Correlated Signals, Data-Vendor Market Structure, and Price Informativeness

Alejandro Lopez-Lira*
University of Florida

April 20, 2026

Abstract

When investors build private signals from shared inputs, signal errors inherit a common component that does not wash out under aggregation. I study a CARA-normal noisy rational expectations economy in which a data vendor chooses the cross-subscriber error correlation ρ of signals sold to adopters, and buyers cannot verify ρ at the subscription stage. Two results govern the economy. First, price informativeness is bounded above by the ceiling $\tau_{\phi,\infty}(\rho) < \tau_\eta/\rho^2$ for every $\rho > 0$: the wisdom-of-crowds limit survives only at $\rho = 0$. Second, a closed-form threshold $\tau_S^*(0) = 2a\tau_\eta/[a + \sqrt{a^2 + 6\tau_z\tau_\eta}]$ on individual-signal precision separates two regimes. When τ_S exceeds the threshold, an Akerlof-style unraveling drives competitive vendors to the cost-minimizing, highest-correlation corner, and price informativeness ranks $\tau_{\phi,\infty}(\rho^{SP}) > \tau_{\phi,\infty}(\rho^M) > \tau_{\phi,\infty}(\rho^C)$: competition reduces informativeness, reversing the standard antitrust intuition for information goods. Below the threshold, the ranking inverts. Mandated disclosure of cross-subscriber error dispersion collapses the competitive outcome to monopoly and improves welfare.

Keywords: data markets, price informativeness, signal correlation, data-vendor market structure, hidden action, wisdom of crowds.
JEL codes: G14, D82, D83, L15.

*Warrington College of Business, University of Florida. Email: alejandroll10@gmail.com. I thank participants at the autonomous theory pipeline for useful discussions.

1 Introduction

Several hedge funds subscribe to the same alternative-data vendor, call the same pretrained language model through the same API, and feed the same satellite-imagery feature extractor into their trading pipelines. Each fund believes it holds a private signal about next-quarter earnings. In one dimension the belief is correct, in that other market participants do not see the fund's particular portfolio weights. In another dimension the belief is wrong: the errors across funds share a common component inherited from the shared production input. When prices aggregate the funds' demands, that common component does not average out. Wisdom of crowds, understood as the vanishing of noise under aggregation, fails.

The direction of the common component of error is not a property of nature. A data vendor that trains one foundation model and resells its outputs to every subscriber produces highly correlated errors. A vendor that fine-tunes a bespoke model per subscriber, injects per-subscriber noise, or curates subscriber-specific data produces nearly idiosyncratic errors. The cross-sectional error correlation $\rho \in [0, \bar{\rho}]$ is an attribute of the vendor's technology, set by a platform's cost-benefit calculation over differentiation. Two features of this calculation shape the equilibrium of the data market and, through it, the informational properties of asset prices. First, ρ is hard for a buyer to verify at the subscription stage: the joint distribution of errors across other subscribers is not contractible on the subscription date. Second, de-correlation costs the vendor: differential fine-tuning, noise injection, and per-subscriber curation eat into platform margins.

The paper asks how the market structure of the data-vendor sector selects ρ in equilibrium, and what the answer implies for price informativeness and welfare. Two results deliver the characterization.

Proposition 3 (threshold characterization). The sign of $d\tau_{\phi,\infty}/d\rho$ is governed by a unique threshold $\tau_S^*(\rho^2; \tau_\eta, \tau_z, a)$ on individual-signal precision. At $\rho = 0$,

$$\tau_S^*(0; \tau_\eta, \tau_z, a) = \frac{2a\tau_\eta}{a + \sqrt{a^2 + 6\tau_z\tau_\eta}},$$

with $\partial\tau_S^*(0)/\partial\tau_\eta > 0$, $\partial\tau_S^*(0)/\partial a > 0$, $\partial\tau_S^*(0)/\partial\tau_z < 0$. Above the threshold, higher ρ reduces informativeness; below, higher ρ raises informativeness.

Theorem 4 (market-structure ranking, main result). Under hidden correlation (Assumption 2), slack buyer participation, unimodality of the monopolist's and planner's objectives, and $\tau_S > \tau_S^*(\bar{\rho}^2; \theta)$ uniformly, constrained-planner, monopoly, and competitive-vendor equilibria rank strictly as

$$\tau_{\phi,\infty}(\rho^{SP}) \; > \; \tau_{\phi,\infty}(\rho^M) \; > \; \tau_{\phi,\infty}(\rho^C).$$

Moving from monopoly to competition in a data-vendor sector with unobservable error correlation strictly reduces equilibrium price informativeness.

The two results are complementary, not redundant. Proposition 3 is the more foundational result: it identifies the single scalar condition that determines whether the standard antitrust intuition for information goods applies or inverts. Theorem 4 composes that characterization with a welfare ranking of vendor market structures and delivers the market-level conclusion. In the regime $\tau_S < \min_y \tau_S^*(y)$, the theorem's ranking inverts and competition raises informativeness. The paper does not claim antitrust inversion unconditionally; it characterizes the conditions under which the inversion holds, the conditions under which it reverses, and the scalar threshold that separates the two regimes.

Mechanism. Three forces structure the equilibrium. (i) Under hidden correlation, a competitive vendor that deviates toward a lower ρ cannot make the deviation visible to buyers at purchase; the deviation raises cost but does not raise willingness-to-pay. Any symmetric PBE therefore has vendors cost-minimizing at the boundary $\rho^C = \bar{\rho}$, with free entry clearing fees at cost. The argument is Akerlof's lemons logic transposed into rational expectations equilibrium: the hidden-action dimension is error correlation rather than product quality, but the unraveling is the same. (ii) A monopoly vendor, whose ρ choice is common knowledge in equilibrium via single-vendor public commitment, partially internalizes buyers' willingness-to-pay for lower ρ through the fee it can charge. The monopolist does not internalize uninformed traders' welfare, yielding $\rho^M \in (0, \bar{\rho})$ above the planner's choice. (iii) A constrained planner internalizes both informed willingness-to-pay and uninformed welfare, yielding $\rho^{SP} < \rho^M$.

Whether a lower ρ raises or lowers informativeness depends on the balance between two signal-aggregation channels: common-factor aggregation (higher ρ concentrates a larger common component, making the nuisance dimension more extractable from the price) and idiosyncratic-noise aggregation (lower ρ allows a law-of-large-numbers average of signals to reveal the fundamental). When τ_S is large, individual signals are informative enough that the idiosyncratic channel dominates and differentiation raises informativeness; when τ_S is small, the common-factor channel dominates and correlation raises informativeness. Proposition 3 is the exact balance point.

A bounded wisdom-of-crowds theorem. Under hidden correlation, large-n_S price informativeness is bounded above by $\tau_{\phi,\infty}(\rho) < \tau_\eta/\rho^2$ for every $\rho > 0$: aggregation does not eliminate the common-factor component of error. The ceiling is the reciprocal of an explicit cubic's positive root, and the classical Grossman-Stiglitz $\tau_{v|P} \to \infty$ limit is recovered only at the corner $\rho = 0$. This provides the quantitative anchor for the market-structure comparison: the monopolist operates strictly below the ceiling, and the competitive corner pushes informativeness further from it.

Relation to the literature. The closest paper is Admati and Pfleiderer [1986], which studies a monopolist information seller in a Kyle-type strategic setting with exogenous signal correlation. Admati and Pfleiderer's monopolist differentiates signals to separate types; they analyze monopoly only, take correlation exogenous, and do not compare market structures. I endogenize ρ as the vendor's technology choice, compare monopoly against competition, and embed the analysis in a

CARA-normal REE with endogenous adoption. Skreta and Veldkamp [2009] show competition among credit-rating agencies produces lower-quality ratings through *buyer-side* ratings shopping. The mechanism is *seller-side* moral hazard on a hidden action. The forces push informativeness in the same direction but operate through orthogonal channels: Skreta and Veldkamp's issuers select rating agencies ex post on generated inflation, whereas the vendors choose hidden ρ ex ante under non-verifiability. Ozsoylev and Walden [2011] and Lou et al. [2019] take signal correlation exogenous in network-based REE settings; Kawanishi [2011] allows investors to choose among exogenous signal types. Cong et al. [2024] study Cournot-style data-market competition with exogenous ρ. Farboodi et al. [2024] value data under a monopoly-seller assumption and do not characterize the cross-structure ranking. Dugast and Foucault [2018, 2025] study precision substitution (raw versus processed signals); the ρ-escalation channel is orthogonal. García and Strobl [2011] derive preference-based complementarity in information acquisition. Pavan et al. [2025] study endogenous information acquisition with independent signals in a partial-equilibrium setting; the buyer-side neutrality result (Theorem 2) is the correlated-signal analogue. The hidden-action mechanism rests on Akerlof [1970], applied here to an REE information-production setting in which hidden quality is error correlation rather than product quality.

Policy implication, scoped. Under Assumption 2, disclosure of cross-subscriber error dispersion (through model-documentation standards, mandated error-dispersion backtests, or third-party audits) restores the monopoly outcome by rendering ρ observable to buyers. Disclosure therefore dominates antitrust concentration as a policy instrument in the regime $\tau_S > \tau_S^*$. Below τ_S^*, the antitrust instrument recovers its classical direction. The paper reports the model's prediction under explicit scope conditions; it does not advocate a general policy recipe. The scoping connects to the Financial Stability Board's 2025 report on monitoring AI-related vulnerabilities in the financial sector [Financial Stability Board, 2025], which identifies model homogenization across financial-sector users as a material concern.

Roadmap. Section 2 presents the model. Section 3 characterizes the buyer-side equilibrium at a fixed technology, proves the bounded wisdom-of-crowds theorem, and establishes buy-side neutrality in the two-regime economy. Section 4 solves the competitive-vendor game under hidden correlation. Section 5 develops Proposition 3 and the informativeness ranking, Theorem 4. Section 6 sketches a three-regime extension with a proprietary-signal channel and states three numerical observations (framed as open conjectures because the analytic proofs are not closed). Section 7 discusses robustness, testable predictions, and policy. Section 8 concludes. Proofs not in the main text are in Appendix A; numerical grids and calibration details are in Appendices B and C.

2 Model

2.1 Primitives

A unit mass of ex-ante identical investors $i \in [0,1]$ have CARA utility with risk-aversion coefficient $a > 0$. A single risky asset in zero net supply pays $v \sim N(0, 1/\tau_v)$ with $\tau_v > 0$; a riskless asset pays gross return 1. Noise trade $z \sim N(0, 1/\tau_z)$ with $\tau_z > 0$ is orthogonal to v. A latent common factor $\eta \sim N(0, 1/\tau_\eta)$ with $\tau_\eta > 0$ is orthogonal to (v, z); η represents the common source of error shared across subscribers to the data vendor's output and is not a payoff-relevant state.

2.2 Signal technology and the data vendor

A single data vendor produces a shared information technology S at per-unit cost $c(\rho)$ and sells it at subscription fee p. Investor i who subscribes receives the signal

$$s_i^S = v + \rho\eta + \sqrt{1-\rho^2}\,\xi_i^S, \qquad \xi_i^S \overset{\text{iid}}{\sim} N(0, 1/\tau_S), \tag{1}$$

independent of (v, z, η) and of the other subscribers' idiosyncratic components. An investor who does not subscribe is *uninformed* (U); she observes only the price.

The correlation parameter $\rho \in [0, \bar{\rho}]$ with $\bar{\rho} \in (0,1)$ is the vendor's technology choice. It indexes the cross-sectional correlation of subscribers' signal errors: with general $\tau_\eta > 0$, $\mathrm{Cov}(s_i^S - v, s_j^S - v) = \rho^2/\tau_\eta$ for distinct i, j, so the within-signal correlation ρ^2/τ_η divided by $\rho^2/\tau_\eta + (1-\rho^2)/\tau_S$ is the unitless measure. Higher ρ means a larger common-factor loading on the shared input. (The formula in earlier drafts that normalized $\tau_\eta = 1$ is a special case; the model does not require this normalization, and I carry τ_η as a free parameter throughout.)

Assumption 1 (Cost of differentiation)**.** The cost function $c : [0, \bar{\rho}] \to \mathbb{R}_{++}$ is C^2, with $c'(\rho) < 0$ and $c''(\rho) > 0$ on $[0, \bar{\rho}]$. Producing a less-correlated signal is more expensive, with increasing marginal cost of differentiation.

All theoretical results in Sections 3-5 use only Assumption 1 together with the local-unimodality Assumption 4 introduced in Section 5. The exponential parametrization $c(\rho) = c_0 e^{-\kappa\rho}$ is used only for numerical calibration in Section 7 and Appendix C. Assumption 1's economic content is that lowering ρ requires per-subscriber fine-tuning, per-subscriber noise injection, or bespoke curation, each of which is costly relative to a single shared model output. The local-unimodality condition (Assumption 4) is strictly stronger than $c''(\rho) > 0$: it requires the composite $\Phi_1(1; \rho) - c(\rho)$ to have a unique interior argmax with local concavity there. A sufficient primitive condition is that the cost-curvature parameter κ exceeds a threshold $\kappa_*(\theta)$ defined implicitly by $d^2(\Phi_1(1;\rho) - c(\rho))/d\rho^2 < 0$ on $[0, \bar{\rho}]$. At canonical parameters, $\kappa_* \approx 1.04$; the calibration $\kappa = 1.5$ used throughout satisfies this with margin (Appendix C.1).

2.3 Hidden correlation

Assumption 2 (Hidden correlation)**.** Buyers observe the subscription fee p at the subscription stage but cannot verify the realized error correlation ρ. Buyers can, in principle, infer the equilibrium distribution of ρ ex post from the cross-sectional dispersion of subscribers' signals, but not at the point of purchase. For the monopolist, ρ is an announced product attribute that becomes common knowledge in equilibrium via single-vendor public commitment (a single long-term entity that contracts on ρ through a one-shot, observable commitment); Assumption 2 bites only in the competitive vendor subgame.

Assumption 2 is the load-bearing restriction. Its natural domain is the novel, opaque, or newly-entered segment of the data-vendor market: foundation-model API wrappers whose training-data curation is not auditable, alternative-data vendors whose cross-subscriber error dispersion has no accumulated track record, and AI-based signal providers whose cross-client output correlation is not documented to subscribers at the contract stage. In long-horizon relationships with observable ex-post dispersion, reputation may attenuate the hidden-action force; Section 7 returns to the reputation-dynamics question.

On the asymmetric commitment between monopoly and competition. The treatment of the monopolist's ρ as publicly committed and the competitive sector's ρ as hidden is not an asserted asymmetry but a structural consequence of the two market structures. A single long-term vendor with a single contract commits to ρ in the ordinary sense: a one-shot entity's single announcement is credible under repeated-interaction reputation and under ordinary contract enforcement on the product attribute. The competitive sector's non-commitment (Assumption 2) is not an assumption about what any individual competitive vendor could announce. It is an assumption about what is collectively credible in an entry game with a continuum of potential entrants: any individual entrant has a unilateral incentive to deviate from an announced ρ to save cost, and no single entrant's announcement is credible to buyers without third-party audit or an external documentation standard. A disclosure policy (Section 7) is precisely the intervention that converts the competitive market from the hidden-action benchmark to the observable-ρ benchmark. Partial observability, in which ρ is revealed through a noisy public signal at the subscription stage, is an intermediate case: it interpolates between the two regimes and is discussed in Section 7.5.

2.4 Timing

The game has four stages.

Stage 0 (vendor). Vendor(s) choose (ρ, p). Under monopoly, the vendor announces a single pair. Under competition, a continuum of potential entrants make simultaneous independent choices; see Section 4.1.

Stage 1 (adoption). Each investor observes p. Under monopoly she also observes the announced ρ; under competition she forms a posterior belief about ρ from the equilibrium strategy profile. She chooses S or U.

Stage 2 (trade). Each investor submits a demand schedule $x_i(P)$ based on her information. The market clears: $\int_0^1 x_i(P)\,di + z = 0$.

Stage 3 (payoff). v realizes; investors consume their wealth.

2.5 Equilibrium concept

At a fixed pair (ρ, p), a *linear NREE with endogenous adoption* is a tuple (n_S, B, C, D) such that: (i) the equilibrium price takes the linear form $P = Bv + Cn_S\eta + Dz$; (ii) each investor's demand is CARA-optimal given her posterior; (iii) the market clears; (iv) $n_S \in [0, 1]$ is the measure of S-adopters; (v) at interior $n_S \in (0, 1)$, the adoption indifference condition $\Phi_1(n_S; \rho) = p$ holds, where Φ_1 is the certainty-equivalent gain from subscribing.

The full-game equilibrium is *perfect Bayesian*: the vendor's (ρ, p) choice, buyers' subscription decisions, and trade strategies are mutual best responses, and buyer beliefs about ρ are consistent with Bayes' rule on the equilibrium path. Off-path beliefs, where they matter, are specified in Section 4.1.

The continuum formulation relies on a cross-sectional law of large numbers. For the *i.i.d.* components I use the construction of Sun [2006]: the integral $\int_0^1 \xi_i^S\,di = 0$ almost surely under the Fubini extension.

2.6 Welfare criterion

The welfare criterion is the standard CARA certainty-equivalent aggregate:

$$W(\rho, n_S) = -\frac{1}{2a}\left[n_S \log V_S(\rho, n_S) + (1 - n_S)\log V_U(\rho, n_S)\right] - n_S\, c(\rho), \tag{2}$$

where V_S and V_U are the posterior conditional variances of v for subscribers and non-subscribers, respectively, derived in Section 3. The fee p is a pure transfer from buyers to the vendor and drops from the aggregate; the vendor's profit $\pi = n_S(p - c(\rho))$ is added back when aggregating across agents, leaving (2). The planner in Section 5 is a *constrained* planner who chooses ρ taking the buyer-side equilibrium as given; the resulting benchmark is constrained-Pareto, not first-best.

2.7 Notation conventions

I write $b_1 \equiv \tau_S/(1 - \rho^2)$ for the effective precision of the private-signal component on v given the common-factor loading, and $b_2 \equiv \tau_z/\mu^2$ for the effective precision of the price on v where μ is the price-to-noise loading derived below. I use $y \equiv \rho^2$ as a change of variable in the threshold analysis. Throughout, $\theta := (\tau_\eta, \tau_z, a)$ collects the structural parameters that enter the threshold τ_S^*.

3 Buyer-side equilibrium and the wisdom-of-crowds ceiling

This section fixes a vendor-stage choice (ρ, p) and characterizes the buyer-side continuation equilibrium. Four results follow: closed-form posterior statistics (Lemma 1), a cubic fixed point for the price-to-noise loading (Lemma 2), existence and uniqueness of the interior adoption equilibrium (Theorem 1), and the bounded wisdom-of-crowds theorem (Theorem 3). A corollary establishes that the equilibrium is not observationally equivalent to Grossman-Stiglitz augmented by a public signal, and a buy-side neutrality theorem (Theorem 2) isolates the externality on the vendor side.

3.1 Posterior statistics

Conjecture a linear price $P = Bv + Cn_S\eta + Dz$. Let $\phi \equiv P/B = v + \lambda\eta + \mu z$ with $\lambda \equiv Cn_S/B$ and $\mu \equiv D/B$. Price precision for an uninformed agent is

$$\tau_\phi = \frac{1}{\lambda^2/\tau_\eta + \mu^2/\tau_z}. \tag{3}$$

An uninformed agent's posterior on v has variance $V_U = 1/(\tau_v + \tau_\phi)$ and mean $m_U = w_U\phi$ with $w_U = \tau_\phi/(\tau_v + \tau_\phi)$.

A subscriber observes (s_i^S, ϕ). Write $b_1 = \tau_S/(1-\rho^2)$ and $b_2 = \tau_z/\mu^2$. The posterior-precision matrix for (v, η) is

$$K = \begin{pmatrix} \tau_v + b_1 + b_2 & \rho b_1 + \lambda b_2 \\ \rho b_1 + \lambda b_2 & \tau_\eta + \rho^2 b_1 + \lambda^2 b_2 \end{pmatrix}. \tag{4}$$

Lemma 1 (Posterior variance and weights)**.** *$K \succ 0$. The subscriber's posterior variance of v is $V_S = K_{22}/\det K$, with posterior-mean weights $w_S^{(s)} = b_1(K_{22} - \rho K_{12})/\det K$ on s_i^S and $w_S^{(\phi)} = b_2(K_{22} - \lambda K_{12})/\det K$ on ϕ.*

Proof. See Appendix A, Section A.1. □

3.2 The price fixed point

Standard CARA aggregation with market clearing and Lemma 1 yield a fixed-point condition for (λ, μ). Define the regularity condition

$$(\mathrm{P}'): \quad \rho^2 < \min\left(1, \frac{4\tau_\eta a^2 n_S^2}{\tau_z \tau_S^2}\right). \tag{5}$$

Lemma 2 (Cubic fixed point)**.** *Fix $n_S \in (0,1)$, $\rho \in [0,1)$, and suppose (5). There is a unique $(\lambda^*, \mu^*) \in \mathbb{R} \times (0, \infty)$ satisfying the price-consistency conditions: $\lambda^* = \rho$, and μ^* is the unique positive root of*

$$\tau_\eta b_1 \mu^3 - (a/n_S)(\tau_\eta + \rho^2 b_1)\mu^2 - (a\rho^2\tau_z/n_S) = 0. \tag{6}$$

Proof. See Appendix A, Section A.2. □

The fixed point $\lambda^* = \rho$ states that the common factor η loads into the price through subscribers' signals exactly at the rate at which it loads into their private information.

Lemma 3 (Comparative statics in subscription mass)**.** *Under* (5)*:* $\partial\tau_\phi/\partial n_S > 0$*,* $\partial V_U/\partial n_S < 0$*,* $\partial V_S/\partial n_S < 0$*.*

Proof. See Appendix A, Section A.3. □

3.3 The equilibrium is not a GS economy with a public signal

A natural referee concern is that the correlated private-signal model collapses to the classical Grossman-Stiglitz economy augmented by a public noisy signal of v. It does not.

Lemma 4 (Not-a-public-signal)**.** *For any* $\rho \in (0,1)$ *and* $n_S \in (0,1]$*, the equilibrium of Section 3.2 is not observationally equivalent to a Grossman-Stiglitz economy augmented by an exogenous noisy public signal of* v*.*

Proof. The observational content is captured by $\mathrm{Cov}(\mathbb{E}[v \mid P], \eta)$. In the present model,

$$\mathrm{Cov}(\mathbb{E}[v \mid P], \eta) = \frac{\tau_\phi}{\tau_v + \tau_\phi} \cdot \frac{\lambda}{\tau_\eta} = \frac{\tau_\phi \rho}{(\tau_v + \tau_\phi)\tau_\eta} > 0,$$

whereas in any Grossman-Stiglitz economy augmented by a public signal of v alone, $\mathrm{Cov}(\mathbb{E}[v \mid P], \eta) = 0$ because η does not enter the augmented price. □

The price reveals information in a genuinely two-dimensional (v, η) span; an auxiliary public signal of v operates only in the v-dimension.

3.4 Existence of the interior adoption equilibrium

Define the subscriber's certainty-equivalent gain from subscribing:

$$\Phi_1(n_S; \rho) := \frac{1}{2a} \log \frac{V_U(n_S, \rho)}{V_S(n_S, \rho)}. \tag{7}$$

Interior adoption equilibria are characterized by $\Phi_1(n_S; \rho) = p$. Existence requires an interior-adoption condition: denote by $\bar{\rho}^{(1)}$ the largest value of ρ for which (5) holds at $n_S = 1$, and use

$$\text{(T1.cond):} \quad \Phi_1(1^-; \rho) < p < \Phi_1(0^+; \rho). \tag{8}$$

Theorem 1 (Existence and uniqueness of interior adoption)**.** *For every* $\rho \in (0, \bar{\rho}^{(1)})$ *and* p *satisfying* (8)*, the equation* $\Phi_1(n_S; \rho) = p$ *has a unique solution* $n_S^{eq}(\rho, p) \in (0,1)$*, continuously differentiable in* (ρ, p)*.*

Proof. See Appendix A, Section A.4. □

3.5 Buy-side neutrality in the two-regime economy

Fix (ρ, c_S) and define the aggregate welfare derivative with respect to n_S:

$$X(n_S;\rho) := n_S \frac{\partial \log V_S}{\partial n_S} + (1-n_S)\frac{\partial \log V_U}{\partial n_S}. \tag{9}$$

The sign of X characterizes the direction of the buyer-side Pigouvian wedge: $X < 0$ means that a marginal increase in subscription raises aggregate CARA certainty equivalent through the precision channel. In the two-regime (shared-versus-uninformed) economy of Section 2, the sign is unambiguous.

Theorem 2 (Buy-side neutrality, two-regime)**.** *In the two-regime economy, for every $\rho \in (0, \bar{\rho}^{(1)})$ and $n_S \in (0,1)$, $X(n_S;\rho) < 0$. The planner uniformly wants more S-adoption than the market supplies at any fixed (ρ, c_S): the equilibrium indifference condition $\Phi_1(n_S;\rho) = c_S$ coincides (up to the monetary transfer) with the planner's first-order condition in n_S* only *when the fee equals marginal cost. No buy-side wedge operates at fixed technology.*

Proof. See Appendix A, Section A.5. □

Theorem 2 is a negative result: it locates the externality. In the two-regime economy, the Pigouvian wedge on adoption is *not* the mechanism through which market structure matters. The mechanism operates entirely through the vendor's technology choice. A proprietary-signal channel, shut down here by assumption, would plausibly restore a classical buy-side wedge on the proprietary margin without disturbing the vendor-side wedge analyzed in Sections 4-5; Section 6 returns to the proprietary-channel wedge.

3.6 The wisdom-of-crowds ceiling

As $n_S \to 1^-$, the cubic (6) limits to

$$\tau_\eta b_1 \mu^3 - a(\tau_\eta + \rho^2 b_1)\mu^2 - a\rho^2\tau_z = 0. \tag{10}$$

Theorem 3 (Bounded wisdom of crowds)**.** *Fix $\rho \in (0,1)$ and $\tau_z > 0$. As $n_S \to 1^-$:*

(a) $\lambda^* \to \rho$;

(b) $\mu^*(n_S;\rho) \to \mu_\infty(\rho;\tau_S) > 0$, the unique positive root of (10);

(c) $\tau_\phi \to \tau_{\phi,\infty}(\rho;\tau_S) = 1/(\rho^2/\tau_\eta + \mu_\infty^2/\tau_z)$;

(d) $\tau_{\phi,\infty}(\rho) \in (0, \tau_\eta/\rho^2)$ strictly;

(e) $\tau_{\phi,\infty}$ is continuous on $(0,1)$, with $\lim_{\rho\downarrow 0}\tau_{\phi,\infty}(\rho) = \tau_z\tau_S^2/a^2$.

Proof. See Appendix A, Section A.6. □

The ceiling $\tau_{\phi,\infty}(\rho) < \tau_\eta/\rho^2$ encapsulates the failure of the wisdom-of-crowds limit: under positive ρ, private signals share a common component that survives aggregation. Only at the corner $\rho = 0$ does the classical $\tau_{v|P} \to \infty$ limit recover, via the limiting value $\tau_z \tau_S^2/a^2 \cdot 1 = \tau_z \tau_S^2/a^2$ which diverges in τ_S. Under any positive ρ, no amount of aggregation can drive price informativeness past τ_η/ρ^2.

Corollary 1 (Data abundance at the buyer margin)**.** *Under* (5) *and* (8)*,* $d\tau^{eq}_{v|P}/dp < 0$*: holding* ρ *fixed, cheaper shared data raises equilibrium price informativeness.*

Proof. See Appendix A, Section A.7. □

Corollary 1 is the standard channel emphasized by Dugast and Foucault [2018]: at a fixed vendor technology, cheaper data raises adoption and informativeness. Section 5 shows that cheaper competitive data can reverse the direction because cost reductions in the competitive regime are achieved by cost-minimizing on differentiation, raising ρ.

4 The competitive-vendor game under hidden correlation

This section solves the vendor stage under Assumption 2 and establishes that competition unravels to the highest-correlation, lowest-cost corner. Section 5 then compares against the monopolist and the constrained planner.

4.1 The competitive game G^C

The competitive game G^C is played by a continuum of potential entrants indexed by $j \in [0, 1]$ (use the Sun [2006] measurability construction for the continuum). Each entrant's strategy is a pair $(\rho_j, p_j) \in [0, \bar{\rho}] \times \mathbb{R}_+$, chosen simultaneously in Stage 0. Following Stage 0, each buyer observes the vector of *fees* $\{p_j\}$ but, by Assumption 2, not the vector of correlations $\{\rho_j\}$. Buyers form a posterior belief about ρ_j from the equilibrium strategy profile and subscribe to at most one vendor (a standing assumption in the symmetric environment of this paper; identical subscriber signals from multiple vendors add no information).

Free-entry protocol. I model the competitive sector as a continuum of potential entrants each paying a vanishing fixed cost $\varepsilon > 0$ to enter. In the limit $\varepsilon \downarrow 0$, the symmetric PBE fee tends to marginal cost $c(\rho)$ via standard Bertrand undercutting at observable fees: any candidate symmetric profile with $p > c(\rho)$ admits an entrant who undercuts and attracts the full subscription mass. Off-path beliefs about ρ_j at fees not on the equilibrium path are *pessimistic* (the buyer assigns probability 1 to $\rho_j = \bar{\rho}$), which is the standard Bertrand-with-product-quality selection and supports the free-entry equilibrium as the symmetric limit. I restrict attention to symmetric perfect Bayesian equilibria throughout.

I require an interiority condition at the boundary.

Assumption 3 (Slack participation at the boundary)**.** $\Phi_1(1; \bar{\rho}) > c(\bar{\rho})$. Equivalently, at the worst technology and full adoption, the subscription remains strictly profitable for buyers at cost-based pricing. This condition is invoked in identical form in Section 4.1, Section 5, and the appendix.

Proposition 1 (Competitive PBE under hidden correlation)**.** *Under Assumptions 1-3 and the buyer-stage regularity (5), in the $\varepsilon \downarrow 0$ free-entry limit, the unique symmetric* pooling *perfect Bayesian equilibrium of G^C satisfies*

$$\rho^C = \bar{\rho}, \qquad p^C = c(\bar{\rho}). \tag{11}$$

Proof sketch (full proof in Appendix A, Section A.8). The argument is a hidden-action unraveling in the style of Akerlof [1970].

Step 1 (hidden-action deviation). Fix any candidate symmetric PBE (ρ^*, p^*) with $\rho^* < \bar{\rho}$. Consider a deviator j who sets $(\rho_j, p_j) = (\bar{\rho}, p^*)$ — charging the same fee but producing at the lowest-cost technology. By Assumption 2, buyers cannot detect ρ_j at the subscription stage; their posterior over ρ_j conditional on observing $p_j = p^*$ is pinned by the equilibrium belief, which under symmetric play assigns mass 1 to $\rho_j = \rho^*$. Buyers therefore subscribe to j in identical proportion to the equilibrium vendor, yielding identical revenue but strictly lower cost. The deviation is strictly profitable by $c(\bar{\rho}) < c(\rho^*)$ from Assumption 1. Any symmetric PBE therefore features $\rho^* = \bar{\rho}$.

Step 2 (zero-profit pricing). Under free entry, any candidate symmetric PBE with $p^* > c(\bar{\rho})$ admits a deviator who sets $(\bar{\rho}, p^* - \varepsilon)$ and attracts all buyers (the lower fee dominates at identical posterior beliefs about ρ). The only price consistent with zero-profit entry is $p^* = c(\bar{\rho})$.

Step 3 (buyer-side continuation). At $(\bar{\rho}, c(\bar{\rho}))$ with Assumption 3, the buyer-side continuation equilibrium is an interior $n_S \in (0, 1)$ (or the corner $n_S = 1$ if Slack holds with strict inequality at full adoption, which it does). Theorem 1 delivers uniqueness.

Off-path beliefs and zero-measure-deviation convention. Step 1 does not invoke D1 or divinity. The hidden-action deviation is invisible by Assumption 2; the buyers' posterior at $p_j = p^*$ is pinned by the equilibrium belief, not by a refinement. I adopt the standard continuum-game convention that unilateral deviations by a single (measure-zero) vendor leave the buyers' posterior distribution of ρ across active vendors unchanged, so buyers' beliefs at on-path fees are anchored to the equilibrium (degenerate-at-ρ^*) distribution. This belief-consistency convention is standard in continuum auction and search games and is sometimes called the "no-single-agent-can-move-beliefs" convention.

Uniqueness among pooling equilibria and ruling out separating equilibria. Under Assumption 2, fees are observable at purchase but ρ is not; the causal channel for belief formation about ρ is therefore the set of possible offers rather than any realized ρ-signal at the subscription stage. Consider a candidate *separating* PBE in which different vendor types post different fees, so that buyer beliefs about ρ_j update on p_j. For the separation to be incentive-compatible, each vendor type must strictly prefer its own fee to any other type's fee given buyers' updating. But the cost of producing at ρ_j is sunk before fees are set and does not affect the vendor's continuation payoff from a given fee-belief pair: conditional on any buyer belief about ρ_j, every vendor type strictly prefers to produce at $\bar{\rho}$ (lowest cost). Hence any candidate separating profile unravels in the same way Step 1 unravels pooling at $\rho^* < \bar{\rho}$: the vendor who was supposed to produce at a low-ρ fee can post the same fee and produce at $\bar{\rho}$, collecting the equilibrium revenue at strictly lower cost. I therefore restrict attention to symmetric pooling PBE; within that class, Steps 1–3 establish uniqueness. Standard refinements such as D1 or the intuitive criterion are not required because the hidden-action argument does not rely on off-path beliefs about ρ_j at on-path fees. □

Proposition 1 isolates the mechanism: hidden correlation transforms the vendor's action into an unobservable quality choice, and Akerlof's lemons logic drives the competitive outcome to the corner.

4.2 What happens if correlation is observable

Relaxing Assumption 2 turns the competitive game into Bertrand-with-observable-product-attribute. The appropriate solution concept is a symmetric subgame-perfect equilibrium with free entry where each vendor chooses (ρ_j, p_j), buyers observe both, and subscription is voluntary. With ρ observable, no vendor wants to post (ρ_j, p_j) with $\rho_j > \rho^M$: at any such profile a deviator posting $(\rho^M, c(\rho^M) + \delta)$ with $\delta > 0$ sufficiently small attracts all buyers because $\Phi_1(1; \rho^M) - c(\rho^M) > \Phi_1(1; \rho_j) - c(\rho_j)$ whenever ρ^M is the monopolist's argmax of $\Phi_1(1; \rho) - c(\rho)$. Symmetrically, no vendor wants $\rho_j < \rho^M$ either: a deviator can imitate and undercut. The unique symmetric free-entry equilibrium therefore has $\rho_j = \rho^M$ for all active vendors, with fees at marginal cost $p = c(\rho^M)$ by the standard Bertrand argument at given quality. Hence $\rho^C = \rho^M$ when ρ is observable, and the ranking of Theorem 4 collapses to $\rho^{SP} < \rho^M = \rho^C$. Assumption 2 is the load-bearing scope of the antitrust inversion: the inversion is a consequence of non-verifiability, not of competition *per se*.

This observation is the basis of the policy implication in Section 7: disclosure of cross-subscriber error dispersion (model-documentation standards, error-correlation audits) breaks Assumption 2 and collapses the competitive outcome to the monopoly outcome.

4.3 Relation to Admati and Pfleiderer and to Skreta and Veldkamp

Admati and Pfleiderer [1986] study a monopolist information seller in a Kyle-type strategic setting with exogenous signal correlation. Their monopolist chooses how much to differentiate signals across types to separate them at the subscription stage, but correlation is a primitive of the model and market structure is never varied. The present paper endogenizes ρ as a vendor technology choice, compares monopoly against competition, and embeds the analysis in a CARA-normal REE with endogenous adoption. The mechanisms do not overlap: Admati and Pfleiderer's monopolist differentiates signals to price-discriminate; the present monopolist chooses ρ to internalize buyer willingness-to-pay against differentiation cost.

Skreta and Veldkamp [2009] show that competition among credit-rating agencies produces lower-quality ratings via issuer-side *ratings shopping*: issuers select ex post among competing agencies on the rating inflation observed, and competition selects for inflationary rating technologies. The present mechanism is seller-side hidden action: vendors choose an action (ρ) that buyers cannot detect, and cost minimization drives the action to the boundary. The forces push in the same direction — competition reduces information quality — through orthogonal channels. Skreta and Veldkamp's analysis does not apply under observable ratings; the present analysis does not apply under observable ρ. In each case the informational opacity is the load-bearing feature, but the opacity operates on different margins.

5 The τ_S threshold and the informativeness ranking

This section develops the two main results: the closed-form threshold characterization (Proposition 3) and the market-structure informativeness ranking (Theorem 4). The latter composes the former with a welfare ranking of vendor market structures (Proposition 4).

5.1 The welfare identity

I start with an algebraic identity that links aggregate welfare to monopoly profit and uninformed welfare.

Proposition 2 (Welfare identity)**.** *For every (ρ, n_S) satisfying the buyer-side equilibrium of Section 3,*

$$W(\rho, n_S) = \pi(\rho, n_S) - \frac{1}{2a} \log V_U(\rho, n_S), \tag{12}$$

where $\pi(\rho, n_S) = n_S[\Phi_1(n_S; \rho) - c(\rho)]$ is the vendor's profit per unit of subscription at the indifference fee.

Proof. See Appendix A, Section A.9. □

The identity (12) decomposes welfare into two terms: vendor profit (which the monopolist maximizes) and uninformed welfare, monotone in $\tau_{\phi,\infty}$ (which neither competitive nor monopoly vendors internalize directly). The wedge between the monopolist and the planner is exactly $-(1/(2a)) \log V_U$, the uninformed-welfare term.

5.2 The τ_S threshold: Proposition 3

Let $y := \rho^2$ and recall the large-n_S cubic

$$F(\mu, y; \tau_S, \theta) := \tau_\eta b_1(y)\mu^3 - a[\tau_\eta + y b_1(y)]\mu^2 - ay\tau_z = 0, \qquad b_1(y) = \tau_S/(1-y). \tag{13}$$

I want to characterize the sign of $d\tau_{\phi,\infty}/d\rho$ in terms of primitive parameters.

5.2.1 Sign equivalence

Differentiating the ceiling $\tau_{\phi,\infty} = 1/(\rho^2/\tau_\eta + \mu_\infty^2/\tau_z)$ in ρ and using implicit differentiation of (13) yields

$$\operatorname{sign}\left(\frac{d\tau_{\phi,\infty}}{d\rho}\right) = -\operatorname{sign}\left(\frac{1}{\tau_\eta} + \frac{2\mu_\infty}{\tau_z} \cdot \frac{d\mu_\infty}{dy}\right). \tag{14}$$

Substituting $d\mu_\infty/dy$ from implicit differentiation of F and clearing positive factors produces the primitive sign-determining function

$$G(y,\mu;\tau_S,\theta)=\tau_z\tau_\eta\tau_S\mu^3(1-y)+2ay\tau_z^2(1-y)^2+2a\tau_\eta\tau_z\mu^2(1-y)^2-2\tau_\eta\tau_S\mu^4(\tau_\eta\mu-a). \quad (15)$$

The first step of the characterization is that G and $d\tau_{\phi,\infty}/d\rho$ have opposite signs.

Lemma 5 (Positivity of $\tau_\eta\mu_\infty - ay$)**.** *At $\mu=\mu_\infty(y;\tau_S)$, $\tau_\eta\mu_\infty > ay$.*

Proof. The cubic F has positive leading coefficient $\tau_\eta b_1 > 0$, so $F(+\infty,y)=+\infty$. Evaluate F at $\mu=0$: $F(0,y)=-ay\tau_z<0$ for $y>0$. Evaluate F at $\mu=ay/\tau_\eta$:

$$F(ay/\tau_\eta,y)=\tau_\eta b_1\cdot\frac{a^3y^3}{\tau_\eta^3}-a[\tau_\eta+yb_1]\cdot\frac{a^2y^2}{\tau_\eta^2}-ay\tau_z=-ay\left(\frac{a^2y}{\tau_\eta}+\tau_z\right)<0$$

for $y>0$. Using $F=0$ to substitute $a[\tau_\eta+yb_1]\mu^2=\tau_\eta b_1\mu^3-ay\tau_z$, one computes $\partial F/\partial\mu=(\tau_\eta b_1\mu^3+2ay\tau_z)/\mu>0$ at every positive zero of F (equation (†)). Combined with $F(+\infty)=+\infty$, $F(0,y)<0$, and $F(ay/\tau_\eta,y)<0$, the cubic has a *unique* positive real root $\mu_\infty(y)$ and that root satisfies $\mu_\infty(y)>ay/\tau_\eta$ strictly; equivalently $\tau_\eta\mu_\infty>ay$. □

Proposition 3 (Threshold characterization)**.** *Fix $\theta=(\tau_\eta,\tau_z,a)>0$ and $\bar\rho\in(0,1)$. Let G be as in* (15) *and $\mu_\infty(y;\tau_S)$ be the positive root of* (13)*.*

(i) Sign equivalence. *$d\tau_{\phi,\infty}/d\rho<0$ if and only if $G(y,\mu_\infty(y;\tau_S);\tau_S,\theta)>0$. Equivalently, $d\tau_{\phi,\infty}/d\rho<0$ if and only if $N(\mu_\infty,y;\theta)>0$, where*

$$N(\mu,y;\theta):=\tau_z(1-y)(3\tau_\eta\mu-2ay)-2\tau_\eta\mu^2(\tau_\eta\mu-a).$$

(ii) Existence and uniqueness. *For each $y\in[0,1)$, there is a unique $\tau_S^*(y;\theta)\in(0,\infty)$ such that $d\tau_{\phi,\infty}/d\rho<0$ if and only if $\tau_S>\tau_S^*(y;\theta)$.*

(iii) Uniform threshold. *Define $\tau_S^*(\bar\rho^2;\theta):=\max_{y\in[0,\bar\rho^2]}\tau_S^*(y;\theta)$, attained by Weierstrass on the compact set $[0,\bar\rho^2]$ by continuity of $\tau_S^*(y;\theta)$ in y. Then $d\tau_{\phi,\infty}/d\rho<0$ uniformly on $[0,\bar\rho^2]$ if and only if $\tau_S>\tau_S^*(\bar\rho^2;\theta)$.*

(iv) Closed form at $\rho=0$.

$$\tau_S^*(0;\theta)=\frac{2a\tau_\eta}{a+\sqrt{a^2+6\tau_z\tau_\eta}}, \quad (16)$$

with $\partial\tau_S^*(0)/\partial\tau_\eta>0$, $\partial\tau_S^*(0)/\partial a>0$, $\partial\tau_S^*(0)/\partial\tau_z<0$.

(v) Primitive sufficient condition. *Let (V**) denote the condition $\tau_\eta\mu_\infty(y;\tau_S)<a+a\tau_z(1-y)^2/(\tau_S\mu_\infty^2)$. (V**) evaluated at $(y,\mu_\infty(y;\tau_S))$ implies $G>0$, and so $d\tau_{\phi,\infty}/d\rho<0$. (V**) is strictly sufficient but not necessary.*

Proof. See Appendix A, Section A.10. □

Three features of Proposition 3 deserve comment.

On the closed form. Setting $y = 0$ in $N = 0$ yields, after substituting $\mu_\infty(0; \tau_S) = a/\tau_S$, the quadratic $3\tau_z\tau_S^2 + 2a^2\tau_S - 2a^2\tau_\eta = 0$, whose positive root gives (16). The closed form is sharp: at canonical $(\tau_\eta, \tau_z, a) = (1, 1, 1)$, $\tau_S^*(0) = 2/(1 + \sqrt{7}) \approx 0.549$. The comparative statics at $\rho = 0$ follow from direct differentiation of (16); their signs confirm the economic intuition. Higher common-factor precision τ_η means a lower effective nuisance variance, which requires a more precise private signal before differentiation pays. Higher risk aversion a strengthens the weight on precision in equilibrium demand, again raising the threshold. Higher noise-trade precision τ_z lowers the threshold because the price is already informative through noise-trade precision and further differentiation pays even at lower τ_S.

On the uniform threshold. The uniform-threshold statement in (iii) is what enters Theorem 4. Numerical work (Appendix C) shows $\tau_S^*(y; \theta)$ is monotone in y at canonical parameters, with the maximum attained at $y = \bar{\rho}^2$; the closed-form lower bound $\tau_S^*(0; \theta)$ is therefore a conservative sufficient condition in the more restrictive form $\tau_S > \tau_S^*(\bar{\rho}^2; \theta)$.

Economic interpretation. Two signal-aggregation channels determine the sign of $d\tau_{\phi,\infty}/d\rho$. Channel (i), common-factor aggregation, says higher ρ concentrates the signals' common component, making η more extractable from the price; this channel raises informativeness in ρ. Channel (ii), idiosyncratic aggregation, says lower ρ allows subscribers' idiosyncratic noise to average out via law of large numbers, revealing v more precisely; this channel lowers informativeness in ρ. When τ_S is small, individual signals are noisy, channel (ii) is weak, and channel (i) dominates: higher correlation raises informativeness. When τ_S is large, individual signals are precise, channel (ii) dominates: differentiation raises informativeness. τ_S^* is the exact balance point.

5.2.2 Monotonicity of the ceiling

Lemma 6 (Informativeness monotone in ρ)**.** *If $\tau_S > \tau_S^*(y; \theta)$ at $y = \rho^2$, then $d\tau_{\phi,\infty}/d\rho < 0$ and $d \log V_U(\rho, 1)/d\rho > 0$.*

Proof. The first claim is Proposition 3(i)-(ii). The second follows because $V_U = 1/(\tau_v + \tau_{\phi,\infty})$ is decreasing in $\tau_{\phi,\infty}$, which is decreasing in ρ under the hypothesis. □

5.3 Unimodality and the vendor's problem

The monopolist's problem is $\max_{\rho\in[0,\bar{\rho}]} \Pi^M(\rho)$ with $\Pi^M(\rho) := \Phi_1(1; \rho) - c(\rho)$. The constrained planner's problem is $\max_{\rho\in[0,\bar{\rho}]} W^M(\rho)$ with $W^M(\rho) := \Pi^M(\rho) - (1/(2a)) \log V_U(1; \rho)$ (using the welfare identity at $n_S = 1$, which is the relevant subgame under Assumption 3).

Assumption 4 (Local unimodality with interior argmax)**.** Π^M and W^M are strictly unimodal in ρ on the compact interval $[0, \bar{\rho}]$, with unique interior argmax $\rho^M \in (0, \bar{\rho})$ for Π^M (satisfying $d^2\Pi^M/d\rho^2|_{\rho^M} < 0$) and unique interior argmax $\rho^{SP} \in (0, \bar{\rho})$ for W^M (satisfying $d^2W^M/d\rho^2|_{\rho^{SP}} < 0$). Unimodality plus local concavity at the argmax is what the proofs of Proposition 4 and Theorem 4 below require.

Assumption 4 is weaker than global strict concavity: the theorems that follow need only a unique interior argmax and local concavity at the argmax, not global concavity of Π^M or W^M. A sufficient primitive condition is that the cost-curvature parameter κ exceeds the threshold $\kappa_*(\theta)$ defined in Section 2.2 (implicitly by $d^2(\Phi_1(1;\rho) - c(\rho))/d\rho^2 < 0$ on $[0, \bar{\rho}]$). At canonical parameters $\kappa_* \approx 1.04$, so the calibration $\kappa = 1.5$ satisfies the condition with margin.

Grid verification. I verify Assumption 4 numerically on a 125-point parameter grid over $(\kappa, a, \tau_S) \in \{0.5, 1, 1.5, 2, 3\} \times \{0.5, 1, 1.5, 2\} \times \{0.5, 1, 1.5, 2\}$ (Appendix B). Unimodality of Π^M holds at 100% of grid points, unimodality of W^M holds at 100% of grid points, local concavity of Π^M at ρ^M holds at 93%, and local concavity of W^M at ρ^{SP} holds at 77%. The 7% and 23% failure rates are confined to the high-risk-aversion corner ($a \geq 2$) where the monopolist's argmax ρ^M pins to the boundary $\bar{\rho}$; at such corner points Proposition 4(c) holds trivially by boundary evaluation and Theorem 4 reads as a boundary statement rather than a first-order statement. Accordingly, I state Proposition 4 and Theorem 4 as *local* statements at any (ρ^M, ρ^{SP}) satisfying the interior FOCs with local concavity (that is, inside the verified subset); on the canonical calibration $\kappa = 1.5$ used throughout the numerical illustration, both conditions hold and the theorems apply in their interior form.

5.4 The welfare ranking

Proposition 4 (Welfare ranking of technology choices)**.** *Under Assumptions 1-4, (5), and $\tau_S > \tau_S^*(\bar{\rho}^2; \theta)$:*

(a) $\rho^C = \bar{\rho}$ (Proposition 1).

(b) $\rho^M \in (0, \bar{\rho})$, characterized by $d\Phi_1(1;\rho)/d\rho\big|_{\rho^M} = c'(\rho^M)$.

(c) $\rho^{SP} < \rho^M$ strictly.

(d) $\rho^{SP} < \rho^M < \rho^C = \bar{\rho}$ strictly.

Proof. See Appendix A, Section A.11. □

The key observation behind (c) is the welfare identity: writing $L(\rho) := W^M(\rho) - \Pi^M(\rho) = -(1/(2a)) \log V_U(1;\rho)$, Lemma 6 under the uniform-threshold hypothesis yields $L'(\rho) < 0$ on $(0, \bar{\rho})$. Therefore $dW^M/d\rho\big|_{\rho^M} = L'(\rho^M) < 0$, and unimodality of W^M with interior argmax forces $\rho^{SP} < \rho^M$. Informally, the planner internalizes uninformed welfare, which is increasing in $\tau_{\phi,\infty}$ and therefore decreasing in ρ; at the monopolist's argmax, the uninformed-welfare slope is strictly negative, so the planner's argmax lies to the left.

5.5 The informativeness ranking: Theorem 4

Theorem 4 (Price-informativeness ranking)**.** *Under Assumptions 1-4, (5), and $\tau_S > \tau_S^*(\bar{\rho}^2; \theta)$ uniformly on $[0, \bar{\rho}^2]$ (sufficient primitive form: (V**), i.e., $\tau_\eta \mu_\infty(y; \tau_S) < a + a\tau_z(1-y)^2/(\tau_S \mu_\infty^2)$,* *holds uniformly on* $[0, \bar{\rho}^2]$*),*

$$\tau_{\phi,\infty}(\rho^{SP}) \; > \; \tau_{\phi,\infty}(\rho^{M}) \; > \; \tau_{\phi,\infty}(\rho^{C}), \qquad \textit{strictly.} \tag{17}$$

Proof. Lemma 6 gives $\tau_{\phi,\infty}$ strictly decreasing in ρ on $[0, \bar{\rho}]$ under the uniform threshold. Proposition 4(d) gives $\rho^{SP} < \rho^{M} < \rho^{C}$. Composing yields (17). □

Theorem 4 is the main result: under hidden correlation and sufficient private precision, competition strictly reduces equilibrium price informativeness relative to monopoly, which in turn reduces it relative to the constrained planner. The direction of the antitrust effect inverts relative to the classical information-goods intuition.

5.6 The low-τ_S regime: a characterization, not a caveat

When $\tau_S < \tau_S^*(\bar{\rho}^2; \theta)$, the hypothesis of Theorem 4 fails in part or in full. Two cases.

Case A: partial failure. If $\tau_S^*(0; \theta) < \tau_S < \tau_S^*(\bar{\rho}^2; \theta)$, there is a subset of $[0, \bar{\rho}^2]$ on which $d\tau_{\phi,\infty}/d\rho > 0$. Over that subset, the monotonicity step of Theorem 4 fails, and the composition with $\rho^{SP} < \rho^{M} < \rho^{C}$ does not yield the full informativeness ranking.

Case B: full inversion. If $\tau_S < \min_{y \in [0, \bar{\rho}^2]} \tau_S^*(y; \theta)$, then $d\tau_{\phi,\infty}/d\rho > 0$ *everywhere* on $[0, \bar{\rho}]$. The full ranking then *inverts*:

- Proposition 4(c) reverses: $\rho^{SP} > \rho^{M}$ strictly (the planner prefers higher ρ because higher correlation raises informativeness and uninformed welfare).
- The corner $\rho^{SP} = \bar{\rho} = \rho^{C}$ becomes possible (the planner may push to the boundary).
- The informativeness ranking becomes $\tau_{\phi,\infty}(\rho^{C}) > \tau_{\phi,\infty}(\rho^{M})$: competition strictly *improves* informativeness relative to monopoly.

This is the low-τ_S regime: very noisy shared signals, where the common-factor aggregation channel dominates and correlation is a desirable feature. At canonical parameters $(\tau_\eta, \tau_z, a) = (1, 1, 1)$, $\tau_S^*(0) \approx 0.549$ and $\tau_S^*(\bar{\rho}^2 = 0.81) \approx 0.70$; the partial-failure region is $\tau_S \in (0.549, 0.70)$, and the full-inversion region is $\tau_S < 0.549$. In the full-inversion region, the standard antitrust direction is restored; the policy prescription flips from disclosure-mandates-to-discipline-competition to antitrust-instruments-to-encourage-competition.

Proposition 3 is therefore a characterization, not a scope caveat. The paper identifies the exact scalar — the primitive threshold $\tau_S^*(\bar{\rho}^2; \theta)$ — that determines which direction competition operates in. At any particular calibration, the sign of $\tau_S - \tau_S^*$ answers the question.

5.7 Numerical illustration

At the canonical calibration $\tau_v = \tau_\eta = \tau_S = \tau_z = a = 1$, $c(\rho) = 0.1 \cdot e^{-\kappa\rho}$, $\kappa = 1.5$, $\bar{\rho} = 0.9$, the equilibrium is:

$$\rho^{SP} \approx 0.02, \qquad \rho^{M} \approx 0.40, \qquad \rho^{C} = 0.90,$$

with price-informativeness values

$$\tau_{\phi,\infty}(\rho^{SP}) \approx 1.00, \qquad \tau_{\phi,\infty}(\rho^{M}) \approx 0.73, \qquad \tau_{\phi,\infty}(\rho^{C}) \approx 0.48.$$

The monopoly-competition informativeness gap is $(0.73 - 0.48)/0.48 \approx 52\%$ of the competitive baseline, and the planner-monopoly gap is $(1.00 - 0.73)/0.73 \approx 37\%$ of the monopoly baseline. The threshold values are $\tau_S^*(0) = 2/(1+\sqrt{7}) \approx 0.549$ and $\tau_S^*(\bar{\rho}^2 = 0.81) \approx 0.70$; the canonical $\tau_S = 1$ is comfortably above both. The Bai et al. [2016] empirical price-informativeness range is 0.6-0.9 in their R^2-based measure; the model's τ_ϕ is conceptually related but not identical to the BPS measure (see Section 7.5), so I treat the comparison as an order-of-magnitude illustration rather than a calibrated policy prediction. Under that caveat, the model's monopoly value 0.73 lies within the BPS range, and a move to competitive vendor provision would push informativeness below 0.5 — a drop of more than 30 percentage points relative to the monopoly baseline. Appendix C verifies that $\kappa = 1.5$ sits inside the primitive sufficient condition (V-prim) with margin.

6 Extension: a proprietary-signal channel

The two-regime model of Sections 2-5 shuts down a proprietary-signal option to isolate the vendor-side externality. Adding a proprietary channel with independent errors raises two natural questions: does the Akerlof-style unraveling of Proposition 1 survive when buyers have a non-hidden-action alternative, and does the absent buy-side wedge of Theorem 2 reappear on the proprietary margin? This section sets up the three-regime environment and records three conjectures whose analytic characterization is left to future work.

6.1 The three-regime environment

Investors choose among three regimes:

- S shared signal, $s_i^S = v + \rho\eta + \sqrt{1-\rho^2}\,\xi_i^S$ with $\xi_i^S \sim N(0, 1/\tau_S)$, fee p_S, vendor cost $c(\rho)$;
- P proprietary signal, $s_i^P = v + \xi_i^P$ with $\xi_i^P \sim N(0, 1/\tau_P)$ independent errors, fee $c_P > 0$ competitive-cost pricing;
- U uninformed.

Adoption measures are (n_S, n_P) with $n_U = 1 - n_S - n_P$. Assumption 2 applies only to the S sector; the P sector has observable independent-error signals.

6.2 Price and posteriors

Proprietary-signal errors are cross-sectionally independent; by the Sun [2006] law of large numbers, $\int \xi_i^P \, di = 0$ almost surely, so proprietary signals do not enter the aggregate demand net of fundamentals. The linear price takes the same form $P = Bv + Cn_S\eta + Dz$ as in Section 3, and $\phi = P/B = v + \lambda\eta + \mu z$ with $\lambda = Cn_S/B$, $\mu = D/B$, and $\tau_{\phi,\infty} = 1/(\lambda^2/\tau_\eta + \mu^2/\tau_z)$. The posterior-precision matrix for an S-subscriber is the K of (4); for a P-subscriber,

$$K^P = \begin{pmatrix} \tau_v + \tau_P + \tau_z/\mu^2 & \lambda\tau_z/\mu^2 \\ \lambda\tau_z/\mu^2 & \tau_\eta + \lambda^2\tau_z/\mu^2 \end{pmatrix}, \tag{18}$$

with $V_P = K^P_{22}/\det K^P$ and weights given by the standard normal-linear conditioning. An interior three-regime equilibrium requires adoption indifference:

$$\Phi_1^S(n_S, n_P; \rho) = p_S, \qquad \Phi_1^P(n_S, n_P; \rho) = c_P, \tag{19}$$

with $\Phi_1^S = (1/(2a))\log(V_U/V_S)$ and $\Phi_1^P = (1/(2a))\log(V_U/V_P)$.

6.3 Three open conjectures

Conjecture 1 (Hidden-action survives on the S channel)**.** *In any three-regime model with Assumption 2 on the S channel and observable errors on the P channel, the competitive S-vendor symmetric PBE still features $\rho^C = \bar{\rho}$. The logic of Proposition 1 — that a buyer cannot detect a unilateral ρ-deviation at the subscription stage — is independent of whether a separate P channel exists.*

Conjecture 2 (Informativeness ranking survives)**.** *There exist parameter regions with non-degenerate three-regime interior equilibrium in which the ranking $\tau_{\phi,\infty}(\rho^{SP}) > \tau_{\phi,\infty}(\rho^M) > \tau_{\phi,\infty}(\rho^C)$ of Theorem 4 holds.*

Conjecture 3 (Buy-side wedge reappears on the P margin)**.** *In any three-regime interior equilibrium with Assumptions 1-4, the Pigouvian wedge on the P-adoption margin,*

$$X^P := \frac{1}{2a}\left[n_S\,\partial_{n_P}\log V_S + n_P\,\partial_{n_P}\log V_P + (1 - n_S - n_P)\,\partial_{n_P}\log V_U\right],$$

is negative: the market under-provides P-adoption relative to the constrained planner. The wedge on the P margin is a classical Grossman-Stiglitz aggregation externality, orthogonal to the vendor-side wedge of Theorem 4.

Conjectures 1-3 are open. Preliminary numerical work at the canonical parameters of Section 5.7 encounters a corner regime in which $n_U = 0$ in equilibrium: at canonical fees, both informed regimes are strictly preferred to uninformed, and the three-regime interior with $n_U > 0$ does not exist. A rigorous attack requires either a parameter range in which participation is tighter so that $n_U > 0$ holds, or a reformulation of the indifference at the corner $n_S + n_P = 1$ as an S-versus-P margin. Appendix D documents the canonical numerical experiment.

6.4 Interpretation

If Conjectures 1-3 hold, the three-regime economy decomposes into two orthogonal welfare wedges.

- The *S-vendor wedge* (Theorem 4), a seller-side moral-hazard wedge on correlation, addressed by disclosure policy that breaks Assumption 2.
- The *P-buyer wedge* (Conjecture 3), a Grossman-Stiglitz aggregation externality on participation, addressed by a classical participation subsidy on proprietary signals.

Two distinct policy instruments address two distinct market failures. The two-regime analysis of Sections 2-5 isolates the first wedge by shutting down the second; a three-regime analysis — if the conjectures hold — would restore the second without disturbing the first. This decomposition is the natural economic interpretation of the extension, and is the best case for the paper's value added relative to the classical literature, which has studied only the buy-side wedge that the proprietary channel would restore.

7 Discussion

7.1 Testable predictions

The model delivers three predictions amenable to empirical test, each conditional on explicit scope.

1. *Competition-informativeness inversion.* Conditional on $\tau_S > \tau_S^*$ and Assumption 2 (opaque vendor market with hidden error correlation), residual price-informativeness variance $\mathrm{Var}(v - \mathbb{E}[v \mid P])$ is higher under competitive vendor provision than under concentrated provision. Identification requires variation in data-vendor market concentration. Candidate sources include natural experiments on vendor consolidation, entry of new AI-based signal providers, and regulatory mergers, combined with controls for time-varying fundamentals volatility. In the low-τ_S regime, the prediction reverses.

2. *Cross-subscriber return dispersion in entry events.* Under Assumption 2, entry of new vendors lowers cross-subscriber return-factor dispersion because the hidden-action argument drives each vendor to $\bar{\rho}$. Tests: pre-post event studies around vendor-entry episodes, measuring the cross-sectional dispersion of subscriber returns net of common factors. Prediction: dispersion declines. Under the low-τ_S regime, this prediction also reverses.

3. *Disclosure restores the monopoly outcome.* Regulatory regimes that break Assumption 2 (mandated model-documentation disclosure, third-party error-dispersion audits, public reporting of cross-subscriber error correlation) should produce informativeness closer to the monopoly benchmark than to the competitive benchmark. Cross-jurisdictional variation in AI-model disclosure rules (e.g., EU AI Act versus U.S. model-risk governance) provides identifying variation.

7.2 Regime identification via the threshold

Proposition 3 converts the scope condition into an identification strategy. Measurement of the sign of $d\tau_{\phi,\infty}/d\rho$ in cross-industry data identifies which side of the threshold a given market sits on. The sign is, in principle, estimable from panel variation in vendor-induced error correlation (e.g., the share of subscribers using common foundation models as a proxy for ρ) against measured price informativeness. A positive estimated slope indicates a low-τ_S regime and reverses the antitrust-inversion policy prescription; a negative slope indicates a high-τ_S regime in which the inversion applies. The sign test is orthogonal to the precision-substitution channel of Dugast and Foucault [2018], which operates at fixed ρ.

7.3 Disclosure versus antitrust

The policy corollary of Theorem 4 and Section 4.2 is that, in the regime $\tau_S > \tau_S^*$ and under Assumption 2, disclosure of cross-subscriber error dispersion dominates antitrust concentration as a policy instrument for informational efficiency. Breaking Assumption 2 collapses ρ^C from $\bar{\rho}$ to ρ^M, restoring the monopoly level of informativeness at competitive prices. Antitrust concentration policy operates through the market-structure margin and is strictly dominated by disclosure in the relevant regime. The point is not a general claim that data-market consolidation is socially beneficial; it is that, under explicit scope, disclosure policy hits the distortion more precisely than antitrust policy does.

The paper does not advocate the general policy claim. The scope conditions are load-bearing: in the low-τ_S regime, Section 5.6 shows the direction reverses, and the antitrust instrument regains its classical orientation. A regulator implementing disclosure policy should first estimate the sign of $d\tau_{\phi,\infty}/d\rho$ in the relevant market; the prescription depends on the sign.

The framing aligns with the Financial Stability Board [2025] report on AI adoption in the financial sector, which identifies model homogenization across subscribers as a source of systemic informational risk. The present model makes precise the mechanism (hidden error correlation leading to competitive cost-minimization on the differentiation margin) and the lever (disclosure of cross-subscriber error dispersion). Antitrust tools and disclosure tools operate on different informational margins; the model says which one to use when.

7.4 Robustness to the cost direction

Assumption 1 imposes $c'(\rho) < 0$: less-correlated signals are costlier. The direction matches the economics of per-subscriber fine-tuning, per-subscriber noise injection, and per-subscriber data curation: each raises cost relative to a single shared model output. In some production technologies the direction could invert (for instance, if a common model is the cheap off-the-shelf benchmark and differentiation is the economic default). In such technologies, the vendor's incentive to minimize cost under Assumption 2 would push ρ toward zero rather than toward $\bar{\rho}$, and the mechanism would attenuate or reverse. The paper's mechanism lives on the joint assumption of hidden correlation and costly differentiation.

7.5 What the paper does not deliver

1. *Reputation dynamics.* The model is static. In repeated data-vendor relationships with observable ex-post signal dispersion, reputation may attenuate the hidden-action force. The paper's analysis applies most directly to new, opaque, or newly-entered segments of the data-vendor market, where reputation has not yet disciplined vendor choices.

2. *Three-regime characterization.* Section 6 sets up the three-regime model and states three numerical observations (labeled conjectures below, as the analytic proofs are open). Preliminary numerical work at canonical parameters encounters a corner regime ($n_U = 0$); a rigorous attack requires either a different parameter range or a reformulation around the S-versus-P indifference.

3. *$\tau_S < \tau_S^*$ regime.* Theorem 4 holds only under $\tau_S > \tau_S^*(\bar{\rho}^2; \theta)$. Proposition 3 characterizes the boundary exactly; Section 5.6 documents the inversion. I do not treat this as a limitation of the theorem — the paper states exactly when each policy conclusion applies — but it does mean that the main result should be read as a characterization rather than an unconditional claim.

4. *Uniform threshold versus closed-form threshold.* The uniform threshold $\tau_S^*(\bar{\rho}^2; \theta)$ can be substantially larger than the closed-form $\tau_S^*(0; \theta)$; at canonical parameters the ratio is ≈ 1.27 $(0.70/0.549)$. For applications where $\bar{\rho}$ is small, the closed-form threshold is a close proxy; for applications in which $\bar{\rho}$ is near 1, the uniform-threshold statement is the operative one and the extra τ_S requirement is not negligible.

5. *Partial observability of ρ.* Assumption 2 is the polar case: ρ is fully hidden at purchase. An intermediate case in which buyers observe a noisy public signal $\tilde{\rho} = \rho + \nu$ at the subscription stage would interpolate between the hidden-action benchmark of Proposition 1 and the observable-ρ benchmark of Section 4.2. The signal-to-noise ratio of $\tilde{\rho}$ determines the degree to which buyer beliefs update on $\tilde{\rho}$, and the competitive equilibrium ρ^C moves continuously from $\bar{\rho}$ (full hidden action) to ρ^M (full observability) as the noise $\mathrm{Var}(\nu)$ falls to zero.

6. *Calibration of τ_ϕ against empirical measures.* The model's τ_ϕ maps conceptually to the informativeness of prices about fundamentals but is not identical to Bai et al. [2016]'s R^2-based measure, which is constructed from earnings-forecasting regressions of prices on future profits. The numerical comparison in Section 5.7 treats the monopoly-versus-competition gap as an order-of-magnitude illustration of informativeness effects, not as a calibrated quantitative policy prediction. A formal mapping from τ_ϕ to the BPS measure would require modeling the earnings-forecasting regression structure, which is outside the scope of the present paper.

7. *Disclosure policy: implementation costs and scope.* The disclosure-dominates-antitrust claim (Section 7.3) is *conditional on Assumption 2*. The model does not incorporate implementation costs of disclosure (audit costs, documented-error gaming, disclosure-induced shifts in the

hidden attribute itself). The policy takeaway is that, under the maintained hypotheses, the disclosure lever hits the hidden-action distortion more precisely than the antitrust lever; it is not a claim that disclosure is first-best.

8. *CARA-normal tractability.* The welfare identity $W = \pi - (1/(2a)) \log V_U$ relies on CARA exponential utility. I conjecture the qualitative ranking extends to small perturbations from CARA; a formal robustness statement under CRRA or Epstein-Zin preferences is future work.

9. *Constrained-Pareto planner.* The planner chooses ρ taking the buyer-side equilibrium as given; the welfare object is constrained-Pareto, not first-best. A first-best planner who also commands market participation and the fee would coincide with the present planner at optimal ρ only when Assumption 3 binds; when slack is interior, the constrained planner and the unconstrained planner diverge.

10. *Interior n_S.* Assumption 3 delivers $n_S = 1$ at the relevant subgames (the participation constraint is non-binding). Interior $n_S < 1$ equilibria are covered by the buyer-side analysis of Section 3 but do not alter Theorem 4's ranking under the maintained hypothesis.

7.6 Comparison to the broader literature on data economics

The paper sits at the intersection of three literatures: noisy rational expectations (Grossman and Stiglitz, 1980, Hellwig, 1980, Verrecchia, 1982, Diamond and Verrecchia, 1981, Admati, 1985), data economics and non-rivalry (Farboodi and Veldkamp, 2020, Jones and Tonetti, 2020, Acemoglu et al., 2022, Farboodi et al., 2024, Abis and Veldkamp, 2024), and information-selling and market structure (Admati and Pfleiderer, 1986, 1988, Skreta and Veldkamp, 2009, Cong et al., 2024, Goldstein et al., 2024).

Relative to the first, the paper introduces a correlated-signal-error structure that breaks the classical aggregation limit (Theorem 3) and a vendor-side technology choice that endogenizes the correlation itself. Relative to the second, the paper embeds data non-rivalry in an asset-pricing model and characterizes its effect on price informativeness as a function of vendor market structure. Relative to the third, the paper identifies a hidden-action mechanism distinct from the ratings-shopping channel of Skreta and Veldkamp and from the Cournot channel of Cong et al., and establishes a market-structure ranking that Admati and Pfleiderer leave unstudied.

A distinct strand studies AI and algorithmic homogenization in financial markets. Dou et al. [2025] model AI traders using a common foundation model and show that reinforcement-learning convergence sustains algorithmic collusion in a Kyle-type strategic setting; their mechanism is an output of simulation rather than a structural object, and the relevant market structure is trader competition in the trading stage. The present paper models the signal correlation as a structural object arising from data-vendor technology choice and characterizes the information-market equilibrium analytically. Kleinberg and Raghavan [2021] define algorithmic monoculture and show it can reduce collective decision quality; the paper is one instantiation of that observation in a rational-expectations pricing environment with an endogenous producer.

Farboodi and Veldkamp [2020] develop a framework for the long-run value of data in which data non-rivalry generates increasing returns to scale and progressive data accumulation. Their focus is on the aggregate growth implications of data accumulation rather than on the cross-sectional properties of price informativeness that arise once data are sold through a vendor market with a hidden technology attribute. The present paper takes the vendor's technology as the object of endogenous choice at a point in time, holding aside the data-accumulation margin. The two papers treat complementary dimensions of the data-economics problem: Farboodi and Veldkamp characterize how data stocks evolve; the present paper characterizes how the market structure of vendors translates data into prices.

8 Conclusion

When investors build private signals from shared information-production inputs, signal errors inherit a common component that survives aggregation. A data vendor's technology choice over error correlation shapes the equilibrium informativeness of asset prices. Under hidden correlation, competition among data vendors drives the correlation parameter to the cost-minimizing boundary through an Akerlof-style lemons unraveling, strictly reducing price informativeness relative to monopoly and to the constrained planner (Theorem 4). The sign of the effect of correlation on informativeness is governed by a unique threshold on individual-signal precision, with a closed form at zero correlation (Proposition 3). Above the threshold, the classical antitrust intuition inverts: concentration aids informational efficiency. Below the threshold, the intuition is restored. Mandated disclosure of cross-subscriber error dispersion breaks the hidden-action channel and restores the monopoly outcome, dominating antitrust concentration as a policy instrument in the regime where the inversion holds. The paper's contribution is the characterization of the conditions under which each direction operates and the closed-form threshold that separates them.

References

Simona Abis and Laura Veldkamp. The changing economics of knowledge production. *Review of Financial Studies*, 37:hhad059, 2024. doi: 10.1093/rfs/hhad059.

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4): 218–256, 2022. doi: 10.1257/mic.20200200.

Anat R. Admati. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica*, 53(3):629–658, 1985.

Anat R. Admati and Paul Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

Anat R. Admati and Paul Pfleiderer. Selling and trading on information in financial markets. *American Economic Review*, 78(2):96–103, 1988.

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

Jennie Bai, Thomas Philippon, and Alexi Savov. Have financial markets become more informative? *Journal of Financial Economics*, 122(3):625–654, 2016.

Lin William Cong, Shiyang Huang, Qiufei Li, and Jingtao Ni. Cournot competition, informational feedback, and real efficiency. Technical report, NBER Working Paper 32944, 2024. URL https://www.nber.org/papers/w32944.

Douglas W. Diamond and Robert E. Verrecchia. Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics*, 9(3):221–235, 1981. doi: 10.1016/0304-405X(81)90026-X.

Winston Wei Dou, Itay Goldstein, and Yan Ji. Ai-powered trading, algorithmic collusion, and price efficiency. Technical report, NBER Working Paper 34054, 2025.

Jérôme Dugast and Thierry Foucault. Data abundance and asset price informativeness. *Journal of Financial Economics*, 130(2):367–391, 2018. doi: 10.1016/j.jfineco.2018.07.004.

Jérôme Dugast and Thierry Foucault. Equilibrium data mining and data abundance. *Journal of Finance*, 80(1):211–258, 2025. doi: 10.1111/jofi.13397.

Maryam Farboodi and Laura Veldkamp. Long-run growth of financial data technology. *American Economic Review*, 110(8):2485–2523, 2020. doi: 10.1257/aer.20171349.

Maryam Farboodi, Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran. Valuing financial data. *Review of Financial Studies*, 37(6):hhae034, 2024. doi: 10.1093/rfs/hhae034.

Financial Stability Board. Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector. Technical report, Financial Stability Board, October 2025. URL https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-related-vulnerabilities-in-the-financial-sector/.

Diego García and Günter Strobl. Relative wealth concerns and complementarities in information acquisition. *Review of Financial Studies*, 24(1):169–207, 2011. doi: 10.1093/rfs/hhq086.

Itay Goldstein, Yan Xiong, and Liyan Yang. Information sharing in financial markets. *Journal of Financial Economics*, 160:103967, 2024. doi: 10.1016/j.jfineco.2024.103967.

Sanford J. Grossman and Joseph E. Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70(3):393–408, 1980.

Martin F. Hellwig. On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22(3):477–498, 1980.

Charles I. Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020. doi: 10.1257/aer.20191330.

Tetsuya Kawanishi. The diversity of information acquisition strategies in a noisy ree model with a common signal and independent signals. In *Recent Advances in Financial Engineering 2010*. World Scientific, 2011.

Jon Kleinberg and Manish Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118, 2021. doi: 10.1073/pnas.2018340118.

Youcheng Lou, Sahar Parsa, Debraj Ray, Duan Li, and Shouyang Wang. Information aggregation in a financial market with general signal structure. *Journal of Economic Theory*, 183:594–624, 2019. doi: 10.1016/j.jet.2019.05.013.

Han N. Ozsoylev and Johan Walden. Asset pricing in large information networks. *Journal of Economic Theory*, 146(6):2252–2280, 2011. doi: 10.1016/j.jet.2011.10.009.

Alessandro Pavan, Savitar Sundaresan, and Xavier Vives. (in)efficiency in information acquisition and aggregation through prices. Technical report, Working paper, 2025. SSRN 4028893.

Vasiliki Skreta and Laura Veldkamp. Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics*, 56(5):678–695, 2009. doi: 10.1016/j.jmoneco.2009.04.006.

Yeneng Sun. The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126(1):31–69, 2006. doi: 10.1016/j.jet.2004.10.005.

Robert E. Verrecchia. Information acquisition in a noisy rational expectations economy. *Econometrica*, 50(6):1415–1430, 1982.

A Proofs

This appendix collects the proofs of results stated in Sections 3-5. Numerical grids and calibration details follow in Appendices B-D.

A.1 Proof of Lemma 1 (posterior variance)

Stacking the subscriber's observations (s_i^S, ϕ) as a vector with linear-normal structure in (v, η), standard normal-linear filtering gives

$$\begin{pmatrix} \mathbb{E}[v \mid s^S, \phi] \\ \mathbb{E}[\eta \mid s^S, \phi] \end{pmatrix} = K^{-1} \left[\begin{pmatrix} b_1 \\ \rho b_1 \end{pmatrix} s^S + \begin{pmatrix} b_2 \\ \lambda b_2 \end{pmatrix} \phi \right],$$

with K the posterior-precision matrix of (v, η) in (4). The posterior variance of v alone is the $(1,1)$-element of K^{-1}: $V_S = K_{22}/\det K$. Positive definiteness $K \succ 0$ follows from $K = \mathrm{diag}(\tau_v, \tau_\eta) + b_1(1,\rho)(1,\rho)^\top + b_2(1,\lambda)(1,\lambda)^\top$: the first summand $\mathrm{diag}(\tau_v, \tau_\eta) \succ 0$ strictly (both $\tau_v > 0$ and $\tau_\eta > 0$), and the remaining two rank-one outer products are positive semidefinite, so the sum is strictly positive definite. The weights $w_S^{(s)}, w_S^{(\phi)}$ are read off from the first row of K^{-1} times the precision vectors.

A.2 Proof of Lemma 2 (cubic fixed point)

CARA demand with linear-normal posterior yields $x_i(P) = (\mathbb{E}_i[v] - P)/(a\,\mathrm{Var}_i[v])$. Market clearing $\int x_i(P)\,di + z = 0$ pins down the linear price coefficients (B, C, D) in terms of (λ, μ) and n_S. Matching coefficients on (v, η, z) on both sides yields $\lambda = \rho$ (the common factor loading in the price inherits from the subscribers' signal loading on η) and the scalar cubic (6) in μ.

Under condition (5), the cubic has a unique positive real root. Write $F_{n_S}(\mu) := \tau_\eta b_1 \mu^3 - (a/n_S)(\tau_\eta + \rho^2 b_1)\mu^2 - (a\rho^2\tau_z/n_S)$. At $\mu = 0$, $F_{n_S}(0) = -a\rho^2\tau_z/n_S < 0$; as $\mu \to \infty$, $F_{n_S}(\mu) \to +\infty$. The discriminant condition (5) ensures that the derivative $F'_{n_S}(\mu) = 3\tau_\eta b_1\mu^2 - 2(a/n_S)(\tau_\eta + \rho^2 b_1)\mu$ has its positive critical point at $\mu_c = (2/3)(a/n_S)(\tau_\eta + \rho^2 b_1)/(\tau_\eta b_1)$, and F_{n_S} is strictly increasing past μ_c. Direct algebraic verification that $F_{n_S}(\mu_c) < 0$ (equivalently (5)) ensures the cubic crosses zero exactly once at $\mu^* > \mu_c$.

A.3 Proof of Lemma 3 (comparative statics in n_S)

Implicit differentiation of (6) in n_S yields $d\mu^*/dn_S = -\partial F_{n_S}/\partial n_S/\partial F_{n_S}/\partial\mu$, with both numerator and denominator signed by the cubic structure. The sign of $\partial\tau_\phi/\partial n_S$ follows from $\tau_\phi = 1/(\rho^2/\tau_\eta + \mu^{*2}/\tau_z)$, decreasing in μ^*: verifying $d\mu^*/dn_S < 0$ under (5) yields $\partial\tau_\phi/\partial n_S > 0$. Then $V_U = 1/(\tau_v + \tau_\phi)$ is decreasing in τ_ϕ, and $V_S = K_{22}/\det K$ is decreasing through the $b_2 = \tau_z/\mu^{*2}$ channel as μ^* falls.

A.4 Proof of Theorem 1 (interior adoption)

The function $\Phi_1(n_S; \rho) = (1/(2a))\log(V_U/V_S)$ is the certainty-equivalent gain from subscribing. By Lemma 3, V_U is strictly decreasing in n_S and so is V_S. The ratio V_U/V_S is strictly decreasing in n_S because the precision channel through τ_ϕ raises subscribers' posterior precision more than uninformeds'. Continuity on $(0,1)$ follows from the C^1-dependence of the cubic's root. Under (8), $\Phi_1(0^+; \rho) > p > \Phi_1(1^-; \rho)$, so the intermediate-value theorem yields a unique $n_S^{eq} \in (0,1)$. Smoothness in (ρ, p) follows from the implicit-function theorem applied to $\Phi_1(n_S; \rho) - p = 0$, with strictly positive Jacobian $\partial\Phi_1/\partial n_S$.

A.5 Proof of Theorem 2 (buy-side neutrality)

We compute $X(n_S; \rho) = n_S\partial\log V_S/\partial n_S + (1 - n_S)\partial\log V_U/\partial n_S$ using Lemma 1 and the fixed point of Lemma 2.

Step 1: V_U and V_S as functions of μ. On the equilibrium path, $\lambda = \rho$ (Lemma 2), so $\tau_\phi = 1/(\rho^2/\tau_\eta + \mu^2/\tau_z)$ and $V_U = 1/(\tau_v + \tau_\phi)$. For V_S, using K in (4) with $\lambda = \rho$ collapses the cross-terms: $K_{12} = \rho(b_1 + b_2)$ and $\det K = (\tau_v + b_1 + b_2)(\tau_\eta + \rho^2 b_1 + \rho^2 b_2) - \rho^2(b_1 + b_2)^2$. Expanding,

$$\det K = \tau_v\tau_\eta + \tau_v\rho^2(b_1 + b_2) + \tau_\eta(b_1 + b_2) + \rho^2(b_1 + b_2)^2 - \rho^2(b_1 + b_2)^2 = \tau_v\tau_\eta + (\tau_\eta + \tau_v\rho^2)(b_1 + b_2).$$

With $K_{22} = \tau_\eta + \rho^2(b_1 + b_2)$, we obtain

$$V_S = \frac{\tau_\eta + \rho^2(b_1 + b_2)}{\tau_v\tau_\eta + (\tau_\eta + \tau_v\rho^2)(b_1 + b_2)}.$$

Step 2: n_S-derivative through μ. Both V_U and V_S depend on n_S only through $\mu^*(n_S; \rho)$, via $b_2 = \tau_z/\mu^2$ and τ_ϕ. Let $D_U := \tau_v + \tau_\phi$ (denominator of V_U) and $D_S := \tau_v\tau_\eta + (\tau_\eta + \tau_v\rho^2)(b_1 + b_2)$ (denominator of V_S). Differentiating $\log V_U = -\log D_U$ and using $d\tau_\phi/d\mu = -2\mu/(\tau_z(\rho^2/\tau_\eta + \mu^2/\tau_z)^2) = -2\mu\tau_\phi^2/\tau_z$,

$$\frac{\partial \log V_U}{\partial \mu} = -\frac{1}{D_U} \cdot \frac{\partial \tau_\phi}{\partial \mu} = \frac{2\mu\tau_\phi^2}{\tau_z D_U} > 0.$$

For V_S: $\partial V_S/\partial b_2 = [\tau_v\tau_\eta\rho^2 - \tau_v\tau_\eta(\tau_\eta + \rho^2(b_1 + b_2))/(b_1 + b_2)] \cdot [\ldots]$; the sign that matters is

$$\frac{\partial \log V_S}{\partial \mu} = \frac{\rho^2}{K_{22}} \cdot \frac{\partial b_2}{\partial \mu} - \frac{\tau_\eta + \tau_v\rho^2}{D_S} \cdot \frac{\partial b_2}{\partial \mu} = \frac{\partial b_2}{\partial \mu} \cdot \left[\frac{\rho^2}{K_{22}} - \frac{\tau_\eta + \tau_v\rho^2}{D_S}\right].$$

With $\partial b_2/\partial\mu = -2\tau_z/\mu^3 < 0$ and $D_S - (\tau_\eta + \tau_v\rho^2)K_{22}/\rho^2 \cdot \rho^2 = \tau_v\tau_\eta(1 - K_{22}/\rho^2 \cdot (\tau_\eta + \tau_v\rho^2)) + \ldots$, a direct expansion gives the bracket equal to $-\tau_v\tau_\eta/(K_{22}D_S) < 0$, so $\partial \log V_S/\partial\mu = (2\tau_z\tau_v\tau_\eta)/(\mu^3 K_{22} D_S) > 0$.

Step 3: the weighted sum. Since both derivatives share the factor $d\mu^*/dn_S < 0$ (Lemma 3), $\partial \log V_U/\partial n_S = (d\mu^*/dn_S) \cdot \partial \log V_U/\partial\mu < 0$ and $\partial \log V_S/\partial n_S = (d\mu^*/dn_S) \cdot \partial \log V_S/\partial\mu < 0$. Writing $g_U := \partial \log V_U/\partial\mu > 0$ and $g_S := \partial \log V_S/\partial\mu > 0$,

$$X(n_S; \rho) = \frac{d\mu^*}{dn_S}\left[n_S g_S + (1 - n_S)g_U\right] = -\frac{\tau_\eta}{\tau_S b_1\mu^2 + a\rho^2\tau_z/\mu} \cdot \left[n_S g_S + (1 - n_S)g_U\right],$$

where the bracket is strictly positive (both $g_S, g_U > 0$ and $n_S \in (0, 1)$) and the prefactor is strictly positive. The whole expression has the form $X = -\tau_\eta^2/(D_1 D_2) < 0$ with $D_1 = D_U(\tau_S b_1\mu^2 + a\rho^2\tau_z/\mu) > 0$, $D_2 = K_{22}D_S > 0$, both strictly positive. Therefore $X(n_S; \rho) < 0$ on $(0, 1) \times (0, \bar{\rho}^{(1)})$.

Interpretation. In the two-regime economy, $\Phi_1(n_S; \rho) = p$ coincides with the planner's first-order condition in n_S at the marginal-cost fee $p = c(\rho)$, leaving no buy-side Pigouvian wedge at fixed technology.

A.6 Proof of Theorem 3 (wisdom-of-crowds ceiling)

(a), (b). As $n_S \to 1^-$, the a/n_S coefficient in (6) limits to a, yielding the limiting cubic (10). The same argument as in Lemma 2 (modulo the vanishing $1/n_S$) gives a unique positive root

$\mu_\infty(\rho;\tau_S) > 0$, and continuity of the root in n_S delivers $\mu^*(n_S;\rho) \to \mu_\infty$. The equality $\lambda^* = \rho$ is preserved.

(c) Substituting $\lambda = \rho$ and $\mu = \mu_\infty$ into (3) gives $\tau_{\phi,\infty}(\rho) = 1/(\rho^2/\tau_\eta + \mu_\infty^2/\tau_z)$.

(d) The strict upper bound $\tau_{\phi,\infty} < \tau_\eta/\rho^2$ is equivalent to $\mu_\infty^2/\tau_z > 0$, which is immediate. The lower bound $\tau_{\phi,\infty} > 0$ follows from μ_∞ finite.

(e) Continuity on $(0,1)$: the cubic (10) has coefficients continuous in ρ; its unique positive root is therefore continuous in ρ (standard implicit-function argument with $\partial F/\partial\mu > 0$). Limit at $\rho \downarrow 0$: (10) reduces to $\tau_\eta\tau_S\mu^3/1 - a\tau_\eta\mu^2 = 0$, so $\mu_\infty(0;\tau_S) = a/\tau_S$. Substituting, $\tau_{\phi,\infty}(0) = 1/(0 + (a/\tau_S)^2/\tau_z) = \tau_z\tau_S^2/a^2$.

A.7 Proof of Corollary 1 (data abundance)

From $\Phi_1(n_S;\rho) = p$ and Lemma 3 with $\partial\Phi_1/\partial n_S < 0$, $dn_S^{eq}/dp < 0$: cheaper data raises n_S^{eq}. Then $d\tau_{v|P}/dp = (\partial\tau_\phi/\partial n_S)(dn_S^{eq}/dp) < 0$ by Lemma 3.

A.8 Proof of Proposition 1 (competitive PBE)

The proof sketch in the main text is complete up to three technical details, filled here.

Technical detail 1 (measurability of the continuum). The continuum of potential entrants is indexed by $j \in [0,1]$ equipped with the Sun [2006] Fubini extension; measurability of buyer subscription demand as a function of the strategy profile follows from standard results on continuum games.

Technical detail 2 (no-refinement invocation). Step 1 of the main-text sketch establishes that the hidden-action deviation is not observable at the subscription stage. Buyers' posterior over the deviator's ρ is pinned by Bayes' rule on the observed fee p_j. Off-path fees $p_j \neq p^*$ are disciplined by the slack-participation condition in Assumption 3; no D1 or divinity refinement is required because the mechanism operates on a hidden action rather than on a signaling dimension.

Technical detail 3 (strict inequality in Step 1). The profitability of the deviation in Step 1 uses $c(\bar\rho) < c(\rho^*)$ with strict inequality: Assumption 1 gives $c'(\rho) < 0$ strictly on $[0,\bar\rho]$, so $c(\bar\rho) - c(\rho^*) < 0$ strictly whenever $\rho^* < \bar\rho$.

A.9 Proof of Proposition 2 (welfare identity)

Start from the aggregate CARA certainty equivalent (2) and the relation $\Phi_1(n_S;\rho) = (1/(2a))\log(V_U/V_S)$:

$$-\frac{1}{2a}[n_S \log V_S + (1-n_S)\log V_U] = -\frac{1}{2a}\log V_U + \frac{n_S}{2a}\log(V_U/V_S) = -\frac{1}{2a}\log V_U + n_S\Phi_1(n_S;\rho).$$

Subtracting the vendor's cost $n_S c(\rho)$ (the fee p is a transfer, which cancels in aggregate; the cost is a real-resource expenditure) yields

$$W = n_S\Phi_1(n_S;\rho) - n_S c(\rho) - \frac{1}{2a}\log V_U = \pi - \frac{1}{2a}\log V_U,$$

with $\pi = n_S[\Phi_1(n_S;\rho) - c(\rho)]$ the vendor's profit at the indifference fee.

A.10 Proof of Proposition 3 (threshold)

Part (i). Differentiating $\tau_{\phi,\infty} = 1/(y/\tau_\eta + \mu_\infty^2/\tau_z)$ in y and using the chain rule with $d\mu_\infty/dy$ from implicit differentiation of $F(\mu, y; \tau_S, \theta) = 0$ yields

$$\frac{d\tau_{\phi,\infty}}{dy} = -\tau_{\phi,\infty}^2 \left[\frac{1}{\tau_\eta} + \frac{2\mu_\infty}{\tau_z}\frac{d\mu_\infty}{dy}\right].$$

Compute $\partial F/\partial\mu = 3\tau_\eta b_1\mu^2 - 2a[\tau_\eta + yb_1]\mu$. Using $F = 0$ to substitute $a[\tau_\eta + yb_1]\mu^2 = \tau_\eta b_1\mu^3 - ay\tau_z$:

$$\frac{\partial F}{\partial\mu} = \frac{\tau_\eta b_1\mu^3 + 2ay\tau_z}{\mu} > 0.$$

Compute $\partial F/\partial y$ using $db_1/dy = b_1/(1-y)$ and $b_1 + y \cdot db_1/dy = \tau_S/(1-y)^2$:

$$\frac{\partial F}{\partial y} = \frac{\tau_S\mu^2(\tau_\eta\mu - a)}{(1-y)^2} - a\tau_z.$$

By implicit-function theorem, $d\mu_\infty/dy = -(\partial F/\partial y)/(\partial F/\partial\mu)$, which after substitution yields G of (15) as the sign-determining function.

On-shell form. Rewrite (10) as $\tau_S\mu^2(\tau_\eta\mu - ay)/(1-y) = a(\tau_\eta\mu^2 + y\tau_z)$, so

$$\tau_S\mu^2 = \frac{a(1-y)(\tau_\eta\mu^2 + y\tau_z)}{\tau_\eta\mu - ay} \qquad (\diamond)$$

(valid by Lemma 5). Substituting ($\diamond$) into G after algebra gives

$$G = \frac{a(1-y)(\tau_\eta\mu^2 + y\tau_z)}{\tau_\eta\mu - ay} \cdot N(\mu, y; \theta), \qquad N(\mu, y; \theta) = \tau_z(1-y)(3\tau_\eta\mu - 2ay) - 2\tau_\eta\mu^2(\tau_\eta\mu - a).$$

Therefore $\mathrm{sign}(G) = \mathrm{sign}(N)$ along the cubic.

Part (ii) (existence and uniqueness). Define $H(\tau_S) := N(\mu_\infty(y;\tau_S), y; \theta)$. Step 1: $H(0^+) \to -\infty$ (direct substitution with $\mu_\infty \to \infty$), $H(\infty) > 0$ (direct substitution with $\mu_\infty \to 0^+$); by intermediate-value theorem, H has at least one zero. Step 2: at any zero μ_0 of N, we have $\mu_0 > a/\tau_\eta$ (from $N = 0$ with $3\tau_\eta\mu - 2ay > 0$). Substituting $\tau_z(1-y) = 2\tau_\eta\mu_0^2(\tau_\eta\mu_0 - a)/(3\tau_\eta\mu_0 - 2ay)$ (from $N = 0$) into $\partial N/\partial\mu$ yields

$$\left.\frac{\partial N}{\partial\mu}\right|_{\mu_0} = \frac{2\tau_\eta\mu_0 \cdot P(\mu_0; y)}{3\tau_\eta\mu_0 - 2ay}, \quad P(\mu_0; y) := -6\tau_\eta^2\mu_0^2 + 3\tau_\eta a(1+2y)\mu_0 - 4a^2y.$$

P is a downward parabola in μ_0 with discriminant $\Delta(y) = 9\tau_\eta^2a^2(1+2y)^2 - 96\tau_\eta^2a^2y = 3\tau_\eta^2a^2(12y^2 - 20y + 3)$. The factor $12y^2 - 20y + 3$ has roots $y = 1/6$ and $y = 3/2$.

Case 1: $y \in [0, 1/6)$. The discriminant $\Delta(y) > 0$, so P has two real roots $\mu_- < \mu_+$. We show

$\mu_0 > \mu_+$ at any zero μ_0 of N. Since $N(\mu_0, y) = 0$ implies $\mu_0 > ay/\tau_\eta$ (from $\tau_z(1-y)(3\tau_\eta\mu_0 - 2ay) = 2\tau_\eta\mu_0^2(\tau_\eta\mu_0 - a) > 0$ requires $3\tau_\eta\mu_0 > 2ay$; combined with Lemma 5 $\mu_0 > ay/\tau_\eta$, strict), and further $\mu_0 > a/\tau_\eta$ (because the left side requires $\tau_\eta\mu_0 > a$, or N cannot balance a positive left against a negative right). Evaluating P at $\mu_0 = a/\tau_\eta$ gives $P(a/\tau_\eta; y) = -6a^2 + 3a^2(1+2y) - 4a^2y = a^2(2y-3) < 0$ for $y < 3/2$. Because P is a downward parabola with $P(a/\tau_\eta; y) < 0$, either $a/\tau_\eta > \mu_+$ (parabola is negative to the right of both roots) or $a/\tau_\eta < \mu_-$ (parabola is negative to the left of both roots). Compute the vertex $\mu_V = a(1+2y)/(4\tau_\eta)$; for $y < 1/6$, $\mu_V < a(1+1/3)/(4\tau_\eta) = a/(3\tau_\eta) < a/\tau_\eta$, so a/τ_η lies to the right of the vertex, hence $a/\tau_\eta > \mu_+$. Since $\mu_0 > a/\tau_\eta > \mu_+$, $P(\mu_0; y) < 0$.

Case 2: $y \in [1/6, 1]$. The discriminant $\Delta(y) \leq 0$, so P has no real roots on $(0, \infty)$. Combined with the downward-parabola orientation and $P(a/\tau_\eta; y) = a^2(2y-3) < 0$, $P(\mu; y) < 0$ on all of $(0, \infty)$.

In both cases $P(\mu_0; y) < 0$ at every zero μ_0 of N, so $\partial N/\partial\mu|_{\mu_0} < 0$. Combined with $d\mu_\infty/d\tau_S < 0$ (from Lemma 5 corollary, shown below), $dH/d\tau_S|_{\text{zero}} > 0$: each zero of H is crossed upward, so there is only one.

The corollary $d\mu_\infty/d\tau_S < 0$ follows from implicit differentiation: $\partial F/\partial\tau_S = \mu^2(\tau_\eta\mu - ay)/(1-y) > 0$ (by Lemma 5), so $d\mu_\infty/d\tau_S = -(\partial F/\partial\tau_S)/(\partial F/\partial\mu) < 0$.

Part (iii) (uniform threshold). Continuity of $\tau_S^*(y;\theta)$ in y follows from the implicit-function theorem applied to $N(\mu_\infty(y;\tau_S^*), y; \theta) = 0$: $\partial N/\partial\mu < 0$ and $d\mu_\infty/d\tau_S < 0$ give $\partial N/\partial\tau_S \neq 0$, so τ_S^* is continuous. Weierstrass on the compact set $[0, \bar\rho^2]$ then yields attainment of the maximum.

Part (iv) (closed form at $\rho = 0$). At $y = 0$, $\mu_\infty(0;\tau_S) = a/\tau_S$ from Theorem 3(e). Substituting into $N = 0$:

$$N(a/\tau_S, 0; \theta) = 3\tau_z\tau_\eta(a/\tau_S) - 2\tau_\eta(a/\tau_S)^2[\tau_\eta(a/\tau_S) - a] = 0.$$

Divide by $\tau_\eta(a/\tau_S)$ and simplify: $3\tau_z\tau_S^2 + 2a^2\tau_S - 2a^2\tau_\eta = 0$, a quadratic in τ_S. The positive root is $\tau_S^*(0) = [-2a^2 + \sqrt{4a^4 + 24\tau_z\tau_\eta a^2}]/(6\tau_z) = [-a^2 + a\sqrt{a^2 + 6\tau_z\tau_\eta}]/(3\tau_z)$. Multiply numerator and denominator by $a + \sqrt{a^2 + 6\tau_z\tau_\eta}$ to obtain the form (16).

The comparative statics follow by direct differentiation with $R := \sqrt{a^2 + 6\tau_z\tau_\eta}$: $\partial\tau_S^*(0)/\partial\tau_\eta = (2a/[a+R]) \cdot [1 - (3\tau_z\tau_\eta)/(R[a+R])]$. Algebra using $R^2 = a^2 + 6\tau_z\tau_\eta$ shows the bracketed term is positive, so $\partial\tau_S^*(0)/\partial\tau_\eta > 0$. Similar direct computation gives $\partial\tau_S^*(0)/\partial a > 0$ and $\partial\tau_S^*(0)/\partial\tau_z < 0$.

Part (v) (primitive sufficient condition). (V**) is the condition $\tau_\eta\mu_\infty(y;\tau_S) < a + a\tau_z(1-y)^2/(\tau_S\mu_\infty^2)$. The only term of G in (15) with potentially unfavorable sign is $-2\tau_\eta\tau_S\mu^4(\tau_\eta\mu - a)$. (V**) rearranges to $\tau_S\mu^2(\tau_\eta\mu - a) < a\tau_z(1-y)^2$, so

$$G > \tau_z\tau_\eta\tau_S\mu^3(1-y) + 2ay\tau_z^2(1-y)^2 + 2a\tau_\eta\tau_z\mu^2(1-y)^2 - 2\tau_\eta\mu^2 \cdot a\tau_z(1-y)^2 = \tau_z\tau_\eta\tau_S\mu^3(1-y) + 2ay\tau_z^2(1-y)^2 > 0.$$

That (V**) is not necessary follows from the numerical example at $y = 0.25$, $\tau_S = 0.7$, canonical θ, where (V**) fails but direct evaluation gives $G > 0$.

A.11 Proof of Proposition 4 (welfare ranking)

(a) From Proposition 1.

(b) Under Assumption 4, Π^M has a unique interior argmax $\rho^M \in (0, \bar{\rho})$ satisfying the first-order condition $d\Phi_1(1;\rho)/d\rho\big|_{\rho^M} = c'(\rho^M)$.

(c) Using the welfare identity, $W^M(\rho) = \Pi^M(\rho) + L(\rho)$ with $L(\rho) := -(1/(2a)) \log V_U(1;\rho)$. Lemma 6 under the uniform-threshold hypothesis gives $d \log V_U/d\rho > 0$, so $L'(\rho) < 0$ on $(0, \bar{\rho})$. Therefore

$$\left.\frac{dW^M}{d\rho}\right|_{\rho^M} = \left.\frac{d\Pi^M}{d\rho}\right|_{\rho^M} + L'(\rho^M) = 0 + L'(\rho^M) < 0.$$

By strict unimodality of W^M (Assumption 4), the unique interior maximizer ρ^{SP} lies to the left of any point where $dW^M/d\rho < 0$. Therefore $\rho^{SP} < \rho^M$.

(d) Combine (a), (b), (c) with $\rho^M < \bar{\rho}$ (interior argmax): $\rho^{SP} < \rho^M < \bar{\rho} = \rho^C$.

B Numerical grid verifications

Numerical verification reported in this appendix uses the scripts in `code/explore/` against the canonical parameters $\tau_v = \tau_\eta = \tau_S = \tau_z = a = 1$, $\bar{\rho} = 0.9$, $c(\rho) = 0.1 \cdot e^{-\kappa\rho}$. The scale $c_0 = 0.1$ is chosen so that the maximum subscription fee in equilibrium is of the order of 10% of the per-investor certainty-equivalent gain from subscription at canonical $\tau_S = 1$; this is a normalization, not a calibration to industry data, and the qualitative comparative statics and the ranking of Theorem 4 are scale-invariant in c_0 (a uniform rescaling of $c(\rho)$ rescales ρ^M and ρ^{SP} continuously but does not flip the ranking).

B.1 Verification of Assumption 4

Over $(\kappa, a, \tau_S) \in \{0.5, 1, 1.5, 2, 3\} \times \{0.5, 1, 1.5, 2\} \times \{0.5, 1, 1.5, 2\}$ (125 grid points):

Property	Fraction
Π^M locally concave at ρ^M	93%
W^M unimodal on $[0, \bar{\rho}]$	100%
W^M locally concave at argmax ρ^{SP}	77%
Welfare ranking $\rho^{SP} < \rho^M < \rho^C$ strict	88%
Failures (corner $\rho^M = \bar{\rho}$ under high a)	12%

Failures of the strict ranking occur exclusively at corner collapses (ρ^M hitting $\bar{\rho}$ under high a) and not at sign reversals of the interior FOC. The qualitative ranking is robust.

B.2 Verification of $\tau_S^*(y;\theta)$ signs for $y > 0$

Grid: $\theta \in \{0.5, 1, 1.5, 2\}^3 \times y \in \{0.05, 0.15, 0.30, 0.50, 0.81\}$, 320 grid points. The signs of the θ-comparative statics at the closed-form value $\tau_S^*(0;\theta)$ extend to the grid:

Sign property	Fraction
$\partial\tau_S^*/\partial\tau_\eta > 0$	100%
$\partial\tau_S^*/\partial a > 0$	100%
$\partial\tau_S^*/\partial\tau_z < 0$	100%

The closed-form sign conclusions of Proposition 3(iv) hold at every grid point. The statement is a numerical observation for $y > 0$; the main results of the paper do not depend on the $y > 0$ extension.

C Canonical calibration details

C.1 (V-prim) at canonical

At $(a, \tau_\eta, \tau_S, \tau_z) = (1, 1, 1, 1)$, the primitive condition (V-prim) reduces to $u + u^5 = 1$ with $u = 1 - y$; numerical root $u^* = 0.7549$, so $\rho_{vp} = \sqrt{1 - 0.7549} = 0.4951$. Condition (V**) holds numerically for $\rho \in [0, \sim 0.65]$, a strictly larger region.

C.2 κ^* threshold and the canonical calibration

Define κ^* as the minimum cost-curvature at which $\rho^M < \rho_{vp}$ at canonical θ:

$$\kappa^*(\theta) := -\frac{1}{c(\rho_{vp})} \cdot \frac{d\Phi_1(1;\rho)}{d\rho}\bigg|_{\rho=\rho_{vp}}.$$

Numerically $\kappa^* \approx 1.04$ at canonical θ. The paper's canonical calibration $\kappa = 1.5$ sits inside the (V-prim) region with margin $\rho_{vp} - \rho^M = 0.4951 - 0.40 = 0.095$.

C.3 Canonical equilibrium magnitudes

At $\kappa = 1.5$:

Quantity	Planner	Monopoly	Competition
ρ	0.02	0.40	0.90
$\tau_{\phi,\infty}$	1.00	0.73	0.48

Monopoly-competition gap: $(0.73 - 0.48)/0.48 = 52\%$ of competitive baseline. Planner-monopoly gap: $(1.00 - 0.73)/0.73 = 37\%$ of monopoly baseline. Against the Bai et al. [2016] empirical R^2-based range $[0.6, 0.9]$, the model's monopoly value 0.73 is central and competitive provision pushes informativeness below 0.5 — an order-of-magnitude illustration (the τ_ϕ-to-BPS mapping is not formal; see Section 7.5).

C.4 Threshold values at canonical

At canonical $(\tau_\eta, \tau_z, a) = (1, 1, 1)$:

$$\tau_S^*(0) = \frac{2}{1+\sqrt{7}} \approx 0.549, \qquad \tau_S^*(\bar{\rho}^2 = 0.81) \approx 0.70.$$

Canonical $\tau_S = 1$ exceeds both; the high-τ_S regime and the full informativeness ranking apply. The low-τ_S inversion of Section 5.6 applies for $\tau_S < 0.549$ (full inversion) and $\tau_S \in (0.549, 0.70)$ (partial).

D Three-regime numerical appendix

The preliminary numerical investigation of Section 6 uses the script `code/explore/three_regime_v9.py` with $\tau_P \in \{0.5, 1.0, 1.5, 2.0, 3.0\}$ and $c_P \in \{0.05, 0.10, 0.15, 0.20\}$ (20 grid points at canonical $\tau_v = \tau_\eta = \tau_z = a = 1$, $\tau_S = 1.5$, $\bar{\rho} = 0.9$, $c_0 = 0.1$, $\kappa = 1.5$).

At canonical parameters, the interior three-regime equilibrium with $n_U > 0$ does not exist: both informed regimes strictly dominate uninformed at competitive-cost pricing, and the equilibrium sits at the corner $n_S + n_P = 1$. Over a 21-point ρ-grid on $[0.02, \bar{\rho}]$, evaluating (i) the competitive $\rho^C = \bar{\rho}$ by the hidden-action argument (Conjecture 1), (ii) the monopolist's ρ^M by direct optimization of $n_S[\Phi_1^S - c(\rho)]$ over (ρ, n_S) with n_P adjusting to satisfy $\Phi_1^P = c_P$, and (iii) the planner's ρ^{SP} by direct maximization of $W = -(1/(2a))[n_S \log V_S + n_P \log V_P + n_U \log V_U] - n_S c(\rho) - n_P c_P$ over ρ at the buyer-side competitive fixed point, the informativeness ranking $\tau_{\phi,\infty}(\rho^{SP}) > \tau_{\phi,\infty}(\rho^M) > \tau_{\phi,\infty}(\rho^C)$ holds on 20/20 grid points.

The numerical result supports Conjecture 2 at the particular parameter values tested but does not prove it analytically. A rigorous proof requires characterizing the three-regime buyer-side fixed point either in an interior parameter range or at the corner $n_S + n_P = 1$ via the S-versus-P indifference. Both approaches are open.