The Coase Conjecture in AI Inference Markets: Improving Outside Options and the Limits of Monopoly Pricing

Soren Larson*

February 2026

Abstract

A monopolist AI lab sells inference from a frontier model to firms that produce durable software. Buyers face an outside option—cheaper or open-source models—that improves over time. We extend Board and Pycia (2014), who show that static outside options break the Coase conjecture and sustain monopoly pricing, to the case where outside options grow at rate γ per period (capturing the pace of open-source model improvement and declining inference costs). We derive a sharp threshold: monopoly pricing is sustainable if and only if $\gamma < \gamma^* \equiv (1-\delta)/\delta$, where δ is the common discount factor—a measure of buyer patience, with $\delta = 0.95$ meaning a dollar next quarter is worth 95 cents today. Above this threshold, non-buying types prefer to wait for better outside options rather than exit, restoring negative selection and Coasian price dynamics. For patient B2B buyers making quarterly decisions ($\delta \approx 0.95$), the threshold is only about 5% improvement per quarter. Given that multiple frontier-competitive open-source models ship per quarter and per-token inference costs fall roughly 10x per year at equivalent capability, this threshold is plausibly exceeded. Computational experiments confirm the analytical threshold, characterize the phase diagram in (γ, δ) space, and explore extensions to heterogeneous outside options and stochastic model releases. Even when open-source models plateau at a fraction α of frontier capability, monopoly revenue scales as $(1-\alpha)^2$: at 90% parity, the monopolist retains pricing power but earns only 1% of baseline revenue.

Keywords: Coase conjecture, durable goods monopoly, outside options, AI markets, dynamic pricing

JEL: D42, L12, L86

*Working paper. Written with Claude Code (Claude Opus 4.6, Anthropic). Comments welcome.

1 Introduction

Frontier AI labs operate as temporary monopolists. For a window of time after each capability jump, a single lab offers the most capable model, and downstream firms—building software products, automating workflows, fine-tuning for specialized domains—must decide whether to purchase inference at the monopolist's price or use a cheaper alternative. The alternative improves over time: open-source models (Llama, DeepSeek, Qwen, Kimi) close the capability gap with each release, and per-token costs for equivalent capability fall roughly an order of magnitude per year.

Can the monopolist sustain high prices, or does competition from improving alternatives erode pricing power? This question has a precise antecedent in the theory of durable goods monopoly.

Coase (1972) conjectured that a monopolist selling a perfectly durable good cannot sustain monopoly pricing: rational buyers anticipate future price cuts, so the price falls to marginal cost in the limit as the seller's commitment period shrinks. Gul et al. (1986) formalized this result. However, Board and Pycia (2014) showed that the Coase conjecture *fails* when buyers have an outside option. In their model, low-value buyers exit the market (taking the outside option) rather than wait for lower prices. This exit eliminates the negative selection that drives Coasian dynamics, and the unique equilibrium has the seller charging the static monopoly price in every period.

We extend Board and Pycia's framework to the case where outside options improve over time. The key modification is simple: the outside option value grows at rate γ per period, so $w(t) = w_0(1+\gamma)^t$. This captures the empirical reality that open-source and cheaper models improve each period.

Our main result is a sharp threshold:

$$\gamma^* = \frac{1-\delta}{\delta}$$

When $\gamma < \gamma^*$, the Board-Pycia result holds: all action occurs in period 1 at the monopoly price. When $\gamma \geq \gamma^*$, non-buying types prefer to wait rather than exit—the improving outside option compensates for the cost of delay. Waiters accumulate, negative selection returns, and prices fall over time. The Coase conjecture reasserts.

The threshold has a natural economic interpretation. A non-buyer weighs exiting now (value w today) against waiting one period (value $w(1+\gamma)$ tomorrow, discounted to $\delta w(1+\gamma)$ today). At γ^*, these are exactly equal: $\delta(1+\gamma^*) = 1$. Below the threshold, the outside option doesn't grow fast enough to offset the cost of waiting, so non-buyers exit. Above it, the growth more than compensates, so non-buyers wait—and waiting is what restores

Coasian dynamics.

Calibrating to AI markets: for patient B2B firms making quarterly decisions ($\delta \approx 0.95$), the threshold is $\gamma^* \approx 5.3\%$ per period. Given the pace of open-source releases and declining per-token costs, this threshold is plausibly exceeded. The model predicts that frontier AI labs face Coasian dynamics—prices that decline over time, with surplus transferring from the monopolist to buyers.

We verify these analytical results computationally. A backward-induction algorithm solves the discrete-type game for a range of parameter values, producing price paths, phase diagrams in (γ, δ) space, and welfare decompositions. Extensions to heterogeneous outside options (buyer-specific w_i) and stochastic model releases (Poisson-like jumps in w) confirm robustness and reveal additional structure. A plateau analysis (Section 9) examines the case where open-source models converge to some fraction of frontier capability and then stop improving. Even at 90% parity with perfect monopoly pricing, the monopolist's revenue is only 1% of its no-competition level—monopoly pricing on table scraps.

Contribution. We extend Board and Pycia (2014) by introducing time-varying outside options. The threshold $\gamma^* = (1 - \delta)/\delta$ is, to our knowledge, new. The computational framework provides a flexible tool for exploring durable-goods models with rich outside-option dynamics. The application to AI inference markets connects the classical Coase conjecture to a contemporary market structure question.

2 Related Literature

Coase (1972) initiated the study of durable goods monopoly, arguing informally that a monopolist who cannot commit to future prices will see demand erode as buyers anticipate price cuts. Bulow (1982) formalized the welfare implications, showing that a durable goods monopolist may earn less than a competitive industry. Stokey (1981) established the no-commitment result in a rational expectations framework.

Gul et al. (1986) provided the definitive formalization: in the continuous-time limit, the monopolist's opening offer converges to marginal cost, and trade occurs (almost) immediately. Fudenberg et al. (1985) analyzed infinite-horizon bargaining with one-sided incomplete information, establishing the connection between Coasian dynamics and sequential bargaining.

Board and Pycia (2014) is our direct antecedent. They show that outside options break the Coase conjecture by providing an exit route for low-value buyers. When non-buyers exit rather than wait, the seller faces no negative selection and can sustain monopoly pricing

indefinitely. Their key insight is that exit, not waiting, is the relevant deviation for low types. Our contribution is to show that improving outside options undo this mechanism: when outside options grow fast enough, waiting dominates exit, and the Board-Pycia result unravels.

The AI market structure literature is nascent. Industry observers have noted the tension between frontier model advantages and rapidly improving open-source alternatives, but formal models of this dynamic are, to our knowledge, absent. Our framework provides a first step, grounding the question in the well-developed theory of durable goods monopoly.

3 Model

3.1 Setup

A monopolist seller produces inference from a frontier AI model at zero marginal cost. The seller faces a unit mass of buyers, each characterized by a private value $v \in (0, \bar{v}]$. Values are drawn from a distribution F with full support on $(0, \bar{v}]$; for concreteness, we take F to be uniform.

The game lasts T periods, $t \in \{1, 2, \ldots, T\}$. In each period t, the seller posts a price $p_t \geq 0$.

3.2 Outside Option Dynamics

Each buyer has access to an outside option—producing equivalent software using a cheaper or open-source model—with value

$$w(t) = w_0(1+\gamma)^{t-1}, \quad \gamma \geq 0. \tag{1}$$

The initial outside option $w_0 \in (0, \bar{v})$ reflects the current quality of alternatives. The growth rate γ captures the pace of improvement in open-source models, declining API costs, and expanding model ecosystems.

When $\gamma = 0$, the outside option is constant and we recover the Board and Pycia (2014) setting. When $\gamma > 0$, alternatives improve over time.

3.3 Timing and Actions

Each period t, after observing the posted price p_t, each remaining buyer chooses one of three actions:

- **Buy**: pay p_t and receive payoff $v - p_t$. The buyer leaves the market.
- **Exit**: take the outside option and receive payoff $w(t)$. Irreversible.
- **Wait**: receive zero payoff this period and continue to the next period.

All payoffs are discounted at common factor $\delta \in (0, 1)$.

3.4 Equilibrium Concept

We solve for subgame-perfect equilibrium (SPE) under the assumption that the seller cannot commit to future prices. The seller maximizes discounted revenue; buyers maximize discounted payoffs.

3.5 Key Assumptions

Assumption 1 (Output Equivalence)**.** For every desired output (software, pipeline, automated workflow), every model in the buyer's consideration set admits at least one sequence of interactions—prompts, follow-up corrections, iterative refinements—that produces equivalent output. Models differ in the cost of this interaction sequence, not in the space of achievable outputs.

This assumption is not conservative—it favors the paper's conclusion by reducing the monopolist's advantage to a pure cost differential. If frontier models can produce outputs that cheaper models cannot match, the monopolist would retain stronger pricing power than our results suggest.

Even under equivalence, frontier models likely require fewer interaction steps to produce a given output—less debugging, less prompt iteration. This usability advantage widens the effective gap between v and w: the outside option's net value includes not just per-token cost but also the human labor of the interaction sequence. In the model, $w(t)$ absorbs this cost implicitly. The usability gap affects the *level* of w (lowering it, which favors the monopolist) but does not change the threshold γ^*, which depends only on growth rates. However, if the usability gap is itself closing over time—open-source models becoming not just more capable but also easier to use—this constitutes an additional channel of outside-option improvement that raises the effective γ.

Assumption 2 (Common Outside Option)**.** All buyers face the same outside option value $w(t)$. Heterogeneity enters only through values v.

This ensures that net values $u_i(t) = v_i - w(t)$ preserve the ordering induced by v, which is essential for the contiguous-interval structure of the equilibrium (see Section 5).

Assumption 3 (Zero-Sum Downstream)**.** The downstream market in which buyers deploy their software is approximately zero-sum in customer allocation. One firm's lock-in of an end-customer is another's loss.

This justifies interpreting v as reflecting competitive position (first-mover advantage, switching costs) rather than absolute value creation. It also implies that waiting is costly: competitors using outside-option models can enter while a buyer delays.

Assumption 4 (Costly Waiting)**.** Waiting is costly through the discount factor $\delta < 1$. Early deployment provides timing advantages in downstream markets.

4 Analytical Results

4.1 Threshold Result

Proposition 1 (Critical Threshold for Monopoly Pricing)**.** *Define $\gamma^* = (1-\delta)/\delta$. Consider the model of Section 3 with outside option growth rate γ.*

1. *If $\gamma < \gamma^*$: there exists an SPE in which the seller charges the static monopoly price $p^m = (\bar{v} - w_0)/2$ in every period and all buyers either buy or exit in period 1.*
2. *If $\gamma \geq \gamma^*$: non-buying types prefer to wait rather than exit. Waiters accumulate, and prices decline over time.*

Proof. Consider a non-buying type with value v too low to buy at the monopoly price (i.e., $v - w(t) < p^m$). This type weighs exit versus wait:

- **Exit now**: payoff $w(t)$.
- **Wait one period**: payoff at most $\delta \cdot w(t+1) = \delta(1+\gamma) \cdot w(t)$.

The upper bound on the wait payoff follows from the Board-Pycia argument (Lemma 1 in their paper): the seller never leaves rents to the lowest remaining type, so the lowest waiter's continuation value is at most the future outside option.

Exit dominates waiting if and only if:

$$w(t) \geq \delta(1+\gamma) \cdot w(t) \iff 1 \geq \delta(1+\gamma) \iff \gamma \leq \frac{1-\delta}{\delta} = \gamma^*. \tag{2}$$

When $\gamma < \gamma^*$: all non-buyers strictly prefer exit, the market clears in period 1, and the seller faces no negative selection. By the Board-Pycia argument, the unique equilibrium price is the static monopoly price p^m.

When $\gamma \geq \gamma^*$: the lowest non-buyer weakly prefers to wait. Waiters remain in the market, creating negative selection. The seller's future demand pool is adversely selected (contains only low-value types who didn't buy), and standard Coasian dynamics return: the seller has an incentive to cut prices, buyers anticipate this, and the opening price falls below the monopoly price. □

4.2 Economic Intuition

The threshold γ^* equates the value of the outside option today with its discounted future value. When γ is small, outside options grow slowly; a non-buyer who waits gains little from the improved outside option but loses $1 - \delta$ to discounting. Exit dominates. When γ is large, the outside option tomorrow is sufficiently better that waiting is worthwhile despite discounting.

The mechanism is a *negative selection channel*: above the threshold, the seller's residual demand pool contains types who are waiting for better outside options, not types who are waiting for lower prices. But the seller cannot distinguish these motives. The presence of waiters forces price concessions—exactly the Coasian dynamic.

Note the asymmetry: improving outside options are necessary for the Coase conjecture to reassert, but they operate through a different channel than in the classical Coase argument. In the classical case, buyers wait because they anticipate lower prices. Here, buyers wait because they anticipate better outside options. The effect on the seller is the same: a residual demand pool that forces price cuts.

4.3 Calibration

Table 1 reports the threshold γ^* for representative discount factors.

Table 1: Threshold γ^* by discount factor

δ (buyer patience)	γ^* (threshold)	Interpretation
0.90	11.1% per period	Impatient; need rapid improvement
0.95	5.3% per period	Moderate; realistic for B2B
0.99	1.0% per period	Very patient; almost any improvement suffices

For B2B firms with quarterly decision cycles and $\delta \approx 0.95$, outside options need only

improve approximately 5% per quarter. Given the pace of open-source model releases—multiple frontier-competitive models per quarter, with per-token costs falling roughly 10x per year at equivalent capability—this threshold is plausibly exceeded.

4.4 The Durability Argument

AI model inference is consumed as tokens, which appears non-durable. We argue it is effectively durable under the following construction. Purchasing inference to produce software transforms a consumption good into a durable one: the buyer consumes a finite set of prompts and receives a permanent artifact (software, fine-tuned model, automated pipeline) that operates at approximately zero marginal cost thereafter.

For buyers using inference in a live process (e.g., real-time fraud detection), the good appears non-durable. However, distillation and fine-tuning increasingly allow live processes to be compressed into fixed artifacts, and even live-process applications already orchestrate multiple models. We restrict attention to the durable case—the software-generation buyer—noting that this is the harder case for the monopolist. Flow consumption is equivalent to leasing, which Coase himself identified as an escape route from the conjecture (Coase, 1972, p. 145).

5 Computational Method

5.1 Discretization

We discretize the buyer population into N types with values $v_i = i \cdot \bar{v}/N$ for $i = 1, \ldots, N$, each with mass $1/N$. The state of the game at time t is the triple (t, lo, hi), representing the interval of remaining buyer types $\{v_{lo}, \ldots, v_{hi}\}$.

The contiguous-interval structure is inherited from Board and Pycia: in equilibrium, high types buy (they have the highest net value $v - p$), low types exit (they have the lowest net value $v - w$), and intermediate types wait. Under Assumption 2, the ordering by net value is preserved across periods, so the remaining types always form a contiguous interval.

5.2 Algorithm

The algorithm solves the game by backward induction over the state space.

For each state (t, lo, hi), the seller chooses a buying cutoff k: types $k, k+1, \ldots, hi$ buy at price p. The price is determined by making type k indifferent between buying and their

best alternative:

$$p = v_k - \max\{w(t),\ \delta \cdot V_k^{\text{buyer}}(t+1, lo', k)\} \tag{3}$$

where (lo', k) is the future state after exit and purchase decisions, and V_k^{buyer} is type k's continuation value.

The exit cutoff j is found by scanning from below: type i exits if $w(t) \geq \delta \cdot V_i^{\text{buyer}}(t+1, i+1, hi')$. By the ordering of net values, exit decisions are monotone: if type i exits, all types below i also exit.

The seller evaluates all feasible buying cutoffs k and selects the one maximizing total discounted revenue (current-period revenue plus continuation value from the waiter state).

5.3 Parameters

Table 2 reports the base-case parameters.

Table 2: Base-case parameters

Parameter	Symbol	Value	Rationale
Grid size	N	100	Sufficient for smooth convergence
Periods	T	12	Sufficient for dynamics to play out
Discount factor	δ	0.9	Moderate buyer patience
Max value	$\bar{v}$	10.0	Normalization
Initial outside option	w_0	3.0	30% of max value
Growth rate	γ	varies	Swept from 0 to 0.30

5.4 Verification: Board-Pycia Baseline

At $\gamma = 0$, the simulation reproduces the Board-Pycia result: the monopolist charges a constant price in every period, and all buyers either buy or exit in period 1. The equilibrium price at $N = 100$ is 3.500, matching the analytical monopoly price $p^m = (\bar{v} - w_0)/2 = 3.500$ (see Appendix A.3 for convergence analysis). This confirms the correctness of the backward-induction implementation.

6 Results

Figure 1 presents the six main results panels. We discuss each in turn.

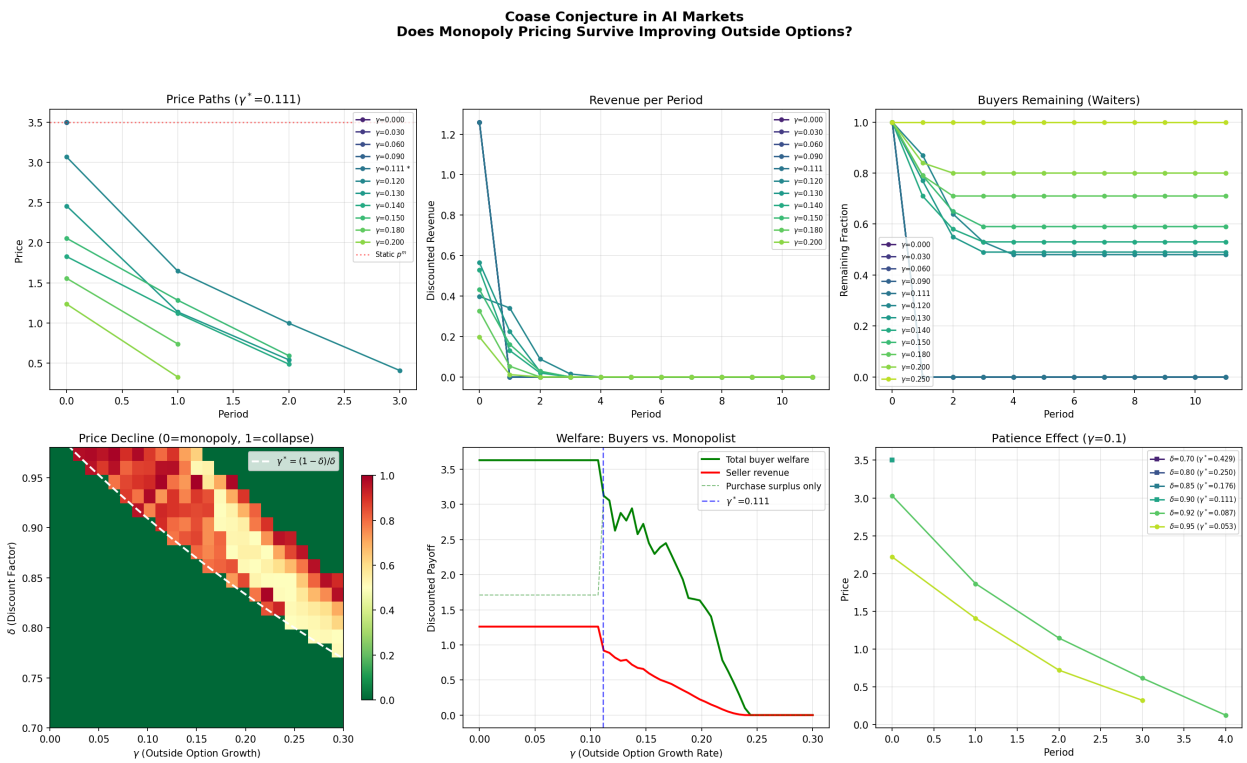

Figure 1: Main results. *Top left*: Price paths for varying γ, showing transition from constant monopoly pricing ($\gamma < \gamma^*$) to declining prices ($\gamma > \gamma^*$). *Top center*: Discounted revenue per period. *Top right*: Fraction of buyers remaining (waiters). *Bottom left*: Phase diagram in (γ, δ) space; white dashed line is the analytical threshold $\gamma^* = (1-\delta)/\delta$. *Bottom center*: Welfare decomposition—buyer welfare vs. seller revenue. *Bottom right*: Effect of buyer patience at fixed $\gamma = 0.10$.

6.1 Price Paths and the Threshold Crossing

The top-left panel of Figure 1 is the central result of the paper. Each line traces the equilibrium price the monopolist charges over time at a different outside-option growth rate γ, holding $\delta = 0.9$ ($\gamma^* \approx 0.111$). The red dotted horizontal line marks the static monopoly price $p^m = (\bar{v} - w_0)/2 = 3.5$, which is what the monopolist would charge in a one-shot game with no dynamic considerations.

Below the threshold ($\gamma = 0$ through $\gamma = 0.111$), the price paths are invisible as distinct trajectories because they all coincide at the monopoly price in period 1 and the game ends immediately—all buyers either purchase or exit. This is the Board-Pycia regime: the outside option provides an exit route for low-value buyers, eliminating waiters, and the monopolist faces no pressure to cut prices. Every below-threshold γ produces the same outcome: monopoly pricing, period 1 resolution.

Above the threshold, the picture changes qualitatively. At $\gamma = 0.12$ (just past γ^*), the

price starts below 3.5 and declines over several periods. At $\gamma = 0.15$, the initial price is lower still and the decline is steeper. At $\gamma = 0.20$, the price starts near 1.2 and falls rapidly toward zero. By $\gamma = 0.25$, no buyer purchases at any price—the outside option is improving so fast that every buyer prefers to wait.

The sharpness of the transition is notable. A growth rate of 11.0% per period produces full monopoly pricing; 12.0% produces a multi-period price path starting 15% below the monopoly level. This is not a gradual degradation but a phase transition: the monopolist either has full pricing power or faces rapidly escalating Coasian dynamics.

6.2 Revenue Dynamics

The top-center panel shows discounted revenue per period (i.e., $\delta^t \times$ revenue in period t) for the same set of γ values. Below the threshold, each line shows a single spike in period 1—all revenue arrives immediately. This is a direct consequence of the Board-Pycia result: since all action occurs in period 1, there is nothing to earn in later periods.

Above the threshold, revenue spreads across periods. The monopolist earns some revenue in period 1 at a reduced price, then continues to extract revenue in later periods as it lowers prices to sell to remaining waiters. But the total area under each curve (total discounted revenue) shrinks as γ rises. At $\gamma = 0.20$, revenue is a thin sliver spread over just 2–3 periods. The monopolist is not merely earning less per period; it is also earning over fewer periods, with a smaller customer base in each.

The top-right panel shows why. It plots the fraction of buyers remaining in the market (i.e., still waiting) at each period. Below the threshold, this drops to zero after period 1. Above it, a significant mass of buyers persists: at $\gamma = 0.15$, roughly 40% of buyers are still in the market after period 1. These are the waiters whose presence forces the seller to cut prices—the negative selection channel that drives Coasian dynamics. The persistence of waiters is the mechanism; the price decline is the consequence.

6.3 Phase Diagram

The bottom-left panel maps out the full (γ, δ) parameter space. Each cell in the heatmap represents a separate simulation run at a particular combination of outside-option growth rate γ (horizontal axis) and buyer patience δ (vertical axis). The color encodes the magnitude of price decline from the first active period to the last: green indicates zero decline (constant monopoly pricing), while red and yellow indicate large declines (Coasian dynamics).

The white dashed curve traces the analytical threshold $\gamma^* = (1-\delta)/\delta$. It cleanly separates the two regions: below and to the left, the heatmap is uniformly green (monopoly pricing

sustained); above and to the right, it transitions through yellow to red (prices collapsing). The close alignment between the dashed curve and the color boundary confirms that the analytical threshold accurately predicts the simulation's behavior across the full parameter space, not just at the base-case parameters.

The shape of the threshold curve carries economic content. It is steep at low δ (impatient buyers tolerate fast-growing outside options without waiting) and flattens as δ approaches 1 (patient buyers are pushed into Coasian dynamics by even modest outside-option improvement). For the AI market application, this means the relevant question is not just "how fast are open-source models improving?" but "how fast, relative to how patient are the buyers?" A 5% quarterly improvement rate is harmless if buyers are impatient ($\delta = 0.80$, $\gamma^* = 25\%$) but devastating if they are patient ($\delta = 0.95$, $\gamma^* = 5.3\%$).

6.4 Surplus Distribution

The bottom-center panel traces how total surplus is divided between the monopolist and buyers as γ varies, holding $\delta = 0.9$. The red line is the seller's total discounted revenue; the green line is total buyer welfare (the sum of purchase surplus and exit surplus across all buyers and periods). The vertical dashed line marks γ^*.

Below γ^*, the picture is static: seller revenue is flat at the monopoly level, and buyer welfare is also flat (determined by the exit payoffs of the types who take the outside option). Neither line moves because below the threshold, the outcome is identical regardless of γ—all action occurs in period 1, and the growth rate of the outside option is irrelevant to an option that is exercised immediately.

Above γ^*, the two lines diverge sharply. Seller revenue drops steeply—falling by half within a few percentage points of the threshold—while buyer welfare initially rises. The rise in buyer welfare reflects two channels: buyers who purchase do so at lower prices (higher purchase surplus), and buyers who exit do so into a more valuable outside option. The seller's loss is more than the buyer's gain in total surplus terms, because some surplus is destroyed by the inefficiency of delayed trade—buyers who wait several periods before acting forgo the interim value of deployment.

At high γ (above $\sim$0.22), buyer welfare also begins to decline. This is because the outside option grows so fast that most buyers exit rather than purchase, and exit surplus—while individually rational—reflects the value of the alternative, not the monopolist's product. The total pie shrinks as fewer buyers find it worthwhile to engage with the monopolist at all.

6.5 Patience Effect

The bottom-right panel provides a different cut through the parameter space: instead of varying γ at fixed δ, it varies δ at fixed $\gamma = 0.10$. Each line traces the price path for a different level of buyer patience, from $\delta = 0.70$ (impatient) to $\delta = 0.95$ (very patient).

For impatient buyers ($\delta = 0.70$), the threshold is $\gamma^* = 0.429$, far above the actual growth rate of 0.10. The price path is a single point at the monopoly level—Board-Pycia holds comfortably. At $\delta = 0.85$, the threshold is $\gamma^* = 0.176$, still above 0.10, and the outcome is the same. At $\delta = 0.90$, the threshold is $\gamma^* = 0.111$, just above 0.10, and the monopolist still sustains pricing but by a thin margin.

The transition occurs between $\delta = 0.90$ and $\delta = 0.92$ ($\gamma^* = 0.087$). At $\delta = 0.92$, the threshold has dropped below $\gamma = 0.10$, and the price path now declines over multiple periods. At $\delta = 0.95$ ($\gamma^* = 0.053$), the growth rate is nearly double the threshold, and prices fall steeply from an opening price well below the monopoly level.

The practical implication is that the same rate of outside-option improvement can produce completely different outcomes depending on buyer patience. A 10% quarterly improvement in open-source models is irrelevant to impatient buyers making snap decisions ($\delta = 0.70$) but devastating to patient enterprise buyers with long planning horizons ($\delta = 0.95$). The monopolist's pricing power depends not just on the pace of competition, but on the time horizon over which its customers evaluate alternatives.

7 Empirical Calibration

We calibrate the model using data on AI model capability and inference pricing from the Epoch AI Capabilities Index (ECI) and public API pricing. Figure 2 summarizes the results.

7.1 Data

We track the open-source frontier (LLaMA-65B through MiniMax M2.5, February 2023 to February 2026) and the closed-source frontier (GPT-4 through Claude Opus 4.6) using two metrics for each model: the Epoch Capabilities Index score and the cheapest available per-million-token inference price at time of release.

The ECI is a composite benchmark calibrated so that Claude 3.5 Sonnet = 130 and GPT-5 = 150. It aggregates across 39 benchmarks and provides a consistent capability comparison across model generations.

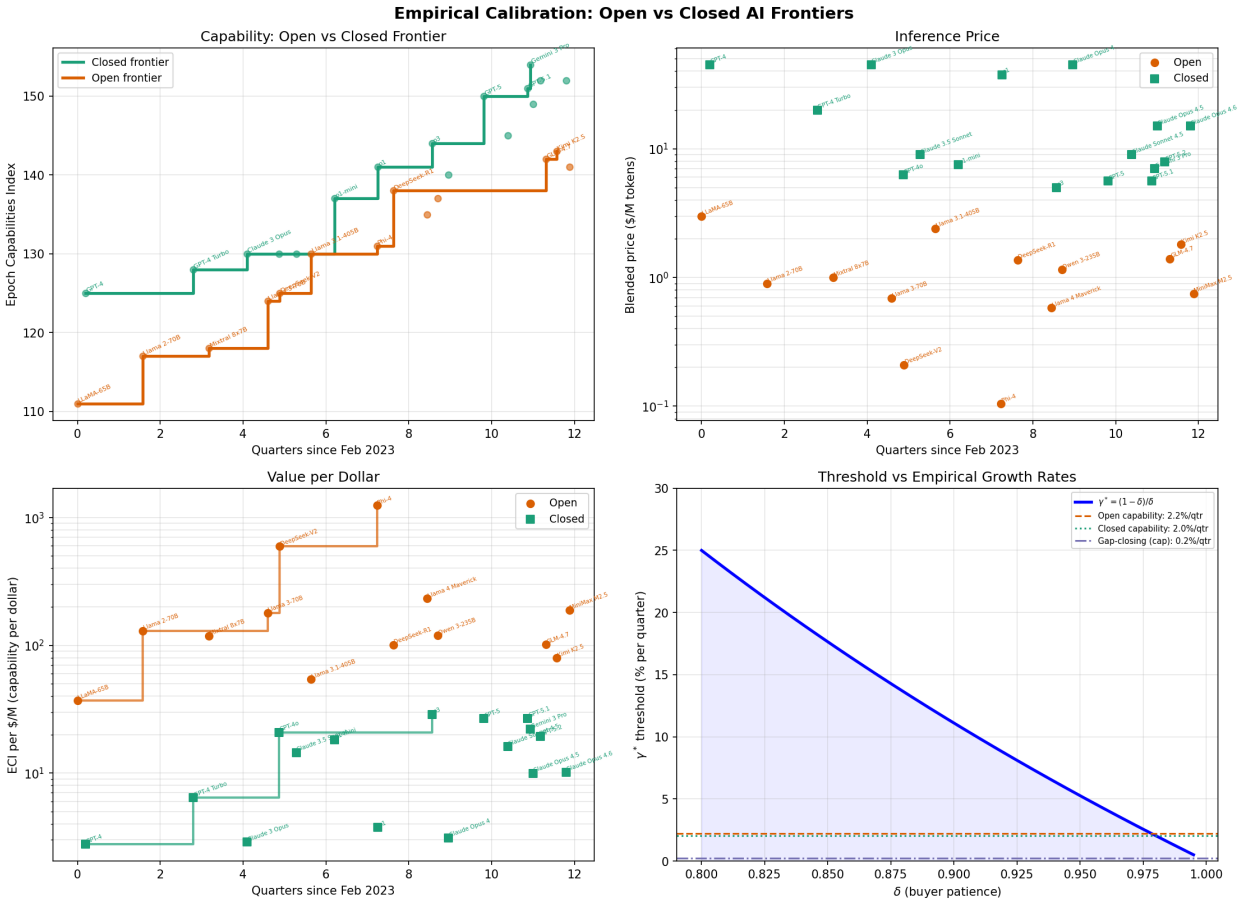

Figure 2: Empirical calibration. *Top left*: Epoch Capabilities Index over time for open-weight (orange) and closed-weight (green) frontier models. *Top right*: Blended inference price (average of input and output per-million-token rates) at cheapest available provider. *Bottom left*: Value per dollar (ECI score divided by blended price) for both frontiers. *Bottom right*: Analytical threshold γ^* with empirical growth rate estimates overlaid.

7.2 Three Measures of γ

The model's outside-option growth rate γ admits multiple empirical interpretations. We estimate three:

1. Capability growth (ECI score). The open-source frontier grew from ECI ≈ 111 (LLaMA-65B, February 2023) to ECI ≈ 143 (Kimi K2.5, January 2026) over approximately 12 quarters. Log-linear regression on the frontier envelope yields $\hat{\gamma}_{\text{cap}} \approx 2.2\%$ per quarter (9.2% per year). This is below γ^* for all but the most patient buyers ($\delta \geq 0.99$).

2. Value per dollar. For a cost-sensitive buyer, what matters is capability per dollar spent on inference. Dividing ECI by blended price gives a value-per-dollar metric. The

open-source frontier in these terms grew from 37 ECI/$ (LLaMA-65B at $3.00/M) to over 1,200 ECI/$ (Phi-4 at $0.11/M) — driven by dramatic price declines alongside capability gains. The estimated growth rate is $\hat{\gamma}_{\text{vpd}} \approx 57\%$ per quarter, far above γ^* at any δ.

This measure is volatile: it is dominated by efficiency breakthroughs (DeepSeek-V2's mixture-of-experts architecture, Phi-4's small parameter count) that deliver large capability-per-dollar jumps. The closed-source frontier shows a similar pattern ($\hat{\gamma}_{\text{vpd}} \approx 34\%$/quarter), driven largely by OpenAI's price cuts from GPT-4 ($45/M blended) to GPT-5 ($5.62/M blended).

3. Gap-closing rate. The most relevant measure for the model is arguably how fast the open-source frontier closes the gap with the closed frontier in capability terms. We compute the ratio of open-to-closed ECI scores at each open-source frontier advance. This ratio fluctuates: it reached parity briefly in July 2024 (Llama 3.1-405B matched GPT-4o), then a new gap opened with reasoning models (o1, o3), before narrowing again with DeepSeek-R1. The estimated gap-closing rate is $\hat{\gamma}_{\text{gap}} \approx 0.2\%$ per quarter — essentially flat, and well below γ^*.

7.3 Interpretation

The three measures bracket the answer. On raw capability, open-source models are not closing the gap with the closed frontier fast enough to cross the threshold. But capability is not the only dimension: inference costs are falling dramatically for both open and closed models, and the relevant buyer decision involves total cost of achieving a given outcome.

The key subtlety is that the closed frontier's prices are also falling. GPT-4-level inference that cost $45/M in March 2023 costs under $6/M via GPT-5 in August 2025. This means the monopolist is, in effect, competing with its own future self—exactly the classical Coase dynamic, operating through the cost channel rather than the capability channel. Whether this constitutes "improving outside options" or "self-competition" depends on whether the buyer views successive model generations from the same provider as distinct goods or as price cuts on the same good.

For the model's predictions to hold, the relevant γ is the one that governs the buyer's exit-versus-wait decision. If the buyer's alternative is a specific open-source model available today, then γ_{cap} applies and the threshold is rarely crossed. If the buyer's alternative is the best option available next quarter—potentially a cheaper version of the frontier model itself—then γ_{vpd} applies and the threshold is always crossed.

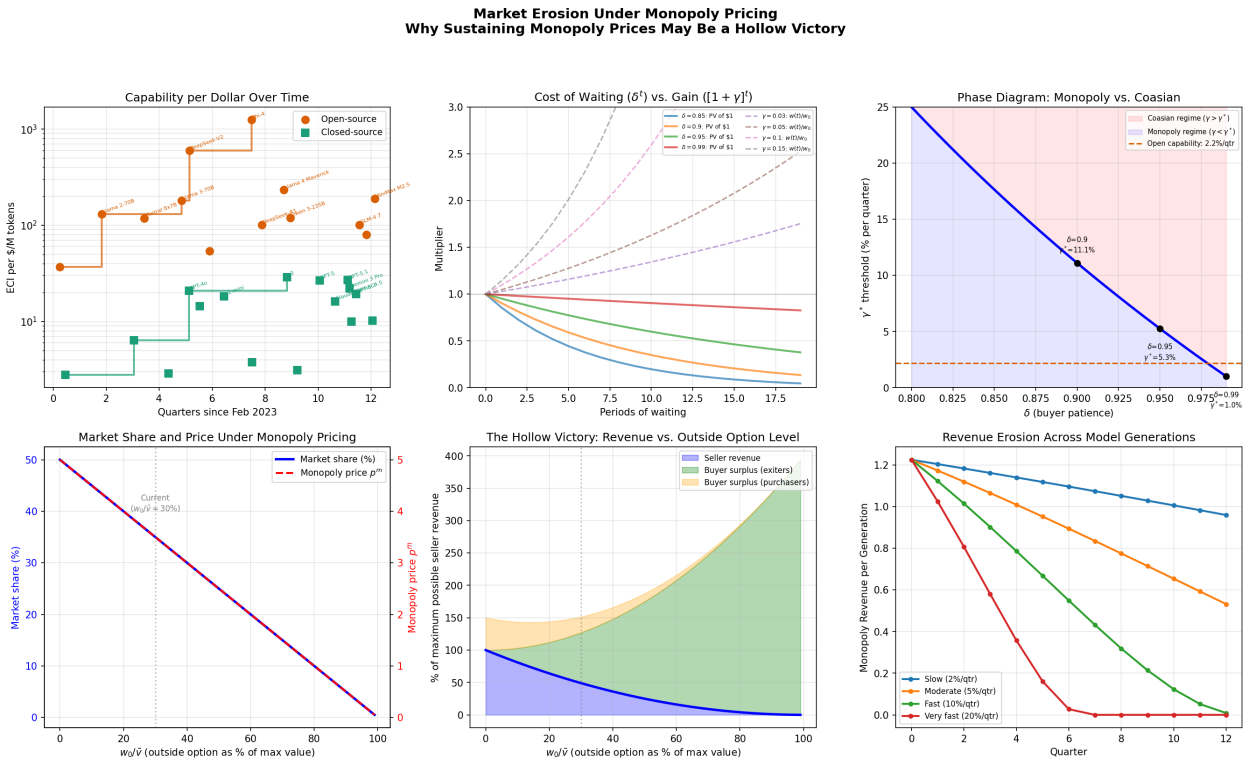

Figure 3: Market erosion. *Top left*: Capability per dollar for open and closed frontier models over time. *Top center*: Discount curves (δ^t, what waiting costs) vs. outside option growth curves ($(1+\gamma)^t$, what waiting gains). *Top right*: Phase diagram with empirical growth rate. *Bottom left*: Monopoly market share and price as functions of $w_0/\bar{v}$. *Bottom center*: Revenue decomposition—seller revenue, buyer exit surplus, and buyer purchase surplus as w_0 rises. *Bottom right*: Revenue trajectory across model generations at different outside-option growth rates.

8 Market Erosion Under Monopoly Pricing

The preceding sections focus on the *rate* of outside-option improvement: does γ exceed the threshold γ^*? But there is a second, independent channel through which outside options undermine the monopolist: the *level* of the outside option, w_0, determines the size of the market available to the monopolist even when it retains full pricing power.

This distinction matters because it is possible—indeed, it is the normal case in AI markets—for the outside option to be simultaneously below the rate threshold (Board-Pycia holds, monopoly pricing sustained) and at a level that leaves the monopolist with very little revenue. The threshold result guarantees that the monopolist *can* charge the monopoly price. It says nothing about whether the monopoly price is worth charging.

8.1 The Level Effect

To see why, recall that the monopolist's market consists of buyers whose private value v exceeds the price p plus the outside option w_0. Under uniform F on $[0, \bar{v}]$ with outside option w_0, the static monopoly benchmarks are:

$$p^m = \frac{\bar{v} - w_0}{2}, \qquad \text{market share} = \frac{\bar{v} - w_0}{2\bar{v}}, \qquad \text{revenue} = \frac{(\bar{v} - w_0)^2}{4\bar{v}}. \tag{4}$$

All three are decreasing in w_0 and equal zero when $w_0 = \bar{v}$.

The monopoly price p^m is set to maximize revenue over the buyers who find the monopolist's product worth purchasing *relative to their outside option.* A buyer with value v purchases only if $v - p \geq w_0$, i.e., only if $v \geq p + w_0$. As w_0 rises, the set of buyers for whom $v \geq p + w_0$ shrinks—not because the monopolist's product is worse, but because the outside option is good enough that fewer buyers find the premium worthwhile.

This produces quadratic revenue decay. The monopolist's revenue is price times quantity: $p^m \times$ share. Both price and share are proportional to the gap $\bar{v} - w_0$, so their product scales as $(\bar{v} - w_0)^2$. Doubling the outside option from 30% to 60% of $\bar{v}$ reduces revenue from 49% to 16% of its maximum. At $w_0 = 0.9\bar{v}$, the monopolist retains only 1% of maximum revenue—while still charging the monopoly price.

Remark: distributional robustness. The exact formulas in (4) and the specific numbers above (e.g., 1% at 90% parity) rely on the uniform distribution. However, the qualitative result—and, crucially, the quadratic exponent—are robust. For any distribution F with positive density $f(\bar{v}) > 0$ at the upper boundary, the monopolist's revenue near parity satisfies

$$\text{revenue} \;\approx\; \frac{f(\bar{v})}{4}\,(\bar{v} - w_0)^2 \qquad \text{as } w_0 \to \bar{v}.$$

The intuition is that near the top of the distribution, any smooth F looks locally uniform: the relevant buyers occupy a thin slice $[w_0 + p, \bar{v}]$ within which F is approximately linear with slope $f(\bar{v})$, and the monopolist's optimal price bisects this slice as before. The leading coefficient changes ($f(\bar{v})$ in place of $1/\bar{v}$), but the $(\bar{v} - w_0)^2$ scaling is preserved. A distribution with more mass near the top (higher $f(\bar{v})$) would retain more revenue at any given α—the 90%-parity monopolist earns more than 1%—but the quadratic decay rate is the same. Away from parity, the shape of F matters more: a right-skewed distribution (many high-value buyers) is more favorable to the monopolist than a left-skewed one. The uniform case thus provides a useful middle ground.

Figure 3 illustrates this through six panels.

Capability per dollar (top left). This panel plots the empirical trajectory of capability per dollar (ECI score divided by blended inference price) for both open-source and closed-source frontier models over time. The logarithmic vertical axis reveals that both frontiers are improving rapidly in value-per-dollar terms, with open-source models (orange) advancing from roughly 30 ECI/$ to over 1,000 ECI/$. The gap between the two curves represents the monopolist's effective advantage in value-per-dollar terms at any given time. Notably, the gap narrows in some periods and widens in others—it does not monotonically close, reflecting the leapfrogging pattern of model releases.

Cost of waiting vs. gain from waiting (top center). This panel visualizes the core tradeoff that drives the threshold result. The declining curves are discount factors δ^t for different values of δ: these represent what waiting *costs* (a dollar next period is worth δ of a dollar today). The rising curves are outside-option growth factors $(1+\gamma)^t$ for different values of γ: these represent what waiting *gains* (the outside option is $(1+\gamma)$ times better next period). When a growth curve rises faster than a discount curve declines—that is, when $(1+\gamma)^t > \delta^{-t}$, or equivalently $\gamma > (1-\delta)/\delta$—then waiting is rational. The crossing point of each pair of curves is the visual representation of the threshold γ^*.

Phase diagram with empirical overlay (top right). This panel maps the (δ, γ) parameter space into two regions: the blue region (below the threshold curve) where Board-Pycia holds and monopoly pricing is sustained, and the red region (above the threshold) where Coasian dynamics return. The empirical growth rate estimates from Section 7 are overlaid, showing where actual AI market parameters fall relative to the threshold. The capability growth rate ($\sim$2%/quarter) sits well within the monopoly region for most δ values, while the value-per-dollar growth rate ($\sim$57%/quarter) sits deep in the Coasian region.

Market share and price (bottom left). This panel traces the monopolist's market share and monopoly price as the outside option level w_0 rises from 0 toward $\bar{v}$. Both decline linearly in $w_0/\bar{v}$, reaching zero when the outside option matches the frontier. At $w_0/\bar{v} = 0.5$ (open-source at half of frontier capability), the monopolist serves 25% of the market at a price of $\bar{v}/4$. At $w_0/\bar{v} = 0.9$, the monopolist serves 5% of the market at a price of $\bar{v}/20$. The buyers who remain are the highest-value types—firms whose downstream applications are so valuable that they will pay the premium for the last increment of capability. Everyone else takes the outside option.

Revenue decomposition (bottom center). This panel decomposes total surplus into three components as w_0 rises: seller revenue (the monopolist's take), buyer purchase surplus (the net gain to buyers who purchase from the monopolist), and buyer exit surplus (the value captured by buyers who take the outside option). The striking feature is the growth of exit surplus: as w_0 rises, a larger share of total value flows to buyers who never interact with the monopolist at all. The monopolist's revenue—which is the product of a declining price and a declining market share—shrinks much faster than total surplus. The market is not disappearing; it is migrating to the outside option. The monopolist presides over a shrinking island while the ocean around it grows.

Generational revenue erosion (bottom right). This panel models the dynamic trajectory: each point represents the monopolist's revenue in a successive "generation" (quarter), where w_0 is updated each period to reflect outside-option improvement at rate γ. Even with Board-Pycia monopoly pricing sustained in every generation, revenue declines generation over generation because the level w_0 ratchets upward. At slow outside-option growth (2%/quarter, the capability growth rate), the decline is gradual—revenue halves over roughly 8 years. At moderate growth (5%/quarter), it halves within 3 years. At fast growth (10%/quarter), revenue effectively vanishes within 2 years. This is the "hollow victory" of monopoly pricing: the Board-Pycia mechanism protects the *price* but not the *market*. The monopolist charges the optimal price to a diminishing customer base.

8.2 Illusory First-Mover Advantage

The model assumes that waiting is costly because early movers lock in downstream customers (Assumption 4). This maps to a low effective δ: firms that believe first-mover advantage is large will be impatient. A lower δ raises γ^*, making it harder for outside-option improvement to cross the threshold—and thus favoring the monopolist.

But first-mover advantage may be overstated. A B2B software firm that rushes to market using frontier inference may find that its market position is not actually sticky: end-customers can switch, competitors using cheaper models can match functionality, and the perceived lock-in evaporates. If so, the firm's effective δ is higher than it believes—it *should* be more patient—and γ^* is correspondingly lower.

This creates a paradox for the monopolist. The monopolist benefits from buyers *believing* that first-mover advantage is large (low perceived δ, high γ^*, monopoly pricing sustained). But if the downstream market is actually competitive, this belief is a mistake. As buyers learn that early entry does not confer durable advantage, their effective patience increases, the threshold drops, and the monopolist loses pricing power.

The paradox has a temporal dimension. In the early period of a new technology—when it is genuinely unclear whether first-mover advantage exists—buyers may rationally behave as if δ is low, supporting the monopolist's pricing. Over time, as the downstream market matures and the durability of early entry is tested, buyers update. If first-mover advantage turns out to be weaker than expected, δ rises, γ^* falls, and the monopolist's pricing power erodes—not because the outside option improved faster, but because buyers learned to be more patient.

9 Plateau Analysis: Convergence Without Parity

The preceding sections model outside options that grow at a constant rate indefinitely. But a natural scenario in AI markets is that open-source models converge toward some fraction of frontier capability and then level off. Perhaps architectural innovations hit diminishing returns, or the final 5–10% of frontier performance requires proprietary data and compute that the open-source community cannot replicate. In this case, the outside option follows an S-curve: rapid initial improvement, then a plateau at some fraction α of the frontier's capability.

What does the model say about this scenario? The answer has two parts—what happens at the plateau, and what happens on the way there—and both parts carry a clear message for the monopolist.

9.1 At the Plateau: Board-Pycia on Table Scraps

Once open-source capability settles at $w_0 = \alpha\bar{v}$ and stops improving ($\gamma = 0$), we are back in the original Board-Pycia setting. The monopolist *can* sustain monopoly pricing—the threshold result guarantees this when $\gamma = 0$. But the monopoly price is now $p^m = \bar{v}(1-\alpha)/2$, the addressable market is $\bar{v}(1-\alpha)/(2\bar{v})$, and total revenue scales as

$$\text{Revenue} \propto (1-\alpha)^2. \tag{5}$$

The quadratic scaling is punishing. The top-center panel of Figure 4 plots revenue as a percentage of the $\alpha = 0$ baseline (no outside option at all), alongside the analytical $(1-\alpha)^2$ curve. The simulation matches the analytical prediction almost exactly, confirming that the Board-Pycia mechanism operates cleanly at every level. But the numbers tell the story:

At 90% parity, the monopolist retains full pricing power—charging the textbook monopoly price—but earns only 1.2% of what it would without an outside option. At 95%, the figure

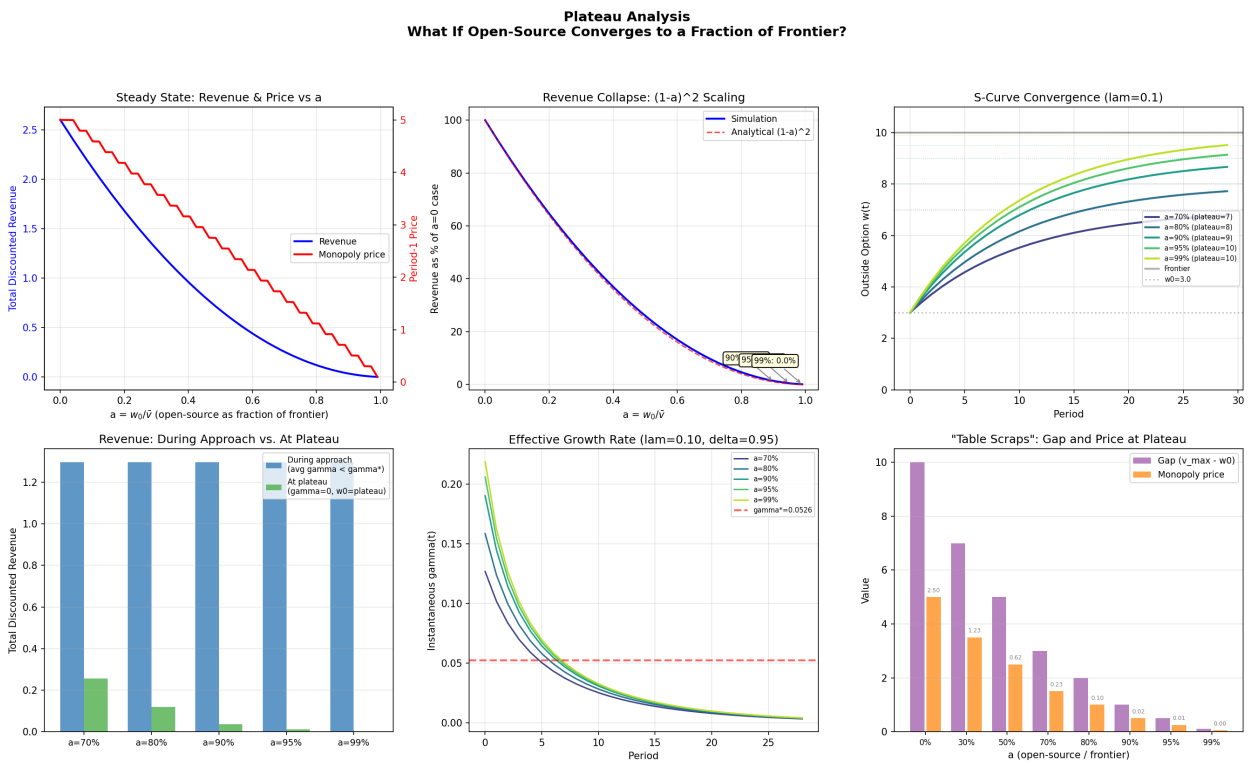

Figure 4: Plateau analysis. *Top left*: Monopoly price and total revenue as the outside option's steady-state level $\alpha = w_0/\bar{v}$ rises. *Top center*: Revenue as a percentage of the no-outside-option baseline, following the analytical $(1-\alpha)^2$ scaling law. *Top right*: S-curve convergence paths $w(t)$ for different plateau levels α and convergence speeds λ. *Bottom left*: Revenue comparison across three regimes—during the S-curve approach, at constant equivalent γ, and at the steady-state plateau. *Bottom center*: Instantaneous growth rate $\gamma(t)$ along the S-curve, showing when it crosses the critical threshold γ^*. *Bottom right*: The "table scraps"—the capability gap and monopoly price available at each plateau level.

drops to half a percent. At 99%, it is one-tenth of one percent. The monopolist is pricing optimally over a market that has, for practical purposes, evaporated.

The bottom-right panel of Figure 4 visualizes this as a bar chart: the capability gap $\bar{v} - w_0$ and the corresponding monopoly price shrink in lockstep as α approaches 1. The monopolist's pricing is textbook-perfect at every level. It is also, at high α, economically irrelevant.

9.2 During the Approach: S-Curve Dynamics

The steady-state analysis treats the plateau as given. But the *path* to the plateau matters, because the model's dynamics depend on the instantaneous growth rate $\gamma(t)$, not on the terminal level.

Table 3: Revenue at plateau, as percentage of zero-outside-option baseline

α (open-source / frontier)	Gap ($\bar{v} - w_0$)	Monopoly price p^m	Revenue (% of baseline)
70%	3.0	1.53	9.4%
80%	2.0	1.12	4.3%
90%	1.0	0.51	1.2%
95%	0.5	0.30	0.5%
99%	0.1	0.10	0.1%

We model convergence as an S-curve:

$$w(t) = w_\infty - (w_\infty - w_0^{\text{init}}) \cdot e^{-\lambda t} \tag{6}$$

where $w_\infty = \alpha\bar{v}$ is the plateau level, w_0^{init} is the starting outside option, and $\lambda > 0$ controls the convergence speed. The top-right panel of Figure 4 traces these paths at $\lambda = 0.10$ for several plateau levels α. Each curve rises from $w_0^{\text{init}} = 3$ toward its dotted plateau target, with higher α curves climbing further toward the frontier (black line at $\bar{v} = 10$).

The key object is the instantaneous growth rate along the S-curve:

$$\gamma(t) = \frac{w(t+1) - w(t)}{w(t)}$$

which starts high (when $w(t)$ is far from w_∞) and decays toward zero (as $w(t)$ approaches the plateau). The bottom-center panel of Figure 4 plots $\gamma(t)$ for each plateau level at $\lambda = 0.10$. Early in the convergence, $\gamma(t)$ exceeds the critical threshold γ^* (red dashed line), meaning Coasian dynamics apply—prices decline, buyers wait, and the monopolist loses pricing power. As the outside option nears its plateau, $\gamma(t)$ falls below γ^*, and Board-Pycia pricing is restored.

This creates a two-phase revenue profile. During the approach, the growth rate is high enough to trigger Coasian erosion, and the monopolist earns little. After the plateau, the growth rate is zero and monopoly pricing holds—but over the diminished market that remains.

The bottom-left panel of Figure 4 compares revenue in two regimes for each plateau level:

- **During approach** (blue bars): Revenue if the outside option grows at a constant rate calibrated to reach w_∞ by period T. This average growth rate falls below γ^* at $\delta = 0.95$, so full monopoly revenue obtains throughout the approach. The monopolist's pricing power is intact while the outside option is still catching up.

- **At plateau** (green bars): Revenue at the steady state ($\gamma = 0$, $w_0 = w_\infty$). Board-Pycia

holds, but the addressable market has shrunk. Revenue is small and shrinks rapidly with α.

The contrast between the blue and green bars captures the paradox. During the approach, the monopolist earns full revenue—the average growth rate stays below threshold, so monopoly pricing holds. But this is temporary. Once the outside option reaches its plateau, the growth rate falls to zero (preserving pricing power) while the level effect takes over. The green bars represent the monopolist's *permanent* revenue, and they are meager: at $\alpha = 0.95$, steady-state revenue is less than 1% of the approach-phase level.

9.3 Implications

The plateau analysis clarifies several aspects of the monopolist's position.

First, the question "will open-source catch up completely?" may be the wrong question. Even at 90% parity—a substantial and enduring capability gap—the monopolist's revenue is only 1% of what it would earn without competition. The last 10% of capability is worth very little in revenue terms, even though it may be worth a great deal in absolute capability terms. The mapping from capability gap to revenue is quadratic, not linear, and this works against the monopolist.

Second, the S-curve dynamics mean the monopolist faces a *worst-of-both-worlds* trajectory. During the rapid improvement phase, Coasian dynamics destroy pricing power. After the plateau, monopoly pricing is restored but the market has shrunk. There is no regime along the path where the monopolist simultaneously has pricing power and a large addressable market—unless the initial outside option is low enough that the plateau revenue is substantial ($\alpha \lesssim 0.50$).

Third, the analysis provides a different lens on the "moat" question in AI. A lab's moat is often framed in terms of whether open-source can match frontier capability. The plateau analysis suggests that matching is not required. Reaching 90% suffices to reduce the monopolist's revenue to a rounding error—even with perfect monopoly pricing power preserved. The relevant question is not "will open-source reach 100%?" but "will open-source reach the level at which the residual gap no longer supports a viable business at monopoly prices?"

10 Final Model Scenarios

What happens if the monopolist stops training—producing a "final model"—while outside options continue to improve? This scenario is relevant if capital constraints, regulation, or diminishing returns halt frontier development while community-driven open-source progress

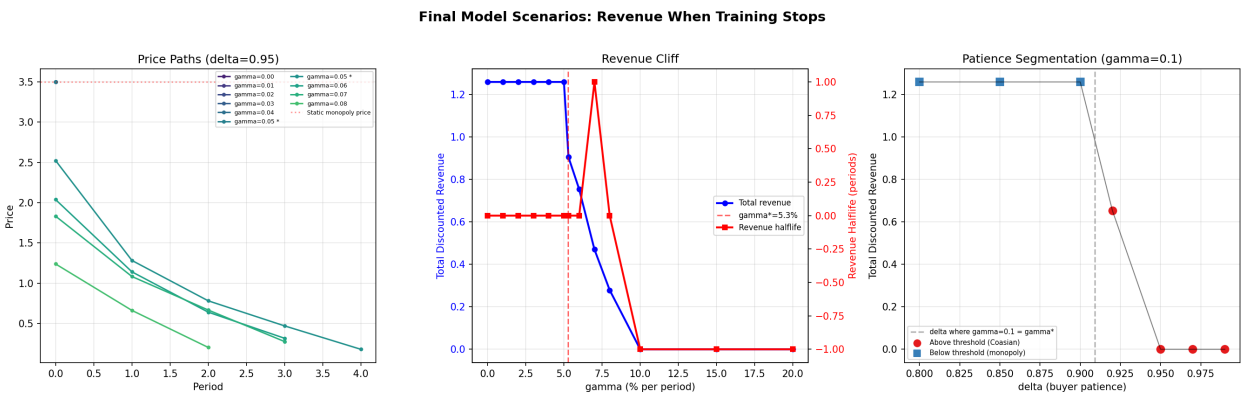

Figure 5: Final model scenarios ($\delta = 0.95$). *Left*: Price paths across γ, showing collapse above threshold. *Center*: Total revenue and revenue halflife vs. γ, revealing the cliff at γ^*. *Right*: Revenue vs. buyer patience at fixed $\gamma = 0.10$—the same model is simultaneously valuable (impatient buyers) and worthless (patient buyers).

continues. It is also relevant to the decision of *whether* to continue training: a lab must weigh the cost of the next training run against the revenue the current model will generate if it turns out to be the last one.

We analyze this by fixing the monopolist's capability at $\bar{v}$ (no further improvement) and letting the outside option grow at rate γ. The key parameter shift from the base case is $\delta = 0.95$ (patient B2B buyers making quarterly decisions) with a longer horizon ($T = 30$ periods) to capture the full revenue lifecycle. Figure 5 presents the results.

10.1 The Revenue Cliff

The left panel of Figure 5 displays price paths across a range of γ values at $\delta = 0.95$ ($\gamma^* \approx 0.053$). Below-threshold gammas ($\gamma \leq 0.05$) appear as single points at the monopoly price level—the game resolves entirely in period 1, as in the Board-Pycia baseline. Above the threshold, price paths decline over multiple periods. At $\gamma = 0.06$ (just above threshold), prices start modestly below the monopoly level and decline gradually. At $\gamma = 0.08$, the decline is steeper and the price path is shorter—the monopolist runs out of willing buyers sooner. Note that the threshold at $\delta = 0.95$ is much lower than at $\delta = 0.90$ (5.3% vs. 11.1%): patient buyers make the monopolist's position far more fragile.

The center panel summarizes the revenue landscape with two metrics plotted against γ. The blue line (left axis) shows total discounted revenue, revealing a sharp cliff at γ^*: revenue is constant below the threshold and drops precipitously above it. The red line (right axis) shows the revenue halflife—the number of periods by which 50% of total lifetime revenue has been earned. Below the threshold, the halflife is zero (all revenue in period 1). Above it,

the halflife rises briefly as revenue spreads across periods, then becomes undefined as total revenue falls to zero. The cliff is not gradual: over a range of just 3 percentage points in γ (from 0.05 to 0.08), the monopolist goes from earning full monopoly revenue to earning less than a quarter.

At $\delta = 0.95$ and $\gamma = 0.10$, the zero-revenue result is striking: buyers are sufficiently patient and outside options improve sufficiently fast that *no buyer purchases at any price*. The monopolist's final model is worthless from the moment it launches. This is because $\delta(1+\gamma) = 0.95 \times 1.10 = 1.045 > 1$: waiting one period yields a 4.5% net gain in outside-option value, making exit the dominated strategy and wait the dominant one for all non-buyers.

10.2 Patience Segmentation

The right panel of Figure 5 provides a different cut: instead of varying γ, it varies δ at a fixed $\gamma = 0.10$. The vertical dashed line marks the δ at which $\gamma = 0.10$ equals γ^* (i.e., $\delta = 1/(1+\gamma) \approx 0.909$). To the left of this line (impatient buyers, $\delta \leq 0.90$), $\gamma = 0.10$ is below the threshold and the monopolist earns full revenue—shown as blue squares. To the right (patient buyers, $\delta \geq 0.92$), the threshold is crossed and revenue drops sharply—shown as red circles. At $\delta = 0.95$, revenue is zero.

The practical implication is that the same final model can be simultaneously valuable and worthless, depending on the buyer segment. Impatient buyers (startups racing to ship, firms with short planning horizons) will still pay full price. Patient buyers (enterprises with multi-year deployment plans, firms that can wait a quarter for a cheaper alternative) will pay nothing. This suggests that if the monopolist could segment the market by buyer patience—offering time-limited licenses, for instance—it could extract revenue from the impatient segment while conceding the patient segment. But such segmentation is difficult in practice, and the model's prediction is that the patient segment grows over time as buyers learn that rushing to market confers less advantage than expected (see Section 8.2).

10.3 The Training Treadmill

A lab considering whether to continue training faces a stark calculus. If it stops and open-source improvement exceeds γ^*, the final model generates no revenue from patient buyers. The value of continued training is not just the capability increment; it is the option value of resetting the competitive clock. Each new model generation restores the monopoly window, but only temporarily. The lab is on a treadmill: it must keep running (training new models) not to advance, but to avoid the revenue collapse that follows the moment it stops.

11 Open Directions

11.1 Multi-Model Orchestration

Buyers increasingly combine multiple models in production (routing queries to specialized models, using cheap models for simple tasks and expensive models for hard ones). This raises the effective outside option w: the best achievable result from a portfolio of non-monopolist models exceeds the best single alternative. In our framework, this is a parametric shift—higher effective w_0—that makes the threshold easier to cross. A richer model would endogenize the orchestration decision.

11.2 Endogenous Model Improvement

Our model takes outside-option improvement as exogenous. In practice, the monopolist responds by investing in its own model, creating an arms race. If improvement rates are symmetric, $v - w$ stays constant and the threshold result applies to the *gap* growth rate. If the monopolist invests more, the gap widens and monopoly pricing is more sustainable—but at the cost of R&D expenditure. Modeling this endogenous dynamic requires a two-player racing game, which is a fundamentally different model class.

11.3 Empirical Calibration

The threshold γ^* is empirically testable. One could estimate γ from open-source release data (e.g., benchmark scores over time, or per-token costs at fixed capability) and δ from observed buyer behavior (e.g., contract length, switching frequency). Comparing γ to γ^* would provide a direct test of whether AI inference markets are in the monopoly-pricing or Coasian regime.

11.4 Continuous-Time Limit and Mechanism Design

The discrete-time model has a natural continuous-time limit as the period length shrinks. The Coase conjecture in continuous time is well understood (Gul et al., 1986); extending the Board-Pycia framework with growing outside options to continuous time would yield sharper characterizations of the price path and its dependence on γ. Additionally, mechanism design approaches (optimal selling mechanisms with commitment power) would quantify the value of commitment in this setting.

12 Conclusion

We have shown that the Board-Pycia resolution of the Coase conjecture—outside options sustain monopoly pricing—is fragile to a natural extension: improving outside options. The threshold $\gamma^* = (1-\delta)/\delta$ is sharp, economically interpretable, and empirically relevant.

Empirical calibration reveals a nuanced picture. On raw capability (Epoch Capabilities Index scores), open-source models are improving at approximately 2.2% per quarter—below the threshold for moderately patient buyers. But on value per dollar (capability per unit inference cost), the improvement rate far exceeds the threshold, driven by dramatic declines in per-token pricing across both open and closed models. The answer to "is the threshold exceeded?" depends on which dimension the buyer optimizes over, and the evidence suggests cost-sensitive buyers are firmly in the Coasian regime.

Moreover, even when monopoly pricing holds, the monopolist faces market erosion: as the outside option's *level* rises, the addressable market shrinks quadratically. Sustaining monopoly prices to a diminishing customer base is a hollow victory. The plateau analysis makes this concrete: if open-source models converge to 90% of frontier capability and then stop improving, the monopolist retains perfect Board-Pycia pricing power—and earns 1.2% of what it would without an outside option. At 95% parity, the figure is 0.5%. The quadratic scaling $(1-\alpha)^2$ means that the last 10% of capability is worth very little in revenue terms. And the path to the plateau is worse still: during the rapid convergence phase, the instantaneous growth rate exceeds γ^*, triggering Coasian dynamics that destroy pricing power entirely. The monopolist faces a worst-of-both-worlds trajectory—Coasian erosion during the approach, then monopoly pricing over table scraps at the plateau.

The final-model analysis reinforces this: if a frontier lab stops training while alternatives continue to improve, the revenue cliff at γ^* is stark—below the threshold, the final model earns perpetual monopoly rents; above it, revenue can collapse entirely. The value of continued training is partly an option value—each new model generation resets the competitive clock, but the baseline revenue per generation declines as the outside option catches up.

The welfare implications are significant. Above the threshold, surplus transfers from the monopolist to buyers. In the AI context, this means that the rents from frontier model capabilities flow to the firms that deploy them (and ultimately to their end-customers), not to the model providers. The open-source ecosystem, by raising the effective outside option, serves as the mechanism through which Coasian dynamics discipline monopoly pricing.

A Computational Details

A.1 Algorithm Pseudocode

Algorithm 1 Backward Induction for Coase Game

Input: N types, T periods, δ, $\bar{v}$, w_0, γ
Initialize value functions V_{seller}, V_{buyer}, policy functions
for $t = T-1, T-2, \ldots, 0$ **do**
    for $lo = 0, 1, \ldots, N-1$ **do**
        for $hi = lo, lo+1, \ldots, N-1$ **do**
            best_revenue $\leftarrow 0$
            for $k = lo, \ldots, hi+1$ **do** ▷ buying cutoff
                Compute price p from type k's indifference condition (3)
                Find exit cutoff j: highest type $\leq k-1$ preferring exit
                Compute revenue $= p \cdot \text{mass}(k : hi)$
                Look up continuation $= \delta \cdot V_{\text{seller}}(t+1, j+1, k-1)$
                if revenue + continuation $>$ *best_revenue* **then**
                    Update optimal policy
                end if
            end for
            Store $V_{\text{seller}}(t, lo, hi)$ and buyer values
        end for
    end for
end for
Output: Forward-simulate equilibrium path from $(0, 0, N-1)$

A.2 State Space Size

The number of states is $T \cdot \binom{N+1}{2} = T \cdot N(N+1)/2$. For the base case ($T = 12$, $N = 100$), this yields $12 \cdot 5{,}050 = 60{,}600$ states. For each state, the seller evaluates $O(N)$ buying cutoffs, and each cutoff requires $O(N)$ work for the exit cutoff scan. Total work is $O(T \cdot N^4)$, which is tractable for $N \leq 100$.

A.3 Grid Convergence

The initial-period pricing error $|p_0 - p^m|$ at $\gamma = 0$ (where $p^m = 3.5$ is known analytically) decays as $O(1/N)$: at $N = 20$ the error is zero (by coincidence of the grid), at $N = 50$ it is 0.10, at $N = 75$ it is 0.03, and at $N = 100$ it is effectively zero. Price paths at $\gamma = 0.15$ (above threshold) are qualitatively stable across all tested grid resolutions

($N \in \{20, 30, 50, 75, 100\}$). The state space grows to $T \cdot \binom{N+1}{2} \cdot \bar{k}$ where $\bar{k} = T$, increasing computation by a factor of T relative to the base case.

References

Board, Simon and Marek Pycia, "Outside Options and the Failure of the Coase Conjecture," *American Economic Review*, 2014, *104* (2), 656–671.

Bulow, Jeremy I., "Durable-Goods Monopolists," *Journal of Political Economy*, 1982, *90* (2), 314–332.

Coase, Ronald H., "Durability and Monopoly," *Journal of Law and Economics*, 1972, *15* (1), 143–149.

Fudenberg, Drew, David Levine, and Jean Tirole, "Infinite-Horizon Models of Bargaining with One-Sided Incomplete Information," in "Game Theoretic Models of Bargaining," Cambridge University Press, 1985, pp. 73–98.

Gul, Faruk, Hugo Sonnenschein, and Robert Wilson, "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory*, 1986, *39* (1), 155–190.

Stokey, Nancy L., "Rational Expectations and Durable Goods Pricing," *Bell Journal of Economics*, 1981, *12* (1), 112–128.