Pricing Foundational Models and Pretraining Data in Knowledge Economy *

Chen He† Yanqing Yang‡ Zibo Zhao§

July 23, 2026

Abstract

We develop a two-layer equilibrium model of foundational AI that links downstream pricing of model capability to upstream procurement of pretraining data through neural scaling laws. Downstream, where AI acts as a productivity floor for knowledge workers, competition exhibits quality-price coupling: small capability advantages allow leaders to limit-price rivals, generating endogenous winner-take-all outcomes. Upstream, we identify a novel reversed information asymmetry: model producers can verify data quality via computational scaling experiments, while data holders observe only noisy signals. We show that informed buyers optimally pool across quality types and cherry-pick underpriced datasets, systematically extracting information rents from data creators. Connecting both layers, we derive a sellable data constraint: the value of data quality is convex in downstream capability targets. Data is effectively commoditized at low capability levels but becomes a bottleneck input at the frontier, explaining the bifurcation of AI data markets and concentration in foundation model production.

Keywords: Foundational Models, AI Market Structure, Neural Scaling Laws, Information Asymmetry, Knowledge Economy

JEL Codes: D43, D82, L13, O33

*All remaining errors are our own.

†, ShanghaiTech University Email: hechen@shanghaitech.edu.cn

‡, ShanghaiTech University Email:yangyanqing@shanghaitech.edu.cn

§Department of Economics, W.P.Carey School of Business, Arizona State University. Email: zzhao203@asu.edu

1 Introduction

The emergence of foundational AI models represents a paradigm shift in the knowledge economy. Unlike previous waves of automation that targeted routine physical tasks, Large Language Models (LLMs) augment and potentially substitute for cognitive labor across a spectrum of expertise. This technological shock raises two interconnected questions of market design that current literature treats in isolation. Downstream, how are these models priced when they serve as a productivity floor for heterogeneous knowledge workers? Upstream, how does the demand for model capability filter back to the market for pretraining data—the essential input whose quality determines the efficiency of neural scaling?

Current economic analysis tends to bifurcate these issues, focusing either on the labor market impacts of AI adoption or the legal and copyright frictions in data acquisition (Acemoglu, 2025; Ide and Talamàs, 2025). Missing from this discourse is a structural link between the engineering reality of AI production and the economic reality of market competition.

We fill this gap by developing a two-layer equilibrium model that connects downstream AI adoption with upstream data procurement through the empirically established Neural Scaling Laws (Hoffmann et al., 2022). These scaling laws imply extreme convexity: as producers target the capability frontier, marginal improvements in data quality become exponentially more valuable. As a result, downstream pricing power and upstream data valuation cannot be analyzed independently.

Our downstream analysis models the Knowledge Economy following Garicano (2000); Garicano and Rossi-Hansberg (2004); Ide and Talamàs (2025), where workers face problems of varying difficulty. We depart from recent literature by conceptualizing current AI not as an autonomous agent, but as a capability augmentation tool. This structure generates a distinctive form of vertical competition. Because low-skill users derive identical benefits from any sufficiently capable model, competition collapses into quality–price coupling: a small capability advantage allows the leader to limit-price rivals and capture the entire low-skill segment. Contrary to standard vertical differentiation models, market segmentation is unstable, and equilibrium dynamics favor winner-take-all outcomes. Crucially, we demonstrate that this intrinsic drive toward firm-level concentration is not an artifact of a specific functional form, but a robust feature of the capability-augmentation framework that persists under both submodular and supermodular technologies.

Upstream, we examine the market for pretraining data under an information structure that operates on a reversed adverse-selection paradigm. In classic asset markets (Akerlof, 1970), sellers hold superior private information about quality. In the pretraining-data market the asymmetry runs in the opposite direction: model producers (the buyers) possess the massive computational infrastructure required to run scaling-law experiments and thereby perfectly verify data quality, while data holders (the sellers) observe only noisy signals of their datasets' intrinsic training value.

This reversal, combined with perfect buyer observability and a procurement rather than sales-market organization, fundamentally alters equilibrium outcomes. Informed buyers optimally pool across quality types to preserve their information advantage and systematically cherry-pick under-priced high-quality datasets, especially in fragmented markets with many sellers. The resulting verification trap extracts information rents from data creators and can stifle long-term data supply. Importantly, we demonstrate that this buyer advantage is structurally robust: the capacity to extract information rents via selective cherry-picking survives even when quality verification incurs explicit computational costs, and when the posted price mechanism is replaced by ex-post Bayesian Nash bargaining where sellers hold strong bargaining power.

The central contribution of the paper is to synthesize these two layers into a unified equilibrium. By embedding neural scaling laws as a production technology, we derive a sellable data constraint that characterizes when high-quality data can command a premium. The constraint implies a bifurcated data market: at low capability targets, data is effectively commoditized, while at the frontier, high-quality data becomes a bottleneck input capable of supporting substantial rents. This mechanism explains the observed industry shift from indiscriminate web scraping toward expensive, curated, and proprietary datasets.

Our analysis yields three main insights. First, the productivity-floor nature of AI drives intrinsic market concentration downstream. Second, reversed information asymmetry allows model producers to exploit data providers, creating a verification trap that may stifle long-term data supply. Third, the economic value of data is not intrinsic but is endogenously determined by the capability targets of the downstream model, creating strong complementarities between market power, scaling efficiency, and input valuation. Altogether, the core downstream concentration and upstream rent extraction dynamics persist under generalized functional forms, explicit verification frictions, and alternative bargaining protocols.

The rest of this paper is structured as follows. Section 2 reviews related literature. Section 3

analyzes the knowledge economy layer under monopoly and duopoly market structures. Section 4 develops the pretraining data market with information asymmetry. Section 5 combines both layers through the power law production technology to endogenize model capability. Section 6 concludes. An Online Appendix accompanies this paper in five parts. Appendix A provides all omitted proofs. Appendix B formally characterizes the two-layer general equilibrium as a fixed-point problem and establishes conditions for its existence. Appendix C extends our downstream analysis to incorporate a general capability augmentation function, showing that the winner-take-all dynamics and entry deterrence strategies remain robust under both submodular and supermodular technologies. Appendix D relaxes the assumption of costless quality verification, introducing explicit screening and verification costs to the upstream market. Finally, Appendix E replaces the posted-price mechanism with a Bayesian Nash bargaining protocol, demonstrating the persistence of buyer information rents under alternative bargaining structures.

2 Related Literature

Our paper relates and contributes to various strands of the literature. First and foremost, we build on the foundational framework of knowledge-based hierarchies initiated by Garicano (2000), who models production as problem-solving where workers with limited knowledge refer exceptions to specialists. Garicano and Rossi-Hansberg (2004, 2006) embed this framework in a setting with heterogeneous agents to study organization and inequality, deriving equilibrium assignment patterns, wage structures, and organizational hierarchies.[1] We adopt their characterization of a knowledge economy where workers draw problems of varying difficulty and can solve those within their knowledge range. Within this stream, the most closely related work is Ide and Talamàs (2025), who integrate AI into the hierarchy framework to study organizational restructuring. While they distinguish between autonomous AI and non-autonomous AI to analyze labor substitution, our framework departs by focusing strictly on the non-autonomous regime where AI functions as a productivity-enhancing tool rather than an independent agent. This modeling choice, which aligns with the current technological reality, fundamentally alters the demand derivation: value flows through worker productivity gains rather than standalone

[1] Other important contributions to this literature include Garicano and Hubbard (2016), Garicano and Rossi-Hansberg (2015), Fuchs et al. (2015), Bloom et al. (2014), Gumpert et al. (2022), and Carmona and Laohakunakorn (2025). See Garicano and Rossi-Hansberg (2015) for a comprehensive survey.

AI output. Consequently, whereas Ide and Talamàs (2025) study occupational choice, we focus on how the knowledge economy determines the willingness to pay for model capability and the resulting derived demand for training data.

Next to this, and closely related to our analysis of the downstream market structure, is the industrial organization literature on vertical quality differentiation. Shaked and Sutton (1982, 1983) originally demonstrated that quality differences relax price competition and that a finiteness property limits the number of firms that can profitably operate in equilibrium.[2] The standard result in this literature is that firms differentiate to segment markets and avoid head-to-head competition. Our model, however, generates a distinct outcome driven by the specific nature of AI as a productivity floor. Because an AI model benefits all workers with knowledge below its capability level identically, consumers in the low-skill segment face identical utility comparisons. We show that this produces pure winner-take-all dynamics rather than differentiated market shares, illuminating why the foundation model industry exhibits intense quality competition where the marginal revenue from quality improvement captures the entire low-skill market segment.

Related to the upstream data market, our analysis contributes to the mechanism design literature on informed principals (Myerson, 1983). Our setting inverts the standard informational structure found in the classical adverse selection framework of Akerlof (1970) or the signaling models following Spence (1973). Conceptually, the informational structure of our upstream market builds on the literature regarding price competition for an informed buyer, most notably Moscarini and Ottaviani (2001). They analyze price competition between two uninformed sellers offering differentiated goods to a single buyer who holds a private, noisy signal about relative product quality. We advance this reversed asymmetry paradigm into a novel procurement setting, introducing two fundamental theoretical departures tailored to the engineering realities of AI.

First, the nature of the information asymmetry is transformed: rather than holding a noisy signal about the sellers' relative qualities, our buyer (the AI producer) possesses perfect observability of data quality via computational scaling experiments, while the sellers (data holders) observe only noisy signals about their own assets. Second, the strategic interaction shifts from a duopoly sales market to a fragmented procurement market. Because our informed buyer faces no incentive to signal its superior information, the equilibrium involves systematic pooling. When

[2]Related contributions include Mussa and Rosen (1978), Gabszewicz and Thisse (1979), Moorthy (1988), and Vandenbosch and Weinberg (1995).

multiple data sellers compete, the buyer leverages its perfect observability to scan the market and cherry-pick hidden gems (high-quality datasets from sellers who received pessimistic private signals and subsequently underpriced their assets). This selection mechanism yields microfoundations for the inefficiencies observed in emerging AI training data markets and for the bifurcation between commoditized low-quality data and bottleneck high-quality data at the capability frontier.

Finally, there is a rapidly growing strand of literature on the production and impact of AI that we connect through our two-layer framework. We endogenize the relationship between data and capability using the neural scaling laws documented by Hoffmann et al. (2022), Kaplan et al. (2020), and subsequent work at Epoch AI.[3] By treating these engineering laws as an economic production function, we link data quality to model capability. This allows us to contribute to the broader debate on AI's labor market effects, where Acemoglu (2025) predicts modest TFP gains while Brynjolfsson et al. (2025) provide experimental evidence of substantial productivity improvements for lower-performing workers.[4] Our functional form is consistent with these findings, specifically the observation that AI adoption has significantly more positive productivity effects on low-skilled workers compared to high-skilled workers. By deriving demand from first principles, we provide the microfoundations for willingness to pay that these aggregate productivity estimates do not capture.

3 Market Structure of the Knowledge Economy

We model an economy where production requires resolving stochastic problems. Following the hierarchy-of-knowledge framework (Garicano, 2000; Garicano and Rossi-Hansberg, 2004), we characterize a population of knowledge workers with heterogeneous expertise $z \in [0, 1]$, distributed according to a cumulative distribution function $G(z)$ with a continuous density $g(z)$.

3.1 The Production Environment

Problems of difficulty $s \in [0, 1]$ arise with density $f(s)$. Assuming a uniform distribution of tasks, $s \sim U[0, 1]$, a worker with expertise z successfully resolves a problem if $z \geq s$. Thus, the

[3] Hoffmann et al. (2022) introduce compute-optimal training schedules, showing that model size and training data should scale proportionally. Sardana et al. (2024) extend these results to incorporate inference costs.

[4] Related empirical work includes Noy and Zhang (2023), Dell'Acqua et al. (2026), Eloundou et al. (2024), Chen et al. (2025), and Alekseeva et al. (2024). Acemoglu and Johnson (2023) discuss the broader question of whether AI will augment or replace human labor.

baseline productivity of worker z is:

$$F(z) = \int_0^z f(s)ds = z. \tag{1}$$

3.2 AI-Augmented Productivity: The Baseline Model

A model producer introduces a foundational model with capability $z_{AI} \in [0, 1]$. The joint capability of a human-AI pair is denoted by $\beta(z, z_{AI})$. In our baseline specification, the AI perfectly substitutes for human knowledge up to its capability limit:

$$\beta(z, z_{AI}) = \max(z, z_{AI}). \tag{2}$$

Although the model solves tasks up to difficulty $s \leq z_{AI}$, we assume it cannot operate as a standalone economic agent. Two factors necessitate this constraint. First, current Large Language Models (LLMs) operate primarily via probabilistic sampling and lack the agency required for autonomous production cycles (Schaeffer et al., 2025). Second, economic production requires a traceable channel of liability; unlike human agents who respond to monetary incentives or legal sanctions, AI models lack the legal and economic standing to internalize the costs of production errors.

Consequently, AI serves strictly as a productivity-enhancing tool activated by a human worker. The model establishes a productivity floor, yielding a total output $y(z, z_{AI}) = \beta(z, z_{AI})$ for agent z. In the baseline set-up we use $y(z, z_{AI}) = \max(z, z_{AI})$ for tractability, we extend the analysis to general $\beta(z, z_{AI})$ consistent with the notion "productivity compression"[5] that is consistent with empirical evidences of productivity of LLM (Chen et al., 2025; Dell'Acqua et al., 2026; Brynjolfsson et al., 2025), we define the notion "productivity compression" formally in Online Appendix C.

3.3 The Productivity Floor: Pricing and Capability Targets

We initially model the AI producer as a monopolist[6], who chooses a subscription price r_{AI} and a capability level z_{AI} to maximize profit. We separate short-run pricing from long-run capability choice.

[5] Fahn et al. (2025) use a similar discrete form of the "productivity compression" to study AI impact on the labor market effect.

[6] Section 3.4 endogenizes this market structure, showing that competition naturally collapses to a monopoly.

Let $C(z)$ denote the computational input required to achieve capability z, with the price of compute normalized to 1. Section 5 generalizes this cost function to incorporate data quality and pricing.

3.3.1 Adoption Decision and Market Demand

An agent z adopts the model if the productivity gain exceeds the subscription cost r_{AI}:

$$\max(z, z_{AI}) - z \geq r_{AI}. \tag{3}$$

For expert workers with $z > z_{AI}$, the model provides no marginal knowledge, resulting in zero gain. For agents with $z \leq z_{AI}$, the adoption condition simplifies to $z \leq z_{AI} - r_{AI}$. We define the marginal adopter as:

$$z^* = z_{AI} - r_{AI}. \tag{4}$$

Workers with $z \leq z^*$ adopt the model. This threshold generates a downward-sloping demand curve: an increase in r_{AI} lowers z^*, pricing out higher-skilled workers.

3.3.2 Optimal Pricing and Quality-Price Coupling

Taking the sunk training cost $C(z_{AI})$ as given, the producer sets r_{AI} to maximize total revenue (TR_{AI}):

$$\max_{r_{AI}} \quad TR_{AI} = \int_0^{z^*} r_{AI} g(t) dt = r_{AI} G(z^*). \tag{5}$$

Throughout this section and Section 5, we assume the revenue function $r \mapsto r\, G(z_{AI} - r)$ has a unique interior maximizer $r^* \in (0, z_{AI})$ satisfying the second-order condition $2g(z^*) - r^* g'(z^*) > 0$. This holds, for instance, whenever G is log-concave on $(0, z_{AI})$.[7] Under this interiority condition, the first-order condition (FOC) with respect to r_{AI} is:

$$\frac{\partial TR_{AI}}{\partial r_{AI}} = \underbrace{G(z^*)}_{\text{Infra-marginal Gain}} - \underbrace{r_{AI} g(z^*)}_{\text{Marginal Loss}} = 0. \tag{6}$$

The optimal price balances two opposing forces: raising the price generates an "infra-marginal

[7] Log-concavity of the CDF holds for a wide class of distributions including the uniform, normal, logistic, and Beta distributions with shape parameters ≥ 1. Since $G(z_{AI} - r)$ is then log-concave in r, the product $r\, G(z_{AI} - r)$ is quasi-concave and has a unique interior maximizer when $G(0) > 0$.

gain" by extracting more revenue from existing adopters, but incurs a "marginal loss" as workers at the threshold z^* exit the market. To analyze how the producer adjusts pricing in response to improved model capability, we apply the Implicit Function Theorem to the FOC. Defining $r^*(z_{AI})$ as the optimal price, we derive the following differential equation:

$$\frac{dr^*}{dz_{AI}} = \frac{g(z^*) - r^* g'(z^*)}{2g(z^*) - r^* g'(z^*)}. \tag{7}$$

Equation (7) governs the quality-price coupling, characterizing the producer's ability to extract surplus from technological advancement. In the benchmark case of a uniform distribution ($g'(z) = 0$), we obtain $dr^*/dz_{AI} = 1/2$.

3.3.3 Equilibrium Model Capability

In the long-run equilibrium, the producer endogenizes the choice of model capability, z_{AI}, by balancing the marginal revenue of quality improvement against the marginal cost of training, $C'(z_{AI})$. Differentiating the total revenue function with respect to z_{AI} using the Leibniz Integral Rule, and substituting the optimal price response (7), yields the optimality condition:

$$\frac{dTR_{AI}}{dz_{AI}} = \underbrace{r^* g(z^*)}_{\text{Market Expansion}} - \underbrace{r^* g(z^*) r'}_{\text{Adoption Erosion}} + \underbrace{r' G(z^*)}_{\text{Rent Extraction}} = C'(z_{AI}). \tag{8}$$

This decomposition highlights the three distinct economic forces driving the producer's investment incentives. The first term, market expansion, represents the revenue gained from satisfying increasingly skilled workers as the capability floor rises. However, as the producer raises the price to reflect this higher quality, adoption erosion occurs, pricing out some marginal users. These losses are mitigated by the rent extraction effect, where the producer captures additional surplus from the entire remaining user base through the higher price.

The sum of these marginal revenue components constitutes the producer's willingness to pay for capability. By equating this marginal benefit to the marginal cost of training $C'(z_{AI})$, this condition provides the fundamental structural link between the downstream value of AI and the upstream derived demand for pretraining data.

3.4 Winner-Take-All Dynamics

We now analyze competition between two producers of foundational models. Consistent with the monopolistic setting, we assume both producers face an identical cost function $C(z)$. In the short run, capability is fixed, and the marginal cost of serving a worker is c (which may be normalized to zero given the abundance of inference compute). In the long run, the marginal cost of improving model capability is denoted by $C'(z)$.

Following the Garicano-style normalization with uniformly distributed tasks, we maintain $F(z) = z$ throughout this section. Let firms $i \in \{1, 2\}$ offer models with capabilities z_1 and z_2 at subscription prices r_1 and r_2. A worker with expertise z evaluates model i according to

$$U_i(z) = \max\{z, z_i\} - r_i.$$

We define the *effective quality* of model i as $\tilde{z}_i \equiv z_i - r_i$. This is the net surplus delivered to any worker whose skill lies below the model's capability floor.

Tie-breaking conventions. If a worker is indifferent between adoption and non-adoption, we assume they do not adopt. If a worker is indifferent between two firms (i.e., equal utility), we assume they split evenly across the two firms. Unless otherwise stated, when we refer to a firm's *market share* in this section, we mean its share *among adopters*.

3.4.1 Symmetric Capabilities

In the symmetric case where $z_1 = z_2 = z$, the models are perfect substitutes. Standard Bertrand logic dictates that price competition drives subscriptions to marginal cost, $r_1 = r_2 = c$. While producers might attempt to coordinate on a monopoly price or split the market, Section 3.4.4 (Proposition 1) shows that this symmetric state is vulnerable to profitable escalation whenever the marginal value of a small capability lead exceeds the marginal training cost.

3.4.2 Quality-Price Coupling and Market Equilibrium

Suppose, without loss of generality, that Firm 1 possesses a technological advantage ($z_1 > z_2$). The competition is characterized by a purely vertical structure. However, unlike standard vertical differentiation models where firms segment the market by income or preference, the productivity-

floor nature of AI generates a distinct winner-take-all dynamic. We establish this through three lemmas.

Lemma 1 (Skill Protection)**.** *Fix $z_1 > z_2$ and prices $r_1 \geq 0$, $r_2 \geq 0$. For any worker type $z \in (z_2, z_1]$, adoption of Model 2 yields utility*

$$U_2(z) = z - r_2 \leq z \equiv U_0(z),$$

where $U_0(z)$ denotes the non-adoption utility. Under the convention that agents do not adopt when indifferent, Model 2 is never chosen on $(z_2, z_1]$.

Lemma 1 establishes that the superior model enjoys a protected market segment consisting of intermediate-skill workers. For a worker with skill $z > z_2$, the inferior model offers no marginal productivity gain over the worker's own capability; adoption is therefore (weakly) dominated by non-adoption at any nonnegative price.

Lemma 2 (Winner-take-all among low-skill workers)**.** *For the segment of workers $z \in [0, z_2)$, utility from model i is*

$$U_i(z) = z_i - r_i \equiv \tilde{z}_i.$$

Hence if $\tilde{z}_i > \tilde{z}_{-i}$ then firm i captures all adopters in this segment. If $\tilde{z}_i = \tilde{z}_{-i}$, then workers are indifferent and market shares are determined by the tie-breaking rule.

Lemma 2 characterizes the commoditized nature of the low-skill market. Workers with $z < z_2$ derive payoffs $z_1 - r_1$ from Firm 1 and $z_2 - r_2$ from Firm 2. Since these utilities are constants, the market does not segment. These workers simply compare effective qualities and universally choose the maximizer.

Lemma 3 (Leader's Pricing Advantage)**.** *If $z_1 > z_2$, Firm 1 can ensure $\tilde{z}_1 > \tilde{z}_2$ while setting $r_1 > c$. Specifically, if Firm 2 sets $r_2 = c$, Firm 1 captures the entire market for all $r_1 \in (c, c + z_1 - z_2)$. At $r_1 = c + z_1 - z_2$, low-skill workers are indifferent ($\tilde{z}_1 = \tilde{z}_2$) and market shares are determined by tie-breaking.*

Lemma 3 confirms that the technological leader can leverage its capability advantage to force the follower out of the market. Even if the follower prices aggressively at marginal cost ($r_2 = c$),

yielding an effective quality of $\tilde{z}_2 = z_2 - c$, the leader can secure the market at a price strictly above marginal cost by setting r_1 to (just) undercut the follower's effective quality.

Building on these lemmas, we derive the market equilibrium. In any equilibrium where both firms have entered, the inferior firm (Firm 2) exerts maximum competitive pressure by pricing at marginal cost, $r_2 = c$. Firm 1 responds by setting its price to undercut Firm 2's effective quality by an arbitrarily small amount, thereby capturing the entire market.

The equilibrium price r_1^* is determined by the tighter of two constraints: the profit-maximizing monopoly price r_1^m or the rival-imposed limit-pricing constraint. Formally,

$$r_1^* = \begin{cases} r_1^m, & \text{if } r_1^m < c + z_1 - z_2, \\ c + z_1 - z_2 - \varepsilon, & \text{if } r_1^m \geq c + z_1 - z_2, \end{cases} \tag{9}$$

where $\varepsilon > 0$ can be chosen arbitrarily small. This result highlights that competition in the foundational model industry acts as a ceiling on markups rather than a mechanism for market sharing. The leader extracts rents proportional to its technological lead, $z_1 - z_2$, effectively converting capability advantages directly into pricing power.

3.4.3 Extension to Portfolio Competition

A natural question is whether producers can mitigate this intense competition by releasing a portfolio of models, $P_i = \{(z_{i,j}, r_{i,j})\}_{j=1}^N$, rather than a single flagship product. We analyze this under two cost structures: independent development costs and costless degradation.

Lemma 4 (Portfolio Collapse)**.** *Let P_i be the portfolio of firm i. The competitive position of firm i is determined solely by the scalar*

$$\tilde{z}_i^* \equiv \max_{(z,r)\in P_i} \{z - r\}.$$

Any model $(z_{i,j}, r_{i,j}) \in P_i$ such that $z_{i,j} - r_{i,j} < \tilde{z}_i^*$ captures zero market share.

Lemma 4 demonstrates that in a vertical market where goods are perfect substitutes for low-skill workers, consumers strictly prefer the option offering the highest net surplus. A firm cannot segment the market by offering inferior options because its own best offer cannibalizes them.

Consequently, the strategic interaction between firms simplifies from competing sets of models to a competition between two scalar values: the maximum effective quality of each firm.

Lemma 5 (Strategic Redundancy)**.** *Suppose the total cost of portfolio P_i is $C(P_i) = \sum_{(z,r)\in P_i} C(z)$ with $C(z) > 0$. For any multi-model portfolio P_i ($|P_i| > 1$), there exists a single-model portfolio $P_i' \subset P_i$ such that $\pi(P_i') > \pi(P_i)$ where π denotes profit.*

Lemma 5 implies that if developing distinct models incurs fixed training costs, maintaining a portfolio is a strictly dominated strategy. Since only the model with the highest effective quality $\tilde{z}_i^*$ captures market share (by Lemma 4), all other models generate zero revenue while incurring positive development costs $C(z)$. A profit-maximizing firm will therefore rationalize its portfolio to the single most competitive model.

Lemma 6 (Flagship Dominance under Free Degradation)**.** *Assume degradation is costless (models can be served at $r \geq c$ regardless of size). Let $z_i^{\max} = \max_j z_{i,j}$ be the flagship capability. Then:*

$$\max_{(z,r)\in P_i} \{z - r\} \leq z_i^{\max} - c.$$

Lemma 6 addresses the case where smaller models are "free" spin-offs of the flagship (e.g., checkpoints or quantized versions). It states that a firm can never improve its competitive position by offering a smaller model. The flagship model priced at marginal cost always delivers the highest possible effective quality. Therefore, the competitive outcome is entirely determined by the interaction of the firms' frontier models, rendering the "model zoo" strategically irrelevant for market capture.

3.4.4 Local Profitable Deviation from Symmetric Capabilities

We now formalize the instability logic behind capability races. Throughout this subsection we maintain the Section 3 normalization $F(z) = z$.

We conceptualize the capability-pricing game as a two-stage process. In the first stage, firms $i \in \{1, 2\}$ simultaneously choose their target capabilities $z_i \in [0, 1]$, incurring a training cost $C(z_i)$. We assume C is continuously differentiable (C^1), strictly increasing, and satisfies $C(0) = 0$. In the second stage, having observed the capability profile (z_1, z_2), firms simultaneously set their subscription prices $r_i \geq c$, thereby playing the pricing subgame analyzed in Section 3.4.2.

If the firms choose a symmetric capability profile (z, z), Bertrand competition drives prices down to marginal cost ($r_1 = r_2 = c$). Under the tie-breaking convention, each firm serves half of the adopter mass, $G(z - c)$. However, because price equals marginal cost, the second-stage operating profit is zero. Consequently, each firm's continuation payoff at the symmetric profile reduces to the sunk training cost:

$$\Pi^{sym}(z) = -C(z). \tag{10}$$

Suppose Firm 1 unilaterally deviates in the first stage to a marginally higher capability $z + \varepsilon$ (with $\varepsilon > 0$), while Firm 2 remains at z. In the subsequent pricing subgame, Firm 1 can strategically post a price of:

$$r_1(\varepsilon) = c + \varepsilon - \varepsilon^2. \tag{11}$$

Given any follower price $r_2 \geq c$ and any worker type $x \leq z - c$, the respective utilities are

$$U_1(x) = (z + \varepsilon) - r_1(\varepsilon) = z - c + \varepsilon^2, \qquad U_2(x) = z - r_2 \leq z - c, \qquad U_0(x) = x \leq z - c.$$

Because $U_1(x) > U_2(x)$ and $U_1(x) > U_0(x)$, every worker with $x \leq z - c$ strictly prefers Firm 1. By adopting this pricing strategy, the deviating firm guarantees itself a demand of at least $G(z - c)$ and an operating profit of at least $(\varepsilon - \varepsilon^2)G(z - c)$. We define the resulting guaranteed continuation payoff as

$$\Pi^{dev}(z + \varepsilon, z) \equiv (\varepsilon - \varepsilon^2)G(z - c) - C(z + \varepsilon). \tag{12}$$

Proposition 1 (Local profitable deviation from symmetry)**.** *Fix $z \in (c, 1)$ with $G(z - c) > 0$. Then*

$$\Pi^{dev}(z + \varepsilon, z) - \Pi^{sym}(z) = \big[G(z - c) - C'(z)\big]\varepsilon + o(\varepsilon). \tag{13}$$

Consequently, if

$$G(z - c) > C'(z), \tag{14}$$

then the symmetric capability profile (z, z) is locally unstable: for all sufficiently small $\varepsilon > 0$, Firm 1 has a profitable unilateral deviation to $z + \varepsilon$ followed by the pricing rule (11)*.*

Proof. Using (10) and (12),

$$\Pi^{dev}(z+\varepsilon, z) - \Pi^{sym}(z) = (\varepsilon - \varepsilon^2)G(z-c) - \big[C(z+\varepsilon) - C(z)\big].$$

Since C is C^1,

$$C(z+\varepsilon) - C(z) = C'(z)\varepsilon + o(\varepsilon).$$

Substituting yields (13). If (14) holds, the leading term is strictly positive, so the deviation payoff is positive for all sufficiently small ε. □

The economic intuition behind this instability is distinct from a discrete transfer of monopoly rents. Rather, the local gain from a capability lead stems from the strictly positive markup $(\varepsilon - \varepsilon^2)$ that the technological leader can extract while still strictly dominating the follower across the entire inframarginal adopter mass $G(z-c)$. As long as this guaranteed operating profit exceeds the marginal cost of capability improvement, symmetric competition cannot be sustained.

In this section, we treated model capability z as an exogenous parameter to isolate the downstream market mechanics. We found that the role of AI as a productivity floor generates strong concentration pressures, formalized in Proposition 1 as a sufficient local-instability condition for symmetric capability configurations. In Section 4, we examine the upstream market for pretraining data, before synthesizing both layers in Section 5 to endogenize the cost function $C(z)$ via the Neural Scaling Laws.

4 Structure of the Pretraining Data Market

In this section, we analyze the upstream market where model producers acquire pretraining data to develop foundational models. While a transparent, liquid market for such data is still in its infancy as of 2026, the prevailing status quo (often characterized as a cycle of unauthorized scraping followed by litigation) is fundamentally unsustainable.[8] We contend that the emergence of a structured pretraining data market is a prerequisite for the long-term health of the foundational model ecosystem.

[8] High-profile legal challenges against firms like OpenAI highlight the mounting friction between data usage for training and intellectual property rights. See, e.g., `https://www.wsj.com/tech/ai/italian-privacy-watchdog-fines-openai-15-5-million-over-chatgpt-data-use-8c5b4d97`.

Because gains in model capability are driven primarily by the volume and quality of training data (Gunasekar et al., 2023; Li et al., 2023), as codified by empirically calibrated scaling laws (Hoffmann et al., 2022), the supply side of this market represents a potential bottleneck. Inefficient market structures may result in a "data famine," where the supply of high-quality pretraining data stagnates, thereby halting technological progress.[9]

Initially, we focus on a model where technological investment is restricted to data-driven scaling.[10] Later, we extend the model to accommodate data quality curation consistent with Subramanyam et al. (2026); Hoffmann et al. (2022), which allows for performance gains independent of training data volume.

A defining characteristic of this market is *information asymmetry*. Model producers (the buyers) can accurately estimate the marginal utility of a dataset by testing scaling-law performance on small samples—a process requiring significant computational infrastructure and technical expertise. Data producers (the sellers), lacking this infrastructure, can only form noisy beliefs regarding the true value of their data. We therefore model this interaction as an *informed principal* problem, where the informed buyer designs the trading mechanism.

4.1 Model Setup

The upstream market consists of a single representative model producer (the buyer) and a set of I discrete data producers (the sellers), indexed by $i \in \{1, \ldots, I\}$. Each seller holds a unique dataset D_i characterized by a quality parameter b_i. In the baseline specification, data quality is binary, $b_i \in \{b_L, b_H\}$, where $b_H > b_L > 0$ captures the marginal contribution of the dataset to the model's training objective. The mapping from data quality to the buyer's ultimate utility is governed by the structural cost function $C(z, b)$ developed in Section 5.

A defining feature of the pretraining data market is a reversal of the standard information asymmetry. While data sellers are familiar with the raw content of their datasets, they typically lack the computational infrastructure required to evaluate their usefulness for training large-scale

[9] Recent trends show model producers, including OpenAI and ByteDance, increasingly hiring subject-matter experts and Ph.D. students to generate or label high-quality data, signaling a shift toward a "bespoke" data economy. See `https://www.theinformation.com/articles/why-a-14-billion-startup-is-now-hiring-phds-to-train-ai-from-their-living-rooms`.

[10] Following the Chinchilla scaling laws, the optimal model size must co-vary with the data size and quality (Hoffmann et al., 2022; Li et al., 2023). Some researchers forecast that high-quality, naturally generated human data may be exhausted within the decade (Villalobos et al., 2024), necessitating a transition to synthetic or expertly curated datasets.

foundation models. In contrast, the model producer can conduct pilot training runs or deploy automated quality classifiers, allowing them to assess data utility with high precision.

We assume that the buyer perfectly observes the true quality b_i of every dataset available in the market. By contrast, seller i does not observe b_i directly but instead receives a noisy signal $s_i \in \{L, H\}$. Signal accuracy is governed by a precision parameter $\gamma \in (1/2, 1)$, such that

$$\Pr(s_i = H \mid b_i = b_H) = \Pr(s_i = L \mid b_i = b_L) = \gamma.$$

There is a common prior $\mu_0 = \Pr(b_i = b_H)$ that any given dataset is of high quality.

Upon observing the signal s_i, the seller updates their belief about the dataset's quality using Bayes' rule. Let μ_H and μ_L denote the posterior beliefs following signals H and L, respectively:

$$\mu_H \equiv \frac{\gamma\mu_0}{\gamma\mu_0 + (1-\gamma)(1-\mu_0)}, \qquad \mu_L \equiv \frac{(1-\gamma)\mu_0}{(1-\gamma)\mu_0 + \gamma(1-\mu_0)}. \tag{15}$$

Standard properties of Bayesian updating imply $\mu_H > \mu_0 > \mu_L$. As $\gamma \to 1$, the seller's information converges to that of the buyer, while $\gamma \to 1/2$ corresponds to complete ignorance.

The buyer's valuation of a dataset depends on its quality and is denoted by $V(b_i)$, with $V_H \equiv V(b_H)$ and $V_L \equiv V(b_L)$, where $V_H > V_L$. Economically, $V(b)$ represents the reduction in computational resources required to reach a fixed capability target. Data provision is costly: each seller incurs a constant marginal cost c_d to prepare, host, and transfer the dataset. To abstract from market breakdown and focus on distributional consequences driven by information asymmetry, we assume $V_L > c_d$. Under this condition, trade is socially efficient for all quality levels, and any failure to trade or surplus extraction arises from strategic interaction rather than a lack of intrinsic value.

4.2 The Single Seller Case

We begin with a bilateral monopoly ($I = 1$) to establish a benchmark. Nature draws the data quality $b \in \{b_L, b_H\}$. The buyer observes b perfectly, while the seller observes only the noisy signal s. The buyer then makes a take-it-or-leave-it price offer p, after which the seller updates their belief and decides whether to accept or reject the offer.

Although prices may in principle convey information, the buyer's objective of minimizing

acquisition costs generates a strong incentive to suppress any informative signaling.

Lemma 7 (Existence of Pooling Equilibrium)**.** *If trade is socially efficient for all quality levels ($V_L \geq c_d$), there exists a Perfect Bayesian Equilibrium in which the buyer offers a pooling price $p^* = c_d$ regardless of the realized quality. The seller accepts this offer for both signals.*

The intuition is straightforward. The price $p^* = c_d$ weakly satisfies the seller's participation constraint for both quality levels, rendering the offer uninformative. As a result, the seller cannot infer data quality from the price and is unable to extract a premium for high-quality data.

Lemma 8 (The Strategy of Obfuscation)**.** *The buyer strictly prefers the pooling equilibrium at $p^* = c_d$ to any separating equilibrium. Any separating outcome requires the buyer to pay a strictly positive information premium for high-quality data.*

To see this, consider a candidate separating equilibrium in which the buyer offers $p_L = c_d$ for low-quality data and $p_H > c_d$ for high-quality data. Buyer profits are then $\pi_{sep}(b_H) = V_H - p_H$ and $\pi_{sep}(b_L) = V_L - c_d$. Because the buyer observes b perfectly, a buyer facing high-quality data can profitably deviate by mimicking the low-quality type and offering c_d. Since $V_H - c_d > V_H - p_H$ for any $p_H > c_d$, all separating equilibria unravel.

The pooling equilibrium persists because $p^* = c_d$ covers the seller's marginal cost. Since the offer is independent of b, it conveys no additional information beyond the seller's private signal; thus the seller's posterior remains $\mu(b_H \mid s) \equiv \mu_s$, based solely on the signal. Crucially, acceptance depends only on $p \geq c_d$, not on beliefs about quality, so the seller accepts at $p^* = c_d$ regardless of s. Even though the seller knows that the dataset is valuable in expectation, the inability to credibly certify quality forces acceptance at the cost floor.

Proposition 2 (Information Rent Extraction)**.** *In the unique equilibrium of the bilateral negotiation, the buyer captures the entire expected social surplus. The buyer's expected information rent is*

$$\mathbb{E}[\pi_B] = \mu_0 V_H + (1 - \mu_0) V_L - c_d.$$

This benchmark highlights a stark power imbalance. When only the buyer can verify data quality, the seller becomes a price taker earning their reservation value. Notably, the seller's signal precision γ plays no role in equilibrium outcomes as long as the buyer retains take-it-or-leave-it power.

4.3 Single Seller with Strategic Bargaining

We now relax the buyer's unilateral pricing power by allowing the seller to set a reserve price prior to trade. Before observing the buyer's offer, the seller commits to a reserve price $\underline{p}(s)$ contingent on their private signal s. This reserve price reflects both the seller's bargaining power $\theta \in [0, 1]$ and their posterior belief about the dataset's value.

A seller observing signal s computes $\mathbb{E}[V(b) \mid s]$ and sets

$$\underline{p}(s) = \theta \mathbb{E}[V(b) \mid s] + (1 - \theta)c_d. \tag{16}$$

Using the posteriors derived above, the reserve prices following high and low signals are

$$\underline{p}_H = \theta[\mu_H V_H + (1 - \mu_H)V_L] + (1 - \theta)c_d, \tag{17}$$

$$\underline{p}_L = \theta[\mu_L V_H + (1 - \mu_L)V_L] + (1 - \theta)c_d. \tag{18}$$

Since $\mu_H > \mu_L$, the seller demands a higher price following a favorable signal.

Lemma 9 (The Value of Pessimistic Signals)**.** *The buyer's information rent, defined as* $\pi_B(b, s) = V(b) - \underline{p}(s)$*, is strictly maximized when high-quality data is paired with a pessimistic signal, that is, when* $(b = b_H, s = L)$*.*

Lemma 9 identifies the precise configuration under which the buyer's informational advantage is most valuable. When the seller receives a pessimistic signal, they infer a low probability that their dataset is of high quality and therefore set a relatively low reserve price. If the dataset is in fact high quality, the buyer acquires a valuable asset at a price reflecting the seller's underestimation rather than its true contribution to model performance.

Crucially, the buyer's rent does not arise solely from the presence of high-quality data, but from the discrepancy between true quality and the seller's belief about that quality. When the seller receives an optimistic signal, their reserve price partially internalizes the dataset's expected value, compressing the buyer's surplus. In contrast, pessimistic signals widen the wedge between the buyer's valuation and the seller's pricing rule, generating a larger informational rent.

This result highlights that the buyer benefits most from *misalignment* between quality and beliefs, rather than from quality per se. High-quality datasets that are correctly recognized as such

are comparatively less profitable than "hidden gems" whose value is obscured by noisy inference on the seller's side.

Proposition 3 (Information Rent and Signal Precision)**.** *Assume the seller prices at posterior expected value ($\theta = 1$) and $\gamma \in (1/2, 1)$. Then the buyer's expected information rent is strictly decreasing in the seller's signal precision γ, and vanishes as $\gamma \to 1$.*

Proposition 3 formalizes how the buyer's informational advantage depends on the accuracy of the seller's signal. Signal precision determines the extent to which the seller's reserve price internalizes true data quality, and therefore the magnitude of the pricing wedge that the buyer can exploit.

In the limit of perfect information, as $\gamma \to 1$, the seller's posterior beliefs converge to the truth, with $\mu_H \to 1$ and $\mu_L \to 0$. In this case, the seller correctly infers the dataset's quality and sets a reserve price that is approximately equal to

$$\underline{p}(s) \approx \theta V(b) + (1 - \theta)c_d.$$

Pricing becomes fully quality-contingent, and the buyer's private information no longer confers an advantage. The informational wedge between valuation and price collapses, and the buyer's information rent vanishes.

At the opposite extreme, as $\gamma \to 1/2$, the signal becomes pure noise. Posterior beliefs no longer depend on the realized signal, so that $\mu_H = \mu_L = \mu_0$. The seller's reserve price is therefore insensitive to true data quality and reflects only the prior expected value. In this regime, high-quality and low-quality datasets are priced identically from the seller's perspective, despite large differences in their true contribution to the buyer's objective.

This insensitivity maximizes the "surprise" gap between the buyer's valuation and the seller's price whenever high-quality data is realized. Because the buyer can perfectly identify quality, they systematically capture surplus from these mispriced high-quality datasets. Crucially, the frequency of high-quality data is fixed at μ_0 regardless of γ; what changes is the seller's expected reserve price conditional on high quality. Lower signal precision reduces the seller's ability to condition the reserve price on true quality, pushing the expected payment toward the prior expected value and thereby widening the buyer's rent on every high-quality transaction.

Taken together, Proposition 3 shows that information accuracy acts as an implicit information tax on the buyer. Improvements in the seller's ability to assess data quality compress pricing errors and reduce rent extraction, while noise and opacity in the upstream market are a direct source of buyer profitability.

4.4 Multiple Sellers and the Selection Advantage

We now extend the analysis to a market with $I \geq 2$ data producers. Each seller i holds a unique dataset D_i with quality b_i, where qualities are independently and identically distributed with $\Pr(b_i = b_H) = \mu_0$. The buyer continues to observe all qualities perfectly, while each seller observes only a private noisy signal. This environment allows the buyer's informational advantage to operate not only through pricing, but also through selection across multiple datasets.

We assume throughout this section that datasets are non-overlapping, so that the buyer's total value is additive and equal to $\sum_i V(b_i)$. This assumption isolates the selection mechanism in its simplest form; Section 5 relaxes it by introducing diminishing returns through AI scaling laws.

Lemma 10 (Systematic Cherry-Picking)**.** *With non-exclusive access to datasets, the buyer maximizes surplus by systematically targeting underpriced data, defined as datasets whose true quality exceeds the seller's expected valuation conditional on their signal* $\mathbb{E}[V(b_i) \mid s_i]$.

Lemma 10 formalizes a simple but powerful mechanism. Because sellers price their data based on posterior beliefs rather than true quality, some datasets are systematically mispriced. In particular, a dataset is underpriced when it is of high quality but is associated with a pessimistic signal ($b_i = b_H, s_i = L$). The probability of this event for any given seller is $\mu_0(1-\gamma)$.

In a market with I sellers, the expected number of such "hidden gems" is $I\mu_0(1-\gamma)$. As the market becomes thick, the buyer can increasingly rely on these underpriced high-quality datasets, bypassing sellers whose optimistic signals lead them to demand higher prices. The buyer's advantage therefore scales with market size, even though the underlying distribution of data quality remains unchanged.

The logic becomes sharper under exclusive licensing, where sellers compete to supply a single dataset.

Lemma 11 (Belief-based competitive pricing benchmark)**.** *(Competitive benchmark.) Under exclusive licensing, we adopt the reduced-form benchmark that each seller posts a signal-contingent*

price equal to its posterior expected valuation:

$$p_i(s_i) \equiv \mathbb{E}[V(B_i) \mid s_i], \qquad s_i \in \{H, L\}.$$

If $V(b_H) > V(b_L)$ *and the signal is informative in the sense that* $\mathbb{P}(B_i = b_H \mid H) > \mathbb{P}(B_i = b_H \mid L)$*, then prices are strictly monotone in the signal:*

$$p_i(H) > p_i(L).$$

Moreover, under this benchmark $p_i(s_i)$ *depends only on the belief induced by* s_i*, not on the realized quality* b_i*.*

Under exclusive access, competition forces sellers to bid based on their own beliefs (signals) rather than on realized quality. Under the benchmark in Lemma 11, each seller posts $p_i(s_i) = \mathbb{E}[V(B_i) \mid s_i]$, and the buyer chooses

$$i^* \in \arg\max_{i \in \{1,\dots,I\}} \Big\{ V(b_i) - p_i(s_i) \Big\}. \tag{19}$$

Proposition 4 (The Selection Advantage)**.** *Maintain the exclusive-licensing case and the benchmark pricing rule in Lemma 11, i.e.* $p_i(s_i) = \mathbb{E}[V(B_i) \mid s_i]$*. Then:*

(i) (Optimal choice) The buyer selects a dataset that maximizes net surplus, $i^* \in \arg\max_i \{V(b_i) - p_i(s_i)\}$*.*

(ii) (Bias toward pessimism) Net surplus is highest when $(b_i, s_i) = (b_H, L)$*.*

(iii) (Comparative statics) The buyer's expected surplus is strictly increasing in I*, satisfies* $\lim_{\gamma \to 1} \mathbb{E}[V(b_{i^*}) - p_{i^*}(s_{i^*})] = 0$*, and can be non-monotone in* γ *for intermediate* (I, γ)*.*

Proof. See Appendix A.2.8.

The selection advantage described in Proposition 4 arises from the interaction of market fragmentation and asymmetric information. When sellers compete for an exclusive license, competition drives each seller's price $p_i(s_i)$ toward their posterior expected valuation $\mathbb{E}[V(b_i) \mid s_i]$. The buyer therefore faces a menu of datasets priced according to heterogeneous beliefs rather than realized quality.

Because the buyer observes all qualities $\{b_i\}_{i=1}^I$, they select the dataset that maximizes the surplus gap $V(b_i) - \mathbb{E}[V(b_i) \mid s_i]$. This gap is a random variable driven by sellers' signal errors. As the number of sellers I increases, the buyer effectively draws more realizations from the distribution of these pricing errors and selects the maximum order statistic. Even if the event $(b_i = b_H, s_i = L)$ (a high-quality dataset priced according to a pessimistic belief) is rare, a sufficiently large market ensures its occurrence with high probability.

As a consequence, the buyer's expected surplus is governed not by the mean level of data quality in the market, but by the dispersion of sellers' belief errors induced by imperfect signals. Greater market thickness increases the likelihood of extreme pricing errors, strictly raising the buyer's expected surplus. The effect of signal precision γ is more subtle: as γ rises, mispricing events become rarer, but conditional on occurring, the price discount on hidden gems grows larger. For sufficiently large I, hidden gems appear with near certainty, and the conditional surplus effect can dominate, making the relationship between γ and buyer surplus non-monotone. In the limit $\gamma \to 1$, however, mispricing vanishes entirely and the buyer's information rent converges to zero.

This mechanism identifies a systematic source of computational rent in fragmented data markets. The buyer's ability to accurately evaluate data quality through computational experimentation allows them to transform cross-sectional variation in seller beliefs into predictable surplus. From a market design perspective, interventions that increase signal precision γ (such as data quality certification or standardized third) such as third party evaluation reduce the variance of pricing errors, compress the surplus gap, and limit the buyer's ability to extract rents through selective acquisition of underpriced datasets.

5 Combined Two-Layer Market

In this section, we endogenize the model producer's capability choice by linking the knowledge economy (Section 3) with the pretraining data market (Section 4). The fundamental link is the *neural scaling law*, which governs how data quality b modulates the efficiency of computational inputs in producing model capability z. This mapping transforms the abstract value function $V(b)$ from our previous analysis into a concrete equilibrium outcome derived from a joint optimization over data acquisition and training scale.

5.1 Scaling Data and Compute: The Production of Capability

A fundamental empirical regularity in foundational model development is that performance follows a predictable power law with respect to scale (Hoffmann et al., 2022). From an economic and managerial perspective, this relationship is best conceptualized not merely as a technical benchmark, but as a *production frontier* where the inputs are computational capital (compute, C) and informational capital (data quality, b), and the output is the downstream market-clearing capability z.

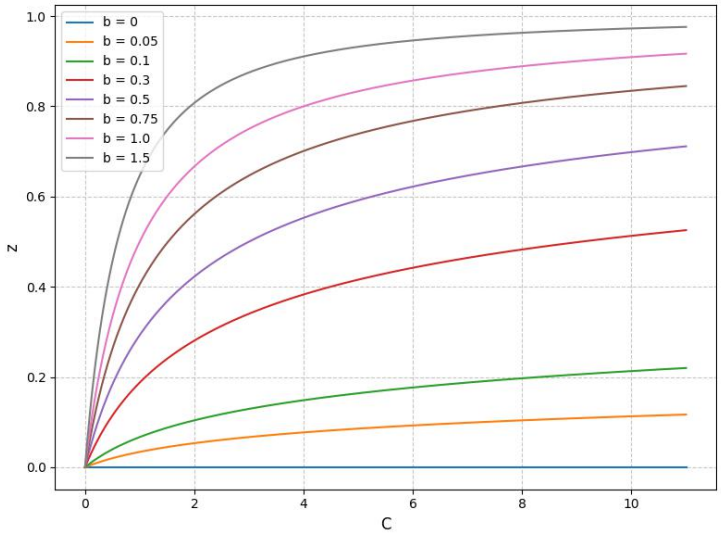
Unlike traditional linear production functions, AI capability production exhibits severely diminishing marginal returns. We formalize this relationship via a capability production function:

$$1 - z = \left(\frac{1}{C+1}\right)^{b}. \tag{20}$$

Given the normalization from Section 3 that $F(z) = z$ (uniform task density), we interpret $z \in [0, 1]$ directly as the fraction of tasks solved. Equation (20) therefore implies the explicit capability mapping:

$$z = 1 - (C+1)^{-b}. \tag{21}$$

Figure 1: Illustration of Power Law ($z = 1 - (C+1)^{-b}$)



Notes: The figure illustrates the diminishing marginal returns of compute C on capability z. Higher data quality b enables the producer to reach the capability frontier at significantly lower compute scales.

The parameter $b \in (0, +\infty)$ captures the *scaling efficiency* of the training set. To understand the model producer's cost structure, we derive the marginal capability gain per unit of compute from (21):

$$\frac{dz}{dC} = b(C+1)^{-(b+1)}. \tag{22}$$

We now define the cost function $C(z, b)$ as the total computational capital C required to reach capability z. By inverting the production relationship (21), we obtain:

$$C(z, b) = (1-z)^{-\frac{1}{b}} - 1. \tag{23}$$

This specification generalizes the cost function $C(z)$ introduced in Section 3 by explicitly incorporating data quality.

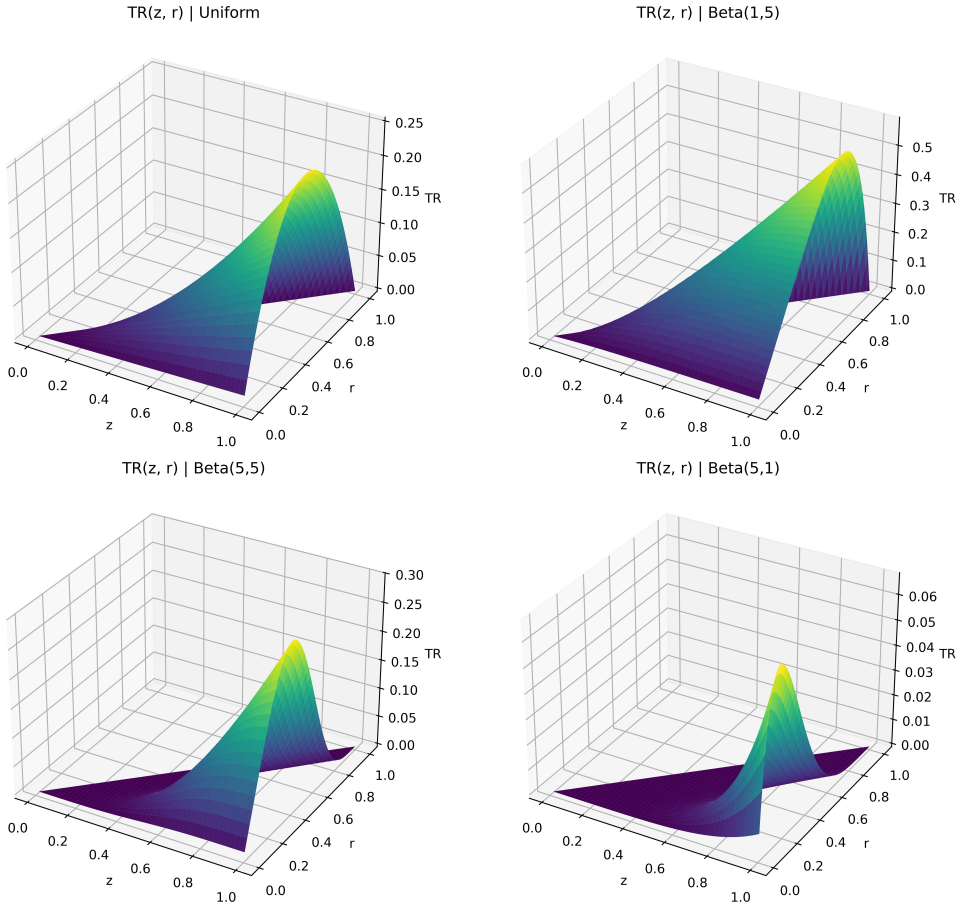
Under this formulation, the marginal cost of advancing the capability frontier, $\frac{\partial C}{\partial z}$, is:

$$\frac{\partial C}{\partial z} = \frac{1}{b}\left(\frac{1}{1-z}\right)^{\frac{1}{b}+1}. \tag{24}$$

Two strategic insights emerge from this formulation. First, the marginal cost is strictly increasing in z, reflecting the escalating difficulty of frontier innovation. Second, the cross-partial $\frac{\partial^2 C}{\partial z \partial b} < 0$ implies that high-quality data serves as a *structural substitute* for compute. As the industry targets higher capability levels ($z \to 1$), the relative value of data quality b grows exponentially. This quality-compute substitution effect means that, at the frontier, a 1% increase in b yields greater cost savings than a substantial increase in C.

5.2 Monopoly Optimization: The Optimal Quality-Capability Bundle

Figure 2: TR Surfaces (Monopoly)



Notes: The surface represents the producer's total revenue as a joint function of capability z and price r. It is easy to see that the short-run implicit function $r^*(z) = \underset{r}{\operatorname{argmax}}\, TR(r, z)$ can be obtained by taking the FOC, and as long as the compute cost $C(z)$ is convex, the long-run $r^*(z) = \underset{r}{\operatorname{argmax}}\, TR(r, z) - C(z)$ is single-valued and can also be obtained by taking the FOC.

The monopolist's total profit is the difference between downstream revenue and the upstream cost of computational capital C. For a fixed data quality b, the producer chooses capability z and price r to maximize:

$$\Pi = \underbrace{\int_0^{z-r} r g(t) dt}_{TR(z,r)} - C(z, b). \tag{25}$$

Assuming an interior optimum exists (guaranteed by the interiority conditions formalized in Assumption 3 of Online Appendix B), the equilibrium model capability z^* and optimal price r^* satisfy two joint first-order conditions: the price-setting condition ($\frac{\partial TR}{\partial r} = 0$), which characterizes the short-run optimal price for a given capability level, and the investment condition ($\frac{dTR}{dz} = \frac{\partial C}{\partial z}$), which balances the marginal revenue from capability improvement against the marginal training cost. The second-order conditions follow from the log-concavity of G (for the pricing problem) and the convexity of $C(z, b)$ in z (for the capability problem).

Drawing on the analysis in Section 3, we establish the quality-price coupling that characterizes how the producer extracts surplus as capability improves:

$$\frac{dr}{dz} = \frac{g(z-r) - rg'(z-r)}{2g(z-r) - rg'(z-r)}. \tag{26}$$

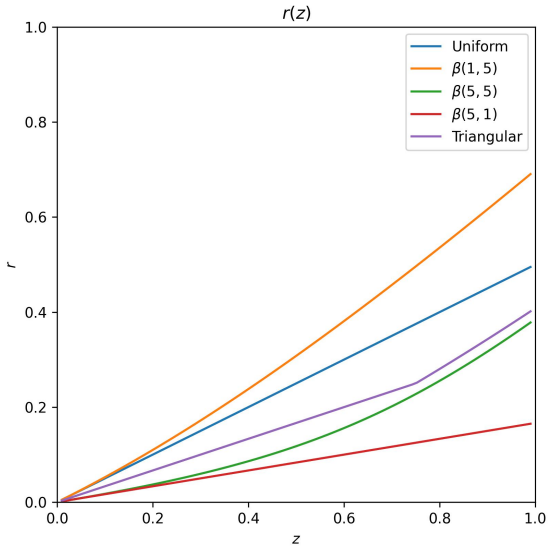
Applying Leibniz's rule to the total revenue function and substituting the optimal price response yields the total marginal revenue with respect to capability:

$$\frac{dTR}{dz} = \underbrace{rg(z-r)\left(1 - \frac{dr}{dz}\right)}_{\text{Expansion \& Erosion}} + \underbrace{\frac{dr}{dz}G(z-r)}_{\text{Rent Extraction}}. \tag{27}$$

The structural equilibrium is therefore defined by two conditions. First, the price-capability constraint: $\int_0^{z-r} g(t)dt = rg(z-r)$. This condition determines the optimal subscription price for any given capability level, balancing the marginal gain from higher prices on infra-marginal workers against the revenue loss from marginal adopters. Second, the capability-cost constraint: $\frac{dTR}{dz} = \frac{\partial C(z,b)}{\partial z}$. This equates the marginal revenue from market expansion and rent extraction to the marginal computational capital required by the scaling law:

$$\frac{dTR}{dz} = \frac{1}{b}\left(\frac{1}{1-z}\right)^{\frac{1}{b}+1}. \tag{28}$$

Figure 3: Price-Capability Constraint (Monopoly)



Notes: This figure illustrates the short-run implicit function $r^*(z) = \underset{r}{\operatorname{argmax}}\, TR(r, z)$ *for different distributions of* $g(\cdot)$.

These dual constraints highlight the fundamental trade-off in the foundational model economy. The producer is constrained not only by the physical limits of compute (C), but also by the distribution of human expertise $g(z)$. As capability z approaches the frontier, the marginal cost of compute escalates sharply, necessitating a concurrent increase in r that eventually erodes the adoption base. Thus, the optimal z^* is endogenously limited by both the scaling law and the skill distribution of the workforce.

5.2.1 The Monopoly Baseline: Scaling Laws and Waste Data

To establish a benchmark for our analysis, we first characterize the equilibrium in a regime where the model producer relies exclusively on an exogenous supply of low-quality "waste data," denoted by D_0. In this scenario, data is treated as a commodity with quality b_0 and unit price p_0. This baseline isolates the effects of the neural scaling law on the producer's capability choice before introducing a strategic data market.

In this regime, the producer's total training expenditure, $\mathcal{C}$, comprises both computational resources (C) and data procurement costs ($p_0 C$). Based on the extension of the cost function

$C(z, b)$ introduced in Section 5.1, the total expenditure required to achieve a capability level z is:

$$\mathcal{C}(z, b_0, p_0) = (1 + p_0)C(z, b_0). \tag{29}$$

Substituting the structural relationship derived from the neural scaling law, where $1 - z = (C + 1)^{-b_0}$, we express the total expenditure as $\mathcal{C}(z, b_0, p_0) = (1 + p_0)[(1 - z)^{-1/b_0} - 1]$. The marginal cost of capability with respect to total expenditure is therefore:

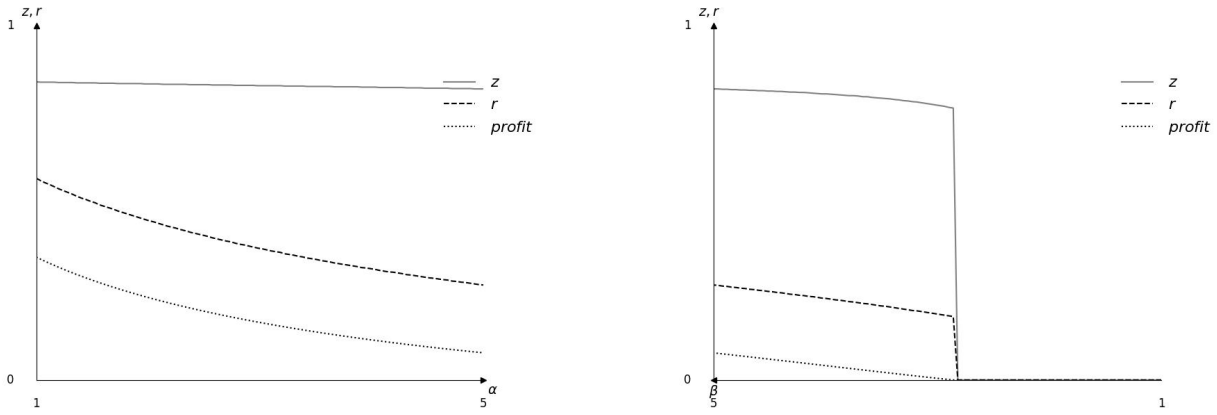
$$MC_z(z, b_0, p_0) = \frac{\partial \mathcal{C}}{\partial z} = \frac{1 + p_0}{b_0} \left(\frac{1}{1 - z} \right)^{\frac{1}{b_0}+1}. \tag{30}$$

This marginal cost function highlights a critical economic tension: as the model approaches the capability frontier ($z \to 1$), the marginal cost grows at an increasing rate, governed by the data quality parameter b_0. Lower data quality (smaller b_0) accelerates this cost explosion, creating a "diminishing returns" trap that limits the equilibrium capability of the AI.

The equilibrium capability z^* and subscription price r^* are determined by the interplay between the producer's cost structure and the distribution of worker skills $g(z)$. We illustrate the key comparative statics below through a numerical evaluation of the equilibrium conditions under specific parametric families (Beta distributions for g and the power-law cost function $C(z, b)$). These results characterize the economic mechanisms at work but are not general propositions; they depend on the functional forms chosen for illustration.

The composition of the labor market dictates the producer's strategic trade-off between capability investment and price extraction. As the distribution $g(z)$ shifts from left-skewed to center-skewed (Figure 4a), the equilibrium capability z^* remains notably inelastic. This suggests that at moderate skill levels, the exponential nature of the training cost curve dominates, leading the producer to respond to shifts in worker composition primarily through price adjustments (r^*) rather than capability gains. Conversely, as the market shifts toward high-skill experts (Figure 4b), we observe a sharp decline in equilibrium capability followed by a market collapse. This underscores a "sophistication trap": when the labor force is highly skilled, the AI capability required to provide marginal value becomes prohibitively expensive under the neural scaling law, eventually rendering the monopolistic provision of AI unviable.

Figure 4: Effect of Worker Skill Distribution $g(z)$ on Equilibrium z, r



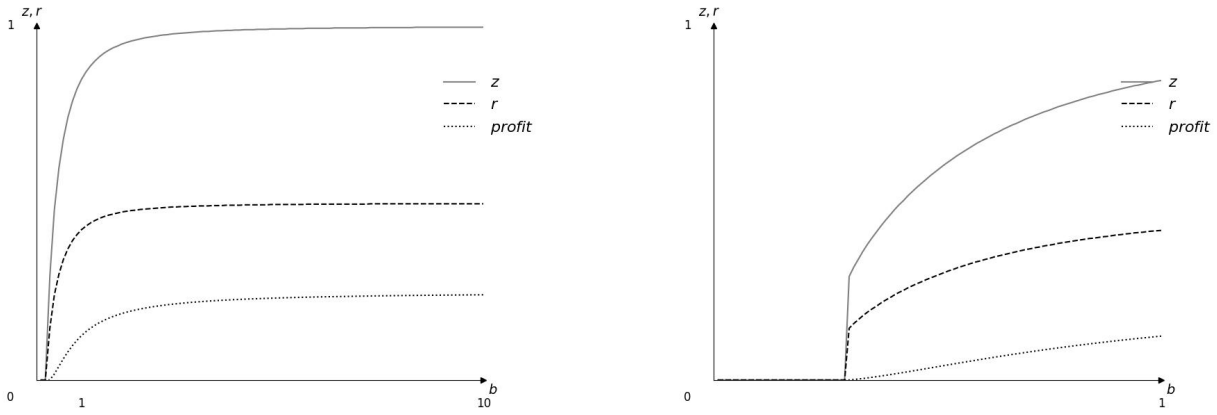
(a): Left-Skewed to Center-Skewed

(b): Center-Skewed to Right-Skewed

Notes: We alter the shape of $g(z)$ by changing the α parameter of a Beta distribution while keeping $\beta = 5$ (left), and changing the β parameter while keeping $\alpha = 5$ (right).

Higher data quality b serves as a structural efficiency gain. As b increases, the producer can achieve higher capability levels at lower marginal costs, leading to a simultaneous increase in equilibrium capability z^* and a more aggressive pricing strategy r^*.

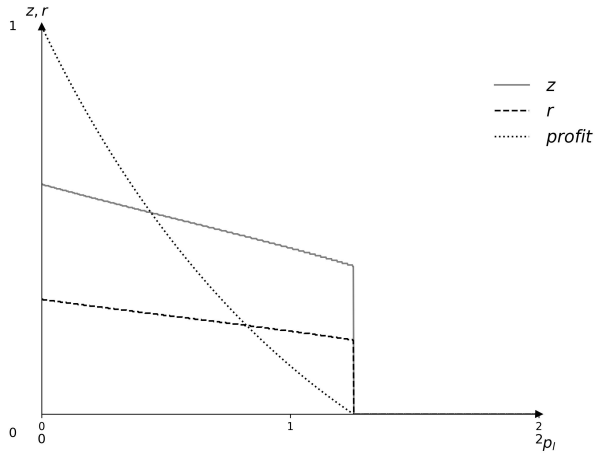
Figure 5: Effect of Data Quality b on Equilibrium z, r



Notes: Results generated using a uniform $g(z)$ and market size $s = 100$.

The price of pretraining data p_0 acts as a tax on capability. Prohibitively high data prices not only reduce the optimal capability level but can lead to market collapse. As shown in Figure 6, once p_0 exceeds a critical threshold, the producer's margin is squeezed to the point where the AI model is no longer viable.

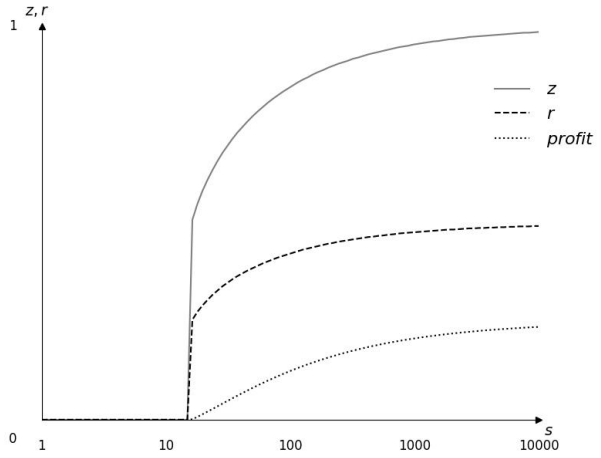
Figure 6: Effect of Data Price p_0 on Equilibrium z, r



Notes: Uniform $g(z)$, $b = 1$, $s = 100$. Higher p_0 increases marginal cost proportionally, reducing z^*, r^*, and profit. Beyond $p_0 \approx 1.25$, the market shuts down.

Finally, we observe a direct scale-capability coupling. In larger markets (higher s), the fixed costs of training are amortized over a larger user base, incentivizing the producer to invest in higher capability z^*. This suggests that AI development is naturally biased toward larger economies or broader linguistic markets.

Figure 7: Effect of Market Size s on Equilibrium z, r



Notes: Results generated using a uniform $g(z)$ and $b = 1$.

5.2.2 Optimal Data Procurement and the Sellability Constraint

We now extend the producer's problem by endogenizing the choice of data quality. Given that a model's architecture and scaling efficiency b are typically fixed prior to pretraining, we assume the producer utilizes a *single* data source throughout the process (i.e., no mid-stream switching). The procurement problem therefore compares *total* expenditures required to reach a target capability.

For a fixed data source (p, b), total expenditure to reach capability z^* is

$$\mathcal{C}(z^*, b, p) = (1+p)\int_0^{z^*} \frac{1}{b}\left(\frac{1}{1-z}\right)^{\frac{1}{b}+1} dz = (1+p)\Big[(1-z^*)^{-1/b} - 1\Big]. \tag{31}$$

The marginal expenditure of capability (the derivative of (31)) is

$$MC_z(z; b, p) \equiv \frac{\partial \mathcal{C}(z, b, p)}{\partial z} = \frac{1+p}{b}\left(\frac{1}{1-z}\right)^{\frac{1}{b}+1}. \tag{32}$$

Let (p_0, b_0) denote the exogenous benchmark of "waste data." A dataset (p_b, b) is *sellable* for target z^* if it weakly reduces total expenditure relative to the baseline:

$$\mathcal{C}(z^*, b, p_b) \leq \mathcal{C}(z^*, b_0, p_0). \tag{33}$$

Equivalently, the seller's unit price must satisfy the total-cost ceiling

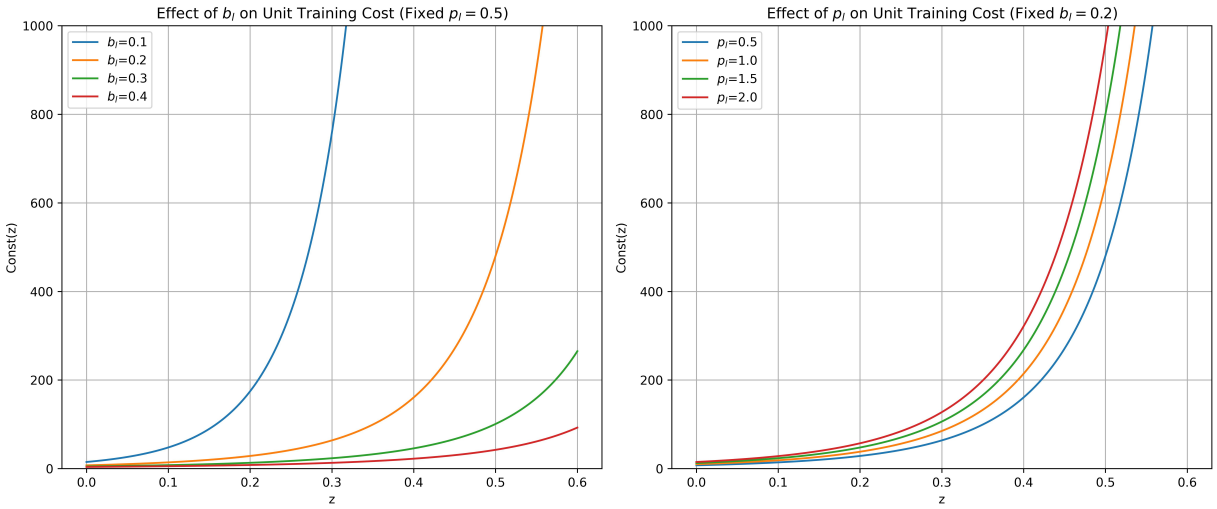
$$p_b \leq p_b^{\max}(b, z^*) \equiv (1+p_0)\frac{(1-z^*)^{-1/b_0} - 1}{(1-z^*)^{-1/b} - 1} - 1. \tag{34}$$

This ceiling coincides with Lemma 13 when the relevant outside option is the baseline dataset.

To build intuition, we define the baseline marginal training cost curve as

$$MC_{base}(z) \equiv MC_z(z; b_0, p_0) = \frac{1+p_0}{b_0}\left(\frac{1}{1-z}\right)^{\frac{1}{b_0}+1}. \tag{35}$$

Figure 8: Baseline Marginal Training Cost Frontier: $MC_{base}(z)$



Notes: The baseline marginal cost curve illustrates how training costs explode near the frontier. Under our no-switching assumption, procurement is governed by total expenditure (31)*, but the steepness of* $MC_{base}(z)$ *explains why total-cost ratios — and hence price ceilings — diverge as* $z \to 1$*.*

Figures 9 and 10 illustrate the iso-surface where the total-cost ceiling (34) binds (treating z as the prospective target capability). The economic implications are twofold. First, at low capability levels ($z \approx 0$), the value of high-quality data is limited because the baseline is still relatively efficient. A first-order expansion yields $(1-z)^{-1/b} - 1 \approx z/b$, so the price ceiling is approximately linear in relative quality.

Second, as the producer targets the frontier ($z \to 1$), the baseline becomes prohibitively inefficient: its total cost scales like $(1-z)^{-1/b_0}$, while its marginal cost explodes at the higher order $(1-z)^{-(1/b_0+1)}$ (from (32)). This "inefficiency explosion" causes the ceiling (34) to rise sharply, allowing high-quality datasets to command large premiums precisely when the buyer trains frontier models.

Figure 9: Sellability Iso-Surface: Quality-Price Relationship over z

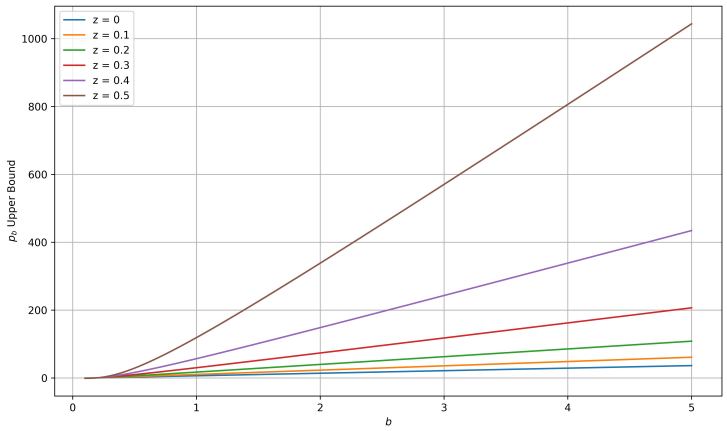
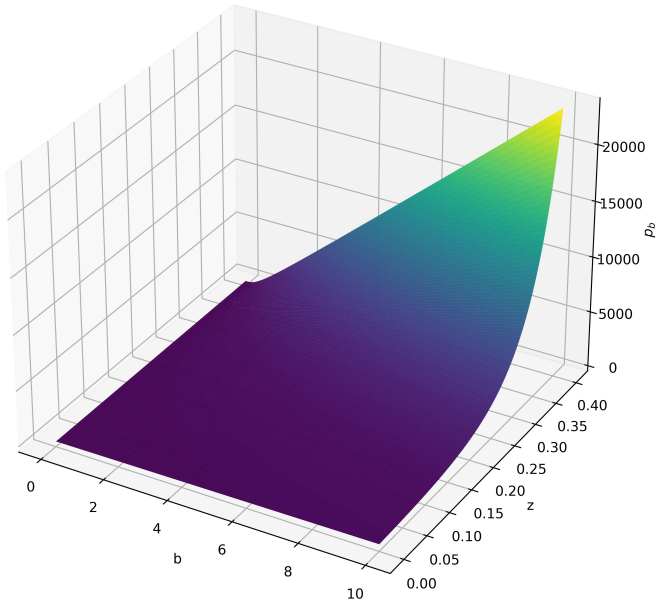


Figure 10: 3D Visualization of the Sellability Frontier



Notes: The surface represents the maximum sellable unit price $p_b^{\max}(b, z)$ implied by (34). Note the sharp increase in the ceiling as z approaches the frontier.

5.2.3 Connecting to the Data Market: Endogenizing $V(b)$

We now close the model by linking the structural production technology dictated by the scaling laws to the data market analysis developed in Section 4. Previously, the buyer's valuation for data

quality b, denoted $V(b)$, was treated as exogenous. We now derive this value endogenously as the *compute-expenditure savings* enabled by higher-quality data.

Consider a model producer targeting a capability z^*. Using baseline "waste data" (p_0, b_0), the total expenditure is $\mathcal{C}(z^*, b_0, p_0) = (1+p_0)[(1-z^*)^{-1/b_0} - 1]$. If the producer instead utilizes data quality $b > b_0$, the *gross value* of this data (the maximum price the producer would pay before reverting to the baseline) is the difference between the baseline cost and the compute capital required under quality b:

$$V(b \mid z^*) = (1+p_0)\left[(1-z^*)^{-1/b_0} - 1\right] - \left[(1-z^*)^{-1/b} - 1\right]. \tag{36}$$

The structural form of the scaling law implies several key properties[11] for the producer's willingness to pay (WTP): (i) *Monotonicity and Complementarity:* The value is strictly increasing in quality ($\frac{\partial V}{\partial b} > 0$) and target capability ($\frac{\partial V}{\partial z^*} > 0$). Moreover, data quality and target capability are strategic complements: $\frac{\partial^2 V}{\partial b \partial z^*} > 0$. (ii) *Accelerated Valuation:* V is convex in the capability target ($\frac{\partial^2 V}{\partial (z^*)^2} > 0$). As $z^* \to 1$, the WTP diverges to infinity, suggesting that high-quality data becomes a non-substitutable bottleneck for frontier models.

The synthesis of the upstream data market and the downstream model market establishes a general equilibrium. By endogenizing the valuation parameters such that $V_L \equiv V(b_L \mid z^*)$ and $V_H \equiv V(b_H \mid z^*)$, we characterize the general equilibrium as a fixed-point problem. Specifically, the optimal capability z^* is determined by the data quality b and price p_b emerging from the information-asymmetry game, while these market outcomes are themselves functions of the valuations $V(b \mid z^*)$ induced by the targeted capability. This coupling underscores how the structural scaling laws of AI training define the information rents and market structures governing the trade of pretraining data.

5.2.4 Fixed-Point Formulation and Existence

Building on this endogenization, we can formalize the general equilibrium as a fixed-point problem. In Online Appendix B, we provide a formal proof of existence for this two-layer equilibrium. Applying Kakutani's Fixed-Point Theorem, we establish that a consistent downstream capability target exists given an upstream equilibrium pricing rule. We explicitly detail the bound-

[11] See Appendix A.5 for proof.

ary conditions required to ensure viability, demonstrating that while standard quasi-concavity is structurally incompatible with the profit function's behavior near the origin, the downstream objective remains robustly single-peaked under standard regularity assumptions.

5.3 Data Market Equilibrium with Multiple Sellers

We now integrate the strategic data market structure with the scaling-law production technology. We consider a market with I data sellers, each holding a dataset D_i, and a single model producer who acts as an informed buyer.

5.3.1 Model Setup and Information Structure

Each seller $i \in \{1, \ldots, I\}$ possesses a dataset characterized by a true quality $b_i \in [\underline{b}, \bar{b}]$ and a unit production cost c_i. Higher-quality data requires more intensive curation, creating a positive correlation between cost and quality:

$$c_i = c(b_i) + \varepsilon_i, \quad c'(b) > 0 \tag{37}$$

where ε_i is a mean-zero noise term. While sellers do not observe b_i directly, they use their private cost c_i to form a Bayesian posterior belief $\mu_i(b|c_i)$. We assume the expected quality $\mathbb{E}[b_i|c_i]$ is strictly increasing in c_i, such that higher costs serve as an informative signal of higher quality.

The model producer (buyer) possesses a significant informational advantage: through preliminary scaling experiments, the producer observes the true quality b_i for all datasets. The producer also observes the posted unit prices p_i and has access to an exogenous baseline of "waste data" (p_0, b_0). The strategic interaction follows a five-stage sequence:

1. Nature draws qualities and costs (b_i, c_i) for all I sellers.
2. Sellers observe private costs c_i and update beliefs $\mu_i(b|c_i)$.
3. Sellers simultaneously post unit prices $p_i(c_i) \geq c_i$.
4. The buyer observes true qualities and prices (b_i, p_i), then selects data and sets target capability z^* to maximize profit.
5. Payoffs are realized based on the final training expenditure and downstream model revenue.

This setup reveals why a competitive data market may not always lead to higher quality. If the noise term ε_i is substantial, a high-cost/low-quality seller might post the same price as a high-cost/high-quality seller. The buyer, observing the true quality, will only purchase from the latter. This mechanism forces sellers into "belief-based" competition, wherein prices reflect not just the intrinsic value of the data, but also the seller's uncertainty about that value.

5.3.2 The Strategic Procurement Problem

Given the menu of prices and qualities, the buyer selects a data source to minimize the total expenditure required to reach the target capability z^*. As established in Section 5.2.2, we assume the producer utilizes a single data source for the duration of the pretraining process, as mixing sources with different scaling efficiencies would necessitate a total re-optimization of the model architecture.

Lemma 12 (The Efficient Procurement Rule)**.** *Given a target capability z^*, the producer selects the data source i^* from the set of available sellers and the baseline $\{0, 1, \ldots, I\}$ to minimize total training expenditure:*

$$i^* = \arg \min_{i \in \{0,1,\ldots,I\}} \mathcal{C}(z^*, b_i, p_i) \tag{38}$$

where $\mathcal{C}(z^*, b, p) = (1 + p)[(1 - z^*)^{-1/b} - 1]$.

Proof. See Appendix A.3.8.

The producer purchases from seller i over the baseline if and only if $(1 + p_i)C(z^*, b_i) < (1 + p_0)C(z^*, b_0)$. This condition highlights that the "sellability" of a dataset is a relative metric: a dataset's value is defined by the computational savings it offers relative to the best available alternative. As the producer targets higher capability frontiers, the baseline cost $C(z^*, b_0)$ increases exponentially, expanding the price-quality space in which third-party sellers can profitably operate.

5.3.3 Value Extraction and the Strategic Price Ceiling

The producer's procurement decision is governed by the surplus generated relative to the best available alternative. We define the producer's net surplus from dataset i at price p_i as:

$$S_i(p_i, z^*) = V(b_i \mid z^*) - p_i \cdot C(z^*, b_i) \tag{39}$$

where $V(b_i \mid z^*)$ represents the gross cost-savings as derived in Section 5.2.3.

Lemma 13 (Maximum Willingness to Pay)**.** *Given a target capability z^*, the maximum unit price the producer will pay seller i relative to the baseline (b_0, p_0) is:*

$$p_i^{\max}(z^*) = (1 + p_0)\frac{(1 - z^*)^{-1/b_0} - 1}{(1 - z^*)^{-1/b_i} - 1} - 1. \tag{40}$$

When competing against a rival seller j, the constraint for seller i tightens to:

$$p_i^{\max}(z^*; p_j, b_j) = (1 + p_j)\frac{(1 - z^*)^{-1/b_j} - 1}{(1 - z^*)^{-1/b_i} - 1} - 1. \tag{41}$$

Proof. See Appendix A.4.

The functional form of the price ceiling reveals that the nature of market competition is endogenously driven by the producer's technological ambition, characterized by two distinct regimes. In the *commodity regime*, where capability targets are low ($z^* \to 0$), the scaling law simplifies via a first-order Taylor expansion, $(1 - z^*)^{-1/b} - 1 \approx z^*/b$. Consequently, the price ceiling converges to a constant ratio, $p_i^{\max} \to (1 + p_0)\frac{b_i}{b_0} - 1$, where quality and price are linearly substitutable and competition is driven by unit costs.

Conversely, as the producer targets the *frontier premium regime* ($z^* \to 1$), the market undergoes a phase transition. Since $b_i > b_0$ implies that the baseline cost grows at a strictly higher order of magnitude than the cost of curated data, the willingness to pay diverges ($\lim_{z^* \to 1} p_i^{\max}(z^*) = +\infty$). In this regime, baseline data becomes prohibitively inefficient, granting high-quality sellers significant structural market power and the ability to extract exponential premiums.

5.3.4 The Seller's Signal-Extraction and Pricing Problem

We now analyze the seller's optimization problem. In contrast to the buyer, who possesses perfect information regarding data quality b_i via pilot testing, seller i operates under information asymmetry. The seller observes only their private marginal cost, c_i, which serves as a noisy signal of the underlying data quality.

Let $\mu(b \mid c_i)$ denote the seller's posterior belief regarding their own data quality b_i, derived via Bayesian updating from the joint distribution of costs and qualities. The seller sets a unit price p_i to maximize expected profit.

Crucially, the seller's probability of being selected depends on the *true* quality b_i, because the buyer observes b_i directly. Therefore, the seller must integrate over all possible realizations of b_i, weighted by their posterior belief $\mu(b \mid c_i)$.

Formally, seller i's problem is:

$$\max_{p_i \geq c_i} \quad \int_{\underline{b}}^{\bar{b}} (p_i - c_i) \cdot \mathcal{C}(z^*(b, p_i), b) \cdot \mathbf{1}_{\{i=i^*(b,p_i,\mathbf{p}_{-i})\}} \, d\mu(b \mid c_i). \tag{42}$$

Here, $\mathcal{C}(z^*, b) = (1 - z^*)^{-1/b} - 1$ is the compute demand conditional on quality b. $\mathbf{1}_{\{i=i^*\}}$ is the indicator function equal to 1 if the specific combination of (p_i, b) is preferred by the buyer over all rivals and the baseline. $d\mu(b \mid c_i)$ is the seller's belief density, linking the observed cost to the unobserved quality.

This formulation underscores the adverse selection risk: the seller wins the auction primarily in states where b is high enough to justify the price p_i. If the seller sets p_i based on an optimistic belief (high $\mu(b \mid c_i)$) but the true b is low, the indicator function becomes 0, and the realized profit is zero.

5.3.5 Equilibrium Analysis

We characterize the Perfect Bayesian Equilibrium (PBE) in symmetric strategies, where each seller i adopts a pricing schedule $p^*(c_i)$. In this environment, the seller must resolve a signal-extraction problem to estimate the buyer's willingness to pay while accounting for the selection bias inherent in the buyer's procurement rule.

To isolate the impact of informational asymmetry, we first consider a single data seller facing a buyer who possesses the baseline quality b_0 as an outside option.

Proposition 5 (Monopoly Pricing with Noisy Signals)**.** *In a single-seller interaction, the equilibrium pricing function $p^*(c)$ satisfies:*

$$p^*(c) = \arg\max_{p \geq c} \int_{\underline{b}}^{\bar{b}} (p - c) \cdot \mathcal{C}(z^*, b) \cdot \mathbf{1}_{\{p \leq p^{\max}(z^*; b)\}} \, d\mu(b \mid c). \tag{43}$$

The equilibrium exhibits the following properties:

(i) ***Monotone Selection****: If the seller's expected-profit objective satisfies the single-crossing condition in (p, c), then there exists an equilibrium pricing schedule $p^*(c)$ that is (weakly)*

increasing in the cost signal c.

(ii) ***Price Smoothing****: Let $X \equiv p^{\max}(z^*; B)$ denote the buyer's true valuation of the dataset (as a function of true quality B), and let $\tilde{p}(c) \equiv \mathbb{E}[X \mid c]$ denote the seller's belief-based valuation. If c is a noisy signal of B, then $\mathrm{Var}(\tilde{p}) < \mathrm{Var}(X) = \mathrm{Var}(p^{\max})$. Since any feasible pricing rule measurable in c (including $p^*(c)$) cannot condition on residual variation in X beyond c, equilibrium prices are necessarily belief-based and cannot fully track realized quality. Moreover, if $p^*(c)$ can be written as $p^*(c) = \psi(\tilde{p}(c))$ for some L-Lipschitz map ψ with $L \leq 1$, then $\mathrm{Var}(p^*) \leq L^2 \mathrm{Var}(\tilde{p}) < \mathrm{Var}(p^{\max})$, so posted prices inherit a quantitative smoothing bound.*

Proof. See Appendix A.3.10.

Remark (Role of regularity conditions). Part (i) is a monotone-comparative-statics result: the single-crossing condition ensures that higher-cost sellers optimally set higher prices. Economically, it implies that the marginal benefit of raising one's price is increasing in the cost signal—a natural property when higher costs indicate higher expected quality and hence a wider range of buyer valuations over which the seller can profitably trade. The condition is standard in mechanism design (Myerson, 1981) and holds whenever the posterior $\mu(b \mid c)$ satisfies the monotone likelihood ratio property (MLRP) in (b, c).

Part (ii) decomposes the price-smoothing result into a qualitative component and a quantitative bound. The qualitative statement that $\mathrm{Var}(\tilde{p}) < \mathrm{Var}(p^{\max})$ follows from the law of total variance and requires only that c is a noisy (non-sufficient) signal of b. The quantitative bound adds a Lipschitz restriction on the map $\tilde{p} \mapsto p^*$; this is a regularity condition on the shape of the seller's profit function, not a primitive of the model. We include it to give a concrete bound on the degree of smoothing, but the economic message that equilibrium prices systematically under-react to true quality holds without it.

The wedge between $p^{\max}(z^*; b)$ and the posted price $p^*(c)$ represents the buyer's information rent. Because prices are conditioned on the noisy signal c rather than on realized quality b, the buyer captures surplus whenever realized quality is high relative to the seller's belief.

With multiple sellers, the buyer does not minimize price, but rather the efficiency-adjusted cost.

Lemma 14 (Efficiency-Adjusted Competition)**.** *Seller i wins the market if and only if they minimize the total training cost $(1+p_i)\mathcal{C}(z^*, b_i)$. Consequently, seller i beats seller j if:*

$$\frac{1+p_i}{1+p_j} < \frac{\mathcal{C}(z^*, b_j)}{\mathcal{C}(z^*, b_i)}. \tag{44}$$

Proof. See Appendix A.3.9.

Proposition 6 (Competitive Equilibrium)**.** *In the symmetric competitive equilibrium:*

(i) ***Strategic Shading****: Sellers shade prices $p^*(c)$ below the monopoly level to mitigate the probability of exclusion.*

(ii) ***Selection-Adjusted Bound****: The winning price $p^*_{i^*}$ is capped by the efficiency of the runner-up:*

$$p^*_{i^*} \leq \min\left\{p^{max}(z^*; b_{i^*}), \min_{j \neq i^*}\left[\frac{(1+p^*_j)C(z^*, b_j)}{C(z^*, b_{i^*})} - 1\right]\right\}. \tag{45}$$

(iii) ***Limit Pricing****: Assuming standard regularity conditions (the induced distribution of equilibrium total costs has a continuous density with a hazard rate bounded away from zero and $\mathcal{C}(z^*, b)$ is bounded away from zero), equilibrium markups vanish as $I \to \infty$, so $p^*(c) - c \to 0$ (pointwise in c). Nevertheless, the buyer's information rent does not generally vanish: even with $p = c$, the buyer can capture surplus from the wedge between the realized quality and the sellers' cost-based pricing.*

Proof. See Appendix A.3.11.

Remark (Economic content vs. regularity conditions in Proposition 6)*.* Parts (i) and (ii) are structural consequences of the buyer's cost-minimization rule (Lemma 14) and require no regularity beyond the model's primitives: the buyer selects the cheapest efficiency-adjusted data source, and this selection disciplines sellers' prices.

Part (iii) adds two regularity conditions (a continuous cost density and a hazard rate bounded away from zero) to obtain a quantitative markup bound that vanishes at rate $O(1/I)$. These conditions are standard in first-price auction theory (see, e.g., the regularity assumptions in the revenue equivalence literature). Economically, a bounded hazard rate ensures that the runner-up's cost is never too far from the winner's, preventing any seller from exercising local monopoly power.

If these conditions are relaxed, markups still decrease in I (by the shading logic of part (i)), but the pointwise convergence to zero may fail at atoms or boundary points of the cost distribution.

The final claim that buyer information rents persist even as markups vanish is a structural result that holds without any regularity condition: it follows from the wedge between the buyer's observed quality b_i and the seller's cost-based price c_i, which is a primitive feature of the reversed information asymmetry.

5.3.6 Information Rent and Market Efficiency

A central implication of asymmetric information in the pretraining data market is that the buyer can extract surplus by exploiting discrepancies between true data quality and sellers' beliefs. We define the buyer's expected information rent as the reduction in total training cost relative to the outside option:

$$\Pi_B = \mathbb{E}\left[\mathcal{C}(z^*, b_0, p_0) - \mathcal{C}(z^*, b_{i^*}, p^*_{i^*})\right],$$

where b_0 and p_0 denote the quality and price associated with the baseline alternative, and $(b_{i^*}, p^*_{i^*})$ correspond to the dataset selected in equilibrium.

Using Lemma 13, define $p_i^{\max}(z^*)$ as the baseline-relative price ceiling for seller i, i.e., $(1 + p_i^{\max}(z^*))C(z^*, b_i) = (1+p_0)C(z^*, b_0)$. Then the buyer's baseline-relative surplus from purchasing i at price p_i can be written as

$$\mathcal{C}(z^*, b_0, p_0) - \mathcal{C}(z^*, b_i, p_i) = (p_i^{\max}(z^*) - p_i)\, C(z^*, b_i),$$

so in equilibrium

$$\Pi_B = \mathbb{E}\big[(p_{i^*}^{\max}(z^*) - p^*_{i^*})\, C(z^*, b_{i^*})\big],$$

with the convention that the term is zero if the baseline option is chosen.

Information rent arises because sellers cannot perfectly condition prices on realized quality. Equilibrium prices therefore reflect expected rather than true quality, allowing the buyer to benefit from favorable deviations between realized and perceived data value. We analyze the buyer's expected information rent in the competitive limit established by Proposition 6(iii), where markups vanish ($p_i^* \to c_i$) and information rents are driven purely by the wedge between true quality and

belief-based pricing. In this limit, the buyer's rent takes the transparent form

$$\Pi_B = \mathbb{E}\left[\left(\max_{1\le i\le I} R_i\right)^+\right], \qquad R_i \equiv \mathcal{C}(z^*, b_0, p_0) - (1+c_i)\, C(z^*, b_i), \tag{46}$$

where $c_i = c(b_i) + \sigma\varepsilon_i$ is the seller's cost, $C(z,b) = (1-z)^{-1/b} - 1$, and $(\cdot)^+ \equiv \max(\cdot, 0)$. Throughout, qualities (b_i) are i.i.d. from F and noise terms (ε_i) are i.i.d. with $\mathbb{E}[\varepsilon] = 0$ and finite variance, independent of the quality draws.

Proposition 7 (Information Rent)**.** *The buyer's expected information rent satisfies the following properties.*

(i) ***Selection Effect.*** Π_B *is weakly increasing in the number of sellers* I*.*

(ii) ***Capability Amplification.*** Π_B *is increasing in the target capability level* z^**.*

(iii) ***Signal Noise Effect.*** *Let* $\nu_1, \ldots, \nu_I$ *be i.i.d. mean-zero random variables, independent of all other primitives. Define the noisier cost signal* $\tilde{c}_i \equiv c_i + \nu_i$ *and the associated rent* $\tilde{R}_i \equiv \mathcal{C}(z^*, b_0, p_0) - (1+\tilde{c}_i)\, C(z^*, b_i)$*. Then*

$$\mathbb{E}\left[\left(\max_i \tilde{R}_i\right)^+\right] \;\ge\; \mathbb{E}\left[\left(\max_i R_i\right)^+\right].$$

That is, degrading the informativeness of sellers' cost signals weakly increases the buyer's information rent.

(iv) ***Quality Dispersion Effect.*** *Adopt the binary quality framework of Section 4, where* $b_i \in \{b_L, b_H\}$ *with common prior* $\mu_0 = \Pr(b_i = b_H)$ *and endogenized valuations* $V_H \equiv V(b_H \mid z^*)$*,* $V_L \equiv V(b_L \mid z^*)$ *from Section 5.2.3.*

(a) *The valuation gap* $\Delta V \equiv V_H - V_L$ *is strictly increasing in any mean-preserving spread of the quality distribution: for any increase in* b_H *and corresponding decrease in* b_L *that preserves* $\mu_0 b_H + (1-\mu_0) b_L$*,* ΔV *strictly increases.*

(b) *The buyer's expected information rent—in both the single-seller setting of Proposition 3 and the multiple-seller setting of Proposition 4—is strictly increasing and linear in* ΔV*.*

Combined, the buyer's rent is strictly increasing in the dispersion of data quality.

Proof. Parts (i)–(ii): see Appendix A.3.12. Part (iii): see Appendix A.3.13. Part (iv): see Appendix A.3.14.

Part (iii) formalizes the intuition that the buyer's informational advantage grows when sellers operate in a foggier information environment. Since $\max(x_1, \ldots, x_I, 0)$ is a convex function, the buyer's rent is a convex functional of the joint distribution of seller rents. Adding independent mean-zero noise to the cost signals creates a mean-preserving spread of each R_i without altering the buyer's observation of true quality, mechanically widening the right tail of the rent distribution from which the buyer selects.

Part (iv) establishes the quality dispersion effect through two steps. First, the compute-cost function $C(z^*, b)$ is strictly decreasing in quality b: better data requires less compute to reach a given capability. Consequently, when we spread the quality distribution (raising b_H, lowering b_L), both margins work in the same direction so that the high-quality option becomes even cheaper to train on, while the low-quality option becomes even more expensive, ensuring the valuation gap $\Delta V = C(b_L) - C(b_H)$ unambiguously widens. Second, in the binary framework of Section 4, the buyer's information rent enters as a linear scaling of ΔV, so widening the quality spread directly magnifies the buyer's ability to extract surplus from mispriced high-quality datasets. Unlike the signal noise effect, which operates through the variance of pricing errors, the quality dispersion effect operates through the level of the mispricing wedge: a wider quality gap means that each "hidden gem" $(b_H, s = L)$ event yields a proportionally larger surplus.

We evaluate allocative efficiency by comparing equilibrium outcomes to the socially optimal allocation. The total surplus generated from the data trade is given by

$$W = TR(z^*) - \mathcal{C}(z^*, b_{i^*}, c_{i^*}),$$

where $TR(z^*)$ denotes revenue generated in the knowledge economy layer and $\mathcal{C}(z^*, b_{i^*}, c_{i^*})$ represents the true resource cost associated with the selected dataset.

In the first-best benchmark, the socially optimal allocation minimizes the efficiency-adjusted training cost $(1 + c_i)C(z^*, b_i)$. By contrast, market selection is governed by prices rather than costs, creating a wedge between private and social incentives.

Proposition 8 (Market Efficiency)**.** *Equilibrium outcomes exhibit the following efficiency distortions:*

(i) ***Selection Distortion****: The market allocates data by minimizing* $(1 + p_i)C(z^*, b_i)$ *rather than the socially optimal criterion* $(1 + c_i)C(z^*, b_i)$.

(ii) ***Exclusion of Efficient Sellers****: High-quality, high-cost sellers may be inefficiently excluded when prices reflect expected rather than realized quality.*

(iii) ***Capability Underinvestment****: If* TR *is concave in* z*, then whenever the winning price satisfies* $p_{i^*} > c_{i^*}$*, the equilibrium capability level* z^* *is inefficiently low.*

Proposition 8 highlights that inefficiency in the pretraining data market does not stem from a failure of trade, but from a misalignment between private and social decision criteria. The buyer selects data sources to minimize expenditure, which depends on posted prices, whereas social efficiency depends on underlying production costs. Because prices reflect sellers' beliefs about data quality rather than realized quality, market selection is governed by expected rather than true efficiency. This wedge can lead the buyer to exclude datasets that are socially efficient but appear expensive due to pessimistic signals, while favoring datasets that are privately cheap but less productive in training. Consequently, the buyer stops increasing z prematurely. Importantly, these distortions persist even under intense competition, as competitive pressure disciplines price levels but does not eliminate informational misranking across sellers. This result underscores that standard competitive forces may be insufficient to restore efficiency when prices embed noisy beliefs rather than costs. The analysis highlights a strategic tension: procurement decisions that minimize short-run training expenditures may inadvertently sacrifice long-run model capability, implying a potential role for internal benchmarking, data auditing, or longer-term contracting mechanisms that better align prices with realized data performance.

5.3.7 Robustness: General Augmentation Function

The monopolistic market structure established in Section 3 relies on the stark productivity-floor technology, $\beta(z, q) = \max\{z, q\}$. A natural theoretical concern is whether the extreme winner-take-all dynamic is merely an artifact of this specific functional form. To address this, we relax this assumption and introduce a general capability augmentation function, $\beta(z, q)$.

Our analysis demonstrates that the intrinsic drive toward firm-level concentration is robust. Specifically, the capability leader can always implement an equilibrium entry-deterrence strategy

under both submodular and supermodular augmentation technologies. However, the exact implementation of this defense—and consequently, the optimal product-line architecture—depends crucially on the cross-partial derivative β_{12}.

Under submodular augmentation ($\beta_{12} \leq 0$), the "portfolio-collapse" property holds: the leader can leverage single-model limit pricing to exclude an inferior rival. Conversely, under supermodular augmentation ($\beta_{12} > 0$), workers exhibit systematic differences in their marginal valuations of capability, rendering the single-model reduction universally sub-optimal. To sustain the winner-take-all outcome, the incumbent must instead deploy a "model family"—segmenting the market by pairing a high-end flagship model with a heavily discounted fighting brand to exclude rivals.

Ultimately, while product-line proliferation may occur under supermodularity, firm-level market concentration remains the robust equilibrium outcome. We refer the reader to Online Appendix C for a formal characterization, including the necessary regularity conditions and proofs of these generalizations.

6 Conclusion

This paper develops a unified economic framework for the foundational model industry, linking the downstream market for AI capabilities with the upstream market for pretraining data. By embedding the empirical regularities of neural scaling laws into a two-layer market model, we uncover the structural forces that will shape the AI economy.

Our analysis yields three central insights. First, in the downstream layer, the nature of AI as a "productivity floor" creates intense winner-take-all dynamics. Because lower-skilled workers view competing models as perfect substitutes, capability leaders can leverage quality-price coupling to limit-price competitors out of the market. Under general augmentation, what remains robust is the winner-take-all market structure and the existence of an equilibrium entry-deterrence strategy. This suggests that the current proliferation of model variants may be a transient phenomenon, eventually collapsing into a structure dominated by a few frontier models, provided that "leap-ahead" capability gains remain feasible.

Second, the upstream data market is defined by a reversal of standard information asymmetry. Unlike traditional lemons markets, here the buyer (model producer) possesses superior information regarding the asset's quality due to their exclusive access to validation compute. We show that this

structure allows model producers to extract significant information rents, systematically "cherry-picking" high-quality data from sellers who underestimate the value of their own assets. This dynamic creates a "sellable data constraint", where high-quality data commands a premium only when the downstream capability target is sufficiently high to render baseline data computationally inefficient.

Finally, the neural scaling law acts as the fundamental transmission mechanism between these two layers. It dictates that as model producers target the capability frontier ($z \to 1$), their willingness to pay for data quality diverges to infinity. This convexity explains the bifurcation observed in the industry: a commoditized market for low-quality web scrapes, coexisting with an increasingly exclusive, high-value market for bespoke, expert-curated data.

For policymakers and managers, our results highlight a critical market friction. The model producer's ability to exploit information asymmetry upstream may stifle the long-term supply of high-quality data by denying data creators their fair share of the surplus. To sustain the data-driven scaling of AI, market design interventions—such as third-party quality certification or public benchmarking infrastructure—are necessary to align incentives and ensure that the scarcity of high-quality human knowledge is accurately reflected in its price.

References

Acemoglu, D. (2025). The simple macroeconomics of ai. *Economic Policy 40*(121), 13–58.

Acemoglu, D. and S. Johnson (2023). *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. PublicAffairs.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Alekseeva, L., J. Azar, M. Giné, S. Samila, and B. Taska (2024). The demand for ai skills in the labor market. *Labour Economics 90*, 102570.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science 60*(12), 2859–2885.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative ai at work. *The Quarterly Journal of Economics 140*(2), 889–942.

Carmona, J. and K. Laohakunakorn (2025). Rosen meets garicano and rossi-hansberg: Stable matchings in knowledge economies. *Working Paper*.

Chen, Y. J., J. Gong, J. Li, and Z. Zhao (2025, October). Better technology, worse motivation: Genai's mediocrity trap. SSRN Working Paper 5208163, SSRN.

Dell'Acqua, F., E. McFowland, E. Mollick, H. Lifshitz, K. C. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science 37*(2), 403–423.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). Gpts are gpts: Labor market impact potential of llms. *Science 384*(6702), 1306–1308.

Fahn, M., J. Li, and C. Sun (2025). Toward a bad job economy: Ai adoption, agency costs, and labor market spillovers. *Working Paper*.

Fuchs, W., L. Garicano, and L. Rayo (2015). Optimal contracting and the organization of knowledge. *The Review of Economic Studies 82*(2), 632–658.

Gabszewicz, J. J. and J.-F. Thisse (1979). Price competition, quality and income disparities. *Journal of Economic Theory 20*(3), 344–359.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*(5), 874–904.

Garicano, L. and T. N. Hubbard (2016). The returns to knowledge hierarchies. *Journal of Law, Economics, and Organization 32*(4), 653–684.

Garicano, L. and E. Rossi-Hansberg (2004). Inequality and the organization of knowledge. *American Economic Review 94*(2), 197–202.

Garicano, L. and E. Rossi-Hansberg (2006). Organization and inequality in a knowledge economy. *The Quarterly Journal of Economics 121*(4), 1383–1435.

Garicano, L. and E. Rossi-Hansberg (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics 7*, 1–30.

Gumpert, A., H. Steimer, and M. Antoni (2022). Firm organization with multiple establishments. *The Quarterly Journal of Economics 137*(2), 1091–1138.

Gunasekar, S., Y. Zhang, J. Aneja, C. C. T. Mendes, A. D. Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi, A. Salim, S. Shah, H. S. Behl, X. Wang, S. Bubeck, R. Eldan, A. T. Kalai, Y. T. Lee, and Y. Li (2023). Textbooks are all you need.

Hoffmann, J., S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. de Las Casas, L. A. Hendricks, J. Welbl, A. Clark, T. Hennigan, E. Noland, K. Millican, G. van den Driessche, B. Damoc, A. Guy, S. Osindero, K. Simonyan, E. Elsen, O. Vinyals, J. W. Rae, and L. Sifre (2022). Training compute-optimal large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA. Curran Associates Inc.

Ide, E. and E. Talamàs (2025). Artificial intelligence in the knowledge economy. *Journal of Political Economy 133*(12), 3762–3800.

Kaplan, J., S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Li, Y., S. Bubeck, R. Eldan, A. D. Giorno, S. Gunasekar, and Y. T. Lee (2023). Textbooks are all you need ii: phi-1.5 technical report.

Moorthy, K. S. (1988). Product and price competition in a duopoly. *Marketing Science 7*(2), 141–168.

Moscarini, G. and M. Ottaviani (2001). Price competition for an informed buyer. *Journal of Economic Theory 101*(2), 457–493.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory 18*(2), 301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research 6*(1), 58–73.

Myerson, R. B. (1983). Mechanism design by an informed principal. *Econometrica 51*(6), 1767–1797.

Noy, S. and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science 381*(6654), 187–192.

Sardana, N., E. Portelance, D. Meunier, W. Liao, M. Yadav, and M. Hutter (2024). Beyond chinchilla-optimal: Accounting for inference in language model scaling laws. *arXiv preprint arXiv:2401.00448*.

Schaeffer, R., J. Kazdan, J. Hughes, J. Juravsky, S. Price, A. Lynch, E. Jones, R. Kirk, A. Mirhoseini, and S. Koyejo (2025). How do large language monkeys get their power (laws)? In *Forty-second International Conference on Machine Learning*.

Shaked, A. and J. Sutton (1982). Relaxing price competition through product differentiation. *The*

Review of Economic Studies 49(1), 3–13.

Shaked, A. and J. Sutton (1983). Natural oligopolies. *Econometrica 51*(5), 1469–1483.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Subramanyam, A., Y. Chen, and R. L. Grossman (2026). Scaling laws revisited: Modeling the role of data quality in language model pretraining. In *The Fourteenth International Conference on Learning Representations*.

Vandenbosch, M. B. and C. B. Weinberg (1995). Product and price competition in a two-dimensional vertical differentiation model. *Marketing Science 14*(2), 224–249.

Villalobos, P., A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbhahn (2024). Position: will we run out of data? limits of llm scaling based on human-generated data. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Online Appendix

A Omitted Proofs

Conventions. (i) If a worker is indifferent between adoption and non-adoption, we assume they do not adopt. (ii) If a worker is indifferent between two firms (equal utility), we assume they split evenly across firms. (iii) In the data-market stage, the seller accepts if and only if $p \geq c_d$. These conventions only matter in knife-edge equality cases.

A.1 Proofs from Section 3: Knowledge Economy

A.1.1 Proof of Lemma 1 (Skill Protection)

Proof. Fix $z \in (z_2, z_1]$. Since $z > z_2$, adopting Model 2 yields utility $U_2(z) = z - r_2$, while non-adoption yields $U_0(z) = z$. Because $r_2 \geq 0$, we have $U_2(z) \leq U_0(z)$ (strict if $r_2 > 0$). Under the convention that agents do not adopt when indifferent, Model 2 attracts no adopters in $(z_2, z_1]$. □

A.1.2 Proof of Lemma 2 (Winner-Take-All Among Low-Skill Workers)

Proof. Fix $z < z_2 \leq z_1$. For each $i \in \{1, 2\}$, adoption yields utility

$$U_i(z) = z_i - r_i \equiv \tilde{z}_i,$$

which is independent of z. Hence all workers in $[0, z_2)$ rank the two firms identically. If $\tilde{z}_i > \tilde{z}_{-i}$ then firm i captures the entire segment; if $\tilde{z}_1 = \tilde{z}_2$, workers are indifferent and market shares are determined by the tie-breaking rule. □

A.1.3 Proof of Lemma 3 (Leader's Pricing Advantage)

Proof. Let Firm 2 set $r_2 = c$. Then $\tilde{z}_2 = z_2 - c$. If Firm 1 sets r_1 such that

$$\tilde{z}_1 \equiv z_1 - r_1 > \tilde{z}_2 = z_2 - c \quad \Longleftrightarrow \quad r_1 < c + z_1 - z_2,$$

then by Lemma 2 Firm 1 captures all adopters in $[0, z_2)$. By Lemma 1, Firm 2 is never adopted on $(z_2, z_1]$, so any adopters there (if any) must choose Firm 1. Since $z_1 > z_2$ implies $z_1 - z_2 > 0$, any $r_1 \in (c,\, c + z_1 - z_2)$ satisfies $r_1 > c$ and the strict inequality above. At $r_1 = c + z_1 - z_2$, we have $\tilde{z}_1 = \tilde{z}_2$ and low-skill workers are indifferent. □

A.1.4 Proof of Lemma 4 (Portfolio Collapse)

Proof. Fix a worker of type z and a model $(z', r') \in P_i$. If $z \geq z'$, then adoption yields utility $z - r' \leq z$, so such a model is (weakly) dominated by non-adoption. Hence any adopted model must satisfy $z' > z$, in which case utility equals the constant $z' - r'$. Therefore, conditional on adopting from portfolio P_i, the worker chooses

$$\arg \max_{(z',r') \in P_i : z' > z} \{z' - r'\},$$

which is (whenever nonempty) attained by any maximizer of $\tilde{z}_i^* \equiv \max_{(z',r') \in P_i} \{z' - r'\}$. Any model with $z' - r' < \tilde{z}_i^*$ is never selected. □

A.1.5 Proof of Lemma 5 (Strategic Redundancy)

Proof. By Lemma 4, only the model $(z_{i,k}, r_{i,k})$ achieving the maximum effective quality $\tilde{z}_i^*$ generates positive revenue. All other models $j \neq k$ in the portfolio earn zero revenue, since they attract no workers. However, each such model incurs a positive development cost $C(z_{i,j}) > 0$.

Define the single-model portfolio $P_i' = \{(z_{i,k}, r_{i,k})\}$. Its revenue is identical to P_i (both portfolios have the same effective quality $\tilde{z}_i^*$), but its cost is strictly lower:

$$C(P_i') = C(z_{i,k}) < C(z_{i,k}) + \sum_{j \neq k} C(z_{i,j}) = C(P_i).$$

Therefore $\pi(P_i') = \text{Rev}(P_i') - C(P_i') = \text{Rev}(P_i) - C(P_i') > \text{Rev}(P_i) - C(P_i) = \pi(P_i)$. □

A.1.6 Proof of Lemma 6 (Flagship Dominance under Free Degradation)

Proof. Consider any model $(z_{i,j}, r_{i,j}) \in P_i$. Since prices are bounded below by marginal cost, $r_{i,j} \geq c$, we have

$$z_{i,j} - r_{i,j} \leq z_{i,j} - c \leq z_i^{\max} - c,$$

where the last inequality follows from $z_{i,j} \leq z_i^{\max}$ by definition. Since this bound holds for every model in the portfolio, it holds for the maximum:

$$\max_{(z,r) \in P_i} \{z - r\} \leq z_i^{\max} - c.$$

Moreover, the upper bound is attained by pricing the flagship model at marginal cost: $(z_i^{\max}, c)$. This means the flagship at marginal cost weakly dominates any portfolio configuration, and the competitive outcome between firms reduces to a comparison of $z_1^{\max} - c$ versus $z_2^{\max} - c$. □

A.2 Proofs from Section 4: Pretraining Data Market

A.2.1 Proof of Lemma 7 (Existence of Pooling Equilibrium)

Proof. Consider the strategy profile: the buyer offers $p^*(b) \equiv c_d$ for all $b \in \{b_L, b_H\}$, and the seller accepts iff $p \geq c_d$. Since $p = c_d$ is uninformative about b, Bayes' rule implies the seller's posterior after observing $p = c_d$ is the signal-posterior $\mu(\cdot \mid s)$.

Seller optimality: accepting yields payoff $c_d - c_d = 0$ and rejecting yields 0, so acceptance is optimal for both signals.

Buyer optimality: given $V_L \geq c_d$ and $V_H > V_L$, for either b the payoff from $p = c_d$ is $V(b) - c_d \geq 0$. Any deviation $p < c_d$ is rejected and yields 0, while any deviation $p > c_d$ yields strictly lower payoff $V(b) - p < V(b) - c_d$. □

A.2.2 Proof of Lemma 8 (The Strategy of Obfuscation)

Proof. Pooling at $p = c_d$ yields expected payoff

$$\mathbb{E}[V(B)] - c_d = \mu_0 V_H + (1 - \mu_0) V_L - c_d.$$

In any separating equilibrium, offers must satisfy $p_H > p_L \geq c_d$ (otherwise the seller rejects), hence $p_H > c_d$ and the buyer's payoff in state b_H is $V_H - p_H < V_H - c_d$. Since the pooling payoff in state b_H equals $V_H - c_d$, the buyer strictly prefers pooling. □

A.2.3 Proof of Proposition 2 (Information Rent Extraction)

Proof. Lemma 7 constructs a pooling PBE with $p^* = c_d$. Lemma 8 rules out separating equilibria (buyer strictly prefers pooling).

In any pooling equilibrium with price $p' > c_d$, the buyer can deviate to $p' - \varepsilon \geq c_d$ and strictly increase payoff, hence the only pooling price is $p^* = c_d$.

At $p^* = c_d$, the seller earns 0 and the buyer earns $V(b) - c_d$ in each state, so

$$\mathbb{E}[\pi_B] = \mathbb{E}[V(B)] - c_d = \mu_0 V_H + (1 - \mu_0)V_L - c_d.$$

□

A.2.4 Proof of Lemma 9 (The Value of Pessimistic Signals)

Proof. For $s \in \{H, L\}$,

$$\underline{p}(s) = \theta \mathbb{E}[V(B) \mid s] + (1 - \theta)c_d.$$

Since $\mu_H > \mu_L$ and $V_H > V_L$,

$$\underline{p}(H) - \underline{p}(L) = \theta(\mu_H - \mu_L)(V_H - V_L) > 0 \quad \Rightarrow \quad \pi_B(b, L) > \pi_B(b, H) \ \forall b.$$

Also, for any s, $\pi_B(b_H, s) - \pi_B(b_L, s) = V_H - V_L > 0$. Hence the state maximizing $\pi_B(b, s)$ is (b_H, L). □

A.2.5 Proof of Proposition 3 (Information Rent and Signal Precision)

Proof. With $\theta = 1$, $\underline{p}(s) = \mathbb{E}[V(B) \mid s]$. For $\gamma \in (1/2, 1)$ we have $\mu_H, \mu_L \in (0, 1)$, hence

$$V_L < \underline{p}(s) < V_H \quad \Rightarrow \quad (V(B) - \underline{p}(s))^+ = \mathbf{1}\{B = b_H\}\,(V_H - \underline{p}(s)).$$

Therefore

$$\Pi_B(\gamma) = \mu_0 \Big[\gamma\big(V_H - \underline{p}(H)\big) + (1 - \gamma)\big(V_H - \underline{p}(L)\big)\Big].$$

Let $\Delta V \equiv V_H - V_L > 0$. Since $\underline{p}(s) = \mu_s V_H + (1 - \mu_s)V_L$ with $\mu_s \equiv \mathbb{P}(B = b_H \mid s)$, we have $V_H - \underline{p}(s) = (1 - \mu_s)\Delta V$.

Using Bayes' rule,

$$1 - \mu_H = \frac{(1 - \gamma)(1 - \mu_0)}{\gamma\mu_0 + (1 - \gamma)(1 - \mu_0)}, \qquad 1 - \mu_L = \frac{\gamma(1 - \mu_0)}{(1 - \gamma)\mu_0 + \gamma(1 - \mu_0)}.$$

Substituting and simplifying yields

$$\Pi_B(\gamma) = \mu_0(1-\mu_0)\,\Delta V \cdot \frac{\gamma(1-\gamma)}{\mu_0(1-\mu_0) + (2\mu_0-1)^2\gamma(1-\gamma)}.$$

Let $x \equiv \gamma(1-\gamma)$ and $g(x) \equiv x/(a+dx)$ with $a \equiv \mu_0(1-\mu_0) > 0$ and $d \equiv (2\mu_0-1)^2 \geq 0$. Then $g'(x) = a/(a+dx)^2 > 0$, while $x'(\gamma) = 1-2\gamma < 0$ on $(1/2, 1)$, so $\Pi_B(\gamma)$ is strictly decreasing in γ. Finally, $\lim_{\gamma\to 1} x = 0$ implies $\lim_{\gamma\to 1} \Pi_B(\gamma) = 0$. □

A.2.6 Proof of Lemma 10 (Systematic Cherry-Picking)

Proof. With non-exclusive access, the buyer chooses a subset $A \subseteq \{1, \ldots, I\}$. Since payoffs are additive across datasets, the objective is

$$\max_A \sum_{i\in A} \big(V(b_i) - \underline{p}_i(s_i)\big),$$

so optimality is characterized pointwise: include i iff $V(b_i) - \underline{p}_i(s_i) \geq 0$, i.e.

$$V(b_i) \geq \underline{p}_i(s_i) = \theta\,\mathbb{E}[V(B_i) \mid s_i] + (1-\theta)c_d.$$

Thus the buyer selects precisely those datasets whose realized value exceeds the seller's signal-contingent valuation ("underpriced" datasets). □

A.2.7 Proof of Lemma 11

Proof. Let $\mu_s \equiv \mathbb{P}(B_i = b_H \mid s)$ for $s \in \{H, L\}$. Under the benchmark,

$$p_i(s) = \mathbb{E}[V(B_i) \mid s] = \mu_s V(b_H) + (1-\mu_s)V(b_L).$$

Hence

$$p_i(H) - p_i(L) = (\mu_H - \mu_L)\big(V(b_H) - V(b_L)\big).$$

By informativeness, $\mu_H > \mu_L$, and by assumption $V(b_H) > V(b_L)$, so $p_i(H) - p_i(L) > 0$. □

A.2.8 Proof of Proposition 4 (The Selection Advantage)

Proof. Under Lemma 11, $p_i(s_i) = \mathbb{E}[V(B_i) \mid s_i]$ and $p_i(H) > p_i(L)$.

(i) The buyer observes $(b_i, p_i(s_i))_{i=1}^{I}$ and chooses $i^* \in \arg\max_i \{V(b_i) - p_i(s_i)\}$.

(ii) If $b_i = b_L$, then $p_i(s_i)$ is a strict convex combination of V_L and V_H, so $V_L - p_i(s_i) < 0$. If $b_i = b_H$, define

$$\Delta_H \equiv V_H - p_i(H), \qquad \Delta_L \equiv V_H - p_i(L).$$

Then $p_i(H) > p_i(L)$ implies $\Delta_L > \Delta_H > 0$, so the maximal net surplus state is (b_H, L).

(iii) Let $\Delta^* \equiv \max\{0, \max_i(V(b_i) - p_i(s_i))\}$. With binary quality/signals,

$$\Delta^* = \begin{cases} \Delta_L, & \exists i : (b_i, s_i) = (b_H, L), \\ \Delta_H, & \nexists\, i : (b_i, s_i) = (b_H, L) \text{ and } \exists\, i : (b_i, s_i) = (b_H, H), \\ 0, & \forall i : b_i = b_L. \end{cases}$$

Let $q_L \equiv \Pr(b_i = b_H, s_i = L) = \mu_0(1-\gamma)$ and $q_H \equiv \Pr(b_i = b_H, s_i = H) = \mu_0\gamma$. Then

$$\mathbb{E}[\Delta^*] = \Delta_L\big(1 - (1-q_L)^I\big) + \Delta_H\big((1-q_L)^I - (1-\mu_0)^I\big).$$

Since $\Delta_L > \Delta_H > 0$ and $q_L, \mu_0 \in (0,1)$,

$$\mathbb{E}[\Delta^*]_{I+1} - \mathbb{E}[\Delta^*]_I = q_L(\Delta_L - \Delta_H)(1-q_L)^I + \mu_0\Delta_H(1-\mu_0)^I > 0,$$

so $\mathbb{E}[\Delta^*]$ is strictly increasing in I. Moreover, as $\gamma \to 1$, $q_L \to 0$ and $\Delta_H = V_H - p_i(H) \to 0$, hence $\mathbb{E}[\Delta^*] \to 0$.[12]

□

[12] Non-monotonicity: for $\mu_0 = 1/2$, $V_H = 2$, $V_L = 1$, $I = 20$, one obtains $\mathbb{E}[\Delta^*](0.6) \approx 0.598 < \mathbb{E}[\Delta^*](0.8) \approx 0.727$, while $\lim_{\gamma\to 1} \mathbb{E}[\Delta^*](\gamma) = 0$.

A.3 Proofs from Section 5: Combined Two-Layer Market

A.3.1 Proof of Lemma 15 (Continuity of $\mathcal{E}$)

Proof. Fix $\omega = (b_i, c_i)_{i=1}^I$. Define the baseline term

$$f_0(\hat{z}) \equiv (1 + p_0)C(\hat{z}, b_0),$$

and for each $i \in \{1, \dots, I\}$ define

$$f_i(\hat{z}, z; \omega) \equiv (1 + p_i^*(c_i; z))\, C(\hat{z}, b_i).$$

By Assumption 1(ii) and continuity of C on $[0, \bar{z}] \times [\underline{b}, \bar{b}]$, each f_i is continuous on $[0, \bar{z}]^2$ (and f_0 is continuous on $[0, \bar{z}]$). Hence

$$\kappa(\hat{z}; z, \omega) = \min\Big\{ f_0(\hat{z}),\ \min_{i \in \{1,\dots,I\}} f_i(\hat{z}, z; \omega) \Big\}$$

is continuous on $[0, \bar{z}]^2$ as the minimum of finitely many continuous functions.

Let $\bar{p} \equiv \max_{1 \le i \le I} \sup_{(c,z) \in [\underline{c}, \bar{c}] \times [0, \bar{z}]} p_i^*(c; z) < \infty$ (Assumption 1(ii)). Since C is continuous and $[0, \bar{z}] \times [\underline{b}, \bar{b}]$ is compact with $\bar{z} < 1$, the bound

$$\bar{C} \equiv \sup_{(\hat{z}, b) \in [0, \bar{z}] \times [\underline{b}, \bar{b}]} C(\hat{z}, b) < \infty$$

holds. Therefore, for all $(\hat{z}, z) \in [0, \bar{z}]^2$ and all ω,

$$0 \le \kappa(\hat{z}; z, \omega) \le \min\{(1 + p_0)C(\hat{z}, b_0),\ (1 + \bar{p})\bar{C}\} \le (1 + \max\{p_0, \bar{p}\})\bar{C}.$$

The right-hand side is integrable (it is a constant), hence dominated convergence implies that $\mathcal{E}(\hat{z}; z) = \mathbb{E}_\omega[\kappa(\hat{z}; z, \omega)]$ is continuous on $[0, \bar{z}]^2$.

Finally, since $i = 0$ is feasible in (48), for every $(\hat{z}, z)$ and ω, $\kappa(\hat{z}; z, \omega) \le (1 + p_0)C(\hat{z}, b_0)$. Taking expectations yields $\mathcal{E}(\hat{z}; z) \le (1 + p_0)C(\hat{z}, b_0)$. □

A.3.2 Proof of Lemma 16 (Kakutani conditions)

Proof. Fix $z \in [0, \bar{z}]$.

(i) By Lemma 15, $\hat{z} \mapsto U(\hat{z}; z)$ is continuous on the compact set $[0, \bar{z}]$. By the Weierstrass theorem, the maximum is attained and $\Phi(z) = \arg\max_{\hat{z} \in [0,\bar{z}]} U(\hat{z}; z)$ is nonempty. Since $\Phi(z)$ is a closed subset of a compact set, it is compact.

(ii) By Assumption 2, the function $\hat{z} \mapsto U(\hat{z}; z)$ has a unique global maximizer on $[0, \bar{z}]$. Hence $\Phi(z)$ is a singleton, and therefore convex.

(iii) Lemma 15 and Assumption 1(i) imply U is continuous on $[0, \bar{z}]^2$. Since the feasible set is the constant compact set $[0, \bar{z}]$, Berge's maximum theorem implies that Φ is upper hemicontinuous; equivalently, Φ has a closed graph.

(iv) Because $TR(0) = 0$ and $C(0, b) = 0$ for all b, we have $\kappa(0; z, \omega) = 0$ and hence $\mathcal{E}(0; z) = 0$. Therefore $U(0; z) = 0$ for all z. By Lemma 15 and Assumption 3(i), for any z,

$$U(\tilde{z}; z) = sTR(\tilde{z}) - \mathcal{E}(\tilde{z}; z) \geq sTR(\tilde{z}) - (1 + p_0)C(\tilde{z}, b_0) > 0,$$

so $0 \notin \Phi(z)$.

Next, $p_i^*(\cdot; z) \geq c$ and Assumption 1(iii) imply $c_i \geq 0$. Together with $b_0 \leq \bar{b}$ and $b_i \leq \bar{b}$, we obtain for every realization ω,

$$\kappa(\bar{z}; z, \omega) = \min\Big\{(1 + p_0)C(\bar{z}, b_0),\ \min_{i \in \{1,\dots,I\}} (1 + p_i^*(c_i; z))C(\bar{z}, b_i)\Big\} \geq C(\bar{z}, \bar{b}).$$

Taking expectations gives $\mathcal{E}(\bar{z}; z) \geq C(\bar{z}, \bar{b})$ and thus

$$U(\bar{z}; z) = sTR(\bar{z}) - \mathcal{E}(\bar{z}; z) \leq s \sup_{t \in [0,\bar{z}]} TR(t) - C(\bar{z}, \bar{b}) < 0$$

by Assumption 3(ii). Since $U(0; z) = 0$, it follows that $\bar{z} \notin \Phi(z)$. □

A.3.3 Proof of Theorem B.1 (Existence of two-layer equilibrium)

Proof. By Lemma 16, the correspondence $\Phi : [0, \bar{z}] \rightrightarrows [0, \bar{z}]$ has nonempty, convex, compact values and a closed graph. Since $[0, \bar{z}]$ is a nonempty, compact, convex subset of $\mathbb{R}$, Kakutani's fixed-point theorem applies and yields an element $z^* \in [0, \bar{z}]$ such that $z^* \in \Phi(z^*)$. Lemma 16(iv) implies $z^* \in (0, \bar{z})$. □

A.3.4 Proof of Proposition 9

Fix (c, z). The seller's expected profit is

$$\Pi(p; c, z) = (p - c) \int_{\underline{b}(p,z)}^{\bar{b}} C(z, b)\, d\mu(b \mid c),$$

where

$$\underline{b}(p, z) \equiv \inf\{b \in [\underline{b}, \bar{b}] : p \le p^{\max}(z; b)\}, \qquad p^{\max}(z; b) = (1 + p_0)\frac{C(z, b_0)}{C(z, b)} - 1.$$

Under the scaling law, $b \mapsto p^{\max}(z; b)$ is continuous and strictly increasing for each $z > 0$, and $\bar{p}(z) \equiv p^{\max}(z; \bar{b}) < \infty$ because $\bar{z} < 1$.

Any price $p < c$ yields weakly negative profit, while any price $p > \bar{p}(z)$ yields zero sales. Hence every maximizer lies in

$$[c, \max\{c, \bar{p}(z)\}],$$

which implies boundedness and $p^*(c; z) \ge c$.

Part (a): monotonicity from MLRP. For fixed z, define the buyer's reservation-value random variable

$$W_z \equiv p^{\max}(z; B),$$

and let $f_z(w \mid c)$ denote its conditional density. Because $b \mapsto p^{\max}(z; b)$ is strictly increasing, Assumption 4 for $B \mid c$ implies the same MLRP property for $W_z \mid c$.

Now let

$$K(z) \equiv (1 + p_0)C(z, b_0).$$

By the definition of $p^{\max}(z; b)$,

$$(1 + p^{\max}(z; b))\, C(z, b) = K(z),$$

so a change of variables from b to $w = p^{\max}(z; b)$ gives

$$Q(p; c, z) = K(z) \int_p^{\bar{p}(z)} \frac{f_z(w \mid c)}{1 + w}\, dw.$$

Define

$$h_z(w \mid c) \equiv \frac{f_z(w \mid c)}{1+w}, \qquad H_z(p,c) \equiv \int_p^{\bar{p}(z)} h_z(w \mid c)\, dw.$$

Since $(1+w)^{-1}$ is positive and independent of c, the family $h_z(\cdot \mid c)$ inherits MLRP from $f_z(\cdot \mid c)$. A standard closure property of MLRP families implies that the upper-tail integral $H_z(p,c)$ is TP_2 in (p,c). Therefore, for any $p' > p$,

$$\frac{H_z(p',c)}{H_z(p,c)}$$

is weakly increasing in c.

For $c < p < p'$, we can write

$$\frac{\Pi(p';c,z)}{\Pi(p;c,z)} = \frac{p'-c}{p-c} \cdot \frac{H_z(p',c)}{H_z(p,c)}.$$

The first factor is strictly increasing in c, and the second is weakly increasing in c, so the ratio is increasing in c. Hence Π satisfies the single-crossing property in (p,c). By the monotone selection theorem, there exists an optimal selection $p^*(\cdot;z)$ that is weakly increasing in c.

Part (b): uniqueness and continuity. Without loss of generality, restrict the seller's action set to

$$A(c,z) \equiv [c, \max\{c, \bar{p}(z)\}],$$

since prices above $\max\{c, \bar{p}(z)\}$ never improve profit relative to $p = c$.

If $c \geq \bar{p}(z)$, then $A(c,z) = \{c\}$ and the unique optimum is $p^*(c;z) = c$.

Now consider the nondegenerate case $c < \bar{p}(z)$. Under Assumption 5, the function $Q(\cdot;c,z)$ is log-concave on $[c, \bar{p}(z)]$. On the open interval $(c, \bar{p}(z))$,

$$\log \Pi(p;c,z) = \log(p-c) + \log Q(p;c,z).$$

The first term is strictly concave and the second is concave, so $\log \Pi(\cdot;c,z)$ is strictly concave. Therefore $\Pi(\cdot;c,z)$ is strictly log-concave on $(c, \bar{p}(z))$, and thus has at most one maximizer there. Since

$$\Pi(c;c,z) = 0 \qquad \text{and} \qquad \Pi(\bar{p}(z);c,z) = 0,$$

the maximizer on $A(c,z)$ is unique.

Finally, by Assumption 1(iii), continuity of the scaling law, and continuity of $p^{\max}(z;b)$ in (z,b), the objective $\Pi(p;c,z)$ is jointly continuous in (p,c,z). The feasible-set correspondence $A(c,z)$ is compact-valued and continuous in (c,z). Since the maximizer is unique for every (c,z), Berge's maximum theorem implies that $p^*(c;z)$ is continuous in (c,z) on $[\underline{c},\bar{c}]\times[0,\bar{z}]$. □

A.3.5 Proof of Lemma 17 (Marginal Cost Explosion)

Proof. Fix $z\in[0,\bar{z}]$ and a realization $\omega=(b_i,c_i)_{i=1}^{I}$. Write

$$\kappa(\hat{z};\omega)=\min_{i\in\{0,1,\dots,I\}} f_i(\hat{z}),\qquad f_i(\hat{z})\equiv(1+p_i)\,C(\hat{z},b_i),$$

where (p_0,b_0) is the baseline and $p_i=p_i^*(c_i;z)$ for $i\geq 1$.

Each f_i is C^2 and strictly convex on $[0,\bar{z})$ with

$$f_i'(\hat{z})=\frac{1+p_i}{b_i}(1-\hat{z})^{-(1/b_i+1)}>0.$$

Since κ is the pointwise minimum of finitely many convex Lipschitz functions, it is Lipschitz on $[0,\bar{z}]$ and differentiable at every $\hat{z}$ where the minimizer is unique. Because the functions $\{f_i\}$ are real-analytic and pairwise distinct (unless $b_i=b_j$ and $p_i=p_j$), any two of them can cross at most finitely many times on $[0,\bar{z}]$. Hence the set of switch points—values of $\hat{z}$ where two or more f_i share the minimum—is finite for each ω, and κ is differentiable outside this finite set. At every regular point $\hat{z}$, if $i^*(\hat{z})=\arg\min_i f_i(\hat{z})$ is unique, then

$$\kappa'(\hat{z};\omega)=f'_{i^*(\hat{z})}(\hat{z}).$$

Lower bound. Since $p_i\geq 0$ for all i, we have $f_i'(\hat{z})=(1+p_i)\,C'(\hat{z},b_i)\geq C'(\hat{z},b_i)$. Moreover, $C'(\hat{z},b)=\frac{1}{b}(1-\hat{z})^{-(1/b+1)}$ is decreasing in b (higher data quality $\Rightarrow$ lower marginal compute), so $C'(\hat{z},b_i)\geq C'(\hat{z},\bar{b})$ for every $b_i\in[\underline{b},\bar{b}]$. Therefore

$$\kappa'(\hat{z};\omega)\geq C'(\hat{z},\bar{b})=\frac{1}{\bar{b}}(1-\hat{z})^{-(1/\bar{b}+1)}$$

at every regular point. Taking expectations yields $\mathcal{E}'(\hat{z};z)\geq C'(\hat{z},\bar{b})$ a.e., which is (52).

Divergence. Since $C'(\hat{z},\bar{b})=(1/\bar{b})(1-\hat{z})^{-(1/\bar{b}+1)}\to+\infty$ as $\hat{z}\uparrow 1$, the lower bound gives

$\mathcal{E}'(\hat{z}; z) \to +\infty$ as $\hat{z} \uparrow \bar{z}$. □

A.3.6 Proof of Proposition 10 (Primitive sufficient condition for a unique interior maximizer: baseline case)

Proof. Since $I = 0$, the current capability choice z is irrelevant and we write

$$U(\hat{z}) = sTR(\hat{z}) - (1 + p_0)C(\hat{z}, b_0).$$

Let $q(\hat{z}) = \hat{z} - r^*(\hat{z})$ denote the worker cutoff induced by the optimal downstream price. The downstream first-order condition is

$$r^*(\hat{z}) = \frac{G(q(\hat{z}))}{g(q(\hat{z}))} = \psi(q(\hat{z})),$$

so

$$\hat{z} = q(\hat{z}) + \psi(q(\hat{z})).$$

Differentiating yields

$$q'(\hat{z}) = \frac{1}{1 + \psi'(q(\hat{z}))} > 0.$$

By the envelope theorem,

$$TR'(\hat{z}) = G(q(\hat{z})),$$

and therefore

$$TR''(\hat{z}) = g(q(\hat{z}))q'(\hat{z}) = \frac{g(q(\hat{z}))}{1 + \psi'(q(\hat{z}))} = M(q(\hat{z})).$$

By assumption, M is nonincreasing and $q(\hat{z})$ is increasing, so TR'' is nonincreasing on $[0, \bar{z}]$.

The baseline cost satisfies

$$C''(\hat{z}, b_0) = \frac{(1/b_0 + 1)}{b_0}(1 - \hat{z})^{-(1/b_0+2)},$$

which is strictly increasing in $\hat{z}$. Hence

$$U''(\hat{z}) = sTR''(\hat{z}) - (1 + p_0)C''(\hat{z}, b_0)$$

is nonincreasing, so U' is concave on $[0, \bar{z}]$.

At the left endpoint,

$$U'(0) = sTR'(0) - (1 + p_0)C'(0, b_0) = -(1 + p_0)/b_0 < 0,$$

because $TR'(0) = G(0) = 0$. At the right endpoint,

$$U'(\bar{z}) \leq s - (1 + p_0)C'(\bar{z}, b_0) < 0,$$

since $TR'(\hat{z}) = G(q(\hat{z})) \leq 1$ for all $\hat{z}$.

By Assumption 3(i), there exists $\tilde{z} \in (0, \bar{z})$ such that

$$U(\tilde{z}) = sTR(\tilde{z}) - (1 + p_0)C(\tilde{z}, b_0) > 0 = U(0).$$

By the mean value theorem, there exists $\xi \in (0, \tilde{z})$ such that $U'(\xi) > 0$. Thus U' is a concave function that is negative at both endpoints and positive somewhere in the interior. Therefore U' crosses zero exactly twice: once at a local minimum of U, and once at a local maximum.

Because $U(0) = 0$, $U(\tilde{z}) > 0$, and Assumption 3(ii) implies

$$U(\bar{z}) \leq s \sup_{t \in [0, \bar{z}]} TR(t) - C(\bar{z}, \bar{b}) < 0,$$

the second zero crossing is the unique global maximizer of U on $[0, \bar{z}]$. It lies in the interior $(0, \bar{z})$. □

A.3.7 Proof of Proposition 11 (Primitive sufficient condition for a unique maximizer: data-market case)

Proof. Fix $z \in [0, \bar{z}]$. Under Assumption 6, for almost every realized menu ω there exists an index $j = j(\omega, z)$ such that

$$\kappa(\hat{z}; z, \omega) = \mathcal{C}_j(\hat{z}; z, \omega) \qquad \forall \hat{z} \in [0, \bar{z}],$$

where $\mathcal{C}_0(\hat{z}; z, \omega) = (1 + p_0)C(\hat{z}, b_0)$ and $\mathcal{C}_i(\hat{z}; z, \omega) = (1 + p_i^*(c_i; z))C(\hat{z}, b_i)$ for $i \geq 1$. Hence

$$\mathcal{E}(\hat{z}; z) = \mathbb{E}_\omega[\mathcal{C}_j(\hat{z}; z, \omega)]\,.$$

In particular, $\mathcal{E}(\cdot; z)$ is C^2, with

$$\partial_{\hat{z}}^2 \mathcal{E}(\hat{z}; z) = \mathbb{E}_\omega \left[\partial_{\hat{z}}^2 \mathcal{C}_j(\hat{z}; z, \omega) \right].$$

Because $C''(\hat{z}, b)$ is increasing in $\hat{z}$ for every b, the function $\partial_{\hat{z}}^2 \mathcal{E}(\hat{z}; z)$ is nondecreasing in $\hat{z}$.

By Proposition 10, the maintained downstream curvature condition implies that TR'' is nonincreasing. Therefore

$$\partial_{\hat{z}}^2 U(\hat{z}; z) = sTR''(\hat{z}) - \partial_{\hat{z}}^2 \mathcal{E}(\hat{z}; z)$$

is nonincreasing, so $\partial_{\hat{z}} U(\cdot; z)$ is concave.

At the left endpoint,

$$\partial_{\hat{z}} U(0^+; z) = -\mathbb{E}_\omega[\partial_{\hat{z}} \mathcal{C}_j(0; z, \omega)] < 0.$$

At the right endpoint,

$$\partial_{\hat{z}} U(\bar{z}; z) \leq s - C'(\bar{z}, \bar{b}) < 0,$$

since $TR'(\hat{z}) \leq 1$ and $\partial_{\hat{z}} \mathcal{E}(\hat{z}; z) \geq C'(\hat{z}, \bar{b})$.

Finally, Lemma 15 and Assumption 3(i) imply that there exists $\tilde{z} \in (0, \bar{z})$ such that

$$U(\tilde{z}; z) \geq sTR(\tilde{z}) - (1 + p_0) C(\tilde{z}, b_0) > 0 = U(0; z).$$

Hence, by the mean value theorem, $\partial_{\hat{z}} U(\cdot; z)$ is positive somewhere in the interior. Since it is concave and negative at both endpoints, it crosses zero exactly twice, and the second crossing is the unique global maximizer of $U(\cdot; z)$ on $[0, \bar{z}]$. □

A.3.8 Proof of Lemma 12 (The Efficient Procurement Rule)

Proof. Fix z^* and a realized menu $\{(b_i, p_i)\}_{i=0}^I$. If the producer purchases from seller i, total training expenditure is

$$\mathcal{C}(z^*, b_i, p_i) = (1 + p_i)\big((1 - z^*)^{-1/b_i} - 1\big).$$

Downstream revenue $TR(z^*)$ is independent of the seller choice. Therefore, conditional on z^*, the producer's profit from choosing i is

$$\Pi_i(z^*) = TR(z^*) - \mathcal{C}(z^*, b_i, p_i).$$

Hence, for any i, j,

$$\Pi_i(z^*) \geq \Pi_j(z^*) \iff \mathcal{C}(z^*, b_i, p_i) \leq \mathcal{C}(z^*, b_j, p_j),$$

so any profit-maximizing choice must minimize $\mathcal{C}(z^*, b_i, p_i)$ over $i \in \{0, 1, \ldots, I\}$. □

A.3.9 Proof of Lemma 14 (Efficiency-Adjusted Competition)

Proof. By Lemma 12, the winning seller minimizes $(1 + p_i)\mathcal{C}(z^*, b_i)$. Thus seller i beats seller j if and only if

$$(1 + p_i)\mathcal{C}(z^*, b_i) < (1 + p_j)\mathcal{C}(z^*, b_j).$$

Since $(1 + p_j)\mathcal{C}(z^*, b_i) > 0$, dividing both sides yields

$$\frac{1 + p_i}{1 + p_j} < \frac{\mathcal{C}(z^*, b_j)}{\mathcal{C}(z^*, b_i)}.$$

□

A.3.10 Proof of Proposition 5 (Monopoly Pricing with Noisy Signals)

Proof. Fix (z^*, c) and let B denote the (true) data quality. In the monopoly case, the buyer purchases from the seller if and only if the total training expenditure using the seller does not exceed the baseline alternative:

$$(1 + p)\,\mathcal{C}(z^*, B) \leq (1 + p_0)\,\mathcal{C}(z^*, b_0).$$

By Lemma 13, this is equivalent to $p \leq p^{\max}(z^*; B)$. Conditional on the signal c (hence posterior $\mu(\cdot \mid c)$), the seller's expected profit from posting $p \geq c$ is

$$\Pi(p, c) = \int_{\underline{b}}^{\overline{b}} (p - c)\,\mathcal{C}(z^*, b)\,\mathbf{1}\{p \leq p^{\max}(z^*; b)\}\,d\mu(b \mid c),$$

which yields the stated maximization problem.

(i) Define the single-crossing condition in (p, c) as: for any $p' > p$, the set $\{c : \Pi(p', c) \geq \Pi(p, c)\}$ is an upper interval. Under this condition, the argmax correspondence $c \mapsto \arg\max_{p \geq c} \Pi(p, c)$ is increasing, so it admits a (weakly) increasing selection $p^*(c)$. (See, e.g., the monotone selection theorem for single-crossing objectives.)

(ii) Let $X \equiv p^{\max}(z^*; B)$ and $\tilde{p}(c) \equiv \mathbb{E}[X \mid c]$. By the law of total variance,

$$\mathrm{Var}(X) = \mathrm{Var}(\mathbb{E}[X \mid c]) + \mathbb{E}[\mathrm{Var}(X \mid c)].$$

If c is a noisy signal of B (i.e., X is not $\sigma(c)$-measurable), then $\mathrm{Var}(X \mid c) > 0$ on a positive-measure set, implying $\mathrm{Var}(\tilde{p}) < \mathrm{Var}(X) = \mathrm{Var}(p^{\max})$.

If, in addition, $p^*(c) = \psi(\tilde{p}(c))$ for some L-Lipschitz map ψ, then for an independent copy $\tilde{p}'$ of $\tilde{p}$ we have

$$\mathrm{Var}(p^*) = \tfrac{1}{2}\,\mathbb{E}\big[(\psi(\tilde{p}) - \psi(\tilde{p}'))^2\big] \le \tfrac{1}{2}\,\mathbb{E}\big[(L\,|\tilde{p} - \tilde{p}'|)^2\big] = L^2\,\mathrm{Var}(\tilde{p}).$$

In particular, if $L \le 1$, then $\mathrm{Var}(p^*) \le \mathrm{Var}(\tilde{p}) < \mathrm{Var}(p^{\max})$. □

A.3.11 Proof of Proposition 6 (Competitive Equilibrium)

Proof. Let $\mathcal{C}_i \equiv \mathcal{C}(z^*, b_i)$ and define the buyer's (efficiency-adjusted) total training cost from seller i as

$$T_i \equiv (1 + p_i)\mathcal{C}_i.$$

The buyer selects any $i^* \in \arg\min_i T_i$.

(i) In competition, posting a higher p_i weakly increases T_i and hence weakly decreases the event $\{i = i^*\}$. Thus the seller's optimal markup is disciplined by the additional constraint that T_i must not exceed rivals' costs. Relative to monopoly (where the only constraint is the baseline ceiling $p^{\max}(z^*; b)$), equilibrium prices are weakly shaded to maintain T_i competitive.

(ii) Fix an equilibrium realization and let i^* be chosen. For any $j \neq i^*$, we must have $T_{i^*} \le T_j$ (otherwise the buyer would strictly prefer j). Rearranging yields

$$p_{i^*} \le \frac{(1 + p_j)\mathcal{C}_j}{\mathcal{C}_{i^*}} - 1.$$

Combining with the feasibility bound $p_{i^*} \le p^{\max}(z^*; b_{i^*})$ gives the stated cap.

(iii) (Limit pricing.) Fix $z^* \in (0, 1)$ and let $\underline{\mathcal{C}} \equiv \min_{b \in [\underline{b}, \bar{b}]} \mathcal{C}(z^*, b) > 0$. Under a symmetric equilibrium strategy, the rivals' total costs $T_j = (1 + p_j)\mathcal{C}_j$ are i.i.d. with CDF F_T. Assume F_T admits a continuous density f_T and hazard rate $h_T(t) \equiv f_T(t)/(1 - F_T(t))$ that is bounded away from zero on the relevant support: $\inf_t h_T(t) \ge \underline{h} > 0$.

Consider a seller with signal c who posts a price $p \geq c$. Conditional on realized quality b (hence $\mathcal{C}(z^*, b)$), the probability of winning is

$$\Pr\{i = i^* \mid b, p\} = \big(1 - F_T((1+p)\mathcal{C}(z^*, b))\big)^{I-1},$$

where ties occur with probability zero under continuity. The seller's expected profit conditional on c is therefore

$$\Pi(p, c) = \mathbb{E}\Big[(p - c)\,\mathcal{C}(z^*, B)\,\big(1 - F_T((1+p)\mathcal{C}(z^*, B))\big)^{I-1} \;\Big|\; c\Big].$$

Differentiating under the expectation (justified by boundedness of $\mathcal{C}$ on $[\underline{b}, \bar{b}]$), the derivative takes the form

$$\frac{\partial \Pi}{\partial p}(p, c) = \mathbb{E}\Big[\mathcal{C}\,W \cdot \big(1 - (p - c)\,\mathcal{C}\,(I-1)\,h_T((1+p)\mathcal{C})\big) \;\Big|\; c\Big],$$

where $\mathcal{C} \equiv \mathcal{C}(z^*, B)$ and $W \equiv (1 - F_T((1+p)\mathcal{C}))^{I-1} > 0$.

If the markup satisfies

$$p - c > \frac{1}{(I-1)\,\underline{\mathcal{C}}\,\underline{h}},$$

then for every realization of B we have $(p - c)\,\mathcal{C}\,(I-1)\,h_T((1+p)\mathcal{C}) \geq (p - c)\,\underline{\mathcal{C}}\,(I-1)\,\underline{h} > 1$, so the term in parentheses is strictly negative. Hence $\partial \Pi / \partial p < 0$ and such a markup cannot be optimal. Therefore, in symmetric equilibrium,

$$0 \leq p^*(c) - c \leq \frac{1}{(I-1)\,\underline{\mathcal{C}}\,\underline{h}} \xrightarrow[I \to \infty]{} 0,$$

which implies limit pricing: $p^*(c) - c \to 0$ pointwise in c.

Finally, vanishing markups do not imply vanishing buyer rents. Define the (baseline-relative) slack

$$S_i \equiv p^{\max}(z^*; b_i) - c_i.$$

If c_i is a noisy signal of b_i, then S_i is non-degenerate and $\Pr(S_i > 0) > 0$. Hence $\mathbb{E}[\max_{1 \leq i \leq I} S_i^+] > 0$ for all I and is (weakly) increasing in I, so the buyer's information rent can remain strictly positive even as equilibrium markups vanish. $\square$

A.3.12 Proof of Proposition 7 (i) and (ii)

Proof. Let the buyer's equilibrium information rent be

$$\Pi_B \equiv \mathbb{E}\big[\mathcal{C}(z^*, b_0, p_0) - \mathcal{C}(z^*, b_{i^*}, p^*_{i^*})\big],$$

where i^* is the equilibrium-selected seller.

(i) (Selection effect.) Let $R_i \equiv \mathcal{C}(z^*, b_0, p_0) - \mathcal{C}(z^*, b_i, p^*_i)$ denote the rent obtainable from seller i (relative to the outside option). Since the buyer chooses $i^* \in \arg\max_i R_i$, we have $\max_{1\le i\le I} R_i \le \max_{1\le i\le I+1} R_i$ pointwise, hence Π_B is weakly increasing in I.

(ii) (Capability amplification.) From Section 5.2.3, $V(b \mid z^*)$ satisfies $\partial_{z^*}V > 0$ and $\partial^2_{bz^*}V > 0$. Therefore, for any pair $b' > b$, the valuation gap $V(b' \mid z^*) - V(b \mid z^*)$ is increasing in z^*, so any given discrepancy between realized and belief-implied quality translates into a larger cost reduction at higher z^*. Hence Π_B is increasing in z^*. □

A.3.13 Proof of Proposition 7(iii) (Signal Noise Effect)

Proof. Define the noisier rent $\tilde{R}_i = R_i - \nu_i\, C(z^*, b_i)$, where ν_i are i.i.d., mean-zero, and independent of $(b_i, \varepsilon_i)_{i=1}^I$. Set $H(x_1, \ldots, x_I) \equiv \max(x_1, \ldots, x_I, 0)$, which is convex. We show $\mathbb{E}[H(\tilde{\mathbf{R}})] \ge \mathbb{E}[H(\mathbf{R})]$ by replacing coordinates one at a time.

For $k = 0, \ldots, I$, define the hybrid vector $\mathbf{R}^{(k)}$ in which sellers $1, \ldots, k$ use the noisier cost (contributing $\tilde{R}_j$) and sellers $k{+}1, \ldots, I$ use the original cost (contributing R_j). Set $\Pi_B^{(k)} \equiv \mathbb{E}\big[H(\mathbf{R}^{(k)})\big]$. We claim $\Pi_B^{(k)} \ge \Pi_B^{(k-1)}$ for each k.

Going from $\mathbf{R}^{(k-1)}$ to $\mathbf{R}^{(k)}$, only coordinate k changes: $R_k \to \tilde{R}_k = R_k - \nu_k C(z^*, b_k)$. Define

$$M_{-k} \equiv \max\big(\{R_j^{(k)}\}_{j\ne k} \cup \{0\}\big),$$

which is identical under both configurations and is independent of $(b_k, \varepsilon_k, \nu_k)$ by the i.i.d. assumption.

Condition on all randomness except ν_k: fix $\omega \equiv \big((b_j, \varepsilon_j)_{j=1}^I, (\nu_j)_{j\ne k}\big)$, so that R_k and M_{-k} are both deterministic and only ν_k remains random. Then $\tilde{R}_k = R_k - \nu_k C(z^*, b_k)$ satisfies $\mathbb{E}[\tilde{R}_k \mid$

$\omega] = R_k$. Since $x \mapsto \max(x,\, M_{-k})$ is convex, Jensen's inequality gives

$$\mathbb{E}\big[\max(\tilde{R}_k,\, M_{-k}) \mid \omega\big] \;\geq\; \max\big(\mathbb{E}[\tilde{R}_k \mid \omega],\, M_{-k}\big) \;=\; \max(R_k,\, M_{-k}).$$

Taking iterated expectations over ω:

$$\Pi_B^{(k)} = \mathbb{E}\big[\max(\tilde{R}_k, M_{-k})\big] \geq \mathbb{E}\big[\max(R_k, M_{-k})\big] = \Pi_B^{(k-1)}.$$

Chaining from $k = 1$ to $k = I$ yields $\Pi_B^{(I)} \geq \Pi_B^{(0)}$, which is the stated inequality. □

A.3.14 Proof of Proposition 7(iv) (Quality Dispersion Effect)

Proof. **Part (a): ΔV is increasing in quality spread.**

The endogenized valuations satisfy $V(b \mid z^*) = (1 + p_0)\big[(1{-}z^*)^{-1/b_0} - 1\big] - \big[(1{-}z^*)^{-1/b} - 1\big]$ (Section 5.2.3), so the valuation gap is

$$\Delta V = V_H - V_L = C(z^*, b_L) - C(z^*, b_H), \tag{47}$$

where $C(z^*, b) = (1{-}z^*)^{-1/b} - 1$.

Consider a mean-preserving spread parameterized by $\eta \geq 0$:

$$b_H(\eta) = b_H + \eta, \qquad b_L(\eta) = b_L - \tfrac{\mu_0}{1-\mu_0}\,\eta,$$

which preserves $\mu_0 b_H + (1 - \mu_0) b_L$. Differentiating (47):

$$\frac{d(\Delta V)}{d\eta} = \underbrace{-\tfrac{\mu_0}{1-\mu_0}\, C_b(z^*, b_L)}_{>\,0} \;-\; \underbrace{C_b(z^*, b_H)}_{<\,0}.$$

Since $C_b(z^*, b) < 0$ for all $b > 0$ and $z^* \in (0, 1)$, both terms are strictly positive:

$$\frac{d(\Delta V)}{d\eta} = \frac{\mu_0}{1 - \mu_0}\,|C_b(z^*, b_L)| + |C_b(z^*, b_H)| > 0.$$

This is strictly positive for any $\mu_0 \in (0, 1)$, $b_H > b_L > 0$, and $z^* \in (0, 1)$. Intuitively, raising b_H reduces compute cost $C(b_H)$ (the numerator of ΔV shrinks), while lowering b_L raises compute

cost $C(b_L)$ (the numerator of ΔV grows). Both margins widen the gap.

Part (b): Buyer rent is linear in ΔV.

Single seller. From Proposition 3 (with $\theta = 1$), the buyer's expected information rent under signal precision $\gamma \in (1/2, 1)$ is

$$\Pi_B(\gamma) = \mu_0(1-\mu_0)\,\Delta V \cdot \frac{\gamma(1-\gamma)}{\mu_0(1-\mu_0) + (2\mu_0 - 1)^2\,\gamma(1-\gamma)}.$$

The coefficient of ΔV is strictly positive for $\gamma \in (1/2, 1)$ and $\mu_0 \in (0, 1)$, so Π_B is strictly increasing and linear in ΔV.

For general bargaining power $\theta \in [0, 1]$, the buyer's rent from state (b_H, s) is $\pi_B(b_H, s) = V_H - \underline{p}(s)$ where $\underline{p}(s) = \theta[\mu_s V_H + (1-\mu_s)V_L] + (1-\theta)c_d$. Substituting $V_H = V_L + \Delta V$ and collecting terms:

$$\begin{aligned} V_H - \underline{p}(s) &= (V_L + \Delta V) - \theta\big[\mu_s(V_L + \Delta V) + (1-\mu_s)V_L\big] - (1-\theta)c_d \\ &= (1 - \theta\mu_s)\,\Delta V + (1-\theta)(V_L - c_d). \end{aligned}$$

Since $V_L \geq c_d$ and $\theta\mu_s < 1$, the constant term $(1-\theta)(V_L - c_d) \geq 0$ and the coefficient of ΔV is $(1 - \theta\mu_s) > 0$, so the expression is strictly increasing in ΔV for each signal s. The expected rent is a positive-weight sum of such terms, hence strictly increasing and affine in ΔV.

Multiple sellers, exclusive license. From the proof of Proposition 4 (Appendix A.2.8), the buyer's expected surplus under the benchmark pricing rule is

$$\mathbb{E}[\Delta^*] = \Delta_L\big(1 - (1-q_L)^I\big) + \Delta_H\big((1-q_L)^I - (1-\mu_0)^I\big),$$

where $\Delta_H = (1-\mu_H)\Delta V$, $\Delta_L = (1-\mu_L)\Delta V$, $q_L = \mu_0(1-\gamma)$. The posteriors μ_H, μ_L and probabilities q_L depend on (γ, μ_0) but not on ΔV. Both Δ_H and Δ_L are proportional to ΔV, and the coefficients $1-(1-q_L)^I$ and $(1-q_L)^I - (1-\mu_0)^I$ are strictly positive for $I \geq 1$, $\gamma \in (1/2, 1)$, and $\mu_0 \in (0, 1)$. Hence $\mathbb{E}[\Delta^*]$ is strictly increasing and linear in ΔV.

Combining (a) and (b): A mean-preserving spread of the binary quality distribution strictly increases ΔV by part (a), and the buyer's information rent is strictly increasing in ΔV by part (b). The composition yields the stated result. □

A.3.15 Proof of Proposition 8 (Market Efficiency)

Proof. (i) Fix z^*. A planner chooses $i^{FB} \in \arg\min_i (1+c_i)C(z^*, b_i)$, while the market chooses $i^* \in \arg\min_i (1+p_i)C(z^*, b_i)$ (Lemma 12). If $p_i \neq c_i$ for some i, the induced rankings can differ.

(ii) In particular, there may exist i, j such that $(1+c_i)C(z^*, b_i) < (1+c_j)C(z^*, b_j)$ but $(1+p_i)C(z^*, b_i) > (1+p_j)C(z^*, b_j)$, implying exclusion of the first-best seller.

(iii) Suppose TR is concave in z so $TR'(z)$ is strictly decreasing. Conditional on choosing seller i, the planner's FOC is

$$s\,TR'(z) = (1+c_i)\,\partial_z C(z, b_i),$$

while the buyer's FOC is

$$s\,TR'(z) = (1+p_i)\,\partial_z C(z, b_i).$$

If $p_{i^*} > c_{i^*}$, then for the selected seller i^* the market equates marginal revenue to a strictly larger marginal cost, hence $z_{\text{market}} < z_{\text{FB}}(i^*)$. Since the planner can always mimic seller i^*, $z_{\text{FB}} \geq z_{\text{FB}}(i^*) > z_{\text{market}}$. □

A.4 Proof of Lemma 13

Proof. Fix z^* and seller i. Indifference between i and the baseline (b_0, p_0) requires

$$(1+p_i)C(z^*, b_i) = (1+p_0)C(z^*, b_0),$$

hence

$$p_i^{\max}(z^*) = (1+p_0)\frac{C(z^*, b_0)}{C(z^*, b_i)} - 1.$$

Competing against seller j replaces (b_0, p_0) by (b_j, p_j), giving

$$p_i^{\max}(z^*; p_j, b_j) = (1+p_j)\frac{C(z^*, b_j)}{C(z^*, b_i)} - 1.$$

With $C(z^*, b) = (1-z^*)^{-1/b} - 1$, these yield the stated closed forms.

As $z^* \to 1$, let $\psi \equiv (1-z^*)^{-1} \to \infty$ so $C(z^*, b) = \psi^{1/b} - 1$. If $b_i > b_0$, then $1/b_i < 1/b_0$ and $\psi^{1/b_0}/\psi^{1/b_i} \to \infty$, implying $p_i^{\max}(z^*) \to \infty$.

As $z^* \to 0$, $(1-z^*)^{-1/b} - 1 = \frac{z^*}{b} + o(z^*)$, so $p_i^{\max}(z^*) \to (1+p_0)\frac{b_i}{b_0} - 1$. □

A.5 Properties of the Value Function $V(b|z^*)$

The value function is defined as:

$$V(b|z^*) = (1+p_0)\left[(1-z^*)^{-1/b_0} - 1\right] - \left[(1-z^*)^{-1/b} - 1\right]$$

Define $\phi(z,b) \equiv (1-z)^{-1/b} - 1$ for notational convenience, so $V = (1+p_0)\phi(z^*, b_0) - \phi(z^*, b)$.

Property 1: Value of baseline data

Claim: $V(b_0|z^*) = p_0 \cdot C(z^*, b_0)$

Proof:

$$\begin{aligned} V(b_0|z^*) &= (1+p_0)\left[(1-z^*)^{-1/b_0} - 1\right] - \left[(1-z^*)^{-1/b_0} - 1\right] \\ &= p_0\left[(1-z^*)^{-1/b_0} - 1\right] \\ &= p_0 \cdot C(z^*, b_0) \end{aligned}$$

□

Property 2: Increasing in quality ($\partial V/\partial b > 0$)

Proof: Since $V = (1+p_0)\phi(z^*, b_0) - \phi(z^*, b)$, and the first term is independent of b:

$$\frac{\partial V}{\partial b} = -\frac{\partial \phi}{\partial b}$$

Let $\psi(b) = (1-z^*)^{-1/b} = \exp\left(-\frac{\ln(1-z^*)}{b}\right)$. Then:

$$\frac{\partial \psi}{\partial b} = \exp\left(-\frac{\ln(1-z^*)}{b}\right) \cdot \frac{\ln(1-z^*)}{b^2}$$

Since $z^* \in (0,1)$, we have $\ln(1-z^*) < 0$, so $\frac{\partial \psi}{\partial b} < 0$.

Therefore:

$$\frac{\partial V}{\partial b} = -\frac{\partial \psi}{\partial b} > 0$$

□

Property 3: Increasing in target capability ($\partial V/\partial z^* > 0$)

Proof: We have:

$$\frac{\partial \phi}{\partial z} = \frac{1}{b}(1-z)^{-(1+b)/b}$$

Therefore:

$$\frac{\partial V}{\partial z^*} = \frac{1+p_0}{b_0}(1-z^*)^{-(1+b_0)/b_0} - \frac{1}{b}(1-z^*)^{-(1+b)/b}$$

Define:

$$R_1(z^*) \equiv \frac{(1+p_0)b}{b_0} \cdot (1-z^*)^{\beta}$$

where $\beta = \frac{1+b}{b} - \frac{1+b_0}{b_0} = \frac{1}{b} - \frac{1}{b_0} = \frac{b_0 - b}{b \cdot b_0}$.

The condition $\frac{\partial V}{\partial z^*} > 0$ is equivalent to $R_1(z^*) > 1$.

At $z^* = 0$: $R_1(0) = \frac{(1+p_0)b}{b_0}$. For $b > b_0$ and $p_0 \geq 0$:

$$R_1(0) = \frac{(1+p_0)b}{b_0} > \frac{b}{b_0} > 1$$

Since $b > b_0$, we have $\beta < 0$, so $(1-z^*)^{\beta}$ is increasing in z^*. Therefore $R_1(z^*) > R_1(0) > 1$ for all $z^* \in (0,1)$. □

Property 4: Strategic complementarity ($\partial^2 V/\partial b\, \partial z^* > 0$)

Proof: Only the term $-\phi(z^*, b)$ depends on b, so

$$\frac{\partial^2 V}{\partial z^*\, \partial b} = -\frac{\partial^2 \phi}{\partial z\, \partial b}(z^*, b).$$

Let $u(z) \equiv -\ln(1-z)$ so that $(1-z)^{-1/b} = e^{u(z)/b}$ for $z \in (0,1)$. Then

$$\frac{\partial \phi}{\partial b}(z,b) = -\frac{u(z)}{b^2} e^{u(z)/b},$$

and since $u'(z) = \frac{1}{1-z}$,

$$\frac{\partial^2 \phi}{\partial z\, \partial b}(z,b) = -\frac{e^{u(z)/b}}{b^2(1-z)}\left(1 + \frac{u(z)}{b}\right) < 0 \qquad \text{for } z \in (0,1),\ b > 0.$$

Therefore $\frac{\partial^2 V}{\partial z^*\, \partial b} > 0$. □

Property 5: Convex in target capability ($\partial^2 V/\partial(z^*)^2 > 0$)

Proof: The second derivative of ϕ with respect to z is:

$$\frac{\partial^2 \phi}{\partial z^2} = \frac{1+b}{b^2}(1-z)^{-(1+2b)/b}$$

Therefore:

$$\frac{\partial^2 V}{\partial (z^*)^2} = \frac{(1+p_0)(1+b_0)}{b_0^2}(1-z^*)^{-(1+2b_0)/b_0} - \frac{1+b}{b^2}(1-z^*)^{-(1+2b)/b}$$

Convexity holds iff:

$$R_2(z^*) \equiv \frac{(1+p_0)(1+b_0)b^2}{(1+b)b_0^2} \cdot (1-z^*)^{\alpha} > 1$$

where $\alpha = \frac{1+2b}{b} - \frac{1+2b_0}{b_0} = \frac{1}{b} - \frac{1}{b_0} = \frac{b_0 - b}{b \cdot b_0}$.

Step 1: Show $R_2(0) > 1$.

At $z^* = 0$:

$$R_2(0) = \frac{(1+p_0)(1+b_0)b^2}{(1+b)b_0^2}$$

Setting $p_0 = 0$ (the binding case), we need to show:

$$\frac{(1+b_0)b^2}{(1+b)b_0^2} > 1$$

This is equivalent to $(1+b_0)b^2 - (1+b)b_0^2 > 0$. Factoring directly:

$$(1+b_0)b^2 - (1+b)b_0^2 = (b-b_0)(b+b_0+b\cdot b_0) > 0$$

since $b > b_0 > 0$ ensures both factors are strictly positive. Therefore $R_2(0) > 1$.

Step 2: Show $R_2(z^*)$ is increasing in z^*.

Since $b > b_0$, we have $\alpha = \frac{b_0 - b}{b \cdot b_0} < 0$. As z^* increases, $(1-z^*)$ decreases, and $(1-z^*)^{\alpha}$ with $\alpha < 0$ increases.

Conclusion: Since $R_2(0) > 1$ and $R_2(z^*)$ is strictly increasing:

$$R_2(z^*) > 1 \quad \text{for all } z^* \in [0, 1)$$

Therefore $\frac{\partial^2 V}{\partial (z^*)^2} > 0$ for all $z^* \in [0,1)$ when $b > b_0$. □

Property 6: Unbounded at frontier ($\lim_{z^* \to 1} V = +\infty$)

Proof: As $z^* \to 1$:

$$V(b|z^*) = (1+p_0)\left[(1-z^*)^{-1/b_0} - 1\right] - \left[(1-z^*)^{-1/b} - 1\right]$$

Both $(1-z^*)^{-1/b_0}$ and $(1-z^*)^{-1/b}$ diverge to $+\infty$. For $b > b_0$, we have $1/b_0 > 1/b$, so the first term dominates:

$$\frac{(1-z^*)^{-1/b_0}}{(1-z^*)^{-1/b}} = (1-z^*)^{1/b - 1/b_0} = (1-z^*)^{(b_0-b)/(b \cdot b_0)} \to +\infty$$

Therefore $V(b|z^*) \to +\infty$ as $z^* \to 1$. □

A.6 Proofs from Appendix C: General Augmentation and Market Structure

A.6.1 Proof of Proposition 12 (General Monopoly Pricing)

Proof. Revenue is $R(r,q) = rG(z^*(r,q))$. Differentiating (53) with respect to r gives

$$\frac{\partial z^*}{\partial r} = -\frac{1}{1 - \beta_1(z^*, q)} < 0.$$

Hence

$$\frac{\partial R}{\partial r} = G(z^*) + rg(z^*)\frac{\partial z^*}{\partial r} = G(z^*) - rg(z^*)\frac{1}{1-\beta_1(z^*,q)}.$$

Setting this equal to zero yields (54). □

A.6.2 Proof of Proposition 13 (Local Comparative Statics)

Proof. Differentiate (55) with respect to q along the optimal path:

$$\frac{dr^*}{dq} = (\beta_1(z^*,q) - 1)\frac{dz^*}{dq} + \beta_2(z^*,q).$$

Differentiate (56):

$$\frac{dr^*}{dq} = \left[-\beta_{11}(z^*, q)\frac{G(z^*)}{g(z^*)} + \big(1 - \beta_1(z^*, q)\big)\frac{g(z^*)^2 - G(z^*)g'(z^*)}{g(z^*)^2}\right]\frac{dz^*}{dq} - \beta_{12}(z^*, q)\frac{G(z^*)}{g(z^*)}.$$

Eliminating dr^*/dq and simplifying using $r^*(q) = \big(1 - \beta_1(z^*, q)\big)G(z^*)/g(z^*)$ yields (58). Substituting (58) back into the derivative of (55) gives (59). □

A.6.3 Proof of Proposition 14 (Envelope Formula for Capability Choice)

Proof. By the envelope theorem,

$$\frac{d}{dq}\big[r^*(q)G(z^*(q))\big] = \frac{\partial}{\partial q}\big[rG(z^*(r, q))\big]\Big|_{r=r^*(q)}.$$

Holding r fixed, (53) implies $\partial z^*/\partial q = \beta_2/(1 - \beta_1)$. Therefore

$$\frac{d}{dq}\big[r^*(q)G(z^*(q))\big] = r^*(q)g(z^*(q))\frac{\beta_2(z^*(q), q)}{1 - \beta_1(z^*(q), q)}.$$

Subtracting $C'(q)$ gives (60). □

A.6.4 Proof of Lemma 18 (Monotonicity of the Gain Differential)

Proof. For $z < q_2 < q_1$, differentiate (62) with respect to z:

$$\Gamma'_{12}(z) = \beta_1(z, q_1) - \beta_1(z, q_2).$$

Applying the fundamental theorem of calculus to the function $t \mapsto \beta_1(z, t)$ on $[q_2, q_1]$ yields

$$\beta_1(z, q_1) - \beta_1(z, q_2) = \int_{q_2}^{q_1} \beta_{12}(z, t)\, dt.$$

Hence the sign of Γ'_{12} is the sign of β_{12} integrated over $[q_2, q_1]$. □

A.6.5 Proof of Proposition 15

Fix $q_1 > q_2 > 0$.

Part (i). Let $z \in (q_2, q_1]$. Since the worker's own capability already exceeds q_2, the weaker model does not raise output, so $\beta(z, q_2) = z$. Hence adopting model 2 at any price $r_2 \geq 0$ yields payoff

$$\beta(z, q_2) - r_2 = z - r_2 \leq z,$$

which is weakly dominated by non-adoption. Therefore model 2 is never chosen.

Part (ii). For $z < q_2$, define

$$\Gamma_{12}(z) \equiv \beta(z, q_1) - \beta(z, q_2).$$

For any $0 \leq z < z' < q_2$,

$$\Gamma_{12}(z') - \Gamma_{12}(z) = \int_z^{z'} \big[\beta_1(t, q_1) - \beta_1(t, q_2)\big]\, dt.$$

Since $q_1 > q_2$ and $\beta_{12} \leq 0$ on $\mathcal{D}$, the mapping $q \mapsto \beta_1(t, q)$ is weakly decreasing for every t, so

$$\beta_1(t, q_1) - \beta_1(t, q_2) \leq 0.$$

Hence $\Gamma_{12}(z') \leq \Gamma_{12}(z)$, i.e. Γ_{12} is weakly decreasing on $[0, q_2)$. Therefore its minimum is attained at the top of the interval:

$$\Gamma_{12}(z) \geq \Gamma_{12}(q_2^-) \qquad \forall z < q_2.$$

Part (iii). Suppose the follower prices at $r_2 = c$ and the leader chooses r_1 with

$$r_1 - r_2 < \Gamma_{12}(q_2^-).$$

Then for every $z < q_2$,

$$\big[\beta(z, q_1) - r_1\big] - \big[\beta(z, q_2) - r_2\big] = \Gamma_{12}(z) - (r_1 - r_2) \geq \Gamma_{12}(q_2^-) - (r_1 - r_2) > 0.$$

So every low-skill worker who would ever choose model 2 strictly prefers model 1.

By part (i), workers in $(q_2, q_1]$ never choose model 2. Workers above q_1 receive no gain from either model and therefore are irrelevant for follower demand. Hence model 2 has zero demand and zero operating profit.

Finally, let r_1^m denote the monopoly price of model 1 when it faces no rival. Then any price

$$r_1^* = \min\{r_1^m,\ c + \Gamma_{12}(q_2^-) - \varepsilon\}$$

with $\varepsilon \in (0, \Gamma_{12}(q_2^-))$ is feasible for the leader and still excludes the follower. □

A.6.6 Proof of Proposition 16

Fix $q_H > q_W > 0$ and define

$$\Gamma_{HW}(z) \equiv \beta(z, q_H) - \beta(z, q_W), \qquad z \in [0, q_W).$$

For any $z \in [0, q_W)$,

$$\Gamma'_{HW}(z) = \beta_1(z, q_H) - \beta_1(z, q_W).$$

Because $q_H > q_W$ and $\beta_{12} > 0$ on $\mathcal{D}$, the mapping $q \mapsto \beta_1(z, q)$ is strictly increasing, so $\Gamma'_{HW}(z) > 0$. Hence Γ_{HW} is strictly increasing on $[0, q_W)$.

Fix any price gap

$$d \in \big(\Gamma_{HW}(0), \Gamma_{HW}(q_W^-)\big).$$

By continuity and strict monotonicity of Γ_{HW}, there exists a unique $\tilde{z} \in (0, q_W)$ such that $\Gamma_{HW}(\tilde{z}) = d$.

Now suppose $r_H - r_W = d$. For any $z < q_W$,

$$\big[\beta(z, q_H) - r_H\big] - \big[\beta(z, q_W) - r_W\big] = \Gamma_{HW}(z) - d.$$

Thus this difference is negative for $z < \tilde{z}$ and positive for $z > \tilde{z}$, yielding the stated sorting pattern. Positive demand for each option additionally requires the corresponding participation constraint against non-adoption. □

A.6.7 Proof of Proposition 17

By construction, the capability-price options available to workers in the accommodated-entry outcome and in the replication outcome are identical:

$$\{(q_H, r_H^A)\} \cup S_E^A.$$

Therefore every worker faces the same utility menu in both cases, and the induced allocation across capability-price options is unchanged up to a measure-zero set of indifferences. This proves part (i).

For part (ii), costless degradation implies that once the incumbent has paid for capability q_H, it incurs no additional fixed training cost from also offering the weaker models in S_E^A. Hence in the replication outcome the incumbent receives operating profit from both its original flagship sales and the sales formerly captured by the entrant:

$$\Pi_I^{\text{rep}} = (r_H^A - c)D_H^A + \sum_{j=1}^{m}(r_j^A - c)D_j^A.$$

By contrast, in the accommodated-entry outcome,

$$\Pi_I^{\text{acc}} = (r_H^A - c)D_H^A.$$

Subtracting yields

$$\Pi_I^{\text{rep}} - \Pi_I^{\text{acc}} = \sum_{j=1}^{m}(r_j^A - c)D_j^A = \pi_E^{\text{op},A},$$

which is (64). This proves part (ii).

Part (iii) follows immediately: the model-family strategy weakly dominates accommodation, and strictly dominates it whenever the accommodated entrant earns strictly positive operating profit. □

A.6.8 Proof of Proposition 18

Fix an entrant portfolio S_E with frontier capability $q_E \in (q_W, q_H]$ and define

$$\pi_E^{\mathrm{op}}(S_E) = \sum_{(q,r)\in S_E} (r-c) D_{(q,r)}.$$

Consider first an entrant option (q, r) with $q \le q_W$. Because the incumbent offers (q_W, c) and $\beta(z, q_W) \ge \beta(z, q)$ for every worker z, any worker who chooses (q, r) must have $r \le c$. Hence such an option contributes weakly negatively to operating profit. Therefore only options with $q > q_W$ can contribute positively.

Now fix any entrant option (q, r) with positive demand and $q \in (q_W, q_E]$. Let z be a worker who chooses this option. Since the worker weakly prefers it to the incumbent weak model (q_W, c),

$$\beta(z, q) - r \ge \beta(z, q_W) - c,$$

so

$$r - c \le \beta(z, q) - \beta(z, q_W) = \int_{q_W}^{q} \beta_2(z, s)\, ds.$$

Because $q \le q_E \le q_H$ and by Assumption 12,

$$r - c \le \int_{q_W}^{q} L_H\, ds \le L_H(q_E - q_W).$$

Therefore every entrant option with positive markup satisfies

$$(r-c) D_{(q,r)} \le L_H(q_E - q_W) D_{(q,r)}.$$

Summing over all entrant options with positive markup, and noting that the total mass of workers served by the entrant is at most one, gives

$$\pi_E^{\mathrm{op}}(S_E) \le L_H(q_E - q_W) \sum_{(q,r)\in S_E} D_{(q,r)} \le L_H(q_E - q_W).$$

This is exactly (65). □

A.6.9 Proof of Corollary 2

Fix any entrant portfolio with frontier capability $q_E \in (q_W, q_H]$. By Proposition 18,

$$\pi_E^{\mathrm{op}}(S_E) \le L_H(q_E - q_W).$$

Total profit is operating profit minus the training cost of the frontier model:

$$\pi_E(S_E) \le L_H(q_E - q_W) - C(q_E).$$

By convexity of C,

$$C(q_E) \ge C(q_W) + C'_+(q_W)(q_E - q_W).$$

Hence

$$\pi_E(S_E) \le L_H(q_E - q_W) - C(q_W) - C'_+(q_W)(q_E - q_W).$$

If $C'_+(q_W) > L_H$, then the right-hand side is strictly negative for every $q_E > q_W$, since $C(q_W) \ge 0$. Therefore every entrant portfolio with frontier capability $q_E \in (q_W, q_H]$ earns strictly negative total profit. □

A.6.10 Proof of Corollary 3

Fix $q_H > 0$. Since C is strictly increasing and $C(0) = 0$, we have $C(q_H) > 0$. By continuity of C, there exists $\bar{q}_W < q_H$ such that for every $q_W \in (\bar{q}_W, q_H)$,

$$C(q_W) > L_H(q_H - q_W).$$

Now fix such a q_W and consider any entrant portfolio with frontier capability $q_E \in (q_W, q_H]$. By Proposition 18,

$$\pi_E^{\mathrm{op}}(S_E) \le L_H(q_E - q_W) \le L_H(q_H - q_W).$$

Since C is increasing,

$$C(q_E) \ge C(q_W).$$

Hence total entrant profit satisfies

$$\pi_E(S_E) \leq L_H(q_H - q_W) - C(q_W) < 0.$$

Therefore every entrant portfolio with frontier capability $q_E \in (q_W, q_H]$ is deterred. □

A.6.11 Proof of Proposition 19

Fix any entrant portfolio S_E. Consider an entrant option (q, r) that receives positive demand from some worker of type z. Since that worker prefers adoption to non-adoption,

$$\beta(z, q) - r \geq z.$$

Therefore

$$r \leq \beta(z, q) - z \leq 1,$$

because $\beta(z, q) \leq 1$ on the normalized domain. It follows that

$$r - c \leq 1 - c.$$

Hence every sold entrant option has markup at most $(1 - c)_+$. Since the total mass of workers served by the entrant is at most one,

$$\pi_E^{\text{op}}(S_E) = \sum_{(q,r)\in S_E} (r - c) D_{(q,r)} \leq (1 - c)_+ \sum_{(q,r)\in S_E} D_{(q,r)} \leq (1 - c)_+,$$

which proves (67).

Now suppose $q_E > q_H$ and $C(q_H) > (1 - c)_+$. Since C is strictly increasing,

$$C(q_E) \geq C(q_H) > (1 - c)_+.$$

Therefore total entrant profit satisfies

$$\pi_E(S_E) \leq (1 - c)_+ - C(q_E) < 0.$$

So every leapfrogging entrant portfolio is unprofitable. □

A.6.12 Proof of Theorem C.1

Fix $q_H > 0$ such that $C(q_H) > (1-c)_+$. By Corollary 3, there exists $q_W < q_H$ such that every entrant portfolio with frontier capability $q_E \in (q_W, q_H]$ earns strictly negative profit when the incumbent offers the weak model (q_W, c).

Now consider any entrant portfolio.

If its frontier satisfies $q_E \in (q_W, q_H]$, it is unprofitable by Corollary 3.

If instead $q_E > q_H$, then Proposition 19 implies

$$\pi_E(S_E) \leq (1-c)_+ - C(q_E) < (1-c)_+ - C(q_H) < 0.$$

Thus no entrant portfolio is profitable once the incumbent offers the two-model menu

$$\{(q_H, r_H), (q_W, c)\}.$$

Therefore the incumbent can sustain the winner-take-all market structure through a model-family entry-deterrence strategy. □

A.7 Proofs from Appendix D: Costly Screening and Verification

A.7.1 Proof of Proposition 20 (Conditional Value of Market Screening)

Proof. By Lemma 19(ii), any blind purchase from an unscreened seller yields expected surplus 0, so the no-screening outside option is normalized to 0.

Suppose the buyer pays K and learns all realized qualities. By Lemma 19(iii), any low-quality seller yields negative realized surplus, so low-quality datasets are never purchased after screening. By Lemma 19(iv), a high-quality seller with signal L yields surplus $\Delta_L = (1-\mu_L)\Delta V$, while a high-quality seller with signal H yields the lower surplus $\Delta_H = (1-\mu_H)\Delta V$.

Therefore the optimal screened purchase rule is: (a) if at least one seller in the L-signal group is high quality, buy one such seller and obtain Δ_L; (b) otherwise, if at least one seller in the H-signal group is high quality, buy one such seller and obtain Δ_H; (c) otherwise do not buy and obtain 0.

Because qualities are conditionally independent within each signal group, the probability of at least one hidden gem among the n_L pessimistic-signal sellers is $1-(1-\mu_L)^{n_L}$. The probability that no hidden gem is present but at least one optimistic-signal seller is high quality equals $(1-\mu_L)^{n_L}[1-(1-\mu_H)^{n_H}]$. Multiplying the corresponding surpluses yields (69). Screening is optimal exactly when its gross value exceeds the fixed cost, i.e. $K < S(n_L, n_H)$. The hidden-gem claim follows because $\Delta_L > \Delta_H$. □

A.7.2 Proof of Corollary 4 (Ex Ante Gross Value of Screening)

Proof. Let A_L denote the event that at least one hidden gem (b_H, L) is present. Because each seller is a hidden gem with probability $q_L = \mu_0(1-\gamma)$, we have $\Pr(A_L) = 1-(1-q_L)^I$. Let A_H denote the event that no hidden gem is present but at least one seller is high quality. Since each seller is high quality with probability μ_0, the event that all sellers are low quality has probability $(1-\mu_0)^I$. Therefore $\Pr(A_H) = (1-q_L)^I - (1-\mu_0)^I$. On A_L the buyer obtains Δ_L after screening, and on A_H the buyer obtains Δ_H. If all sellers are low quality the screened payoff is 0. Taking expectations gives (70). □

A.7.3 Proof of Proposition 21 (Comparative Statics of Screening Value)

Proof. Using (70),

$$\begin{aligned}\bar{S}_{I+1} - \bar{S}_I &= \Delta_L\big[(1-q_L)^I - (1-q_L)^{I+1}\big] + \Delta_H\big[(1-q_L)^{I+1} - (1-q_L)^I - (1-\mu_0)^{I+1} + (1-\mu_0)^I\big] \\ &= q_L(\Delta_L - \Delta_H)(1-q_L)^I + \mu_0\,\Delta_H\,(1-\mu_0)^I > 0,\end{aligned}$$

which proves (i). For (ii), use Lemma 19(iv): $\Delta_L = (1-\mu_L)\Delta V$ and $\Delta_H = (1-\mu_H)\Delta V$. The coefficients depend on (μ_0, γ) but not on ΔV, so $\bar{S}_I$ is affine in ΔV with a strictly positive slope. For (iii), as $\gamma \to 1$ we have $q_L = \mu_0(1-\gamma) \to 0$ and $\Delta_H = (1-\mu_H)\Delta V \to 0$. Hence both terms in (70) converge to zero. □

A.7.4 Proof of Corollary 5 (Self-Financing Screening)

Proof. By Proposition 20, $S(n_L, n_H) = a(n_L, n_H; \mu_0, \gamma)\,\Delta V(z^*)$, where $a(n_L, n_H; \mu_0, \gamma) = (1-\mu_L)[1-(1-\mu_L)^{n_L}] + (1-\mu_L)^{n_L}(1-\mu_H)[1-(1-\mu_H)^{n_H}]$. If $n_L + n_H \geq 1$, then $a > 0$. Since

$\Delta V(z^*) = (1-z^*)^{-1/b_L} - (1-z^*)^{-1/b_H} \to \infty$ as $z^* \to 1$, the gross screening value diverges. Therefore, for any fixed $K > 0$, $K < S(n_L, n_H)$ for all z^* sufficiently close to 1. □

A.7.5 Proof of Proposition 22 (Expected Payoff from Partial Verification)

Proof. If at least one of the k_L verified pessimistic-signal sellers is high quality, the buyer purchases such a seller and obtains realized surplus Δ_L. This event occurs with probability $1-(1-\mu_L)^{k_L}$. If no verified pessimistic-signal seller is high quality, but at least one of the k_H verified optimistic-signal sellers is high quality, then the buyer purchases such a seller and obtains realized surplus Δ_H. Because conditional qualities are independent across groups, this event has probability $(1-\mu_L)^{k_L}[1-(1-\mu_H)^{k_H}]$. If all verified sellers are low quality, the buyer's best outside option is to refrain from purchase or buy blindly, both of which yield payoff 0 by Lemma 19(ii). Subtracting the verification cost $\kappa(k_L+k_H)$ yields (72). □

A.7.6 Proof of Lemma 20 (Marginal Values of Additional Verification)

Proof. Starting from (72),

$$\begin{aligned}\Pi(k_L+1,k_H)-\Pi(k_L,k_H) &= \Delta_L\,\mu_L(1-\mu_L)^{k_L} - \Delta_H\,\mu_L(1-\mu_L)^{k_L}\big[1-(1-\mu_H)^{k_H}\big]-\kappa \\ &= (1-\mu_L)^{k_L}\mu_L\big[(1-\mu_L)-(1-\mu_H)\big(1-(1-\mu_H)^{k_H}\big)\big]\Delta V-\kappa \\ &= (1-\mu_L)^{k_L}\mu_L\big[\mu_H-\mu_L+(1-\mu_H)^{k_H+1}\big]\Delta V-\kappa,\end{aligned}$$

which proves (73). Since $(1-\mu_H)^{k_H+1} \le 1-\mu_H$, the bracket is at most $1-\mu_L$, which gives the bound $\sigma_L\Delta V-\kappa$.

For (74),

$$\Pi(k_L,k_H+1)-\Pi(k_L,k_H) = \Delta_H(1-\mu_L)^{k_L}(1-\mu_H)^{k_H}\mu_H-\kappa = (1-\mu_L)^{k_L}(1-\mu_H)^{k_H}\sigma_H\Delta V-\kappa,$$

which implies the bound $\sigma_H\Delta V-\kappa$ since the multiplicative factors are at most one. □

A.7.7 Proof of Proposition 23 (Threshold for Profitable Verification)

Proof. If $\kappa < \max_{s\in A}\sigma_s\Delta V$, choose one seller from a signal group attaining the maximum. By Proposition 22, $\Pi(1,0)=\sigma_L\Delta V-\kappa$ and $\Pi(0,1)=\sigma_H\Delta V-\kappa$, so at least one of these payoffs is

strictly positive.

Conversely, suppose $\kappa \geq \max_{s \in A} \sigma_s \Delta V$. Start from $(k_L, k_H) = (0, 0)$, where $\Pi(0, 0) = 0$, and add verified sellers one at a time. By Lemma 20, every added L-verification changes payoff by at most $\sigma_L \Delta V - \kappa \leq 0$, and every added H-verification changes payoff by at most $\sigma_H \Delta V - \kappa \leq 0$. Therefore no sequence of added inspections can produce a strictly positive payoff. □

A.7.8 Proof of Lemma 21 (Comparison of Residual Uncertainty)

Proof. Substituting the Bayes posteriors (68) into the residual uncertainty terms yields

$$\sigma_H = \frac{\gamma(1-\gamma)\mu_0(1-\mu_0)}{[\gamma\mu_0 + (1-\gamma)(1-\mu_0)]^2}, \qquad \sigma_L = \frac{\gamma(1-\gamma)\mu_0(1-\mu_0)}{[(1-\gamma)\mu_0 + \gamma(1-\mu_0)]^2}.$$

Taking the difference and simplifying:

$$\sigma_H - \sigma_L = \frac{\gamma(1-\gamma)\mu_0(1-\mu_0)(2\gamma-1)(1-2\mu_0)}{[\gamma\mu_0 + (1-\gamma)(1-\mu_0)]^2[(1-\gamma)\mu_0 + \gamma(1-\mu_0)]^2}.$$

Every factor in the numerator except $(1 - 2\mu_0)$ is strictly positive for $\gamma \in (1/2, 1)$ and $\mu_0 \in (0, 1)$, and the denominator is strictly positive, so $\operatorname{sign}(\sigma_H - \sigma_L) = \operatorname{sign}(1 - 2\mu_0)$. □

A.7.9 Proof of Proposition 24 (Pure Cherry-Picking Region)

Proof. By (74) in Lemma 20, for any feasible (k_L, k_H),

$$\Pi(k_L, k_H + 1) - \Pi(k_L, k_H) = (1-\mu_L)^{k_L}(1-\mu_H)^{k_H}\sigma_H \Delta V - \kappa \leq \sigma_H \Delta V - \kappa \leq 0.$$

Thus verifying any optimistic-signal seller is never profitable once condition (75) holds. Hence every optimizer must satisfy $k_H^* = 0$.

With $k_H = 0$, Proposition 22 reduces to $\Pi(k_L, 0) = \Delta_L[1 - (1-\mu_L)^{k_L}] - \kappa k_L$. The marginal gain from the kth pessimistic-signal verification is $\Pi(k, 0) - \Pi(k-1, 0) = (1-\mu_L)^{k-1}\sigma_L \Delta V - \kappa$. Because this sequence is strictly decreasing in k, the payoff is maximized by taking the largest $k \leq n_L$ for which the marginal gain is still nonnegative, which is exactly (76). Finally, $\kappa < \sigma_L \Delta V$ implies $\Pi(1, 0) > 0$, so $k_L^* \geq 1$. □

B Fixed-Point Formulation and Existence of Two-Layer Equilibrium

To formalize the general equilibrium characterized in Section 5.2.4, we cast the interaction between downstream capability choice and upstream data pricing as a fixed-point problem. Let $C(z,b) \equiv (1-z)^{-1/b} - 1$ for $z \in [0,1)$ and $b \in [\underline{b}, \bar{b}]$. Let

$$TR(z) \equiv \max_{r \in [0,z]} r\, G(z-r), \qquad z \in [0,1),$$

and fix an upper capability bound $\bar{z} \in (0,1)$.

Given an anticipated downstream target $z \in [0,\bar{z}]$, let $p_i^*(\cdot; z)$ denote the upstream equilibrium pricing rule of seller i (as characterized in Section 5.3.3), and let dataset $i = 0$ denote the baseline waste data option characterized by constants (b_0, p_0), where $p_0 \geq 0$ and $b_0 \in [\underline{b}, \bar{b}]$.

For each realization $\omega = (b_i, c_i)_{i=1}^I$, define the minimal training expenditure required to reach capability $\hat{z}$, conditional on upstream prices being anchored by target z:

$$\kappa(\hat{z}; z, \omega) \equiv \min\Big\{(1+p_0)C(\hat{z}, b_0),\ \min_{i \in \{1,\dots,I\}} \big(1 + p_i^*(c_i; z)\big)\, C(\hat{z}, b_i)\Big\}. \tag{48}$$

Define the expected effective cost as

$$\mathcal{E}(\hat{z}; z) \equiv \mathbb{E}_\omega\big[\kappa(\hat{z}; z, \omega)\big], \qquad (\hat{z}, z) \in [0,\bar{z}]^2. \tag{49}$$

Assumption 1 (Regularity).

(i) TR is continuous on $[0,\bar{z}]$.

(ii) For each i, $(c,z) \mapsto p_i^*(c;z)$ is continuous on $[\underline{c}, \bar{c}] \times [0,\bar{z}]$ and bounded, with $p_i^*(c;z) \geq c$ (pointwise).

(iii) (b_i, c_i) has bounded support $[\underline{b}, \bar{b}] \times [\underline{c}, \bar{c}]$ with $\underline{c} \geq 0$, and admits a density.

Part (i) is standard and follows from Berge's Maximum Theorem under the log-concavity of G. Part (ii) represents our core maintained hypothesis: it treats the upstream equilibrium pricing rule as an input to the fixed-point formulation, rather than deriving its existence and regularity

solely from primitives. We discuss which special cases satisfy this condition, and which remain open, in Remark B below. Part (iii) is a standard distributional assumption.

Lemma 15 (Continuity of $\mathcal{E}$)**.** *Under Assumption 1, $\mathcal{E}$ is continuous on $[0,\bar{z}]^2$. Moreover,*

$$\mathcal{E}(\hat{z};z) \le (1+p_0)C(\hat{z},b_0) \qquad \forall(\hat{z},z)\in[0,\bar{z}]^2.$$

Proof. See Appendix A.3.1.

Define the buyer's net payoff

$$U(\hat{z};z) \equiv sTR(\hat{z}) - \mathcal{E}(\hat{z};z), \qquad (\hat{z},z)\in[0,\bar{z}]^2, \tag{50}$$

and the (set-valued) best response correspondence

$$\Phi(z) \equiv \arg\max_{\hat{z}\in[0,\bar{z}]} U(\hat{z};z). \tag{51}$$

Assumption 2 (Unique maximizer in $\hat{z}$). For each $z\in[0,\bar{z}]$, the function $\hat{z}\mapsto U(\hat{z};z)$ has a unique global maximizer on $[0,\bar{z}]$.

In the baseline case ($I=0$), Proposition 10 derives this assumption directly from primitives. In the data-market extension ($I\ge 1$), Proposition 11 provides primitive sufficient conditions under a no-switching restriction; absent that restriction, we maintain the assumption as a reduced-form regularity condition.

Assumption 3 (Viability / interiority).

(i) $\exists\,\tilde{z}\in(0,\bar{z})$ such that $sTR(\tilde{z}) - (1+p_0)C(\tilde{z},b_0) > 0$.

(ii) $C(\bar{z},\bar{b}) > s\sup_{z\in[0,\bar{z}]} TR(z)$.

Lemma 16 (Kakutani conditions)**.** *Under Assumptions 1–3, $\Phi:[0,\bar{z}]\rightrightarrows[0,\bar{z}]$ satisfies:*

(i) $\Phi(z)$ is nonempty and compact for all z;

(ii) $\Phi(z)$ is a singleton for all z, hence convex;

(iii) Φ has a closed graph (equivalently, is upper hemicontinuous);

(iv) $0\notin\Phi(z)$ and $\bar{z}\notin\Phi(z)$ for all z.

Proof. See Appendix A.3.2.

Theorem B.1 (Conditional existence of two-layer equilibrium)**.** *Under Assumptions 1–3—in particular, given an upstream equilibrium pricing correspondence satisfying continuity and boundedness, and given a single-valued downstream best response—there exists* $z^* \in (0, \bar{z})$ *such that* $z^* \in \Phi(z^*)$*.*

Proof. See Appendix A.3.3.

Remark (Scope of the existence result)*.* Theorem B.1 is a conditional fixed-point result: it takes the upstream equilibrium pricing rule as an input and establishes that a consistent downstream capability target exists. We clarify the status of each maintained hypothesis to precisely delineate model primitives from requisite regularity conditions.

Continuity of TR (Assumption 1(i)) follows from Berge's Maximum Theorem applied to the monopolist's pricing problem under the log-concavity of G. The upper bound $\mathcal{E}(\hat{z}; z) \leq (1 + p_0)C(\hat{z}, b_0)$ (Lemma 15) is a direct consequence of the baseline outside option. The viability condition (Assumption 3(i)) ensures that AI production is strictly profitable at some interior capability level, and the cost-dominance condition (Assumption 3(ii)) guarantees that the chosen upper bound $\bar{z}$ is prohibitively expensive; both are parametric inequalities that can be checked for any given primitives (G, b_0, p_0, s).

In the monopoly seller case ($I = 1$), the scaling law yields a bounded feasible price interval and continuity of the seller's objective. Monotonicity and continuity of an optimal pricing rule then follow if the seller's problem satisfies single-crossing in (p, c) and has a unique maximizer for each (c, z); Proposition 9 makes these extra conditions explicit. In the competitive case ($I \geq 2$), the fixed-point theorem continues to treat the upstream pricing rule as an input satisfying Assumption 1(ii); Proposition 6(iii) establishes only the limit-pricing comparative statics under standard auction-theoretic regularity.

In the baseline case ($I = 0$, no strategic data market), Proposition 10 verifies Assumption 2 from primitives, making the best-response $\Phi(z)$ a singleton; Corollary 1 then provides unconditional existence for this case. In the data-market extension ($I \geq 1$), source switching makes $\mathcal{E}$ the expected lower envelope of convex cost curves, which can introduce downward kinks in the realized effective cost. Proposition 11 identifies a primitive sufficient condition—no-switch dominance—under which Assumption 2 again follows from primitives. Absent that restriction, Assumption 2

remains the residual maintained regularity condition on the producer's downstream problem.

Consequently, our existence result operates on two tiers. In the baseline case ($I = 0$), all hypotheses are verified from primitives and Corollary 1 provides unconditional existence. In the general data-market extension ($I \geq 1$), Assumption 2 and the upstream pricing rule serve as maintained inputs to Theorem B.1; Proposition 9 verifies the latter in the monopoly case, while Proposition 11 verifies the former under a no-switching restriction.

We now detail how Assumption 1(ii) can be verified in the monopoly case. Boundedness and continuity of the seller's objective arise from the primitives; monotonicity and continuity of the optimal pricing rule require additional selection regularity, which we state explicitly below. We then turn to Assumption 2 and provide primitive sufficient conditions in the baseline case and, under a no-switching restriction, in the data-market case.

Assumption 4 (Monotone likelihood ratio). The conditional density $\mu(b \mid c)$ satisfies the strict monotone likelihood ratio property (MLRP) in (b, c): for all $c' > c$ and $b' > b$,

$$\mu(b' \mid c')\,\mu(b \mid c) > \mu(b' \mid c)\,\mu(b \mid c').$$

Assumption 4 is a standard primitive condition in mechanism design. In our model, it indicates that higher production costs are a more informative signal of higher data quality—a natural property given the cost-quality relationship $c_i = c(b_i) + \varepsilon_i$ with $c'(b) > 0$. It is satisfied whenever the noise ε_i has a log-concave density (e.g., normal, logistic, or uniform) and $c(\cdot)$ is strictly increasing. MLRP is economically natural, but it is not by itself enough for the fixed-point theorem: to obtain a continuous optimal pricing rule, we also need single-crossing and single-valuedness of the seller's pricing problem. The next proposition makes those requirements explicit.

Assumption 5 (Regular compute-weighted acceptance tail). For each (c, z) with $c < \bar{p}(z) \equiv p^{\max}(z; \bar{b})$, the function

$$Q(p; c, z) \equiv \int_{\underline{b}(p,z)}^{\bar{b}} C(z, b)\, d\mu(b \mid c)$$

is log-concave in p on $[c, \bar{p}(z)]$.

Proposition 9 (Upstream pricing regularity)**.** *Under Assumption 1(iii), Assumption 4, and the scaling law*

$$C(z, b) = (1 - z)^{-1/b} - 1, \qquad \bar{z} < 1,$$

let

$$\Pi(p; c, z) = (p - c)\, Q(p; c, z), \qquad Q(p; c, z) \equiv \int_{\underline{b}(p,z)}^{\bar{b}} C(z, b)\, d\mu(b \mid c),$$

and write $\bar{p}(z) \equiv p^{\max}(z; \bar{b})$ *and* $\bar{p} \equiv \sup_{z \in [0, \bar{z}]} \bar{p}(z) < \infty$.

Then:

(a) For each fixed z*, the seller's objective* $\Pi(p; c, z)$ *satisfies the single-crossing property in* (p, c)*. Consequently, there exists an optimal selection* $p^*(\cdot; z)$ *that is weakly increasing in* c*. Moreover, every optimal price satisfies*

$$p^*(c; z) \in [c, \max\{c, \bar{p}(z)\}] \subseteq [\underline{c}, \max\{\bar{c}, \bar{p}\}],$$

so the pricing rule is bounded and satisfies $p^*(c; z) \geq c$*.*

(b) If Assumption 5 also holds, then for every (c, z) *the seller's pricing problem has a unique maximizer. Hence the optimal pricing rule* $p^*(c; z)$ *is continuous in* (c, z) *on* $[\underline{c}, \bar{c}] \times [0, \bar{z}]$*.*

Proof. See Appendix A.3.4.

Remark. Part (a) isolates the role of MLRP. For a fixed capability target z, the seller is effectively pricing against the buyer's reservation value for the data. A higher observed cost signal c shifts the posterior distribution of that reservation value upward in the likelihood-ratio order, so higher posted prices become relatively more attractive. This delivers monotone pricing.

Part (b) adds a regularity condition on the seller's residual demand. The object $Q(p; c, z)$ is the compute-weighted expected sales volume that survives the buyer's acceptance rule at price p. Log-concavity of $Q(\cdot; c, z)$ implies that as the seller raises the price, the set of quality states that still support trade shrinks in a regular, single-peaked manner. Equivalently, the proportional loss in compute-weighted demand from a marginal price increase rises with the price level. This rules out multiple local optima and yields a unique, continuous pricing rule.

In the monopoly seller case ($I = 1$), Proposition 9 shows that MLRP is sufficient for monotone pricing, while continuity of the pricing rule follows under the additional regularity condition in Assumption 5.

Next, we turn to Assumption 2. The first observation is that quasi-concavity is not the right primitive property: once viability holds, the producer's payoff is generically negative near the

origin but positive at some interior point. The economically relevant requirement is instead that the downstream objective be strictly single-peaked, ensuring a unique best response.

Remark (Why quasi-concavity is too strong). Under our maintained primitives, standard quasi-concavity of the objective $\hat{z} \mapsto U(\hat{z}; z)$ structurally conflicts with viability. Specifically,

$$U(0; z) = 0 \qquad \forall z \in [0, \bar{z}].$$

Moreover, for each realized menu ω,

$$\kappa(\hat{z}; z, \omega) = \hat{z} \min\Big\{ \frac{1 + p_0}{b_0}, \min_{i \in \{1,\dots,I\}} \frac{1 + p_i^*(c_i; z)}{b_i} \Big\} + o(\hat{z}) \qquad \text{as } \hat{z} \downarrow 0,$$

so

$$\mathcal{E}(\hat{z}; z) = \hat{z}\, \mathbb{E}_\omega \left[\min\Big\{ \frac{1 + p_0}{b_0}, \min_{i \in \{1,\dots,I\}} \frac{1 + p_i^*(c_i; z)}{b_i} \Big\} \right] + o(\hat{z})$$

and therefore

$$U'_+(0; z) = -\mathbb{E}_\omega \left[\min\Big\{ \frac{1 + p_0}{b_0}, \min_{i \in \{1,\dots,I\}} \frac{1 + p_i^*(c_i; z)}{b_i} \Big\} \right] < 0.$$

Consequently, $U(\hat{z}; z) < 0$ for all sufficiently small $\hat{z} > 0$. On the other hand, Lemma 15 and Assumption 3(i) imply that there exists $\tilde{z} \in (0, \bar{z})$ such that

$$U(\tilde{z}; z) \geq sTR(\tilde{z}) - (1 + p_0) C(\tilde{z}, b_0) > 0.$$

So $U(\cdot; z)$ is negative near the origin but positive at some interior point, which makes upper contour sets near level 0 disconnected. This is why Assumption 2 is stated as a unique-maximizer condition rather than as quasi-concavity.

The key structural properties for uniqueness are the monotonicity of downstream curvature and the marginal cost explosion near the frontier.

Lemma 17 (Marginal cost explosion)**.** *Under Assumption 1, $\hat{z} \mapsto \mathcal{E}(\hat{z}; z)$ is Lipschitz on $[0, \bar{z}]$ and differentiable a.e. for each z, with*

$$\mathcal{E}'(\hat{z}; z) \geq C'(\hat{z}, \bar{b}) = \frac{1}{\bar{b}} (1 - \hat{z})^{-(1/\bar{b} + 1)} \qquad \text{a.e.} \tag{52}$$

In particular, $\mathcal{E}'(\hat{z}; z) \to +\infty$ as $\hat{z} \uparrow \bar{z}$.

Proof. See Appendix A.3.5.

Proposition 10 (Primitive sufficient condition for a unique interior maximizer: baseline case)**.** *Suppose $I = 0$ and maintain Assumption 3, so that*

$$U(\hat{z}) = sTR(\hat{z}) - (1 + p_0)C(\hat{z}, b_0).$$

Let $r^*(\hat{z})$ solve the downstream pricing problem and define the implied worker cutoff

$$q(\hat{z}) \equiv \hat{z} - r^*(\hat{z}).$$

Write

$$\psi(q) \equiv \frac{G(q)}{g(q)}, \qquad M(q) \equiv \frac{g(q)}{1 + \psi'(q)}.$$

Assume:

(i) G has a C^2 density $g > 0$ on $(0, 1)$ and is log-concave;

(ii) M is nonincreasing on $[0, \bar{z}]$;

(iii) $(1 + p_0)C'(\bar{z}, b_0) > s$.

A convenient primitive sufficient condition for part (ii) is that g is nonincreasing and $\psi = G/g$ is convex. Then U has a unique global maximizer $\hat{z}^* \in (0, \bar{z})$. In particular, for every $z \in [0, \bar{z}]$,

$$\Phi(z) = \{\hat{z}^*\},$$

so Assumption 2 holds in the baseline case.

Proof. See Appendix A.3.6.

Corollary 1 (Unconditional existence: baseline case)**.** *Under $I = 0$, Assumption 3, and the primitive conditions in Proposition 10, there exists $z^* \in (0, \bar{z})$ such that $z^* \in \Phi(z^*)$. For the uniform distribution $G(z) = z$, condition (ii) holds with $M(q) = 1/2$, so existence is unconditional once Assumption 3 is satisfied.*

Proof. With $I = 0$, the upstream pricing rule is trivially the constant (p_0, b_0), so Assumption 1 is automatically satisfied. Proposition 10 shows that $\Phi(z)$ is a nonempty compact singleton for

each z. Lemma 16(iii)–(iv) hold under Assumptions 1 and 3. Kakutani's fixed-point theorem then applies exactly as in the proof of Theorem B.1. □

Assumption 6 (No-switching dominance). For almost every realized menu $\omega = (b_i, c_i)_{i=1}^I$ and every current capability z, assume that a single data source weakly dominates all alternatives across the entire capability domain:

$$\mathcal{C}_0(\hat{z}; z, \omega) \equiv (1 + p_0)C(\hat{z}, b_0), \qquad \mathcal{C}_i(\hat{z}; z, \omega) \equiv (1 + p_i^*(c_i; z))C(\hat{z}, b_i), \quad i = 1, \ldots, I.$$

Specifically, there exists an index $j = j(\omega, z) \in \{0, 1, \ldots, I\}$ such that

$$\mathcal{C}_j(\hat{z}; z, \omega) \leq \mathcal{C}_i(\hat{z}; z, \omega) \qquad \forall \hat{z} \in [0, \bar{z}],\ \forall i \in \{0, 1, \ldots, I\}.$$

Proposition 11 (Primitive sufficient condition for a unique maximizer: data-market case)**.** *Suppose $I \geq 1$. Maintain Assumptions 1 and 3, together with Assumption 6. Assume also the downstream curvature condition in Proposition 10(ii), and that*

$$C'(\bar{z}, \bar{b}) > s.$$

Then for every fixed $z \in [0, \bar{z}]$, the function $\hat{z} \mapsto U(\hat{z}; z)$ has a unique global maximizer on $[0, \bar{z}]$. Hence Assumption 2 holds in the data-market case as well.

Proof. See Appendix A.3.7.

Remark (Data-market case and source switching). While Assumption 6 is restrictive, it isolates the exact friction that precludes a purely primitive-based uniqueness proof in the general data-market case. When the identity of the cheapest effective source changes with $\hat{z}$, the realized lower envelope $\kappa(\hat{z}; z, \omega)$ develops downward kinks at source-switch points. Those kinks destroy the convexity-type regularity needed for a clean single-peakedness theorem. Assumption 6 rules out this source-switching nonconvexity and restores a primitive route to uniqueness. Without it, Assumption 2 remains the maintained regularity condition on the producer's downstream problem.

C General Augmentation and Market Structure

This appendix demonstrates that the monopolistic market structure established in Section 3 is robust to relaxing the baseline productivity-floor technology, $\beta(z, q) = \max\{z, q\}$. We introduce a general augmentation function to delineate the properties of β that preserve the baseline winner-take-all dynamic from those that might facilitate product differentiation. Throughout this appendix, we maintain the normalization $F(z) = z$, the worker-skill distribution G with continuous density $g > 0$ on $(0, 1)$, and the marginal service cost $c \geq 0$ from Section 3. The normalization $F(z) = z$ remains analytically necessary; introducing a nonlinear F would confound workers' relative valuations of quality differences, disentangling them from the cross-partial derivative β_{12}.

C.1 Assumptions

Let $q \in [0, 1]$ denote model capability and let $z \in [0, 1]$ denote worker skill. Assume $\beta : [0, 1]^2 \to [0, 1]$ is continuous on $[0, 1]^2$ and C^2 on $\mathcal{D} \equiv \{(z, q) : 0 < z < q < 1\}$. We impose:

Assumption 7 (Weak augmentation). $\beta(z, q) \geq z$ for all (z, q).

Assumption 8 (Capability ceiling). $\beta(z, q) = z$ for all $z \geq q$.

Assumption 9 (Monotonicity in capability). $\beta_2(z, q) > 0$ on $\mathcal{D}$.

Assumption 10 (Diminishing gain in worker skill). $\beta_1(z, q) < 1$ on $\mathcal{D}$.

Assumption 8 formally implements the "skill protection" mechanism established in Section 3: a model with capability q provides no marginal productivity gain to workers whose skill already exceeds q.

C.2 Monopoly Under General β

Define the gain from model q as

$$\Delta(z, q) \equiv \beta(z, q) - z.$$

Fix a capability level q such that $\Delta(0, q) > 0$. By Assumption 8, $\Delta(z, q) = 0$ for $z \geq q$; by Assumption 10, $\Delta(\cdot, q)$ is strictly decreasing on $(0, q)$. Hence, for every $r \in (0, \Delta(0, q))$, there

exists a unique marginal adopter $z^*(r,q) \in (0,q)$ satisfying

$$\Delta(z^*(r,q),q) = r. \tag{53}$$

Demand is therefore $G(z^*(r,q))$, mirroring the baseline case in Section 3.

Proposition 12 (General monopoly pricing)**.** *Fix q with $\Delta(0,q) > 0$. If $r^*(q)$ is an interior revenue-maximizing price and $z^*(q)$ is the corresponding marginal adopter, then*

$$r^*(q) = \big(1 - \beta_1(z^*(q),q)\big)\frac{G(z^*(q))}{g(z^*(q))}. \tag{54}$$

The optimal monopoly capability and price emerge as the solution to the following system:

$$r = \beta(z,q) - z, \tag{55}$$

$$r = \big(1 - \beta_1(z,q)\big)\frac{G(z)}{g(z)}. \tag{56}$$

Proposition 13 (Local comparative statics)**.** *Suppose $q \mapsto (r^*(q), z^*(q))$ is a C^1 interior solution to (55)–(56) on an open interval, and suppose*

$$2\big(1 - \beta_1(z^*,q)\big)g(z^*) - r^*(q)g'(z^*) - \beta_{11}(z^*,q)G(z^*) \neq 0. \tag{57}$$

Then

$$\frac{dz^*}{dq} = \frac{\beta_2(z^*,q)g(z^*) + \beta_{12}(z^*,q)G(z^*)}{2\big(1 - \beta_1(z^*,q)\big)g(z^*) - r^*(q)g'(z^*) - \beta_{11}(z^*,q)G(z^*)}, \tag{58}$$

and

$$\frac{dr^*}{dq} = \beta_2(z^*,q) - \big(1 - \beta_1(z^*,q)\big)\frac{dz^*}{dq}. \tag{59}$$

In the interior region $z^* < q$, these formulas naturally collapse to the Section 3 expressions under the baseline $\beta(z,q) = \max\{z,q\}$, where $\beta_1 = 0$, $\beta_2 = 1$, and $\beta_{11} = \beta_{12} = 0$.

Proposition 14 (Envelope formula for capability choice)**.** *Let $\Pi(q) = r^*(q)G(z^*(q)) - C(q)$. Then*

$$\Pi'(q) = r^*(q)g(z^*(q))\frac{\beta_2(z^*(q),q)}{1 - \beta_1(z^*(q),q)} - C'(q). \tag{60}$$

Hence, the long-run capability choice satisfies

$$r^*(q)g(z^*(q))\frac{\beta_2(z^*(q),q)}{1-\beta_1(z^*(q),q)} = C'(q). \tag{61}$$

C.3 Duopoly: The Strategic Role of β_{12}

Fix two capabilities $q_1 > q_2 > 0$ and define the low-skill gain differential

$$\Gamma_{12}(z) \equiv \beta(z,q_1) - \beta(z,q_2), \qquad z \in [0,q_2). \tag{62}$$

Given $F(z) = z$, workers with skill levels below q_2 evaluate model 1 against model 2 based exclusively on $\Gamma_{12}(z)$ and the price differential $r_1 - r_2$.

Lemma 18 (Monotonicity of the gain differential)**.** *For $z \in [0,q_2)$,*

$$\Gamma'_{12}(z) = \beta_1(z,q_1) - \beta_1(z,q_2) = \int_{q_2}^{q_1} \beta_{12}(z,t)\,dt.$$

Hence:

1. *if $\beta_{12} \leq 0$ on $\mathcal{D}$, then Γ_{12} is decreasing;*
2. *if $\beta_{12} \geq 0$ on $\mathcal{D}$, then Γ_{12} is increasing.*

C.3.1 Market Structure Versus Product-Line Geometry

Our baseline analysis in Section 3 yielded two logically distinct conclusions: a *winner-take-all market structure* and a *portfolio-collapse property* that reduces a firm's optimal product line to a single flagship model. Under a general augmentation function β, these two phenomena diverge. While the concentrated market structure and the viability of an entry-deterrence strategy remain robust, the sign of β_{12} dictates the strategic implementation of market defense: under submodularity, a single superior model suffices to exclude a weaker rival; under supermodularity, sustaining the winner-take-all outcome may require the incumbent to deploy a model family—pairing a high-end flagship with a weak model acting as a fighting brand.

C.3.2 Submodularity Preserves Single-Model Winner-Take-All

The following proposition demonstrates that under submodularity, the superior model can capture the entire market by deploying a single-model limit-pricing strategy. Although the monopolistic defense against rivals remains robust, the stricter portfolio-collapse property is not universally guaranteed.

Proposition 15 (Submodularity preserves single-model winner-take-all)**.** *Assume $\beta_{12} \leq 0$ on $\mathcal{D}$. Then for any $q_1 > q_2 > 0$:*

(i) ***Skill protection.*** *For every $z \in (q_2, q_1]$, model 2 gives no gain, so a worker never adopts model 2 at any price $r_2 \geq 0$.*

(ii) ***Worst-case low-skill worker.*** *On $[0, q_2)$, the willingness to pay for the better model is minimized at $z = q_2^-$:*

$$\Gamma_{12}(z) \geq \Gamma_{12}(q_2^-) \qquad \forall z < q_2,$$

where $\Gamma_{12}(z) \equiv \beta(z, q_1) - \beta(z, q_2)$.

(iii) ***Limit pricing and rival exclusion.*** *If the inferior firm prices at marginal cost $r_2 = c$, then the superior firm captures all low-skill workers whenever*

$$r_1 - r_2 < \Gamma_{12}(q_2^-).$$

In particular, the leader can set

$$r_1^* = \min\{r_1^m,\ c + \Gamma_{12}(q_2^-) - \varepsilon\} \tag{63}$$

for any $\varepsilon \in (0, \Gamma_{12}(q_2^-))$, and ensure that every adopting worker chooses model 1. The follower earns zero operating profit.

Remark (Limits of Submodularity). Proposition 15 preserves the market structure of Section 3: a capability leader can exclude an inferior rival via single-model limit pricing. However, submodularity does not strictly guarantee portfolio collapse. An incumbent firm may still optimally deploy multiple in-house models to segment its own user base, even if a single superior model is sufficient to deter entry. Thus, submodularity preserves rival exclusion, but not necessarily portfolio

collapse.

C.3.3 Supermodularity Preserves Winner-Take-All Through a Model-Family Strategy

Supermodularity ($\beta_{12} > 0$) implies that workers exhibit systematic differences in their marginal valuations of capability. Consequently, the single-model reduction breaks down, as a lone flagship may no longer serve the entire market optimally. Importantly, this relates to product-line geometry rather than firm-level market structure. The subsequent results establish that supermodularity still preserves the winner-take-all market structure, albeit through a different mechanism: an incumbent can leverage a model family to internalize the low-end segment and subsequently deter entry.

Assumption 11 (Costless degradation and convex training cost). In the remainder of this subsection, we assume a firm that has paid $C(q_H)$ for a frontier model of capability q_H can costlessly deploy any weaker model $q \leq q_H$. The training cost function $C : [0, 1] \to \mathbb{R}_+$ is continuous, strictly increasing, convex, and satisfies $C(0) = 0$.

Assumption 12 (Bounded marginal capability gains). For every $q_H < 1$,

$$L_H \equiv \sup\{\beta_2(z, q) : 0 \leq z < q \leq q_H\} < \infty.$$

Proposition 16 (Relative-preference segmentation under supermodularity)**.** *Assume $\beta_{12} > 0$ on $\mathcal{D}$ and fix $q_H > q_W > 0$. Define*

$$\Gamma_{HW}(z) \equiv \beta(z, q_H) - \beta(z, q_W), \qquad z \in [0, q_W).$$

Then Γ_{HW} is strictly increasing on $[0, q_W)$. Consequently, for any price gap $d \in \big(\Gamma_{HW}(0), \Gamma_{HW}(q_W^-)\big)$ there exists a unique $\tilde{z} \in (0, q_W)$ such that $\Gamma_{HW}(\tilde{z}) = d$, and

$$\begin{cases} \beta(z, q_W) - r_W > \beta(z, q_H) - r_H, & z < \tilde{z}, \\ \beta(z, q_H) - r_H > \beta(z, q_W) - r_W, & z > \tilde{z}, \end{cases} \qquad \text{whenever } r_H - r_W = d.$$

This two-model menu effectively segments consumers based on relative preferences: lower-skill workers naturally sort to the weak model, while higher-skill workers prefer the flagship. Securing

positive demand for both segments remains subject to standard participation constraints against non-adoption.

Proposition 17 (Model-family replication dominates accommodating a non-leapfrogging entrant)**.** *Assume costless degradation and fix the incumbent's flagship capability q_H. Consider the benchmark subgame in which the incumbent offers only the flagship. Suppose that the resulting equilibrium accommodates an entrant portfolio*

$$S_E^A = \{(q_j^A, r_j^A)\}_{j=1}^m \qquad \textit{with} \qquad q_j^A \le q_H \textit{ for every } j,$$

and let r_H^A denote the incumbent's flagship price in that equilibrium. Let D_H^A denote demand for the incumbent flagship and let D_j^A denote demand for entrant option j.

Consider instead the counterfactual no-entry outcome in which the incumbent keeps its flagship at price r_H^A and also offers the entrant's former menu, $S_I^{\text{rep}} = \{(q_H, r_H^A)\} \cup S_E^A$. Then:

(i) *the set of capability-price options available to workers is unchanged, so aggregate demand assigned to each option is unchanged up to a measure-zero indifference set;*

(ii) *the incumbent's total profit rises by exactly the entrant's former operating profit,*

$$\Pi_I^{\text{rep}} - \Pi_I^{\text{acc}} = \pi_E^{\text{op},A}, \tag{64}$$

where $\pi_E^{\text{op},A} \equiv \sum_{j=1}^m (r_j^A - c) D_j^A$;

(iii) *therefore, a model-family strategy weakly dominates the outcome "flagship only + accommodated non-leapfrogging entrant," and strictly dominates it whenever the entrant would otherwise survive with positive operating profit.*

Remark (Single-model entrant special case). If the accommodated entrant uses only one model (q_E^A, r_E^A), then Proposition 17 reduces to the fighting-brand comparison: the incumbent can introduce a single weak model $q_W = q_E^A$ at price $r_W = r_E^A$, thereby raising profit by $(r_E^A - c)D_E^A$.

Proposition 18 (Frontier-specific bound on non-leapfrogging entrant operating profit)**.** *Assume Assumption 12. Fix the incumbent's flagship capability $q_H > 0$, and let the incumbent additionally release a weak model $q_W < q_H$ at price c. For any entrant portfolio S_E, let $q_E \equiv \sup\{q : (q, r) \in$*

$S_E\}$ *denote the entrant's frontier, and write*

$$\pi_E^{\mathrm{op}}(S_E) \equiv \sum_{(q,r)\in S_E} (r-c)D_{(q,r)}$$

for operating profit net of service cost. If $q_E \in (q_W, q_H]$*, then*

$$\pi_E^{\mathrm{op}}(S_E) \le L_H\,(q_E - q_W). \tag{65}$$

Corollary 2 (Convex-cost deterrence criterion)**.** *Assume Assumptions 11 and 12. If*

$$C'_+(q_W) > L_H, \tag{66}$$

then every entrant portfolio with frontier capability $q_E \in (q_W, q_H]$ *earns strictly negative total profit.*

Corollary 3 (Existence of a deterring fighting brand)**.** *Assume Assumptions 11 and 12. Fix* $q_H > 0$*. Then there exists* $\bar{q}_W < q_H$ *such that every* $q_W \in (\bar{q}_W, q_H)$ *deters every entrant portfolio with frontier capability* $q_E \in (q_W, q_H]$*.*

Remark (Strategic Implications of Convexity)*.* Corollary 2 provides a tractable local sufficient condition for entry deterrence: if the marginal training cost at q_W exceeds the maximal marginal operating gain from raising capability above q_W, all non-leapfrogging entrants are deterred. Corollary 3 relies purely on continuity and monotonicity to show that an incumbent can strategically place a weak model sufficiently close to the frontier to extract all residual operating profit necessary for an entrant's survival.

Remark (The Strategic Value of Supermodularity)*.* The operating-profit bound in Proposition 18 relies on the capability ceiling, monotonicity in q, and bounded β_2. Supermodularity, however, is what renders the weak model strategically viable for the incumbent. As shown in Proposition 16, supermodularity induces genuine relative-preference heterogeneity. This allows the weak model to capture lower-skill workers while preserving the flagship's high-end appeal. Thus, rather than unraveling the winner-take-all dynamic, supermodularity simply necessitates a multi-model product-line strategy to sustain it.

Proposition 19 (Global bound on leapfrogging entrant operating profit)**.** *For any entrant portfolio S_E, operating profit satisfies*

$$\pi_E^{\mathrm{op}}(S_E) \leq (1-c)_+. \tag{67}$$

Consequently, if $q_H > 0$ satisfies $C(q_H) > (1-c)_+$, then every entrant portfolio with frontier capability $q_E > q_H$ earns strictly negative total profit.

Theorem C.1 (Supermodularity preserves firm-level winner-take-all)**.** *Assume $\beta_{12} > 0$ on $\mathcal{D}$, Assumptions 11 and 12, and fix an incumbent flagship capability $q_H > 0$ such that $C(q_H) > (1-c)_+$. Then there exists some weak model $q_W < q_H$ such that, if the incumbent offers the menu $\{(q_H, r_H), (q_W, c)\}$ for any flagship price r_H, no entrant portfolio is profitable. Hence, supermodularity preserves the winner-take-all market structure at the firm level, implemented via a model-family entry-deterrence strategy.*

Theorem C.2 (Winner-take-all market structure under general augmentation)**.** *Under the maintained assumptions of this section, both submodularity and supermodularity preserve the winner-take-all market structure:*

(i) *if $\beta_{12} \leq 0$, Proposition 15 yields a single-model limit-pricing strategy that excludes an inferior rival;*

(ii) *if $\beta_{12} > 0$ and Assumptions 11 and 12 hold, then Theorem C.1 establishes a model-family strategy that deters all entrant portfolios.*

Therefore, the concentrated market structure and the existence of an equilibrium entry-deterrence strategy remain robust to the sign of β_{12}. What is sensitive to the functional form is the portfolio-collapse lemma: under supermodularity, the incumbent may require a nondegenerate model family to enforce the winner-take-all outcome.

Remark (Interpretation). The central insight is not that supermodularity restores a stable, differentiated oligopoly; rather, industry concentration persists irrespective of the sign of β_{12}. The distinction lies strictly in the product-line architecture required to defend the market: submodularity permits defense via a single superior model, whereas supermodularity may necessitate a flagship augmented by a fighting brand. The non-robust element is thus the single-model reduction underlying the portfolio-collapse logic of Section 3, rather than the winner-take-all nature of the competition itself.

D Costly Screening and Verification in the Upstream Data Market

The preceding analysis of the upstream data market (Section 4) assumes the buyer observes data quality costlessly through scaling experiments. In practice, compute-optimal pilot runs are resource-intensive. This appendix develops two extensions in which the buyer's informational advantage is no longer free, demonstrating that the hidden-gem mechanism and cherry-picking equilibrium survive under explicit cost structures.

Both extensions maintain the competitive pricing benchmark of Lemma 11: each seller posts $p_i(s_i) = \mathbb{E}[V(B_i) \mid s_i]$. Under this benchmark, the seller's price depends only on their own posterior, so the pricing rule is invariant to the buyer's evaluation decision. We work throughout within the binary-quality framework of Section 4 ($b_i \in \{b_L, b_H\}$, signals $s_i \in \{L, H\}$ with precision $\gamma \in (1/2, 1)$) and the exclusive-licensing setting of Section 4.4.

D.1 Maintained Primitives

Under the competitive benchmark, seller posteriors are

$$\mu_H \equiv \frac{\gamma\mu_0}{\gamma\mu_0 + (1-\gamma)(1-\mu_0)}, \qquad \mu_L \equiv \frac{(1-\gamma)\mu_0}{(1-\gamma)\mu_0 + \gamma(1-\mu_0)}, \tag{68}$$

with $\mu_H > \mu_0 > \mu_L$. We define the surplus terms

$$\Delta_L \equiv V_H - p(L) = (1-\mu_L)\Delta V, \qquad \Delta_H \equiv V_H - p(H) = (1-\mu_H)\Delta V,$$

where $\Delta V \equiv V_H - V_L > 0$. Since $\mu_H > \mu_L$, we have $\Delta_L > \Delta_H > 0$.

Lemma 19 (Price and surplus identities)**.** *Under the competitive benchmark $p(s) = \mathbb{E}[V(B) \mid s]$:*

(i) $p(H) > p(L)$.

(ii) *For either signal $s \in \{L, H\}$, blind purchase yields zero expected surplus:* $\mathbb{E}[V(B) \mid s] - p(s) = 0$.

(iii) *Low-quality datasets are overpriced relative to realized value:* $V_L - p(s) = -\mu_s \Delta V < 0$.

(iv) *High-quality datasets yield positive net surplus:* $V_H - p(L) = \Delta_L = (1-\mu_L)\Delta V$ *and* $V_H - p(H) = \Delta_H = (1-\mu_H)\Delta V$, *with* $\Delta_L > \Delta_H > 0$.

Proof. Direct computation from $p(s) = \mu_s V_H + (1-\mu_s)V_L$. See Appendix A.7.1 for details.

D.2 A Fixed-Cost Screening Technology

The buyer may pay a lump-sum cost $K > 0$ to learn the realized qualities of all sellers. The maintained pricing benchmark remains unchanged.

Timing. Nature draws $(B_i, s_i)_{i=1}^I$, sellers post $p_i(s_i)$, the buyer observes all prices and therefore all signals, the buyer chooses whether to pay K to screen the market, and then the buyer either purchases one dataset or does not buy.

After observing prices the buyer knows the realized counts $n_L \equiv \#\{i : s_i = L\}$ and $n_H \equiv \#\{i : s_i = H\}$, with $n_L + n_H = I$. Conditional on the signal profile, seller qualities remain independent with $\Pr(B_i = b_H \mid s_i = L) = \mu_L$ and $\Pr(B_i = b_H \mid s_i = H) = \mu_H$.

Proposition 20 (Conditional value of market screening)**.** *Fix a realized signal profile with counts* (n_L, n_H). *The buyer's expected gross surplus from paying the screening cost and then choosing optimally is*

$$S(n_L, n_H) = \Delta_L\big[1 - (1-\mu_L)^{n_L}\big] + (1-\mu_L)^{n_L}\,\Delta_H\big[1 - (1-\mu_H)^{n_H}\big]. \tag{69}$$

Without screening, the buyer's best payoff is 0. *Hence, conditional on the observed prices, screening is optimal if and only if* $K < S(n_L, n_H)$. *Moreover, conditional on screening, if at least one hidden gem* $(B_i, s_i) = (b_H, L)$ *is present, then the buyer chooses such a seller.*

Proof. See Appendix A.7.1.

The screening value (69) has a cascade structure: the buyer first searches for hidden gems in the L-signal group (earning the higher surplus Δ_L), and only falls back to the H-signal group (earning Δ_H) if no hidden gem is present.

Corollary 4 (Ex ante gross value of screening)**.** *Let* $q_L \equiv \mu_0(1-\gamma)$ *denote the ex ante probability*

that a given seller is a hidden gem. The ex ante gross value of the screening technology is

$$\bar{S}_I \equiv \mathbb{E}\big[S(N_L, N_H)\big] = \Delta_L\big[1-(1-q_L)^I\big] + \Delta_H\big[(1-q_L)^I - (1-\mu_0)^I\big]. \tag{70}$$

If the fixed fee K must be paid before the buyer observes prices, screening is ex ante optimal if and only if $K < \bar{S}_I$.

Proof. See Appendix A.7.2.

Proposition 21 (Comparative statics of screening value)**.** *The ex ante gross screening value $\bar{S}_I$ has the following properties:*

(i) *It is strictly increasing in the number of sellers I.*

(ii) *It is affine, hence strictly increasing, in ΔV.*

(iii) *It satisfies $\lim_{\gamma\to 1} \bar{S}_I = 0$.*

Proof. See Appendix A.7.3.

Corollary 5 (Self-financing screening near the frontier)**.** *Suppose the valuation wedge is endogenized through $\Delta V(z^*) = (1-z^*)^{-1/b_L} - (1-z^*)^{-1/b_H}$, with $b_H > b_L > 0$, so that $\Delta V(z^*) \to \infty$ as $z^* \to 1$. Fix any realized signal profile with $(n_L, n_H) \neq (0,0)$ and any fixed screening cost $K > 0$. Then there exists $\underline{z}(K, n_L, n_H) \in (0,1)$ such that screening is optimal for every $z^* > \underline{z}(K, n_L, n_H)$.*

Proof. See Appendix A.7.4.

Remark. The fixed-cost screening technology is the cleanest extension because it preserves the paper's original Proposition 4 logic *conditional on screening*: after paying K, the buyer still chooses hidden gems whenever they exist. The only new object is the screening threshold.

D.3 A Per-Dataset Verification Cost

This subsection studies a stricter robustness model. The buyer can verify individual sellers at cost $\kappa > 0$ per dataset after observing prices and therefore signals.

Timing. Nature draws $(B_i, s_i)_{i=1}^I$, sellers post prices, the buyer observes all prices, then chooses a subset of sellers to verify. Verification perfectly reveals the realized quality of each inspected seller and costs κ per inspected seller. After verification the buyer purchases at most one dataset, or does not buy.

Fix a realized signal profile with counts (n_L, n_H). Let $k_L \in \{0, 1, \ldots, n_L\}$ and $k_H \in \{0, 1, \ldots, n_H\}$ denote the numbers of verified sellers from the pessimistic and optimistic signal groups. Define the residual uncertainty terms

$$\sigma_L \equiv \mu_L(1-\mu_L), \qquad \sigma_H \equiv \mu_H(1-\mu_H). \tag{71}$$

Proposition 22 (Expected payoff from partial verification)**.** *For any (k_L, k_H), the buyer's expected net payoff relative to no verification is*

$$\Pi(k_L, k_H) = \Delta_L\big[1-(1-\mu_L)^{k_L}\big] + \Delta_H(1-\mu_L)^{k_L}\big[1-(1-\mu_H)^{k_H}\big] - \kappa(k_L+k_H). \tag{72}$$

Proof. See Appendix A.7.5.

Lemma 20 (Marginal values of additional verification)**.** *For any feasible (k_L, k_H),*

$$\Pi(k_L+1, k_H) - \Pi(k_L, k_H) = (1-\mu_L)^{k_L}\mu_L\big[\mu_H - \mu_L + (1-\mu_H)^{k_H+1}\big]\Delta V - \kappa \le \sigma_L \Delta V - \kappa, \tag{73}$$

$$\Pi(k_L, k_H+1) - \Pi(k_L, k_H) = (1-\mu_L)^{k_L}(1-\mu_H)^{k_H}\sigma_H \Delta V - \kappa \le \sigma_H \Delta V - \kappa. \tag{74}$$

Proof. See Appendix A.7.6.

Proposition 23 (Threshold for profitable verification)**.** *Let $A \equiv \{s \in \{L, H\} : n_s > 0\}$ be the set of signal groups present in the realized market. Then*

$$\max_{0 \le k_L \le n_L,\, 0 \le k_H \le n_H} \Pi(k_L, k_H) > 0 \quad \Longleftrightarrow \quad \kappa < \max_{s \in A} \sigma_s \Delta V.$$

Equivalently, some verification is privately profitable if and only if the cost of one inspection is below the largest available one-draw information value.

Proof. See Appendix A.7.7.

Lemma 21 (Comparison of residual uncertainty)**.** *The residual uncertainty terms satisfy*

$$\sigma_H > \sigma_L \iff \mu_0 < 1/2, \qquad \sigma_H = \sigma_L \iff \mu_0 = 1/2, \qquad \sigma_H < \sigma_L \iff \mu_0 > 1/2.$$

Proof. See Appendix A.7.8.

Proposition 24 (Pure cherry-picking region)**.** *Suppose $n_L \geq 1$ and*

$$\sigma_L > \sigma_H, \qquad \sigma_H \Delta V \leq \kappa < \sigma_L \Delta V. \tag{75}$$

Then every optimal verification policy (k_L^*, k_H^*) satisfies $k_H^* = 0$ and $k_L^* \geq 1$. Moreover, the optimal number of pessimistic-signal sellers to verify is

$$k_L^* = \max\big\{k \in \{1, \ldots, n_L\} : (1-\mu_L)^{k-1}\sigma_L \Delta V \geq \kappa\big\}. \tag{76}$$

Proof. See Appendix A.7.9.

Corollary 6 (Non-empty pure cherry-picking region)**.** *If $\mu_0 > 1/2$, then $\sigma_L > \sigma_H$ by Lemma 21. Therefore, whenever $n_L \geq 1$, the interval $[\sigma_H \Delta V, \sigma_L \Delta V)$ is non-empty, and every κ in this interval supports pure cherry-picking in the sense of Proposition 24.*

Corollary 7 (Pure verification region for optimistic signals)**.** *If $n_H \geq 1$, $\sigma_H > \sigma_L$, and $\sigma_L \Delta V \leq \kappa < \sigma_H \Delta V$, then every optimal verification policy satisfies $k_L^* = 0$ and $k_H^* \geq 1$. The optimal number of optimistic-signal sellers to verify is the largest $k \in \{1, \ldots, n_H\}$ satisfying $(1-\mu_H)^{k-1}\sigma_H \Delta V \geq \kappa$.*

Remark (Allocation priority and finite-market caveat). The per-dataset-cost model yields a sharper but more demanding robustness result than the fixed-cost model. For any feasible (k_L, k_H) with $k_H \geq 1$ and $k_L < n_L$, a swap calculation shows

$$\Pi(k_L + 1, k_H - 1) - \Pi(k_L, k_H) = (1-\mu_L)^{k_L}(\mu_H - \mu_L)\big[\mu_L - (1-\mu_H)^{k_H}\big]\Delta V. \tag{77}$$

When $\mu_0 > 1/2$, one can verify that $\mu_L > 1 - \mu_H$ and hence $\mu_L > (1-\mu_H)^{k_H}$ for all $k_H \geq 1$, so (77) is strictly positive. Thus, holding the total number of verifications fixed, the buyer always prefers to allocate inspections to the L-signal group first.

This observation sharpens the interpretation of Proposition 24: pure cherry-picking is globally optimal on the interval $[\sigma_H \Delta V,\, \sigma_L \Delta V)$, and L-sellers always receive inspection priority, but mixed verification (with $k_H^* > 0$) can arise once the L-signal group is exhausted.

E Bayesian Nash Bargaining Extension

This appendix keeps the main-text allocation and information structure of Section 5.3 fixed and alters only the transfer rule. In the baseline, sellers post prices ex ante and the buyer selects the seller minimizing efficiency-adjusted total training cost. Here we instead suppose that, after the buyer identifies a candidate seller i, the final transfer can be renegotiated via Bayesian Nash bargaining. The key maintained restriction is that the seller bargains on the basis of its posterior expectation, not on the basis of realized data quality, unless the buyer's bargaining move fully reveals that quality.

Fix a target capability $z^* \in (0,1)$ and a seller i. Let

$$C_i \equiv C(z^*, B_i) = (1 - z^*)^{-1/B_i} - 1$$

denote the realized compute demand associated with seller i's data quality. Let the buyer's best alternative total expenditure under the main-text posted-price menu be

$$T_{-i}(z^*) \equiv \min \left\{ (1 + p_0) C(z^*, b_0),\ \min_{j \neq i} (1 + p_j^*(c_j; z^*)) C(z^*, B_j) \right\}.$$

Equivalently, the realized unit-price ceiling at which the buyer is just indifferent between contracting with seller i and reverting to its best alternative is

$$p_i^{cap}(z^*) \equiv \frac{T_{-i}(z^*)}{C_i} - 1 = \min \left\{ p_i^{\max}(z^*),\ \min_{j \neq i} \left[\frac{(1 + p_j^*(c_j; z^*)) C(z^*, B_j)}{C_i} - 1 \right] \right\}.$$

Let h_i denote seller i's public information set at the bargaining stage. By construction h_i contains at least c_i, and it may also contain public information conveyed by the buyer's approach decision. The seller does not observe B_i directly unless it is revealed by h_i. The private-bargaining benchmark is the special case $h_i = c_i$.

Assumption 13 (Bayesian Nash bargaining protocol). Conditional on h_i, buyer and seller bargain

over a unit price p according to

$$\max_{p \geq c_i} \; \mathcal{B}_i(p \mid h_i; z^*)^{1-\theta_i} \, \mathcal{S}_i(p \mid h_i; z^*)^{\theta_i},$$

where $\theta_i \in [0,1]$ is the seller's bargaining weight,

$$\mathcal{B}_i(p \mid h_i; z^*) \equiv E[(p_i^{cap}(z^*) - p)C_i \mid h_i], \qquad \mathcal{S}_i(p \mid h_i; z^*) \equiv (p - c_i)E[C_i \mid h_i].$$

Trade occurs ex post if and only if $p \leq p_i^{cap}(z^*)$; otherwise the buyer exercises its outside option.

Proposition 25 (Closed-form Bayesian Nash bargain)**.** *Define*

$$K_i(h_i; z^*) \equiv E[C_i \mid h_i], \qquad M_i(h_i; z^*) \equiv E[p_i^{cap}(z^*)C_i \mid h_i],$$

and the compute-weighted posterior expected ceiling

$$\widetilde{p}_i(h_i; z^*) \equiv \frac{M_i(h_i; z^*)}{K_i(h_i; z^*)}.$$

If $\widetilde{p}_i(h_i; z^*) \geq c_i$, then the bargaining problem has a unique solution

$$p_i^{BNB}(h_i; z^*) = (1 - \theta_i)c_i + \theta_i \widetilde{p}_i(h_i; z^*).$$

Moreover, the conditional expected gains from trade

$$\Delta_i(h_i; z^*) \equiv E[(p_i^{cap}(z^*) - c_i)C_i \mid h_i] = M_i(h_i; z^*) - c_i K_i(h_i; z^*)$$

are split according to

$$\mathcal{B}_i(p_i^{BNB} \mid h_i; z^*) = (1 - \theta_i)\Delta_i(h_i; z^*), \qquad \mathcal{S}_i(p_i^{BNB} \mid h_i; z^*) = \theta_i \Delta_i(h_i; z^*).$$

If $\widetilde{p}_i(h_i; z^*) < c_i$, then the bargaining program admits no price that yields nonnegative interim surplus to both parties, so disagreement obtains.

Proof. Fix (h_i, z^*) and suppress arguments. Since $z^* \in (0,1)$ and B_i has bounded support on $(0, \infty)$, we have $C_i > 0$ almost surely and hence $K_i = E[C_i \mid h_i] > 0$. The buyer's interim

expected surplus can be written as

$$\mathcal{B}_i(p \mid h_i; z^*) = M_i - pK_i,$$

while the seller's interim expected payoff is

$$\mathcal{S}_i(p \mid h_i; z^*) = (p - c_i)K_i.$$

Thus the bargaining problem is equivalent to

$$\max_{p \in [c_i, M_i/K_i]} (M_i - pK_i)^{1-\theta_i}((p - c_i)K_i)^{\theta_i},$$

where the upper bound enforces the buyer's interim participation constraint $M_i - pK_i \geq 0$. If $M_i/K_i < c_i$, the feasible set is empty and disagreement obtains.

Now suppose $M_i/K_i \geq c_i$. Since $K_i > 0$, maximizing the Nash product is equivalent to maximizing its logarithm,

$$(1 - \theta_i)\log(M_i - pK_i) + \theta_i \log(p - c_i) + \theta_i \log K_i,$$

on the open interval $(c_i, M_i/K_i)$. This objective is strictly concave because

$$\frac{\partial^2}{\partial p^2}\Big[(1 - \theta_i)\log(M_i - pK_i) + \theta_i \log(p - c_i)\Big] = -\frac{(1 - \theta_i)K_i^2}{(M_i - pK_i)^2} - \frac{\theta_i}{(p - c_i)^2} < 0.$$

Hence there is a unique maximizer, characterized by the first-order condition

$$-\frac{(1 - \theta_i)K_i}{M_i - pK_i} + \frac{\theta_i}{p - c_i} = 0.$$

Rearranging gives

$$\theta_i(M_i - pK_i) = (1 - \theta_i)K_i(p - c_i),$$

so that

$$p = \frac{\theta_i M_i + (1 - \theta_i)c_i K_i}{K_i} = (1 - \theta_i)c_i + \theta_i \frac{M_i}{K_i} = (1 - \theta_i)c_i + \theta_i \widetilde{p}_i.$$

This proves the closed form.

For the surplus-splitting identities, substitute the optimal price into the interim payoffs:

$$\mathcal{B}_i(p_i^{BNB} \mid h_i; z^*) = M_i - p_i^{BNB} K_i = M_i - \big[(1-\theta_i)c_i + \theta_i \widetilde{p}_i\big] K_i = (1-\theta_i)(M_i - c_i K_i) = (1-\theta_i)\Delta_i,$$

and similarly

$$\mathcal{S}_i(p_i^{BNB} \mid h_i; z^*) = (p_i^{BNB} - c_i)K_i = \theta_i(\widetilde{p}_i - c_i)K_i = \theta_i(M_i - c_i K_i) = \theta_i \Delta_i.$$

This completes the proof. □

Corollary 8 (Reduction to Section 4.3)**.** *In the binary single-seller environment of Section 4.3, $C_i \equiv 1$, $p_i^{cap}(z^*) = V(B)$, and $h_i = s$. Hence*

$$\widetilde{p}_i(s) = E[V(B) \mid s], \qquad p_i^{BNB}(s) = (1-\theta)c_d + \theta E[V(B) \mid s],$$

which coincides with the bargaining rule in Section 4.3.

Proof. With a single seller and unit quantity, the buyer's outside-option ceiling is exactly the realized valuation $V(B)$. Therefore

$$M_i(s) = E[V(B) \mid s], \qquad K_i(s) = 1,$$

so Proposition 25 yields

$$p_i^{BNB}(s) = (1-\theta)c_d + \theta E[V(B) \mid s].$$

This is the rule used in Section 4.3. □

Proposition 26 (Comparison with the baseline posted-price rule)**.** *Fix the same history (h_i, z^*) and let $p_i^*(c_i; z^*)$ denote the main-text posted price. Whenever $\widetilde{p}_i(h_i; z^*) \neq c_i$, define the baseline extraction coefficient*

$$\lambda_i(h_i; z^*) \equiv \frac{p_i^*(c_i; z^*) - c_i}{\widetilde{p}_i(h_i; z^*) - c_i}.$$

Then

$$p_i^{BNB}(h_i; z^*) - p_i^*(c_i; z^*) = \big(\theta_i - \lambda_i(h_i; z^*)\big)\big(\widetilde{p}_i(h_i; z^*) - c_i\big).$$

Hence the difference between the bargaining appendix and the baseline is entirely driven by the

mapping from the posterior expected ceiling into price: the appendix fixes the seller's extraction share at θ_i, whereas the baseline makes the extraction share endogenous through the markup-versus-selection tradeoff.

Proof. By Proposition 25,

$$p_i^{BNB}(h_i; z^*) = (1 - \theta_i)c_i + \theta_i \widetilde{p}_i(h_i; z^*).$$

By the definition of $\lambda_i(h_i; z^*)$,

$$p_i^*(c_i; z^*) = (1 - \lambda_i(h_i; z^*))c_i + \lambda_i(h_i; z^*)\widetilde{p}_i(h_i; z^*).$$

Subtracting the two expressions yields

$$p_i^{BNB}(h_i; z^*) - p_i^*(c_i; z^*) = \big(\theta_i - \lambda_i(h_i; z^*)\big)\big(\widetilde{p}_i(h_i; z^*) - c_i\big),$$

as claimed. □

Proposition 27 (Persistence of buyer information rents)**.** *Let the buyer's realized gain from trade under the Bayesian Nash bargain be*

$$\pi_{B,i}^{BNB} \equiv \big[p_i^{cap}(z^*) - p_i^{BNB}(h_i; z^*)\big]^+ C_i.$$

Then

$$\pi_{B,i}^{BNB} = \Big[\big(p_i^{cap}(z^*) - \widetilde{p}_i(h_i; z^*)\big) + (1 - \theta_i)\big(\widetilde{p}_i(h_i; z^*) - c_i\big)\Big]^+ C_i.$$

Consequently:

1. *$E[\pi_{B,i}^{BNB} \mid h_i]$ is weakly decreasing in θ_i.*
2. *If $\Pr\big(p_i^{cap}(z^*) > \widetilde{p}_i(h_i; z^*) \mid h_i\big) > 0$, then*

$$E[\pi_{B,i}^{BNB} \mid h_i] > 0$$

even when $\theta_i = 1$.

Hence strong seller bargaining power eliminates the buyer's ordinary share of expected surplus, but not the residual information rent generated by the gap between realized willingness to pay and the seller's posterior expectation. In particular, any residual uncertainty that makes $p_i^{cap}(z^*)$ non-degenerate conditional on h_i is sufficient for strictly positive buyer rent at $\theta_i = 1$.

Proof. Substituting Proposition 25 into the realized gain from trade gives

$$p_i^{cap}(z^*) - p_i^{BNB}(h_i; z^*) = p_i^{cap}(z^*) - \big[(1-\theta_i)c_i + \theta_i \widetilde{p}_i(h_i; z^*)\big] = \big(p_i^{cap}(z^*) - \widetilde{p}_i(h_i; z^*)\big) + (1-\theta_i)\big(\widetilde{p}_i(h_i; z^*) - c_i\big),$$

which proves the decomposition.

For part (i), the term inside the square brackets is affine and weakly decreasing in θ_i for every realization. Since the map $x \mapsto x^+$ is weakly increasing, $\pi_{B,i}^{BNB}$ is weakly decreasing in θ_i state by state. Taking conditional expectations preserves the inequality.

For part (ii), set $\theta_i = 1$. Then the ordinary bargaining-surplus term disappears and

$$\pi_{B,i}^{BNB} = \big[p_i^{cap}(z^*) - \widetilde{p}_i(h_i; z^*)\big]^+ C_i.$$

Because $C_i > 0$ almost surely, this quantity is strictly positive on the event $\{p_i^{cap}(z^*) > \widetilde{p}_i(h_i; z^*)\}$. If that event has positive conditional probability given h_i, the conditional expectation is strictly positive as well. □

Remark (Connection to the fixed-point theorem). Theorem B.1 treats the upstream pricing rule as an input satisfying continuity and boundedness. Accordingly, the same fixed-point argument applies to the bargaining appendix whenever $(c_i, z) \mapsto p_i^{BNB}(h_i; z)$ is continuous and bounded. By Proposition 25, this holds whenever $(c_i, z) \mapsto \widetilde{p}_i(h_i; z)$ is continuous and bounded.

Relative to the main-text baseline, this appendix preserves the buyer's procurement logic and the paper's information structure, but replaces ex ante strategic shading with ex post posterior-based surplus splitting. In the baseline, competition lowers prices because sellers fear exclusion. In the bargaining appendix, competition lowers prices because it improves the buyer's disagreement point. The hidden-gem mechanism survives because, absent full revelation, the seller bargains against the posterior object $\widetilde{p}_i(h_i; z^*)$ rather than the realized ceiling $p_i^{cap}(z^*)$.