .727357 > q_2.$$

Here the algebraic merger occurs after claimant composition changes, so the within-domain dominance equation is no longer the relevant policy boundary.

The second case of Proposition 6 occurs at

$$(q_1, q_2, v_1, v_2, d, \mu, \varepsilon, \rho) = (.20, .75, 1, .80, .30, .062, .25, 1),$$

where the unique stable score-priority ACE is $s^* = .7600$. Its flows are $S_1 = .01240$, $S = .05890$, $D_f = .03752$, and $D_r = .00938$. Thus failed reach applicants receive safety seats, but $s^* > q_2$ and $S > D_f$: all safety rejects are also served. The main-round safety cutoff is .0100; the aftermarket safety cutoff is zero. Abundant supply has reached the lower-scoring safety rejects, who set the aftermarket cutoff.

Computational method. For uniform noise, admitted mass is piecewise quadratic in the cutoff, so the solver computes market clearing exactly. Given $\kappa(s^*)$, it evaluates the incentive function from the flow identities while tracking reach- and safety-tier vacancy values separately. Bisection refines stable upcrossings and unstable downcrossings; welfare and dispersion depth are evaluated directly from equilibrium flows. The accompanying code (`aftermarket_solve.py`) reproduces all tables.