Complex Capital Markets

Sergei Glebkin, Bryan Kelly, Semyon Malamud, Matias Palmunen, and Mohammad Pourmohammadi*

This version: November 5, 2026

Abstract

We propose the Complex Market Hypothesis: substantial return predictability arises from many individually weak signals. We model a computational arms race in which investors choose model scale and learn from limited data. Random matrix theory yields an essentially unique linear equilibrium and characterizes the distribution of model sizes. Prices transmit information from larger models to smaller ones, allowing both to coexist. Cheaper computation can raise aggregate spending and widen performance gaps across funds. Even when every investor observes all signals determining payoffs, uncertainty about their predictive coefficients sustains a sizable equity premium.

*Sergei Glebkin is at INSEAD, CEPR, and FTG. Bryan Kelly is at AQR Capital Management, Yale School of Management, and NBER. Semyon Malamud is at the Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. Matias Palmunen is at the Swiss Finance Institute, EPFL. Mohammad Pourmohammadi is at Yale School of Management. Kanji Suzuki and Junwei Yang provided excellent research assistance. We thank Caio Almeida, Peter Bossaerts, Daniel Chen, Pierre Collin-Dufresne, Victor DeMiguel, James Dow, Maryam Farboodi, Paul Fontanier, Zhiguo He, Christopher Hennessy, Peter Kondor, Andreas Neuhierl, Anna Pavlova, Lasse Heje Pedersen, Indira Puri, Stephen Schaefer, Alp Simsek, and Motohiro Yogo, as well as seminar participants at London Business School and Princeton University, for helpful comments. AQR Capital Management is a global investment management firm that may or may not apply investment techniques or methods of analysis similar to those described herein. The views expressed here are those of the authors and not necessarily those of AQR. Semyon Malamud gratefully acknowledges financial support from the Swiss Finance Institute and the Swiss National Science Foundation, Grant 100018-228042. This work was also supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID abs25 on Alps. Anthropic's Claude was used to assist with parts of the mathematical derivations, numerical checks, coding, and editing; the authors are solely responsible for all content and any errors.

1 Introduction

Stock returns may be difficult to forecast because predictability is dispersed across many weak relationships. Each contributes little on its own, yet together they can explain substantial variation in returns. We call this view the *Complex Market Hypothesis*. It allows for a few strong predictors but emphasizes information that small forecasting models leave unexploited and larger ones can recover.

Empirical evidence motivates this hypothesis. Simple models have had limited success in predicting returns, both at the aggregate market level and in the cross-section.[1] Kelly et al. (2024) establish conditions under which out-of-sample forecast accuracy and portfolio performance improve with model complexity, even when parameters outnumber observations. In their U.S. equity market application, richer nonlinear transformations of a fixed predictor set deliver these gains (see also Kelly et al., 2025b; Kelly and Malamud, 2025). Complexity can thus arise from richer mappings between existing economic variables and returns, whose estimation requires highly parameterized specifications. For example, Timmermann and Vulicevic (2026) train neural networks with up to one million parameters on monthly U.S. stock returns. As Kelly et al. (2026) explain, a network of this size is a model with one million weak signals.

The investment industry appears to be responding to these opportunities. Quantitative firms invest in larger datasets, more sophisticated forecasting methods, and the infrastructure needed to train and operate them. Table 5, adapted from Timmermann and Vulicevic (2026) and updated to November 2026, collects public statements by eight major quantitative firms. They report petabytes of data and computing clusters with up to tens of thousands of GPUs and hundreds of thousands of CPU cores. These disclosures illustrate how technological advances have made complex models economically feasible and fueled a computational arms race. As computing costs fall further, will the largest models dominate, or can smaller ones remain viable? Will prices become so informative that the gains financing this investment disappear? How will these advances affect aggregate risk and expected returns?

We develop an equilibrium theory of this race. Investors choose how many signals to acquire and estimate their predictive relationships from limited historical data. Observing

[1] Welch and Goyal (2008) find that many proposed equity-premium predictors fail to outperform the historical mean out of sample. In the cross-section, Gu et al. (2020) report a monthly out-of-sample R^2 of 0.16% for a linear model using size, value, and momentum, compared with 0.40% for a neural network evaluated on the same sample and against the same benchmark. Kelly and Xiu (2023) review the broader evidence on forecasting gains from complex machine-learning models.

a signal does not reveal how it predicts payoffs: cheaper computation can broaden access while leaving investors uncertain about its predictive content.

Consider an asset with payoff

$$d = \beta' S,$$

where S contains N predictive signals and β their unknown coefficients.[2] Each signal contributes little individually, but their joint contribution is economically significant. Investors are identical ex ante, have constant absolute risk aversion, and learn β through Bayesian updating from T historical observations of signals and payoffs at dates τ, generated by $d_\tau = \beta' S_\tau$. Even a long history may be insufficient to estimate numerous weak relationships precisely. Assuming that investors know β would remove the uncertainty central to our analysis.

In the classical statistical setting, N is fixed and T is large. To capture the presence of many predictors, we instead assume that both N and T are large. Their ratio, $c = N/T$, is the complexity ratio: the number of predictive relationships relative to the length of the history. Investors choose model scale $q = N_1/N$ by paying to acquire N_1 signals. Signal sets are nested: a larger model includes every signal used by a smaller one. At full scale, $q = 1$, an investor observes every signal determining the payoff in this simplified specification. Yet access to S need not reveal $\beta' S$ when β is unknown.

Our main technical contribution is a framework that combines parameter learning with inference from prices: *complex rational expectations equilibrium* (CREE). With few exceptions (Breon-Drish, 2015; Albagli et al., 2024; Avdis and Glebkin, 2026), rational expectations models with learning from prices are linear and Gaussian; even the exceptions assume a known joint distribution of signals and payoffs. Estimating predictive coefficients produces non-Gaussian forecasts and nonlinear demands, requiring investors to extract information from prices that are themselves nonlinear. In our case, the inference problem is also high-dimensional: investors must estimate both the N predictive coefficients and a signal covariance matrix with order N^2 unknown entries. We use random matrix theory to show that these nonlinearities vanish asymptotically, yielding explicit formulas that combine Bayesian updating, inference from prices, market clearing, and endogenous model scale. With a fixed signal set, the equilibrium has a closed form and is unique. Greater complexity can raise or lower participation; at high complexity, more weak-signal predictability can reduce it

[2]For exposition, we suppress strong signals and residual payoff noise, both of which are allowed in the full model.

(Proposition 3). When a positive mass is informed, CREE is also the limit of exact equilibria of regularized finite economies (Appendix L).

Whereas Kelly et al. (2024); Kelly and Malamud (2025); Didisheim et al. (2024) study the benefits of complexity taking return dynamics as given, we let investors choose model scale while learning from prices that reflect competing forecasts. Their choices therefore depend jointly on computational costs and the information revealed in equilibrium.

Within the linear class, an equilibrium exists and is unique. Under mild cost conditions, whenever investors acquire information, a positive mass chooses the largest model in the market (Theorem 5). If an additional condition on primitives holds, sufficiently low fixed costs also support a continuum of smaller models below this group; our calibration satisfies the condition. Investors are identical ex ante: they have the same preferences, prior, and learning technology and can acquire signals from the same universe. Yet they endogenously choose heterogeneous models, which Montiel Olea et al. (2022) and Al-Najjar (2009) take as given. Those using smaller models supplement their forecasts with information conveyed by trades based on larger ones, saving on computation while recovering some of the predictive content they did not acquire directly. Even with this information sharing through prices, the virtue of complexity of Kelly et al. (2024) survives in equilibrium: among the scales in use, larger models leave less payoff risk after inference from prices and earn higher gross Sharpe ratios.

Falling fixed computing costs alter this structure in stages. At prohibitive levels, no investor acquires a complex model. Once entry becomes worthwhile, a group forms at a common large scale. Further reductions initially expand this group and can then support smaller models alongside it. Participation broadens while the largest scale remains unchanged.

Lower variable costs make larger models worthwhile. Participation and concentration can respond nonmonotonically: cheaper acquisition encourages entry, but more informative prices make reliance on others' forecasts more attractive. The largest model may therefore grow even as fewer investors acquire signals. When both fixed and variable costs are sufficiently low, every investor acquires the full signal set, eliminating differences in access.

We use a calibration to compare fund performance across equilibria with different costs. Reductions in fixed costs intensify competition as entrants reveal more information through prices. Gross Sharpe ratios fall first at the largest scale and, once smaller models enter, at the bottom of the distribution.[3] From high levels, lower variable costs benefit investors at the

[3]Consistent with this pattern, Chen et al. (2026) find that AI hedge funds outperformed non-AI hedge funds by about 6% per year in earlier years, and that this gap became statistically indistinguishable from zero after 2017.

top, who improve performance by expanding their models. The smallest models lose ground because their signals are observed by every larger competitor and become increasingly visible in prices. The performance gap therefore widens. As variable costs fall further, it narrows because prices reveal more of the largest models' information. It vanishes once costs are low enough that every investor acquires the full signal set.

Cheaper computation can also raise aggregate spending.[4] Near the entry threshold, our calibration shows that lower fixed costs attract enough investors to more than offset the saving per participant. Reductions in variable costs instead induce expansion of the largest models. A tenfold decline in the variable cost coefficient from a high level raises the largest model's annual compute bill from 257 to 405 basis points and aggregate spending from 38 to 54 basis points, despite falling participation. Spending by the smallest models barely changes.

An apparent paradox is that computing expenditure can rise substantially while predictability barely improves. Lower fixed costs leave the largest model's forecasting accuracy unchanged, while lower variable costs improve it by expanding the model scale. Yet in our calibration, even when every investor observes all signals and there is no residual payoff noise, annual residual volatility remains 16.1%, compared with unconditional payoff volatility of 16.9%.

The resolution is that information-based trading profits depend on the *predictable* component of returns. An improvement that is tiny relative to total uncertainty can be substantial relative to the predictable component and generate a large performance gain. At the baseline, the largest and smallest acquired models leave annual residual volatilities of 16.28% and 16.35%, respectively, yet their annual gross Sharpe ratios are 0.42 and 0.30. A small forecasting edge thus supports a roughly 40% performance advantage and can justify substantial compute spending. The higher gross performance compensates for higher computation costs.

The computational arms race can thus transform the investment industry and eliminate differences in signal access without eliminating aggregate risk. In our calibration, the equity premium remains about 4.7% even when computing costs are low enough for every investor to acquire all signals. The payoff is fully determined by these signals and their coefficients, with no additional noise. Investors observe the signals but do not know the coefficients that

[4]We know of no aggregate series for funds' computing expenditures, but related measures have risen. Neudata (2026) estimate that investment managers spent at least $2.5 billion on alternative data in 2024, after average annual growth of 21% since 2020. Torbey et al. (2025) report that asset managers' technology costs rose by 9% in 2024, compared with 7% for total costs. In April 2026, Jane Street committed about $6 billion to the cloud platform of one compute provider (CoreWeave, 2026).

map them into payoffs, making parameter uncertainty the sole remaining source of aggregate payoff risk.

2 Literature review

We formalize the Complex Market Hypothesis in a complex rational expectations equilibrium.[5] The framework connects two theoretical literatures. The first studies noisy rational expectations equilibria, in which investors have asymmetric information and learn from prices (e.g., Grossman and Stiglitz, 1980; Admati, 1985; Wang, 1993).[6] Classical models are static, linear, and Gaussian; extensions allow dynamics, nonlinearities, or non-Gaussian distributions (e.g., Wang, 1993; Breon-Drish, 2015; Avdis and Glebkin, 2026; Albagli et al., 2024). The economy's structure and parameters, however, are generally common knowledge.

The second examines parameter learning while abstracting from the feedback between prices and investors' inferences about private information. This literature offers explanations for the equity premium, excess volatility, tail risk, return predictability, and behavioral "anomalies" (e.g., Gennotte, 1986; Timmermann, 1993; Pástor and Veronesi, 2003; Lewellen and Shanken, 2002; Brav and Heaton, 2002; Weitzman, 2007). Pástor and Veronesi (2009) and Veldkamp (2011) survey this work. Recent contributions study slow learning when inference is hindered by nonlinearity, high dimensionality, or persistent state dynamics (e.g., Breon-Drish, 2015; Collin-Dufresne et al., 2016; Martin and Nagel, 2022; Farmer et al., 2024; Ashwin et al., 2025; Moll, 2026). Related work incorporates machine learning and artificial intelligence into models of learning in financial markets (Dou et al., 2025, 2026; Ross et al., 2026).

CREE unifies these two forms of learning. Random matrix theory simplifies the nonlinear dependence of equilibrium prices on investors' parameter estimates. In the limit, forecasts are Gaussian and demands and prices are linear in those forecasts, while parameter uncertainty continues to generate payoff risk.

Two closely related papers study how high-dimensional learning affects equilibrium asset prices. Martin and Nagel (2022) show that Bayesian learning can generate in-sample return predictability even when investors use available information optimally and returns are unpredictable out of sample. Davis and Lopez-Lira (2026) show that estimation uncer-

[5] Complexity here is statistical: investors face many weak signals relative to the length of the available history. We do not study cognitive complexity or bounded rationality (Gabaix and Graeber, 2024; Puri and Veldkamp, 2026; Molavi et al., 2024), or complexity generated by interactions among heterogeneous agents, as in Bouchaud (2008, 2013, 2021).

[6] For surveys, see Vives (2008) and Brunnermeier (2001).

tainty generates pricing errors that do not diversify away as the number of assets grows, undermining the diversification argument behind arbitrage pricing theory. Martin and Nagel assume that investors observe the same information. Davis and Lopez-Lira allow heterogeneous information sets but specify them exogenously. Our main question concerns the equilibrium interaction between model scale and market efficiency. Both papers take the scale of investors' information as given. In CREE, investors' model choices shape the information incorporated into prices, while price informativeness changes the gains from acquiring a larger model. These costly choices also generate information heterogeneity among investors who are identical ex ante.

Earlier work examines non-Bayesian rules for learning forecasting relationships. Bray (1982) analyzes a repeated Grossman and Stiglitz (1980) market in which uninformed traders regress returns on prices by least squares, ignoring how their learning changes that relationship. Marcet and Sargent (1989) extend convergence results for least-squares learning to settings with private information, while Routledge (1999) studies traders who imitate successful strategies and experiment. These papers ask whether such rules converge to a rational expectations equilibrium. In Banerjee and Green (2015), Bayesian investors learn from prices and dividends about a latent trader type that determines whether other traders' signals are informative. Our Bayesian investors estimate a high-dimensional predictive model whose parameters remain uncertain in equilibrium.

CREE is also, to our knowledge, the first rational expectations equilibrium in which Bayesian investors learn from prices while knowingly using misspecified models that omit signals. Investors in classical noisy rational expectations models likewise miss some signals, but know how these covary with those they observe, so their models remain correctly specified. Kelly et al. (2024) study misspecified forecasting models outside equilibrium, and Cerreia-Vioglio et al. (2026) examine decision makers who recognize their specification errors. Existing equilibrium analyses generally assume that agents do not account for misspecification (Esponda and Pouzo, 2016; Bastianello and Fontanier, 2025; Ellis et al., 2026). Doing so is typically intractable: optimal use of the acquired signals requires knowing their covariances with the omitted ones, and these are too numerous to estimate from the data. In CREE, random matrix theory reduces this problem to a few parameters of the covariance prior. This result makes models of every scale tractable, allowing us to study equilibrium competition among them.

Our results also help explain the virtue of complexity documented in empirical asset pricing, where highly parameterized machine-learning models deliver substantial gains in

return prediction (e.g., Gu et al., 2020). Beginning with Kelly et al. (2024); Didisheim et al. (2024), several papers measure how performance varies with model size. Timmermann and Vulicevic (2026) estimate the performance limit as computation grows. This evidence motivates the many weak signals in CREE.

3 The model

We consider a one-period economy in the spirit of Grossman and Stiglitz (1980). There are two assets: a risk-free asset, which is the numeraire and has a gross return of one, and a risky asset with price P and payoff d. Investors trade at date 0, and d is realized at date 1.

3.1 Signals, beliefs, and data

Our Complex Market Hypothesis states that predictability is dispersed across many weak relationships, possibly alongside a few strong ones. Assumption 1 formalizes it: the payoff d is predicted by n strong signals and N weak signals. We study this economy with both the signal dimension and the available history large. Assumption 1 also states the investors' common prior over the unknown parameters. We take this prior to be correctly specified: the true parameters are a draw from it. This is without loss for our results on information acquisition and model scale, which investors choose ex ante under the prior.

Assumption 1 (Complex Market Hypothesis) *There are two types of signals, n strong signals $s \in \mathbb{R}^n$ and N weak signals $S \in \mathbb{R}^N$, such that*

$$d \;=\; b's \;+\; \beta' S \;+\; \varepsilon\,, \tag{1}$$

for vectors $b \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^N$. The residual $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is independent of the Gaussian signals $\binom{s}{S} \sim \mathcal{N}(0, \operatorname{diag}(\psi, \Psi))$, where $\psi \in \mathbb{R}^{n\times n}$ and $\Psi \in \mathbb{R}^{N\times N}$ are positive-semidefinite.[7] *The weak-signal parameters (β, Ψ) are unknown, and investors share a common prior under which they are independent, with*

$$\beta \;\sim\; \mathcal{N}\!\left(0,\; \frac{\sigma_\beta^2}{N}\, I_N\right), \qquad \Psi \;\sim\; \mathcal{W}^{-1}\big(V_0 N\, I_N,\; (1+\kappa)N\big), \tag{2}$$

[7] We assume that strong and weak signals are uncorrelated for convenience. The assumption can be relaxed at the cost of more complicated expressions for the equilibrium quantities.

for constants $\sigma_\beta^2 > 0$, $V_0 > 0$, *and* $\kappa > 0$, *where* I_N *is the* $N \times N$ *identity matrix and* $\mathcal{W}^{-1}(\Psi_0, \nu_\Psi)$ *is the inverse-Wishart distribution with scale matrix* Ψ_0 *and* ν_Ψ *degrees of freedom (Appendix A gives its density).*[8] *The number* n *of strong signals and the strong-signal parameters* (b, ψ) *are fixed and known.*[9]

Under the prior (2), each weak-signal coefficient is small, $\beta_i = O_p(N^{-1/2})$, consistent with the standard definition of weak signals (Belkin et al., 2020; Giglio et al., 2021; Shen and Xiu, 2024; Liao et al., 2026), whereas the strong-signal coefficients b are fixed. This split reflects the predictive structure of most real-world economic relationships: a few strong signals alongside a potentially vast number of weak ones (Giannone et al., 2021).

The data available before trading are a *history* of T past observations generated by the same law,

$$d_\tau \;=\; b' s_\tau \;+\; \beta' S_\tau \;+\; \varepsilon_\tau\,, \qquad \tau = 1, \ldots, T, \tag{3}$$

where, given the parameters, the $(s_\tau, S_\tau, \varepsilon_\tau)$ are independent draws from the distribution of (s, S, ε) in Assumption 1, independent of the current draw. Symbols without the index τ refer to the current draw. Stacking the coefficients as $\boldsymbol{\beta} = \binom{b}{\beta} \in \mathbb{R}^{n+N}$ and the signals as $\boldsymbol{\mathcal{S}} = \binom{s}{S} \in \mathbb{R}^{n+N}$, the current payoff is $d = \boldsymbol{\beta}' \boldsymbol{\mathcal{S}} + \varepsilon$. Bold symbols stack the history: $\mathbf{s} = (s_1, \ldots, s_T)' \in \mathbb{R}^{T \times n}$, $\mathbf{S} = (S_1, \ldots, S_T)' \in \mathbb{R}^{T \times N}$, $\boldsymbol{\mathcal{S}} = [\mathbf{s}, \mathbf{S}] \in \mathbb{R}^{T \times (n+N)}$, $\mathbf{d} = (d_1, \ldots, d_T)' \in \mathbb{R}^T$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)' \in \mathbb{R}^T$. In this notation, (3) reads $\mathbf{d} = \boldsymbol{\mathcal{S}} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

The length T of the history is the second dimension of our analysis. The classical econometric setting takes the sample size T to be much larger than the number of parameters, formally $T \to \infty$ with N fixed. The history then reveals the parameters exactly, and parameter uncertainty disappears, as is standard in rational-expectations models. We instead consider the case in which T and N are both large and comparable, formalized as the limit $N, T \to \infty$ with $N/T \to c > 0$; we call c the *complexity ratio.*[10]

[8] Degrees of freedom proportional to N and a scale $V_0 N I_N$ that grows with N keep the prior mean $E[\Psi] = \frac{V_0 N}{\kappa N - 1} I_N$ bounded as $N \to \infty$ and ensure that the prior is nondegenerate for large N.

[9] For informed investors, taking (b, ψ) as known is without loss. Because n is fixed, they can estimate (b, ψ) from the history at the $\sqrt{T}$ rate, and their payoff posterior has the same limit as with a known strong block (Proposition B.1 in Appendix B.3). Uninformed investors do not observe the history. That they know (b, ψ) is an assumption.

[10] Requiring N/T to converge to a positive constant is not a major restriction. With both T and N large, there are three cases. If T is much larger than N, the history reveals the parameters, as in the classical limit. If N is much larger than T, the history is uninformative, and beliefs are determined by the prior. Only the proportional case keeps both the data and the prior relevant, and the other two can be approached as its limits $c \to 0$ and $c \to \infty$. Parameter uncertainty persists in this case, as in high-dimensional Bayesian

Unless explicitly considering a boundary case, the analytical derivations maintain $V_0 > 0$, $\sigma_\varepsilon^2 > 0$, and $\sigma_\eta > 0$. Appendix J uses the zero-noise case $\sigma_\varepsilon^2 = 0$ with $c > 1$. The proofs do not cover this case, and Section 4.3 defines it by continuity in (12).

3.2 Investors

Following Grossman and Stiglitz (1980), the economy is populated by three classes of agents:

- *Informed* investors have CARA preferences with risk aversion $\gamma > 0$. They observe the history $(\boldsymbol{\mathcal{S}}, \mathbf{d})$, the current signals (s, S), and the price. As in Grossman and Stiglitz (1980), the price adds nothing to what they already know, so their information set is $\mathcal{F}_I = \sigma(\boldsymbol{\mathcal{S}}, \mathbf{d}, s, S)$.
- *Uninformed* investors have the same preferences but no access to the data, current or past, and learn only from the price: $\mathcal{F}_U = \sigma(P)$.
- *Noise* traders supply $\sigma_\eta \eta$ shares of the asset inelastically, where $\eta \sim \mathcal{N}(0,1)$ is independent of everything else and σ_η controls the size of noise trading.

Informed and uninformed investors form a unit continuum, a fraction λ of which is informed. Both are rational Bayesians: they share the common prior of Assumption 1 and maximize expected utility conditional on their information sets.

As in Grossman and Stiglitz (1980), the fraction λ is determined by the cost of information. To become informed, an investor pays a cost $\chi > 0$: the expense of building the infrastructure to collect, clean, and process the history and to extract the current signals.

3.3 Equilibrium

We define equilibrium directly in the limiting economy as $N, T \to \infty$ with $N/T \to c$, which we call *the limit* from now on.[11] The limiting economy has a limiting probability space $(\Omega, \mathcal{A}, \mathbb{P}_\infty)$, a limiting payoff d^∞, and limiting informed information set $\mathcal{F}_I^\infty$; E^∞ denotes expectation under $\mathbb{P}_\infty$. Table 6 in Appendix L lists all probability laws used in the paper. Appendix B specifies these objects. We first take the fraction λ of informed investors as given and then determine it by information acquisition.

learning, where posteriors need not concentrate on the truth (Diaconis and Freedman, 1986; Ritov et al., 2014). We show how this uncertainty affects prices and acquisition incentives.

[11] Unless stated otherwise, deterministic limits hold in probability. Random limits hold in distribution, and almost surely on the coupled probability space of Appendix B.2. On that space, functions of the price p, such as demand schedules, converge uniformly in p on compact sets.

Definition 1 (Complex rational expectations equilibrium) *Fix $\lambda \in [0,1]$. For each $p \in \mathbb{R}$, let $x_I(p)$ be $\mathcal{F}_I^\infty$-measurable and $x_U(p)$ deterministic, and let P be $\mathcal{F}_I^\infty \vee \sigma(\eta)$-measurable. Then x_I, x_U, and P form a complex rational expectations equilibrium (CREE) if the following conditions hold.*

(i) Informed optimality. For every realization of the informed investors' information and each $p \in \mathbb{R}$,[12]

$$x_I(p) \;\in\; \operatorname*{argmax}_{x\in\mathbb{R}} \; E^\infty\left[-e^{-\gamma x(d^\infty - p)} \,\middle|\, \mathcal{F}_I^\infty\right].$$

(ii) Uninformed optimality. For each $p \in \mathbb{R}$,

$$x_U(p) \;\in\; \operatorname*{argmax}_{x\in\mathbb{R}} \; E^\infty\left[-e^{-\gamma x(d^\infty - p)} \,\middle|\, P = p\right].$$

(iii) Market clearing. For every realization of the informed investors' information and the supply shock,

$$\lambda\, x_I(P) \;+\; (1-\lambda)\, x_U(P) \;=\; \sigma_\eta\, \eta\,.$$

An equilibrium consists of demand schedules $x_I(p)$ and $x_U(p)$, which specify the amount an informed or uninformed investor is willing to buy or sell at price level p, and a price function P, such that both schedules are optimal and the market clears. Informed demand also depends on the informed investors' information; we suppress this argument. We restrict the price to depend on the state of the economy only through the informed investors' information and the supply shock. Since the supply shock is independent of the payoff and of $\mathcal{F}_I$, the price carries no information about the payoff beyond $\mathcal{F}_I$, which is why $\mathcal{F}_I$ above omits the price. We focus on CREE in which the uninformed demand x_U is continuous and nonincreasing in the price.[13] From now on, CREE means a CREE in this class.

As in Grossman and Stiglitz (1980), the fraction λ adjusts until no investor gains by switching between being informed and uninformed. Given a CREE at λ, the ex-ante expected utilities of an informed and an uninformed investor are

$$V_I \;=\; e^{\gamma\chi}\, E^\infty\left[-e^{-\gamma x_I(P)(d^\infty - P)}\right], \qquad V_U \;=\; E^\infty\left[-e^{-\gamma x_U(P)(d^\infty - P)}\right], \tag{4}$$

[12] The conditional expectation is the version specified in Appendix B, the limit of the finite Bayesian posteriors along the observed information.

[13] Rational expectations models can also admit discontinuous equilibria; see Pálvölgyi and Venter (2015).

where the factor $e^{\gamma\chi}$ accounts for the cost of information. Both are evaluated under $\mathbb{P}_\infty$, so investors value information under their common limiting beliefs.

Definition 2 (Overall equilibrium) *An overall equilibrium is a fraction $\lambda \in [0,1]$ and a CREE (x_I, x_U, P) at λ in which each investor chooses whether to acquire information so that its ex-ante utility in (4) is maximized.*

Under the nondegeneracy conditions of Appendix L, the positive-mass equilibria of Theorems 1 and 2 are limits of finite REEs under regularized laws, with vanishing covariance and price-support regularizations. The construction includes binary information acquisition and recovers the limiting prices, portfolios, and informed fractions. It does not cover the zero-informed case, the endogenous scales of Section 5, or the zero-noise calibration of Appendix J.

Unless a finite economy is explicitly under consideration, we suppress the superscript ∞ on limiting random variables and information sets and write E, Var, and Cov for moments under $\mathbb{P}_\infty$.

4 Equilibrium characterisation

We take the informed fraction $\lambda \in (0,1]$ as given and characterise the CREE of Definition 1 in four steps. First, we solve the informed investors' Bayesian problem and derive the limit of their demand. Second, market clearing shows that the price reveals the same signal in every CREE. Third, we solve the uninformed investors' signal-extraction problem. Finally, optimality determines the uninformed demand, which is linear, and market clearing gives the price.

4.1 Learning by informed investors

Beyond the residual ε, the payoff (1) is uncertain for two reasons. *Fundamental uncertainty* is uncertainty about the current signals s and S, the *fundamentals* that drive the payoff. *Model uncertainty* is uncertainty about the predictive coefficients β that map the fundamentals into the payoff. The uninformed face both. The informed observe the fundamentals and so face only model uncertainty, which they reduce by learning β from the history. This learning is tractable: the posterior of β is Gaussian, as Lemma 1 below shows.

Throughout this subsection, N and T are finite: s, S, and the history $(\boldsymbol{\mathcal{S}}, \mathbf{d})$ are those of the economy with N weak signals and T past observations, and the posteriors are exact.

Section 4.3 takes the limit. Define the shrinkage parameter

$$z = \frac{\sigma_\varepsilon^2}{\sigma_\beta^2} \tag{5}$$

and the sample covariance matrix

$$\hat{\Psi} = \frac{1}{T}\,\mathbf{S}'\mathbf{S} \;\in\; \mathbb{R}^{N\times N}. \tag{6}$$

We write $c_N = N/T$ for the finite ratio and I for the $N \times N$ identity matrix.

Lemma 1 (Bayesian updating) ***(i) Coefficients.*** *The informed investors' posterior for the weak-signal coefficient is Gaussian,*

$$\beta \mid \mathcal{F}_I \;\sim\; \mathcal{N}(\hat{\beta}, \hat{\Sigma}_\beta),$$

where, writing $\mathbf{d}^\perp = \mathbf{d} - \mathbf{s}\,b$ *for the payoff history net of the known strong component,*

$$\hat{\beta} = T^{-1}(zc_N I + \hat{\Psi})^{-1}\mathbf{S}'\mathbf{d}^\perp \tag{7}$$

and

$$\hat{\Sigma}_\beta = \mathrm{Var}[\beta \mid \mathcal{F}_I] = T^{-1}\sigma_\varepsilon^2(zc_N I + \hat{\Psi})^{-1}\,. \tag{8}$$

(ii) Payoff. *Correspondingly, the payoff is conditionally Gaussian,*

$$d \mid \mathcal{F}_I \;\sim\; \mathcal{N}\Big(b's + \hat{\beta}'S,\; \sigma_\varepsilon^2 + S'\hat{\Sigma}_\beta S\Big)\,.$$

4.2 Optimal demand of informed investors

Lemma 1 makes the informed investors' problem tractable. Their payoff is conditionally Gaussian, so with CARA preferences, maximizing $E[-e^{-\gamma x(d-p)} \mid \mathcal{F}_I]$ reduces to a standard mean–variance problem. Its solution is

$$x_I(p) \;=\; \frac{E[d \mid \mathcal{F}_I] - p}{\gamma \mathrm{Var}[d \mid \mathcal{F}_I]} \;=\; \frac{b's + \hat{\beta}'S - p}{\gamma\big(\sigma_\varepsilon^2 + S'\hat{\Sigma}_\beta S\big)}\,. \tag{9}$$

4.3 Limiting demand of informed investors

The demand (9) is nonlinear in the signals, which complicates inference from the price for the uninformed. Its numerator contains the forecast $M_N = b's + \hat{\beta}'S$, a linear form in the N current signals with coefficients $\hat{\beta}$ estimated from the history. The forecast is Gaussian given the history and Ψ, but its variance $b'\psi b + \hat{\beta}'\Psi\hat{\beta}$ is random, so it is not Gaussian unconditionally. Its denominator, $\sigma_\varepsilon^2 + S'\hat{\Sigma}_\beta S$, is a quadratic form in the same signals, and the matrix $\hat{\Sigma}_\beta$ of this form is itself a nonlinear function of the signal history. The slope of the demand is therefore random and correlated with the forecast.

Random matrix theory removes both nonlinearities in the limit. The quadratic form $S'\hat{\Sigma}_\beta S$ averages over signal dimensions, as the law of large numbers averages over observations in the classical setting with N fixed and T large. It converges in probability to a deterministic value (Lemmas I.1 and I.6 and Proposition I.4 in Appendix I).

The forecast needs no central limit theorem. Given (β, Ψ), the payoff is exactly Gaussian with variance $b'\psi b + \beta'\Psi\beta + \sigma_\varepsilon^2$. This variance converges in probability to the unconditional payoff variance

$$\sigma_d^2 \;=\; b'\psi b + \sigma_\beta^2 \frac{V_0}{\kappa} + \sigma_\varepsilon^2$$

by the same two lemmas, so the payoff is Gaussian in the limit. By Lemma 1, the payoff is the forecast plus an error that is Gaussian given $\mathcal{F}_I$, and the variance of this error has a deterministic limit. Hence the forecast is Gaussian in the limit as well. Lemma 2 gives both limits in closed form.

Lemma 2 (Equilibrium informed demand) *In the limit, the informed demands (9) converge in distribution, and almost surely on the coupled space of Appendix B.2, to*

$$x_I(p) \;=\; \frac{\tilde{S} - p}{\gamma\sigma_I^2(c)}\,, \tag{10}$$

where $\tilde{S}$ is the limit of the forecast $b's + \hat{\beta}'S$, and

$$\sigma_I^2(c) \;=\; \sigma_\varepsilon^2 \frac{Z_*}{zc}\,,$$

with $Z_* > zc$ the larger root of

$$(1 + c\kappa)\, Z_*^2 + c\big[V_0(1-c) - z(2 + c\kappa)\big] Z_* - c^2 z\,(V_0 - z) \;=\; 0\,.$$

The payoff d, the forecast $\tilde{S}$, and the supply shock η are jointly Gaussian, with

$$d \;=\; \tilde{S} + e\,, \qquad \tilde{S} \;\sim\; \mathcal{N}\big(0, \sigma_d^2 - \sigma_I^2(c)\big), \qquad e \;\sim\; \mathcal{N}\big(0, \sigma_I^2(c)\big), \tag{11}$$

where e is independent of $\tilde{S}$ and η. The demand (10) is the informed demand of every CREE.

Lemma 2 is the central technical result of the paper. It turns a nonlinear filtering problem into a Gaussian one, with residual risk in closed form. The effective ridge Z_* solves the Marchenko–Pastur fixed-point equation for the sample covariance of the signals (Marčenko and Pastur, 1967; Silverstein and Bai, 1995; Rubio and Mestre, 2011). Estimation error raises the ridge penalty from zc, the value implied by the prior, to Z_*, and this raises the residual risk to $\sigma_I^2(c) = \sigma_\varepsilon^2 Z_*/(zc)$. Because the prior spectrum is itself an inverse Marchenko–Pastur law, the fixed-point equation reduces to a quadratic with an explicit root (Appendix B.1). Random matrix theory is a standard tool in finance for covariance estimation (Ledoit and Wolf, 2017; Bun et al., 2017) and for the out-of-sample performance of complex models (Kelly et al., 2024; Didisheim et al., 2024). To our knowledge, Lemma 2 is the first use of the Marchenko–Pastur equation to characterize an equilibrium with learning from prices.

Lemma 2 requires $\sigma_\varepsilon^2 > 0$. At $\sigma_\varepsilon^2 = 0$, the ridge z is zero and the sample covariance $\hat{\Psi}$ is singular for $c > 1$, so the proof does not apply. We define this case by continuity, with the limit $N, T \to \infty$ taken first:

$$\sigma_I^2(c)\Big|_{\sigma_\varepsilon=0} \;:=\; \lim_{\sigma_\varepsilon \downarrow 0} \sigma_I^2(c) \;=\; \sigma_\beta^2 V_0\, \frac{(c-1)^+}{1+c\kappa}\,. \tag{12}$$

Indeed, $\sigma_I^2(c) = \sigma_\beta^2 Z_*/c$ for $z > 0$, and the larger root Z_* is continuous in z. At $z = 0$ the quadratic has the roots 0 and $cV_0(c-1)/(1+c\kappa)$. The limit is positive if and only if $c > 1$.

4.4 Learning by uninformed investors

The uninformed learn from the price. To see what the price reveals, substitute the informed demand (10) into market clearing,

$$\lambda\, \frac{\tilde{S} - P}{\gamma \sigma_I^2(c)} \;+\; (1-\lambda)\, x_U(P) \;=\; \sigma_\eta \eta\,,$$

and rearrange:

$$s_P \;\equiv\; \tilde{S} - \nu(c)\,\eta \;=\; P - \frac{(1-\lambda)\gamma\sigma_I^2(c)}{\lambda}\,x_U(P)\,, \qquad \nu(c) \;=\; \frac{\gamma\sigma_\eta\sigma_I^2(c)}{\lambda}\,. \tag{13}$$

The right-hand side depends only on the price and on the uninformed investors' own schedule, so the uninformed can recover s_P from the price. The price thus conveys a noisy signal of the informed forecast. Since $d = \tilde{S} + e$ with e independent of $\tilde{S}$ and η (Lemma 2), the uninformed solve a Gaussian signal-extraction problem, which the next lemma solves.

Lemma 3 (Uninformed posterior) *In every CREE, $d \mid P \sim \mathcal{N}\big(\alpha(c)\,s_P,\ \sigma_U^2(c)\big)$, where*

$$\alpha(c) \;=\; \frac{\tau_P}{\tau_S+\tau_P}\,, \qquad \sigma_U^2(c) \;=\; \sigma_I^2(c) + \frac{1}{\tau_S+\tau_P}\,, \qquad \tau_S \;=\; \frac{1}{\sigma_d^2-\sigma_I^2(c)}\,, \qquad \tau_P \;=\; \frac{1}{\nu(c)^2}\,.$$

4.5 Equilibrium

Under the posterior of Lemma 3, CARA optimization gives the uninformed demand

$$x_U(p) \;=\; \frac{\alpha(c)\,s_P(p) - p}{\gamma\sigma_U^2(c)}\,.$$

By (13), $s_P(p)$ itself depends on $x_U(p)$, so this is a linear equation for $x_U(p)$ at each price. Its solution is linear in p: linearity of the equilibrium is a consequence of the equilibrium conditions, not an assumption. Market clearing then determines the price.

Theorem 1 (Complex rational expectations equilibrium) *Let $\lambda \in (0,1]$ and $\sigma_\eta > 0$. There is a unique CREE. In it, the demand schedules are*

$$x_I(p) \;=\; \frac{\tilde{S}-p}{\gamma\sigma_I^2(c)}\,, \qquad x_U(p) \;=\; -\vartheta\,p\,, \qquad \vartheta \;=\; \frac{\lambda\big(1-\alpha(c)\big)}{\gamma\big[\lambda\sigma_U^2(c) + (1-\lambda)\alpha(c)\sigma_I^2(c)\big]} \;>\; 0\,, \tag{14}$$

and the price is

$$P \;=\; a(c)\big(\tilde{S} - \nu(c)\eta\big)\,, \qquad a(c) \;=\; \frac{\lambda\sigma_U^2(c) + (1-\lambda)\alpha(c)\sigma_I^2(c)}{\lambda\sigma_U^2(c) + (1-\lambda)\sigma_I^2(c)}\,, \tag{15}$$

with $\sigma_I^2(c)$ as in Lemma 2, $\nu(c)$ as in (13), and $\alpha(c)$ and $\sigma_U^2(c)$ as in Lemma 3.

4.6 Overall equilibrium

We now determine the informed fraction by information acquisition, as in Definition 2. In the CREE of Theorem 1, the payoff and the price are jointly Gaussian. As in Grossman and Stiglitz (1980), each investor's ex-ante expected utility is then proportional to its posterior standard deviation, so acquiring information is preferred exactly when $\sigma_U(c) > e^{\gamma\chi}\sigma_I(c)$. The uninformed standard deviation depends on the informed fraction, since more informed trading makes the price more informative; we write $\sigma_U(c;\lambda)$, with $\sigma_U(c;0) = \sigma_d$ when the price reveals nothing. The informed fraction is therefore

$$\lambda \;=\; \begin{cases} 0\,, & \sigma_U(c;0) \le e^{\gamma\chi}\sigma_I(c)\,, \\ 1\,, & \sigma_U(c;1) \ge e^{\gamma\chi}\sigma_I(c)\,, \\ \lambda^*\,, & \text{otherwise, where } \sigma_U(c;\lambda^*) = e^{\gamma\chi}\sigma_I(c)\,. \end{cases}$$

Substituting $\sigma_I(c)$ from Lemma 2 and $\sigma_U(c;\lambda)$ from Lemma 3 gives a closed form.

Theorem 2 (Overall equilibrium) *Let $\sigma_\eta > 0$ and $\chi > 0$. There is a unique overall equilibrium. Let $R_I^2(c) = 1 - \sigma_I^2(c)/\sigma_d^2$ be the share of payoff variance that the informed forecast explains, and $\zeta = 1 - e^{-2\gamma\chi}$ the minimum share at which acquiring information pays. The informed fraction is*

$$\lambda(c) \;=\; \begin{cases} 0\,, & R_I^2(c) \le \zeta\,, \\ \min\left\{\gamma\sigma_\eta\sigma_d\sqrt{\dfrac{\big(1-R_I^2(c)\big)\big(R_I^2(c)-\zeta\big)}{\zeta\, R_I^2(c)}},\; 1\right\}, & R_I^2(c) > \zeta\,, \end{cases}$$

and for $\lambda(c) > 0$ its CREE is that of Theorem 1.

Information is acquired exactly when the informed forecast explains more than the share ζ of the payoff variance, and, on the interior margin, the informed fraction rises with the noise-trading scale σ_η, which keeps the price from revealing the forecast.

4.7 Complexity and information acquisition

How does the number of investors who adopt complex models respond to the economy's primitives—how variable the signals are, how costly information is, how much noise trading there is? We show that the answer depends critically on complexity c, and we characterize

it through two polar regimes. As $c \to 0$, the informed learn the weak-signal coefficients β exactly: there is no model uncertainty, and the economy reduces to the classical setting of Grossman and Stiglitz (1980). As $c \to \infty$, the history is uninformative about the weak signal block, so beliefs about it are determined by the prior alone. Complexity thus governs whether the signals S are *usable:* knowing β tells the informed how to use them, while without it they fall back on the prior and leave the signals' predictive content unexploited. As a result, the comparative statics of information acquisition can differ in sign between the two regimes (Table 1).

On the interior acquisition margin, Theorem 2 can be written as

$$\lambda^2 = \frac{\gamma^2 \sigma_\eta^2 \sigma_d^2}{\zeta} \left(1 + \zeta - R_I^2 - \frac{\zeta}{R_I^2}\right), \qquad \zeta < R_I^2 < 1. \tag{16}$$

Thus, the informed R^2 is the key endogenous quantity determining entry, and we characterize how it behaves in the two regimes.

R_I^2 is the explained variance divided by the total variance, $\sigma_d^2 = b'\psi b + \sigma_\beta^2 V_0/\kappa + \sigma_\varepsilon^2$, containing the contributions of the strong block, the weak block, and the residual ε. At zero complexity the informed estimate β exactly and eliminate both the strong and weak contributions, leaving only ε, so the explained variance is $b'\psi b + \sigma_\beta^2 V_0/\kappa$. At high complexity they cannot estimate β and the signals S are not usable, so only the strong block is explained, $b'\psi b$, and the weak contribution becomes residual risk. Dividing by σ_d^2 gives

$$R_I^2(0) = \frac{b'\psi b + \sigma_\beta^2 V_0/\kappa}{b'\psi b + \sigma_\beta^2 V_0/\kappa + \sigma_\varepsilon^2}, \qquad R_I^2(\infty) = \frac{b'\psi b}{b'\psi b + \sigma_\beta^2 V_0/\kappa + \sigma_\varepsilon^2}. \tag{17}$$

Here the arguments 0 and ∞ denote the limits $c \downarrow 0$ and $c \to \infty$; in particular, $\sigma_I^2(0) = \sigma_\varepsilon^2$. How the informed fraction responds to the primitives then differs between the two regimes, as the following proposition shows.

Proposition 3 (Complexity and information acquisition) *Let $A_s := b'\psi b$ and $A_w := \sigma_\beta^2 V_0/\kappa$ be the strong and weak explainable variance. At the two complexity limits, λ depends on the signal parameters $(b, \psi, \sigma_\beta, V_0, \kappa)$ only through A_s and A_w. On interior acquisition margins:*

(i) At every complexity level, λ increases in the noise-trading scale σ_η and in strong explainable variance A_s, and decreases in the acquisition cost χ.

(ii) In weak explainable variance A_w, λ increases at $c = 0$. Let $A_s > 0$. At $c \to \infty$, the margin is interior only if $0 \leq A_w < \frac{1-\zeta}{\zeta} A_s - \sigma_\varepsilon^2$. On this range, λ increases for $A_w < A_w^*$ and decreases for $A_w > A_w^*$, where $A_w^* = \frac{1-\zeta}{2\zeta} A_s - \sigma_\varepsilon^2$. If $A_w^* \leq 0$, λ decreases in A_w throughout.

(iii) In c, λ is hump-shaped if $R_I^2(\infty) < \sqrt{\zeta} < R_I^2(0)$, increasing if $R_I^2(\infty) \geq \sqrt{\zeta}$, and decreasing if $R_I^2(0) \leq \sqrt{\zeta}$.

Table 1 collects the signs.

Table 1: Information acquisition at low and high complexity

Increase in	Low complexity $c = 0$	High complexity $c \to \infty$
Noise-trading scale σ_η	$\uparrow$	$\uparrow$
Acquisition cost χ	$\downarrow$	$\downarrow$
Strong explainable variance A_s	$\uparrow$	$\uparrow$
Weak explainable variance A_w	$\uparrow$	$\cap$

Complexity $c = N/T$:
$\cap$ if $R_I^2(\infty) < \sqrt{\zeta} < R_I^2(0)$;
$\uparrow$ if $R_I^2(\infty) \geq \sqrt{\zeta}$; $\downarrow$ if $R_I^2(0) \leq \sqrt{\zeta}$.

Notes: Entries give the response of the informed fraction λ at interior equilibria, holding other primitives fixed. $\uparrow$ and $\downarrow$ denote monotone responses; $\cap$ denotes a hump-shaped response.

The contrast between A_s and A_w is the main result. Strong explainable variance is usable at every complexity level: a larger A_s gives the informed forecast more to predict without adding to residual risk, so it always raises acquisition. Weak explainable variance is usable only to the extent that β can be learned. At low complexity, the informed estimate β is exact, so A_w is pure predictable variation and raises acquisition. At high complexity, the weak coefficients cannot be learned, so the same A_w becomes residual risk. Residual risk pulls acquisition two ways: it makes informed demand less sensitive to the forecast and leaves more noise in the price, which encourages entry, but it also shrinks the proportional risk reduction that acquisition delivers, which discourages it. At high complexity λ therefore first rises and then falls in A_w, whereas at low complexity it only rises.

5 Scaling in complex markets

We now endogenize the scale of investors' forecasting models. The complexity c of the full signal universe remains fixed, but investors choose how many signals to acquire. This choice balances the value of additional private information against its acquisition cost and the information conveyed by prices.

We order the weak signals and allow investors to acquire nested subsets, as under a data provider's subscription tiers. An investor choosing scale q acquires the historical and current observations of the strong signals and the first N_1 weak signals, where $N_1/N \to q \in (0, 1]$, at cost $\chi(q)$. Investors choose scale before observing either the acquired data or the equilibrium price. Alternatively, they can remain uninformed and learn only from the price at no cost.

The nested structure also arises when investors can choose which signals to acquire. Appendix G allows arbitrary signal subsets in the limiting Gaussian economy. When acquisition costs per unit of forecast variance are nondecreasing, every linear equilibrium of that economy is nested and coincides with the equilibrium characterized here. We discuss the intuition following Theorem 5.

Model scale has a natural interpretation in nonlinear forecasting. Suppose that payoffs satisfy

$$d_{t+1} = f_*(X_t) + \varepsilon_{t+1}, \tag{18}$$

where $X_t \in \mathbb{R}^D$ contains the underlying predictors and f_* is an unknown nonlinear function. Many machine-learning methods approximate this function through an expansion in nonlinear features:

$$f_*(X_t) \approx \sum_{k=1}^{N} \beta_k S_{k,t}, \qquad S_{k,t} = f_k(X_t). \tag{19}$$

A complex payoff function can require a rich expansion, with predictive content dispersed across many individually weak terms. In this case, nonlinear prediction gives rise to the many weak signals in our model, and scaling consists of expanding the set of features used to forecast payoffs (Kelly et al., 2024).[14]

[14]Gains from expanding the feature set also arise in stochastic discount factor estimation. Didisheim et al. (2024) increase the number of nonlinear factors from 36 to 360,000; their largest model attains an out-of-sample Sharpe ratio of 3.7. Kelly et al. (2025a) report a Sharpe ratio of 3.8 for a transformer with about 100,000 parameters and 4.6 for one with about one million. Their linear transformer is equivalent to a regression with about 2.3 million parameters.

Neural networks provide a concrete example. In the neural tangent kernel, or lazy-training, regime, a network's evolution during training is well approximated by its linearization around the initial parameters (Jacot et al., 2018; Lee et al., 2019; Chizat et al., 2019). Changes in the forecast are therefore linear combinations of gradient features, with one feature for each parameter (Kelly et al., 2026).[15] Training estimates the coefficients on these features, and early stopping provides regularization closely related to ridge regression (Ali et al., 2019). Widening a one-hidden-layer network while retaining the existing neurons and their initialization expands the span of these features.[16] This construction parallels model expansion in our economy: the posterior mean in Lemma 1 is a ridge estimator based on payoff history net of the strong component, and increasing scale adds weak signals to that regression.

The available history also constrains which nonlinear components investors can learn. Under the high-dimensional designs studied by Ghorbani et al. (2021) and Misiakiewicz (2022), kernel methods learn progressively higher-degree polynomial components as the sample grows. A history of order D permits learning linear components, whereas quadratic components become learnable at order D^2. Away from these transition scales, a broad class of suitably regularized kernels converges to the same low-degree projection of the target function. Limited network width can impose an additional approximation constraint. Thus, different nonlinear specifications can recover a common set of predictive components, with higher-degree components becoming accessible as the history grows.[17] These results motivate our representation of model scale as an expanding set of signals whose predictive content must be estimated from a limited history.

As in Definition 1, we define equilibrium directly in the limiting economy. Write $\mathcal{F}_q$ for the information acquired at scale q: the strong signals, the first qN weak signals, and the history of these signals and of the payoff. Larger models nest smaller ones, so $\mathcal{F}_q \subseteq \mathcal{F}_{q'}$ for $q \leq q'$. Appendix D defines these sets rigorously. An investor at scale q also learns from

[15] Publicly disclosed language models are far larger than the return-forecasting networks with up to one million parameters of Timmermann and Vulicevic (2026). Kimi K3 has 2.8 trillion total parameters, and DeepSeek-V4-Pro has 1.6 trillion (Kimi Team, 2026; DeepSeek-AI, 2026). GLM-5 has 744 billion parameters, and Mistral Large 3 has 675 billion (GLM-5 Team, 2026; Mistral AI, 2025).

[16] Under the usual inverse-square-root width normalization, widening rescales the existing gradient features by a common factor. This rescaling preserves their span but changes their effective regularization when the numerical ridge parameter is held fixed.

[17] At the critical scales $T \asymp D^\ell$, estimation of the degree-ℓ component remains sensitive to the kernel and effective regularization. These transitions can generate successive peaks in test error, depending on regularization and the signal-to-noise ratio (Misiakiewicz, 2022).

the price: the price can reflect the forecasts of investors who scaled further, so it carries information the investor did not acquire. An uninformed investor observes only the price.

Let $\lambda \in [0,1]$ be the fraction of informed investors and let F be the cumulative distribution function of scales among them. Thus $F(q)$ is the fraction of informed investors whose scale is at most q, and $\lambda\, dF(q)$ gives their population shares. The previous section is the special case in which F is a unit point mass at $q = 1$: every informed investor uses the full model. We first take (λ, F) as given.

Definition 3 (CREE with heterogeneous scales) *Fix (λ, F). For each $p \in \mathbb{R}$, let $x_q(p)$, the demand of an informed investor at scale q, be $\mathcal{F}_q$-measurable, let $x_U(p)$ be deterministic, and let P depend only on the information collectively acquired and the supply shock. Then $\{x_q\}_{0<q\le 1}, x_U$, and P form a CREE at (λ, F) if the following conditions hold.*[18]

(i) Informed optimality. For every $q \in (0,1]$,

$$x_q(p) \in \operatorname*{argmax}_{x\in\mathbb{R}} E\left[-e^{-\gamma x(d-p)} \,\middle|\, \mathcal{F}_q,\ P = p\right].$$

(ii) Uninformed optimality.

$$x_U(p) \in \operatorname*{argmax}_{x\in\mathbb{R}} E\left[-e^{-\gamma x(d-p)} \,\middle|\, P = p\right].$$

(iii) Market clearing.

$$\lambda \int_{(0,1]} x_q(P)\, dF(q) + (1-\lambda) x_U(P) = \sigma_\eta \eta.$$

The history (3) contains no past prices. With heterogeneous scales, a past price could reveal information about omitted signals. Appendix E shows that excluding past prices is exact when scaled trading begins after the last history date. The omission has a negligible effect in the limit in two cases: scaled trading occurs at only T_w history dates with $T_w/T \to 0$, or it occurs at every date but few investors take part. Both cases arguably fit the current computational boom: it began only a few years ago, and until recently very few funds traded on scaled models. Figure 7 illustrates the timeline.

[18] Appendix D.4 specifies the measurability, integrability, and conditional-belief conventions.

For a CREE at (λ, F), define the ex-ante values

$$V(q) = e^{\gamma\chi(q)} E\left[-e^{-\gamma x_q(P)(d-P)}\right], \quad 0 < q \leq 1, \qquad V_U = E\left[-e^{-\gamma x_U(P)(d-P)}\right]. \tag{20}$$

Both are evaluated before observing the data or the price, integrating over their realizations under the equilibrium price rule. The value $V(q)$ is defined at every scale, including those outside the support of F, where it is the value of a deviation.

Definition 4 (Overall equilibrium with endogenous scales) *An overall equilibrium with endogenous scales is a scale distribution (λ, F) together with a CREE at (λ, F) in which every investor chooses its scale, and whether to acquire at all, so that its ex-ante utility in (20) is maximized.*

We focus on linear equilibria, in which the price is a linear function of the informed investors' forecasts and the supply shock. We impose the following restrictions on the acquisition cost.

Assumption 2 (Acquisition costs) *The cost is $\chi(q) = \chi_0 + \chi_1(q)$ with a fixed component $\chi_0 \geq 0$ and a scale component $\chi_1 \in C^2([0,1])$ satisfying $\chi_1(0) = \chi_1'(0) = 0$, $\chi_1'(q) > 0$ for $q > 0$, and $\chi_1''(q) \geq 0$.*

Our leading example is the quadratic cost $\chi(q) = \chi_0 + \bar{\chi} q^2$, which satisfies Assumption 2 and is supported by the calibration in Appendix J.[19]

The fixed cost χ_0 represents the overhead required to operate a forecasting model at any scale, including a model based on strong signals alone, represented by the endpoint $q = 0$.[20] This overhead may include the costs of maintaining trading infrastructure and employing researchers and strategists, which we assume are independent of model scale. The variable cost $\chi_1(q)$ captures the additional expense of acquiring and processing weak signals and

[19] Convexity of χ_1 follows from ordering signals by cost. Let the weak signals carry acquisition costs $\kappa_1, \ldots, \kappa_N \geq 0$, and let investors add them cheapest first. Each additional signal then costs at least as much as the previous one. In the continuum, the signal added at scale q costs the q-quantile $F^{-1}(q)$ of the cost distribution F across signals, so $\chi_1(q) = \int_0^q F^{-1}(u)\, du$ is convex because a quantile function is nondecreasing. Assumption 2 further requires $F^{-1}(0) = 0$: the cheapest signals are free. The quadratic case is as special case with linear marginal cost, $\chi_1'(q) = 2\bar{\chi} q$, or equivalently with uniform signal costs on $[0, 2\bar{\chi}]$, with $\bar{\chi}$ the cost of the full signal universe. Appendix G states the cost in this per-signal form and gives the condition under which cheapest-first acquisition is the equilibrium outcome, and Appendix J.5 motivates the curvature from training compute.

[20] Because the number n of strong signals is fixed, a model using only these signals has zero limiting scale under our asymptotic normalization. Including all predictors in the definition of scale, $(n + N_1)/(n + N)$, yields the same limit as $N_1/N \to q$.

training and operating larger models. Uninformed investors pay neither cost component and learn only from prices.

We characterize the equilibrium scale distribution through two residual-risk curves, defined for every $q \in (0, 1]$:

$$R(q) = \mathrm{Var}[d \mid \mathcal{F}_q], \qquad R_P(q) = \mathrm{Var}[d \mid \mathcal{F}_q, P]. \tag{21}$$

The pre-price risk $R(q)$ is the payoff variance left after acquiring scale q. It depends only on the information primitives. The post-price risk $R_P(q)$ is the payoff variance left after acquiring scale q and also observing the price. It depends on the scale distribution (λ, F) through the price. We obtain R by applying the limiting-risk calculation of Lemma 2 to the first qN signals rather than all N, with parameters adjusted for the omitted signals. Appendix D.1 derives these adjustments and proves the following. Recall from Proposition 3 the weak explainable variance $A_w = \sigma_\beta^2 V_0/\kappa$. In the finite economy, write $\mathcal{F}_q^N$ for the information in the payoff history and in the current and past observations of the strong signals and the first N_1 weak signals.

Lemma 4 (Pre-price residual risk across scales) *For $q \in [0, 1]$, let $R(q)$ be the larger root in r of*

$$(1 + c\kappa)r^2 + \Big[\kappa A_w\big(1 - c - \tfrac{1-q}{\kappa+q}\big) - \sigma_\varepsilon^2(2 + c\kappa)\Big]r - \Big[\sigma_\varepsilon^2 + \kappa A_w \tfrac{1-q}{\kappa+q}\Big](\kappa A_w - \sigma_\varepsilon^2) = 0. \tag{22}$$

If $N_1/N \to q \in (0, 1]$, then, in the limit,

$$\mathrm{Var}[d \mid \mathcal{F}_q^N] \xrightarrow{p} R(q).$$

The function R is continuous and decreasing on $[0, 1]$, with $R(0) = \sigma_\varepsilon^2 + A_w$ and $R(1) = \sigma_I^2(c) > \sigma_\varepsilon^2$.

As in Kelly et al. (2024), a scale-q investor uses a misspecified model, with $N_1 < N$ signals. Using the N_1 signals optimally appears to require knowing how they covary with the omitted signals. A remarkable random-matrix cancellation removes this requirement. The observed block of an inverse-Wishart matrix is again inverse-Wishart and independent of the regression of the omitted signals on the observed ones, and the omitted risk has a deterministic limit (Lemmas D.1 and D.2). Apart from the scale of the signals, the

adjustment for misspecification therefore depends only on the three parameters A_w, σ_ε^2, and κ, and the Marchenko–Pastur equation again reduces to a quadratic.[21]

We now turn to the risk remaining after observing the price, which in equilibrium has a simple form on the support of F, pinned down by indifference across scales. Every acquired scale must deliver the same ex-ante utility; otherwise an informed investor would switch to a better one. As in Grossman and Stiglitz (1980), ex-ante utility is proportional to the standard deviation of the unforecast payoff, so, net of the acquisition cost, $V(q) \propto -e^{\gamma\chi(q)}\sqrt{R_P(q)}$, with R_P in (21). Constancy of V on the support therefore gives

$$e^{\gamma\chi(q)}\sqrt{R_P(q)} = e^{\gamma\chi(q_H)}\sqrt{R_P(q_H)}.$$

At the largest acquired scale q_H the price adds no information, which pins down the constant.

Lemma 5 (Post-price residual risk in equilibrium) *Under Assumption 2, in a linear overall equilibrium with $\lambda > 0$, let q_H be the largest scale in the support of F and $R_H = R(q_H)$. Then R and R_P in (21) satisfy*

$$R_P(q) < R(q) \quad \text{for } 0 < q < q_H, \qquad R_P(q) = R(q) \quad \text{for } q_H \le q \le 1, \tag{23}$$

and, for every q in the support of F,

$$R_P(q) = R_H\, e^{2\gamma[\chi_1(q_H)-\chi_1(q)]}. \tag{24}$$

Lower residual risk gives better gross performance.

Corollary 4 (Gross Sharpe ratio across scales) *Under Assumption 2, in a linear overall equilibrium with $\lambda > 0$, let $\pi_q = x_q(P)(d-P)$ be the trading profit at scale q and let $S_P = \mathrm{Var}[d-P]$. For every q in the support of F,*

$$\frac{E[\pi_q]}{\sqrt{\mathrm{Var}[\pi_q]}} = \sqrt{\frac{S_P - R_P(q)}{2S_P - R_P(q)}}, \tag{25}$$

and this ratio is strictly increasing in q on the support of F.

All investors learn from the same price, but a larger model still earns a higher Sharpe ratio before acquisition costs. After these costs, every occupied scale gives the same ex-ante

[21] In (22), c is the limit of N/T, not of N_1/T. The scale q enters only through $(1-q)/(\kappa+q)$. The value $R(0)$ is the risk of an investor who observes only the strong signals.

utility. The calibration adds a positive mean supply; the proof of Corollary 4 in Appendix F.1 gives a sufficient condition for the ratio to increase with scale in that case.

The shape of the post-price risk R_P in (21) determines the scale distribution F. How fast $R_P(q)$ falls as q rises measures how much the price reveals about models larger than q, and hence how many investors run them. If many investors scale beyond q and the price reveals their forecasts almost perfectly, acquiring scale q adds little and R_P is nearly flat there; if few do, the price reveals little and R_P falls steeply. Market clearing makes this precise and recovers F from R_P; the precise inversion takes some algebra and is given in Appendix F. The next lemma recovers a distribution with a continuous density on an interval and a possible atom at its upper endpoint. Theorem 5 below establishes when this form arises.

Lemma 6 (Recovering the scale distribution from residual risk) *Consider a linear overall equilibrium with $\lambda > 0$ whose scale distribution F has support $[q_L, q_H]$, with $0 < q_L < q_H \leq 1$, a continuous density on (q_L, q_H), and no atom except possibly at q_H. For $q_L \leq q < q_H$, the scale CDF is*

$$F(q) = 1 - \frac{\gamma\sigma_\eta}{\lambda} \frac{R_P(q)\sqrt{\frac{R'(q)}{R'_P(q)}} - R(q)}{\sqrt{R(q) - R_P(q)}} \exp\left\{ -\frac{1}{2} \int_q^{q_H} \frac{\left[\sqrt{-R'(s)} - \sqrt{-R'_P(s)}\right]^2}{R(s) - R_P(s)}\, ds \right\}. \quad (26)$$

Outside this interval, $F(q) = 0$ for $q < q_L$ and $F(q) = 1$ for $q \geq q_H$. The jump at the top is

$$1 - F(q_H-) = \frac{\gamma\sigma_\eta}{\lambda} \frac{R(q_H)}{-R'(q_H)} \sqrt{\frac{R''(q_H) - R''_P(q_H)}{2}}, \quad (27)$$

where $R''_P(q_H)$ is taken on the occupied interval. Normalization $F(q_L) = 0$ determines λ.

The preceding lemmas determine a candidate scale distribution once its endpoints are known. For this candidate to be an equilibrium, the recovered CDF must be nondecreasing, its top mass must be well defined, and informed entry must be optimal. In particular, a positive top mass requires $R''(q_H) - R''_P(q_H) > 0$ in (27); a nondegenerate interval also requires positive density and enough informed investors to populate it. The theorem gives conditions under which equilibrium has this form or reduces to a single acquired scale or no acquisition.

Theorem 5 (Equilibrium with endogenous scales) *Under Assumption 2, a linear overall equilibrium exists and is essentially unique.*[22] *Suppose additionally that* $\chi_1 \in C^3([0,1])$, $\chi_1''(q) > 0$, *and* $\chi_1'''(q) < 2\sqrt{2\gamma}\,[\chi_1''(q)]^{3/2}$ *for* $0 < q < 1$.[23] *Assume also that* $\sigma_\varepsilon^2 \le \sigma_\beta^2 V_0$ *and* $\chi_1'(1) > -R'(1)/(2\gamma R(1))$. *Then the largest acquired scale, whenever entry occurs, is the unique* $q_H \in (0,1)$ *satisfying*

$$-\frac{R'(q_H)}{2\gamma R(q_H)} = \chi_1'(q_H).$$

Writing $R_H = R(q_H)$, *no investor acquires information if and only if*

$$\chi_0 \ge \chi_0^{\text{entry}} := \frac{1}{2\gamma}\log\frac{\sigma_d^2}{R_H} - \chi_1(q_H). \tag{28}$$

Below this threshold, either F *is a unit point mass at* q_H, *or its support is* $[q_L, q_H]$ *for a unique* $0 < q_L < q_H$, *with a continuous, strictly positive density on* (q_L, q_H) *and a point mass at* q_H. *In the latter case,* R_P *in* (21) *and* F *are given by Lemmas 5 and 6, and the absolute population mass at the top is*

$$\lambda_H := \lambda[1 - F(q_H-)] = \gamma\sigma_\eta \frac{R_H}{-R'(q_H)}\sqrt{\frac{R''(q_H) - R_P''(q_H)}{2}} > 0.$$

The pair (q_L, λ) *is uniquely determined by* $F(q_L) = 0$, *the population bound* $0 < \lambda \le 1$, *and the ex-ante participation conditions* $V(q_H) \ge V_U$ *and* $(1-\lambda)\big[V(q_H) - V_U\big] = 0$.

The participation conditions require ex-ante indifference between acquiring information and remaining uninformed when $0 < \lambda < 1$, and a weak preference for acquisition when $\lambda = 1$.

Quadratic variable cost, $\chi(q) = \chi_0 + \bar{\chi}q^2$ with $\bar{\chi} > 0$, is our leading example going forward. The largest model equates the marginal reduction in pre-price residual risk, measured in certainty-equivalent units, to marginal acquisition cost.[24] Its owner already knows every forecast reflected in the price; investors with smaller models combine cheaper subscriptions

[22] The informed fraction λ and population measure $\lambda\,dF$ are unique, and the price and optimal demands are unique up to almost-sure equivalence.

[23] These curvature inequalities hold for quadratic costs because $\chi_1'''(q) = 0$.

[24] Kadan and Manela (2019); Farboodi et al. (2022) link the value of information to sufficient statistics estimated from financial data. Our equilibrium delivers a natural sufficient statistic for relative information values: acquisition cost. Endogenous model-scale choice implies that differences in computing expenditure across scales used in equilibrium equal differences in the certainty-equivalent value of information beyond what prices reveal.

with information from the price. Appendix F proves the results and gives the scalar test separating the two positive-entry cases, the endpoint equation for (q_L, λ), and the equilibrium price and demands.

Flexible information acquisition. The equilibrium in Theorem 5 remains an equilibrium when investors can acquire arbitrary subsets of signals in the limiting Gaussian economy. Consider a trader contemplating a bundle that skips some cheaper signals and acquires more expensive ones. Evaluate this bundle by adding its selected signals in increasing cost order. For each selected signal within the occupied interval, compare the trader with an equilibrium investor whose bundle ends at that signal. For that investor, who has acquired all preceding signals, the marginal benefit of this signal equals its acquisition cost. The deviating trader has less preceding information. Because signals are complements at the equilibrium price, the missing information lowers the marginal benefit of that same signal. The gaps make it harder to extract information from price and leave greater residual payoff risk. The signal's acquisition cost, however, is unchanged. Thus adding the later signal to the incomplete bundle yields a weakly negative net marginal benefit.

Appendix G handles signals above q_H and below the smallest occupied cutoff, and gives the full proof. It also shows that every linear equilibrium under flexible acquisition is nested, so no additional linear equilibria arise.

6 The computational arms race and market structure

To understand how the computational arms race may transform financial markets, we study changes in the cost of acquiring and processing information and in the amount of data available for learning. We maintain the quadratic cost $\chi(q) = \chi_0 + \bar{\chi} q^2$, whose two parameters act on the equilibrium differently, and vary them separately.[25] The figures throughout use the baseline calibration in Appendix J. Its period is one month, its history has $T_0 = 120$ months, and its signal universe has complexity $c = N/T_0 = 1000$; premia, volatilities, Sharpe ratios, and costs are reported in annual units. A model of scale q has complexity $X = cq = N_1/T$,

[25]The hardware cost per floating-point operation offers one measure of variable cost. Hobbhahn and Besiroglu (2022) study 470 GPUs released from 2006 to 2021. Floating-point operations per second per dollar double about every 2.5 years, and about every 2.1 years for GPUs used in machine learning. At these rates, the cost of an operation falls tenfold in seven to eight years. Maslej et al. (2025) report that hardware costs decline by 30% per year and that the cost of inference at the level of GPT-3.5 fell more than 280-fold from November 2022 to October 2024.

the number of weak signals it acquires per month of history, and we show scales on this axis when the occupied range spans more than a decade.

The third change we study is the length of the data history. We vary T around its calibrated value, holding the signal universe $N = cT_0$, the prior parameters $(\sigma_\beta^2, V_0, \kappa)$, and the cost $\chi(q)$ of each scale fixed, so that only the complexity ratio $c = N/T$ moves; nothing is refitted. The history is common to all investors and is not a choice. As $T \to \infty$, every model's complexity falls to zero, and the economy converges to one in which investors know the predictive coefficients of the signals they acquire.

Our first results concern the distribution of model scales: who acquires information, how large a model they choose, and how investors spread across scales.

6.1 Scale distribution

Figure 1 illustrates how falling fixed acquisition costs change the distribution of model scales. A speculative reading of this transition is as a past and a present of quantitative investment, with a possible future beyond the last panel. At fixed costs above about 2,300 basis points of fund capital per year, no investor acquires a complex model. This no-entry regime can stand for an earlier era in which the computation required by today's models was effectively infeasible. The relevant horizon is roughly four decades: Renaissance Technologies was founded in 1982.[26]

The point mass in panel (a) represents an earlier stage with high fixed costs, in which advanced forecasting is concentrated in a small group of specialized funds: 7.6% of investors at a fixed cost of 750 basis points. Once acquisition becomes worthwhile, entrants all choose the same large model, with about 5,000 weak signals. Further reductions in fixed cost initially enlarge this group without changing its model scale, until it holds 15.5% of investors at 87 basis points. Panel (b) lies just past that point: at 20 basis points the group at the top is unchanged, and the first smaller models, with 1,800 to 5,000 weak signals, hold 4.5% of investors.

Panel (c) is the baseline, calibrated to present-day survey data, and gives a stylized present. The group using the largest model remains, like the firms in Table 5, but additional investors enter with smaller models. They combine their own forecasts with information revealed by prices. The market therefore supports a wider range of forecasting operations, from about 175 weak signals to the 5,000 of the largest model. A possible future lies beyond the baseline: still lower fixed costs extend that range toward smaller scales.

[26] See Renaissance Technologies' company history.

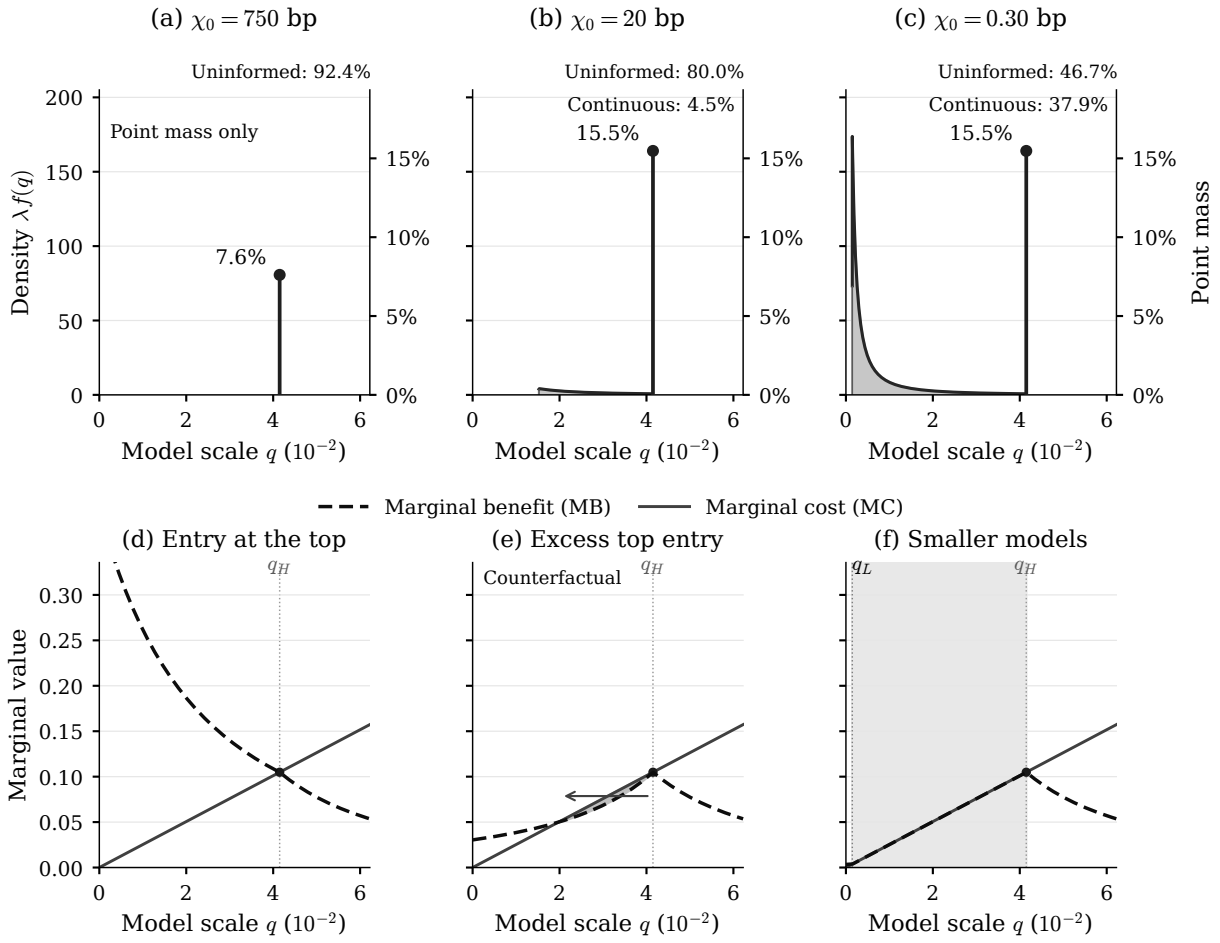

Figure 1: Scale distributions and the marginal value of model scale as fixed acquisition costs fall. Each column has one fixed cost χ_0, in annual basis points of fund capital: 750 in panels (a) and (d), 20 in panels (b) and (e), and the baseline 0.30 in panels (c) and (f); all other primitives are at the calibrated values of Appendix J, and the horizontal axis is the model scale q in units of 10^{-2}. Top row: the shaded area is the population density $\lambda f(q)$ of informed investors (left axis), the stem at q_H is the mass λ_H at the largest scale (right axis), and each panel prints the uninformed share and the share in the continuum. Bottom row: dashed curves are the marginal benefit of scale $\mathrm{MB}(q) = -R_P'(q)/(2\gamma R_P(q))$, with $R_P(q)$ in (21) the residual payoff variance after observing the price, and solid lines the marginal cost $\mathrm{MC}(q) = 2\bar{\chi}q$, both in monthly fractions of fund capital per unit of q; the axis ends at about three times $\mathrm{MB}(q_H)$. Panel (d) is an equilibrium with every informed investor at q_H; panel (e) is a counterfactual that keeps every informed investor at q_H at the cost of panel (b), and the arrow marks a profitable reduction in scale, ending where the continuum of panel (b) begins; panel (f) is the baseline, in which MB and MC coincide on the shaded support $[q_L, q_H]$.

The bottom row of Figure 1 explains the mechanics of this transition through the marginal benefit of expanding a model (MB) and its marginal research cost (MC). The same MC curve applies throughout because the fixed cost does not affect the cost of an additional signal.

In panel (d), investors enter at the largest scale q_H, where MB crosses MC from above. Reducing model scale would sacrifice more in information value than it would save in research costs, making the deviation unprofitable. As more investors enter, their trading makes the common forecast more visible in prices, lowering MB below q_H. The crossing remains at q_H: an investor at that scale already knows every forecast reflected in the price, so further price revelation leaves its marginal information benefit unchanged.

This expansion of the top group cannot continue indefinitely. Panel (e) shows what would happen if entry remained concentrated at q_H after the top mass reached its limit, at the fixed cost of panel (b). MB would fall below MC just to the left of q_H. An investor could then gain by choosing a smaller model: the saving in research costs would exceed the loss in information value. This is a counterfactual allocation, not an equilibrium.

In equilibrium, smaller scales become occupied: the thin continuum of panel (b) at the same cost, and the support of the baseline in panel (f). MB and MC coincide over the occupied interval: the price adjusts so that investors are indifferent among these models. Further reductions in fixed cost extend the interval downward, while leaving the population density at existing scales and the absolute top mass unchanged. The shaded interval in panel (f) shows the baseline support. The following proposition formalizes the result, with $\rho(q) = \lambda f(q)$ denoting the population density, where $f = F'$ on (q_L, q_H), and $\lambda_H = \lambda[1 - F(q_H-)]$ the absolute top mass.

Proposition 6 (Fixed acquisition costs and entry) *Suppose $\chi(q) = \chi_0 + \bar{\chi} q^2$, with $\bar{\chi} > -R'(1)/(4\gamma R(1))$, and $\sigma_\varepsilon^2 \leq \sigma_\beta^2 V_0$. Holding all other primitives fixed, a fall in χ_0 weakly raises participation λ and leaves the largest acquired scale q_H unchanged whenever acquisition occurs. If $q_L < q_H$ and $\lambda < 1$, it lowers q_L and raises λ, while leaving the density $\rho(q)$ at existing scales and the top mass λ_H unchanged. Once $\lambda = 1$, further reductions have no effect on the scale distribution.*

Smaller scales need not become occupied. Under the assumptions of the proposition, they are occupied at some positive fixed cost if and only if $\Lambda_H < \lim_{\chi_0 \downarrow 0} \lambda_{\text{one}}$. Here Λ_H is the top-mass threshold in (64), and λ_{one} is the informed mass of the single-scale candidate in (72), capped at one. Otherwise, all informed investors use q_H at every fixed cost below the entry threshold. The baseline satisfies the condition, with $\Lambda_H = 0.155$ and a limit of

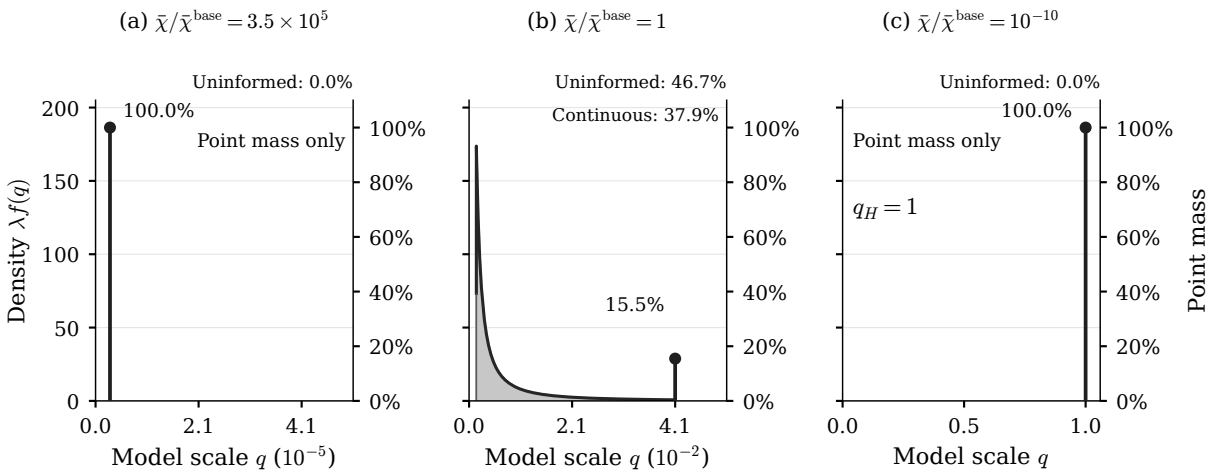

Figure 2: Scale distributions as variable acquisition costs fall. Panels differ only in the variable-cost coefficient $\bar{\chi}$, given in the titles as a multiple of its baseline value; all other primitives are at the calibrated values of Appendix J. Shading, stems, and the printed shares are as in the top row of Figure 1. The horizontal axis is the model scale q on a linear scale whose unit is 10^{-5} in panel (a), 10^{-2} in panel (b), and one in panel (c), so panel (a) is laid out like panel (b) and its atom sits at the left edge. In panel (a), 30% of investors hold a single vanishing model and the rest stay uninformed; in panel (c), every investor holds the full model, $q_H = 1$.

0.182. Smaller scales enter once the fixed cost falls below 87 basis points, where the top group reaches Λ_H.

A fall in the variable-cost coefficient $\bar{\chi}$ has different effects from a fall in the fixed cost χ_0. Lower fixed costs draw investors into the market without changing the largest model; lower variable costs make larger models worthwhile. The investor at the top already knows every forecast reflected in prices, so its marginal information benefit is determined by private signals alone. Reducing $\bar{\chi}$ lowers marginal research cost and shifts the optimal scale upward. Figure 2 illustrates this expansion: at 5.2×10^5 times the baseline coefficient, in panel (a), every informed investor holds the same model of scale $q_H = 2.0 \times 10^{-6}$, a vanishing fraction of the signal universe; at the baseline, in panel (b), the largest model has about 5,000 weak signals; and in panel (c) it reaches $q_H = 1$, encompassing all 120,000 weak signals in the calibrated universe. Recall that $q = 0$ corresponds to a model based on strong signals alone; the near-zero scale in panel (a) effectively corresponds to this case.

Participation and concentration, however, need not change monotonically. Across panels (a)–(c), the informed share goes from 30% to one half and then to one. Two forces govern this response: cheaper computation encourages acquisition, while larger models reveal more

information through prices and make it more attractive to rely on others' forecasts. Along the whole range of $\bar{\chi}$ in Figure 4, the informed share falls from one to one half as the coefficient falls to about ten times its baseline value, and returns to one as it falls further. The largest model therefore expands even when the informed population shrinks. Panel (a) lies past the range drawn there: above about 1.7×10^5 times the baseline coefficient the continuum disappears and every informed investor holds the same model, above about 4.6×10^5 times participation falls below one, and above 5.3×10^5 times no one acquires. In this last band the price reveals almost nothing, so the fixed cost of 0.30 basis points alone decides entry. The value of the vanishing model is about twice its variable cost, so the surplus left to cover the fixed cost shrinks with the model and reaches 0.30 basis points at the entry bound. In panel (c), the full model costs only the fixed cost, and the price cannot reveal the last part of what it forecasts, so every investor acquires.

Panel (c) suggests a possible future of computational "abundance." When both fixed and variable acquisition costs are sufficiently low, every investor acquires the full signal set, $q = 1$. Everyone then learns everything that the available signals can reveal. In the calibrated economy, with the fixed cost at its baseline value, this outcome arises when $\bar{\chi}$ is at most $-R'(1)/(4\gamma R(1))$, about 1.1×10^{-4} times its baseline value; above this bound, $q_H < 1$. Full acquisition still leaves residual payoff risk $R(1) > 0$, because parameter uncertainty persists in the proportional limit. The following proposition formalizes the increase in model scale and the eventual acquisition of all signals.

Proposition 7 (Variable acquisition costs and model scale) *Suppose* $\chi(q) = \chi_0 + \bar{\chi} q^2$, *with* $\bar{\chi} > 0$. *Holding all other primitives fixed, a fall in* $\bar{\chi}$ *weakly raises the largest acquired scale* q_H *whenever acquisition occurs. For sufficiently small positive* $\bar{\chi}$, $q_H = 1$ *whenever acquisition occurs. If* $\chi_0 < \log[\sigma_d^2/R(1)]/(2\gamma)$, *then, for all sufficiently small positive* $\bar{\chi}$, *acquisition occurs and every informed investor chooses* $q = 1$. *If both cost parameters are sufficiently small, every investor acquires the full model.*

6.2 Residual uncertainty

Cheaper computation and longer histories make more information accessible, but how much payoff uncertainty do they remove? Figure 3 reports residual volatility, $\sqrt{R_P(q)}$ with R_P in (21), for the largest and smallest acquired models along the three paths: the fixed cost, the variable-cost coefficient, and the length of the data history. It also shows an oracle who observes the same signals as the largest model and knows the true coefficients and

covariance matrix. Evaluated at the equilibrium scale q_H, its residual variance is $R^{\text{or}}(q_H) = \sigma_\beta^2 V_0(1-q_H)/(\kappa+q_H)$, by Lemma D.2 and the calibration's restriction $\sigma_\varepsilon^2 = 0$. Its residual risk is the payoff variance of the signals it does not observe, and it depends on costs or on the history only through the equilibrium scale q_H. The oracle still faces uncertainty from omitted signals when $q_H < 1$, but its residual risk vanishes at $q_H = 1$.[27] The largest model already observes all information entering the price, so its residual variance is $R_P(q_H) = R(q_H)$; the gap from the oracle reflects uncertainty about how to use the acquired signals. At the baseline, residual volatility is 16.3% for both the largest and the smallest model, against 7.3% for the oracle and a payoff volatility of 16.9%. Of the payoff variance, the largest model explains 7.5%, the signals it does not acquire account for 18.6%, and uncertainty about the coefficients of the signals it does acquire accounts for the remaining 73.9%.

Lower fixed costs induce widespread entry, but leave payoff uncertainty unresolved. The largest model retains the same scale, leaving both its residual risk and the oracle benchmark unchanged wherever acquisition occurs. Because information sets are nested, entrants using smaller models do not expand the information collectively acquired. Only the smallest model's residual risk moves in panel (a), and by less than a tenth of a percentage point: it equals the largest model's at the single-scale boundary of 87 basis points and rises as smaller models enter. At sufficiently low fixed costs, every investor acquires information, yet residual volatility remains above 16%, even after learning from prices.

Lower variable costs allow investors to use larger models, reducing residual uncertainty more substantially. At sufficiently low costs, all investors acquire the full signal set and face the same residual risk. Yet panel (b) shows that their common residual volatility, $\sqrt{R(1)}$, remains 16.1%, while the oracle's falls to zero. The gap reflects persistent uncertainty about the coefficients that determine how signals predict the payoff. At the other end of the path, at 10^4 times the baseline coefficient, the largest model has only twelve weak signals, and the funds' residual volatility of 16.9% is close to the oracle's 16.8%: almost all uncertainty is then about signals that no one acquires.

A longer history is the one change that reduces uncertainty about the coefficients themselves. In panel (c), residual volatility falls from 16.9% with one year of data to between 12.4% and 12.7% with a hundred years, while the oracle's falls from 10.4% to 5.0% as the

[27] Uncertainty about β can be quantitatively important. Chen et al. (2025) report a 95% lower confidence bound of nearly 20% on population R^2 for monthly U.S. market returns, compared with only 1–2% out of sample for their forecasting benchmarks. Population R^2 measures the predictability attainable when the true predictive relationship is known, corresponding to knowing β given the observed signals in our setting. Even models with few parameters can face severe learning constraints; see the analysis of "dark matter" in asset-pricing models by Chen et al. (2024).

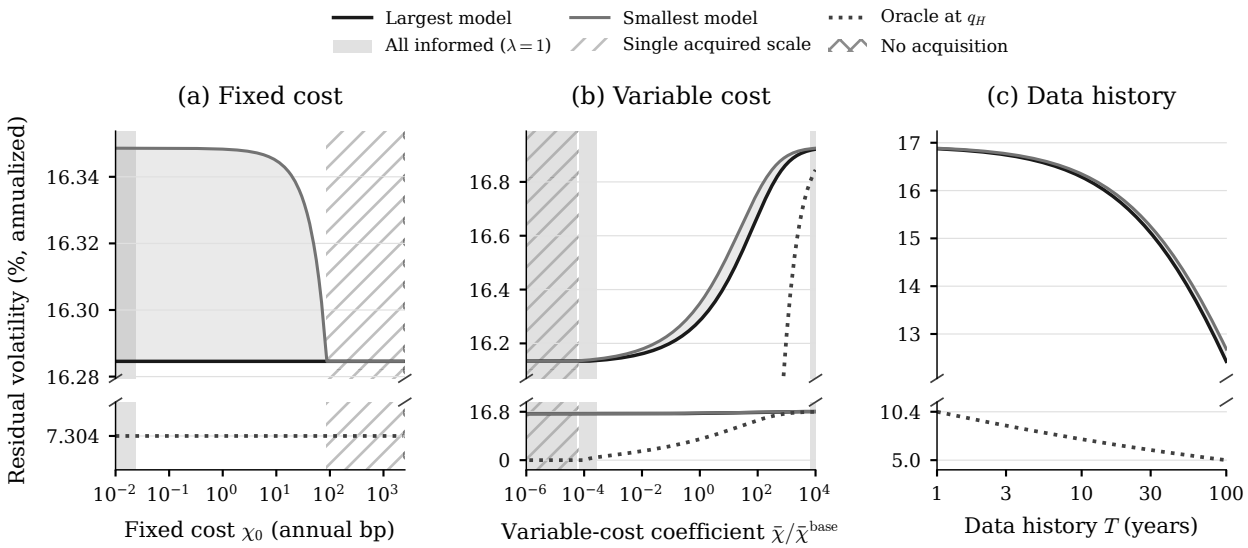

Figure 3: Residual volatility along the three paths. Each panel plots $\sqrt{R_P(q)}$, with R_P in (21), in annualized percent, for the largest (black) and smallest (grey) acquired models and for an oracle that observes the largest model's signals with known coefficients (dotted). The horizontal axes, on logarithmic scales, are the fixed cost χ_0 in annual basis points of fund capital (a), the variable-cost coefficient $\bar{\chi}$ as a multiple of its baseline value (b), and the length of the data history T in years (c); all other primitives are at the calibrated values of Appendix J, and along the history path only $c = N/T$ changes. Grey shading marks the costs at which every investor is informed, diagonal hatching a single acquired scale, and cross-hatching no acquisition.

largest model grows from about 1,900 to 11,400 weak signals. The oracle's residual falls only because the largest model grows; the oracle itself has nothing to learn. More data reduce the uncertainty about every coefficient, and the gap closes only as $T \to \infty$. Cheaper compute expands access to data without lengthening the history available to learn from it, leaving model uncertainty as a key source of priced risk.

7 Fund performance and compute expenditure

The preceding section establishes how acquisition costs affect entry and model scale across parameter values satisfying the propositions' assumptions. The effects on participation, expenditure, fund performance, and prices depend more closely on the other primitives. To discipline these comparisons, we use the calibration in Appendix J throughout. The patterns in this section and the quantitative comparisons that follow are therefore properties

of this calibrated economy, rather than general comparative-static results. We interpret χ_0 primarily as the cost of setting up compute infrastructure and $\bar{\chi}q^2$ as the expense of training and operating a model at scale q; we refer to their sum as *compute expenditure*. Figures 4 and 5 collect the results along the three paths of Section 6, the first for market structure and expenditure, the second for fund performance and prices.

7.1 Participation and model scale

The first two rows of Figure 4 follow the informed share λ, the top mass λ_H, and the scales q_H and q_L of the largest and smallest acquired models along the three paths. They put the snapshots of Section 6.1 on a common axis.

Lower fixed costs change who participates, not what the largest model is. Along the fixed-cost path, q_H and λ_H do not move until the continuum of smaller models disappears at 87 basis points, as Proposition 6 states. Participation rises from zero at the entry threshold of about 2,300 basis points to 6% at 1,000, 53% at the baseline, and one below 0.02 basis points. The smallest model shrinks from the largest model's 5,000 signals at the single-scale boundary to 175 at the baseline and 49 at 0.01 basis points.

Lower variable costs change the largest model, and participation responds in both directions. The largest model grows from 12 signals at 10^4 times the baseline coefficient to 5,000 at the baseline and to the full universe below 1.1×10^{-4} times. Participation falls from one to one half at about ten times the baseline coefficient and returns to one under the full model; the top mass falls from 42% to 11% at about 60 times and then rises to one. Two forces are at work: cheaper computation encourages acquisition, while larger models reveal more information through prices and make it more attractive to rely on others' forecasts. Above about ten times the baseline coefficient the second force dominates, so the largest model expands while the informed population shrinks; below it, the first force dominates.

A longer history has the same effect on participation as cheaper variable cost. From one to a hundred years of data, the largest model grows from about 1,900 to 11,400 signals and the informed share falls from 89% to 27%, while the smallest model stays near 175 signals. A more precise top forecast in the price makes free-riding more attractive.

7.2 Compute expenditure

The other side of this response is the compute bill. The third row of Figure 4 shows the annual expenditure of the largest and smallest acquiring funds, in basis points of their capital. Because F is conditional on acquisition, aggregate expenditure as a fraction of total fund

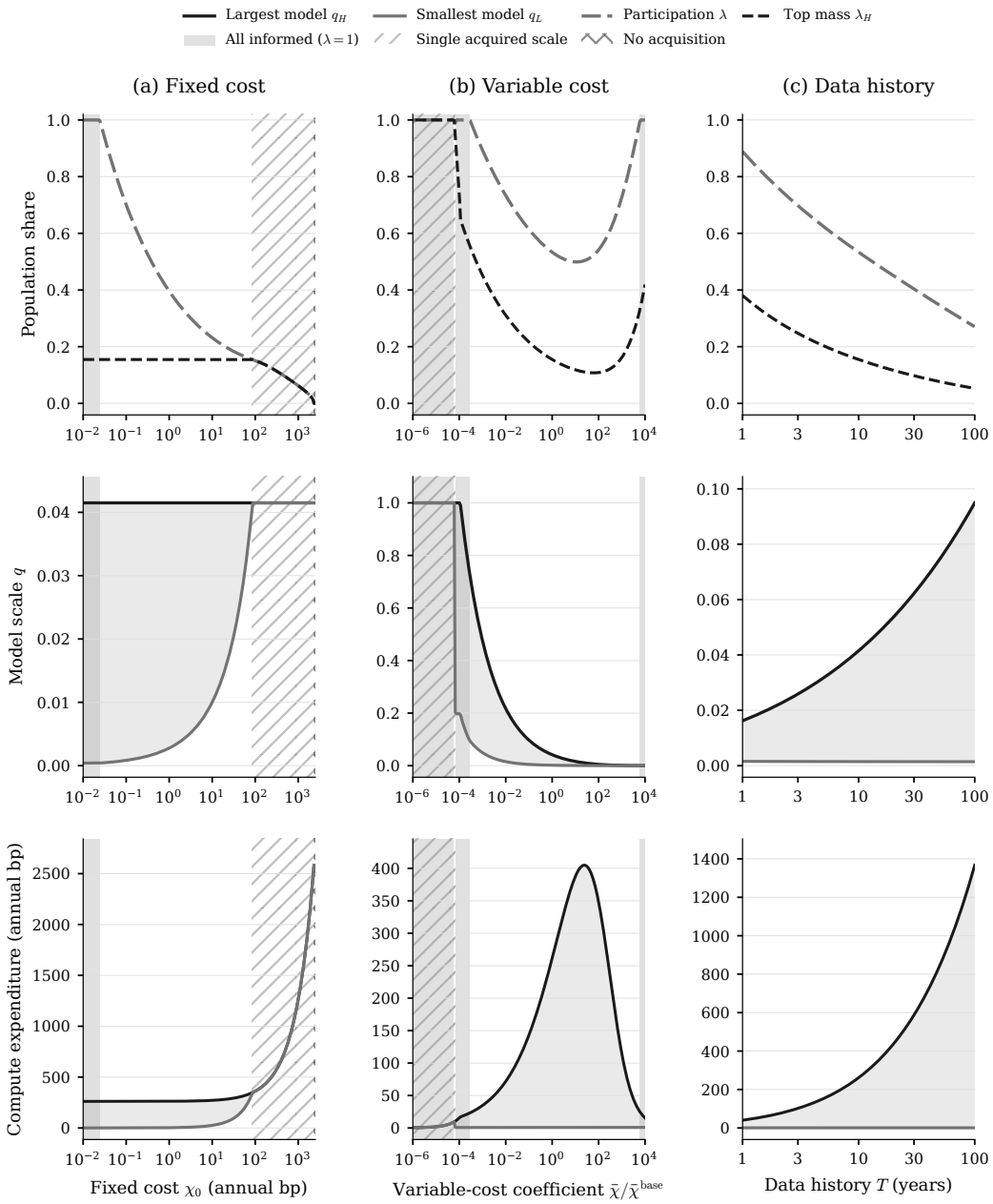

Figure 4: Participation, model scale, and compute expenditure along the three paths. Columns and shading are as in Figure 3. Rows plot participation λ and the top mass λ_H; the scales q_H and q_L of the largest and smallest acquired models, with the occupied support shaded; and the annual compute expenditure of the largest and smallest acquiring funds in basis points of their capital. At the baseline, $\lambda = 53\%$, $\lambda_H = 15\%$, $q_H = 0.041$, $q_L = 0.0015$, and the two funds spend 261 and 0.6 basis points.

capital is

$$\lambda \int_{(0,1]} \chi(q)\, dF(q) = \lambda\chi_0 + \lambda\bar{\chi} \int_{(0,1]} q^2\, dF(q),$$

which we report in annual basis points of total capital.

Lower fixed cost makes a given model cheaper for every fund, which is intuitive. Aggregate spending can nevertheless rise: just below the no-entry threshold, reducing fixed cost brings enough funds into the market to outweigh the saving per fund, and aggregate spending rises from zero at 2,300 basis points to 79 basis points at about 1,000. This is a spending analogue of the rebound mechanism associated with Jevons' paradox (Jevons, 1865, Chapter VII). The effect is not monotone over the whole range. Below 1,000 basis points the saving per fund dominates: the largest fund's bill, which is mostly fixed cost at the top of the range, falls from 1,270 to 261 basis points, aggregate spending falls to 47 basis points at the baseline, and both stay there as participation rises to one. The smallest fund's bill falls from 0.6 to 0.03 basis points as its model shrinks.

A fall in variable cost instead induces a strong expansion of compute spending at the top. Reducing $\bar{\chi}$ tenfold, from 250 to 25 times its baseline value, raises the largest model's budget from 257 to 405 basis points as its model grows from about 310 to 1,240 signals, while the smallest model's budget stays at 0.6 basis points. Aggregate spending, obtained by integrating fund expenditure over the equilibrium population as above, rises from 38 to 54 basis points even though participation falls from 60% to 50%. Here the rebound comes from more spending by the large-model funds rather than from entry. The expansion has a limit: the largest model's budget peaks at 25 times the baseline coefficient, and below that the cost saving dominates. From the baseline, a further tenfold reduction lowers the largest model's budget from 261 to 143 basis points and aggregate spending from 47 to 35 basis points, and under the full model every fund pays little more than the fixed cost.

A longer history raises the compute bill without any change in costs. From one to a hundred years of data, the largest model's expenditure rises from 40 to 1,370 basis points as its model grows, the smallest fund's stays at 0.6 basis points, and aggregate spending rises from 18 to 83 basis points despite falling participation.

7.3 Fund performance

The first row of Figure 5 reports annual gross Sharpe ratios of trading profits, before compute expenditure and fees. Each ratio divides unconditional expected trading profits by their

standard deviation. The upper and lower curves follow the largest and smallest acquired models in each equilibrium, and the band between them spans all acquiring funds.

A fall in fixed cost has a competitive effect, but only while entrants trade on the largest model's information. Entry makes investors' information more visible in prices and erodes the trading opportunities available to each fund. From the entry threshold down to the single-scale boundary at 87 basis points, every entrant uses the largest model, and its gross Sharpe ratio falls from 0.96 to 0.46 even though its scale stays unchanged. Below the boundary, entrants use smaller models whose signals the largest model already holds, so their trades reveal little that the price does not already contain: the largest model's ratio falls to 0.42 at the baseline and stays there as participation rises to one. At the bottom of the scale distribution, the move toward smaller models reduces performance: the smallest fund's ratio falls from 0.46 at the boundary to the uninformed fund's 0.30.

Lower variable costs first widen and then close the performance gap. From 10^4 to about 25 times the baseline coefficient, funds with the largest models expand them and improve their gross performance, from 0.32 to 0.48, while funds using the smallest models stay near 0.30. The signals in a small model are also observed by every investor using a larger one, so this information is widely represented in trading and increasingly revealed by prices. The performance gap therefore widens at first. As costs fall further, price revelation erodes even the largest models' informational edge: their ratio falls to 0.42 at the baseline, and the gap narrows. The gap disappears in the full-model limit, when all investors acquire the same information at $q = 1$ and every fund earns 0.30.

A longer history widens the gap without any change in costs. From one to a hundred years of data, the largest model's ratio rises from 0.33 to 0.71 as its model grows, while the smallest model's falls from 0.31 to 0.23 with the premium. The smallest model's signals are a subset of the largest model's, and the price reveals them.

Figure 6 shows the cross-section behind these ratios at the baseline. Panel (a) plots the gross Sharpe ratio of a fund that acquires the first qN signals and trades against the equilibrium price, which it is too small to move, as a function of its scale. The ratio is flat at the uninformed fund's 0.30 up to about $q = 10^{-2}$, rises to 0.42 across the occupied support, and reaches 0.62 at the full model. The virtue of complexity of Kelly et al. (2024) survives in equilibrium: a larger model earns a higher gross Sharpe ratio, as Corollary 4 states, and the ratio keeps rising beyond the largest model anyone uses. Panel (b) gives the population distribution: 47% of investors are uninformed at 0.30, 38% spread over the occupied support,

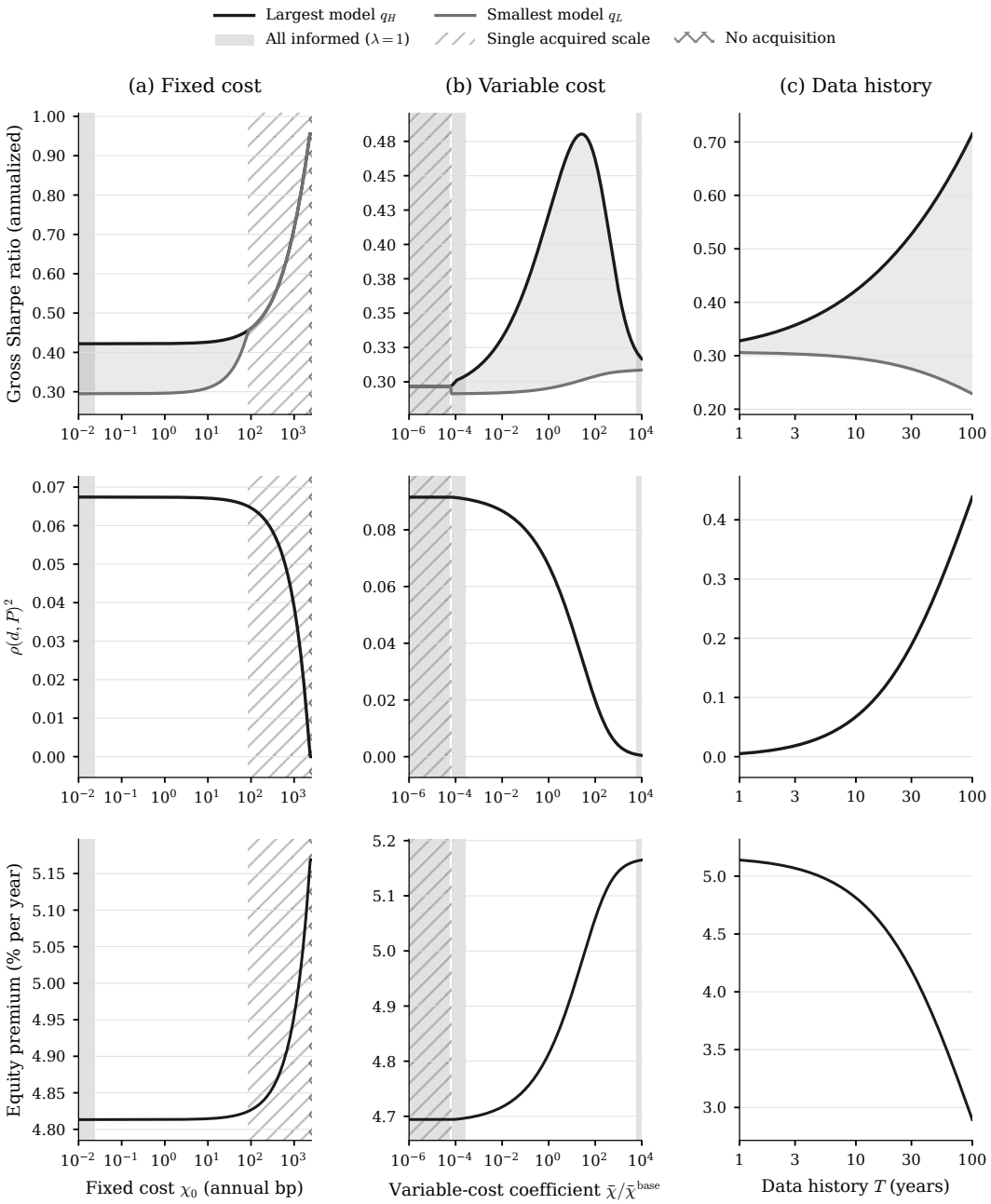

Figure 5: Fund performance, price informativeness, and the equity premium along the three paths. Columns and shading are as in Figure 4. The first row plots the annualized gross Sharpe ratios of the largest (black) and smallest (grey) acquired models, with the range across acquiring funds shaded. The second row plots price informativeness $\rho(d, P)^2 = \text{Var}(E[d \mid P])/\sigma_d^2$, the share of payoff variance that the price explains. The third row plots the equity premium $E[d - P]$ in percent per year. At the baseline the ratios are 0.42 and 0.30, the price explains 6.7% of payoff variance, and the premium is 4.81%.

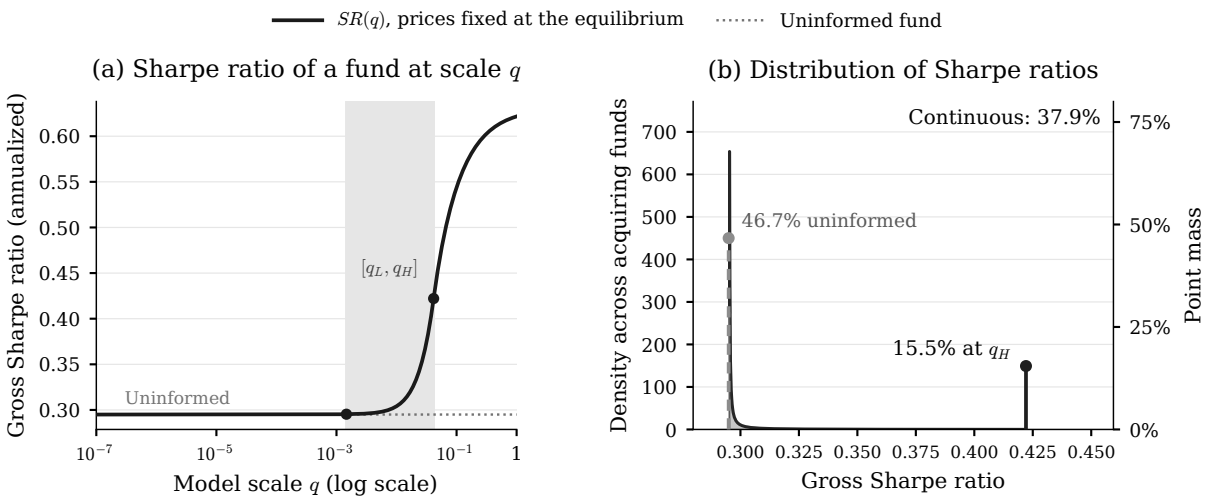

Figure 6: The Sharpe ratio of a fund at each model scale. Panel (a) plots the annualized gross Sharpe ratio $\mathrm{SR}(q)$ of a fund that acquires the first qN signals and trades against the equilibrium price, which it is too small to move, on a logarithmic scale axis; the shaded band marks the occupied scales $[q_L, q_H]$, with dots at its ends, and the dotted line is the uninformed fund. Panel (b) plots the population density of Sharpe ratios across informed investors with scales in (q_L, q_H) (left axis) and, as stems (right axis), the point masses λ_H at the largest scale and $1-\lambda$ of uninformed investors. All primitives are at the calibrated values of Appendix J.

and 15% hold the largest model at 0.42. After compute costs, every occupied scale gives the same ex-ante utility.

Residual volatility and gross performance move by different orders of magnitude. At the baseline, the largest and smallest acquired models leave residual volatilities of 16.28% and 16.35% (Figure 3), a difference too small to detect from returns, while their gross Sharpe ratios in Figure 5 are 0.42 and 0.30. The reason is that the Sharpe ratio is highly elastic with respect to residual variance. A fund's squared Sharpe ratio is, up to a bounded factor, the variance it forecasts plus the squared premium, relative to its residual variance: $[S_P - R_P(q) + \delta^2]/R_P(q)$, with $\delta = E[d-P]$ the premium and $S_P = \mathrm{Var}[d-P]$ the variance of the excess payoff. Per period, both terms in this ratio are of the order of the squared monthly Sharpe ratio of the excess payoff, $s = \delta^2/S_P = 0.0072$, so a change in $R_P(q)$ that is negligible relative to $R_P(q)$ is large relative to the ratio. Proposition H.1 in Appendix H.4 gives exact bounds; the following corollary states their first-order form for the calibrated economy, in which per-period Sharpe ratios are small.

Corollary 8 (Sharpe ratios and residual risk in the calibrated economy) *In the calibrated economy of Appendix J, let $\delta = E[d-P]$ be the premium and $S_P = \mathrm{Var}[d-P]$ the variance of the excess payoff, so that $s = \delta^2/S_P$ is its squared per-period Sharpe ratio. For a fund at scale q, let $\mathrm{SR}(q) = E[\pi_q]/\sqrt{\mathrm{Var}[\pi_q]}$ be the gross Sharpe ratio of its trading profit $\pi_q = x_q(P)(d-P)$, as in Corollary 4, and let*

$$\Delta = \frac{R_P(q_L)}{R_P(q_H)} - 1 = e^{2\gamma[\chi_1(q_H)-\chi_1(q_L)]} - 1$$

be the proportional residual gap between the smallest and the largest acquired model. Then, for every scale q on the support, with S_P and δ held fixed in the derivative,

$$1 + \mathrm{SR}(q)^2 \approx \frac{S_P + \delta^2}{R_P(q)}, \qquad -\frac{\partial \log \mathrm{SR}(q)}{\partial \log \sqrt{R_P(q)}} \approx \frac{1 + \mathrm{SR}(q)^2}{\mathrm{SR}(q)^2},$$

and the residual-volatility gap and the Sharpe gap between the two ends of the support are

$$\log \frac{\sqrt{R_P(q_L)}}{\sqrt{R_P(q_H)}} = \frac{1}{2}\log(1+\Delta), \qquad \log \frac{\mathrm{SR}(q_H)}{\mathrm{SR}(q_L)} \approx \frac{1}{2}\log\Big(1 + \frac{\Delta}{s}\Big).$$

Each $\approx$ is an equality up to a relative error bounded in Appendix H.4; at the baseline the three errors are at most 0.03%, 7%, and 3%.

Sharpe ratios change substantially with small improvements in forecasting ability. The two gaps in the corollary have the same form, $\frac{1}{2}\log(1+x)$, with $x = \Delta$ for residual volatility and $x = \Delta/s$ for the Sharpe ratio: the same residual gap enters next to one in the first and next to s in the second. By (24), $\Delta \approx 2\gamma\chi_1(q_H)$ is twice risk aversion times the top fund's variable spend per period, 0.0079 at the baseline calibration, while $s = 0.0072$. The residual-volatility gap is therefore 0.4%, while the Sharpe gap is $\frac{1}{2}\log 2.09 = 0.37$, a ratio of 1.45 to first order and 1.43 exactly. Locally, the elasticity in the corollary says that one percent less residual volatility is worth $(1+\mathrm{SR}^2)/\mathrm{SR}^2$ percent more Sharpe ratio, about $1/s$ for the smallest acquired model; the exact value, $1/u(q) - 2\kappa(q)$ by part (ii) of Proposition H.1, lies between 64 and 135 across the occupied scales of Figure 6.

The calibrated edges are of the size quantitative funds report. Zuckerman (2019) reports Robert Mercer of Renaissance Technologies on the firm's success rate: "We're right 50.75 percent of the time ... but we're 100 percent right 50.75 percent of the time." A hit rate of 50.75% corresponds to a forecastable share of 0.06% of payoff variance, and the calibrated

top fund's share of 0.79% to a hit rate of 52.8%.[28] Edges of this size leave residual volatility almost untouched, while Sharpe ratios still separate funds by an order of magnitude more.

7.4 Price informativeness and the equity premium

Cheaper compute makes more information accessible, but how much of it reaches prices, and how much aggregate risk does it remove? The second row of Figure 5 reports price informativeness, the share of payoff variance that the price explains to an investor who observes nothing else,

$$\rho(d,P)^2 = \frac{\mathrm{Var}(E[d \mid P])}{\sigma_d^2} = 1 - \frac{R_U}{\sigma_d^2}, \qquad R_U = \mathrm{Var}[d \mid P].$$

The third row reports the equity premium. As in the calibration, mean asset supply is one, so

$$E[d-P] = \frac{\gamma}{\lambda \displaystyle\int_{(0,1]} R_P(q)^{-1}\, dF(q) + (1-\lambda) R_U^{-1}},$$

with the residual risks $R_P(q)$ in (21): the premium is γ times the harmonic mean of residual risk across investors, so it falls when any group of investors faces less of it.

Lower fixed costs induce widespread entry but add little to prices. At the baseline, the price explains 6.7% of payoff variance, 90% of the 7.5% that the largest model explains. Because information sets are nested, entrants using smaller models do not expand the information collectively acquired, and price informativeness stays at 6.7% as participation rises to one. It falls only above the baseline, to 3.9% at 1,000 basis points and to zero at the entry threshold, where few funds acquire information and the price reveals less of what they know. The equity premium follows: 5.17% without acquisition, 4.96% at 1,000 basis points, and 4.81% from the baseline down to full participation, far above zero throughout.

Lower variable costs make prices more informative because larger models bring more information to the market. Price informativeness rises from below 0.1% at 10^4 times the

[28] A fund at scale q forecasts the share $\sigma(q) = [S_P - R_P(q)]/S_P$ of the variance of the excess payoff, and the correlation between its forecast and the excess payoff is $\rho = \sqrt{\sigma(q)}$. For a centered bivariate normal pair with correlation ρ, Sheppard's orthant formula gives $P(\text{forecast and excess payoff have the same sign}) = \frac{1}{2} + \arcsin(\rho)/\pi$, which we take as the hit rate. A hit rate of 50.75% thus gives $\rho = \sin(0.0075\pi) = 0.024$ and $\sigma = \rho^2 = 0.06\%$; the top fund's $\sigma = 0.79\%$ gives $\rho = 0.089$ and a hit rate of 52.8%. The conversion ignores the premium, which raises sign agreement for every fund, and treats a per-trade edge as a per-period edge on one asset.

baseline coefficient to 6.7% at the baseline and 9.2% under the full model, where the price reveals nearly everything the full signal set can explain. The equity premium falls from 5.17% to 4.69% along the same path. Under the full model it equals $\gamma R(1)$, so it remains substantial even when all investors are informed: the residual risk from unknown coefficients is priced.

A longer history makes prices far more informative than cheaper compute does. From one to a hundred years of data, price informativeness rises from 0.5% to 44% and the equity premium falls from 5.14% to 2.89%. With free entry, every acquiring fund's certainty equivalent stays at the uninformed fund's, so the gain from more data goes into the price rather than to the funds.

8 Conclusion

We propose the Complex Market Hypothesis and formalize it through Complex Rational Expectations Equilibria, a framework in which investors learn a high-dimensional predictive model and choose how much information to acquire. Random matrix theory makes Bayesian learning, inference from prices, and market clearing tractable while preserving uncertainty about predictive coefficients and signal covariances. The framework determines model scale, information acquisition, and prices jointly.

Competition can sustain different model scales among otherwise identical investors. Investors using smaller models save on compute by combining their forecasts with information larger models reveal through prices. Equilibrium can therefore combine a group using the largest model with a continuum of smaller models. Falling fixed costs broaden participation without changing the largest scale. Lower variable costs encourage larger models, but participation need not rise because prices reveal more information.

In the calibration, lower fixed costs encourage entry and erode gross fund performance. From high levels, lower variable costs favor the largest models and widen performance gaps; at still lower costs, price revelation narrows these gaps. Compute expenditure can rise as costs fall through entry or model expansion. Investors remain indifferent ex ante across occupied scales after compute costs.

Cheaper computation does not remove the statistical limits of learning. With fixed costs at their calibrated value, sufficiently low variable costs lead every investor to acquire all signals. Yet residual volatility remains about 16.1% and the equity premium about 4.7%, while an oracle who knew the predictive coefficients would face no residual risk.

Full acquisition does not lengthen the history available to learn those coefficients. Model uncertainty remains priced even when compute costs no longer constrain acquisition.

References

Admati, Anat R, "A noisy rational expectations equilibrium for multi-asset securities markets," *Econometrica*, 1985, *53* (3), 629–657.

AIMA and SS&C, "Casting the Net: How Hedge Funds Are Using Alternative Data," Technical Report, Alternative Investment Management Association and SS&C 2020.

Al-Najjar, Nabil I, "Decision makers as statisticians: Diversity, ambiguity, and learning," *Econometrica*, 2009, *77* (5), 1371–1401.

Albagli, Elias, Christian Hellwig, and Aleh Tsyvinski, "Information aggregation with asymmetric asset payoffs," *The Journal of Finance*, 2024, *79* (4), 2715–2758.

Ali, Alnur, J Zico Kolter, and Ryan J Tibshirani, "A continuous-time view of early stopping for least squares regression," in "The 22nd International Conference on Artificial Intelligence and Statistics" PMLR 2019, pp. 1370–1378.

Ashwin, Julian, Paul Beaudry, and Martin Ellison, "Neural network learning for nonlinear economies," *Journal of Monetary Economics*, 2025, *149*, 103723.

Avdis, Efstathios and Sergei Glebkin, "CHILE," 2026. Working Paper.

Bai, Zhi-Dong and Yong-Qua Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," in "Advances In Statistics," World Scientific, 2008, pp. 108–127.

Bai, Zhidong and Jack W Silverstein, *Spectral analysis of large dimensional random matrices*, Vol. 20, Springer, 2010.

___ **and Wang Zhou**, "Large sample covariance matrices without independence structures in columns," *Statistica Sinica*, 2008, pp. 425–442.

Banerjee, Snehal and Brett Green, "Signal or noise? Uncertainty and learning about whether other traders are informed," *Journal of Financial Economics*, 2015, *117* (2), 398–423.

Bastianello, Francesca and Paul Fontanier, "Expectations and learning from prices," *The Review of Economic Studies*, 2025, *92* (3), 1341–1374.

Belkin, Mikhail, Daniel Hsu, and Ji Xu, "Two models of double descent for weak features," *SIAM Journal on Mathematics of Data Science*, 2020, *2* (4), 1167–1180.

Bouchaud, Jean-Philippe, "Economics Needs a Scientific Revolution," *Nature*, 2008, *455* (7217), 1181.

___ , "Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges," *Journal of Statistical Physics*, 2013, *151* (3–4), 567–606.

___ , "Radical Complexity," *Entropy*, 2021, *23* (12), 1676.

Brav, Alon and J. B. Heaton, "Competing theories of financial anomalies," *The Review of Financial Studies*, 2002, *15* (2), 575–606.

Bray, Margaret, "Learning, estimation, and the stability of rational expectations," *Journal of Economic Theory*, 1982, *26* (2), 318–339.

Breon-Drish, Bradyn, "On existence and uniqueness of equilibrium in a class of noisy rational expectations models," *The Review of Economic Studies*, 2015, *82* (3), 868–921.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei, "Language Models Are Few-Shot Learners," *arXiv preprint arXiv:2005.14165*, 2020.

Brunnermeier, Markus K, *Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis, and Herding*, Oxford: Oxford University Press, 2001.

Bun, Joël, Jean-Philippe Bouchaud, and Marc Potters, "Cleaning Large Correlation Matrices: Tools from Random Matrix Theory," *Physics Reports*, 2017, *666*, 1–109.

Cerreia-Vioglio, Simone, Lars Peter Hansen, Fabio Maccheroni, and Massimo Marinacci, "Making decisions under model misspecification," *The Review of Economic Studies*, 2026, *93* (2), 892–925.

Chen, Hui, Winston Wei Dou, and Leonid Kogan, "Measuring "Dark Matter" in Asset Pricing Models," *The Journal of Finance*, 2024, *79* (2), 843–902.

Chen, Shuang, Clemens Sialm, and David X. Xu, "The Growth and Performance of Artificial Intelligence in Asset Management," Working Paper 35273, National Bureau of Economic Research 2026.

Chen, Zhimin, Bryan Kelly, and Semyon Malamud, "Limits To (Machine) Learning," *arXiv preprint arXiv:2512.12735*, 2025.

Chizat, Lénaïc, Edouard Oyallon, and Francis Bach, "On Lazy Training in Differentiable Programming," in "Advances in Neural Information Processing Systems," Vol. 32 2019, pp. 2937–2947.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer, "Parameter learning in general equilibrium: The asset pricing implications," *American Economic Review*, 2016, *106* (3), 664–698.

CoreWeave, "Jane Street Signs $6 Billion AI Cloud Agreement with CoreWeave," Press release April 2026.

Davis, Carter and Alejandro Lopez-Lira, "Complexity Breaks Arbitrage Pricing Theory," 2026. Working paper 2026-05, Charles A. Dice Center for Research in Financial Economics, The Ohio State University.

DeepSeek-AI, "DeepSeek-V4: Towards Highly Efficient Million-Token Context Intelligence," *arXiv preprint arXiv:2606.19348*, 2026.

Diaconis, Persi and David Freedman, "On the consistency of Bayes estimates," *The Annals of Statistics*, 1986, pp. 1–26.

Didisheim, Antoine, Shikun Barry Ke, Bryan T Kelly, and Semyon Malamud, "APT or "AIPT"? the surprising dominance of large factor models," Technical Report, National Bureau of Economic Research 2024.

Dou, Winston Wei, Itay Goldstein, and Yan Ji, "AI-powered trading, algorithmic collusion, and price efficiency," Working Paper 34054, National Bureau of Economic Research 2025.

___ , ___ , Zigang Li, and Liyan Yang, "The Limits of AI Trading," 2026. Working paper, SSRN 5866104.

Eldering, Jaap, "Lecture Notes on Invariant Manifolds," Lecture notes, ICMC-USP 2016. Theorems 4.4 and 5.2.

Ellis, Andrew, Michele Piccione, and Shengxing Zhang, "Equilibrium Information Aggregation under Machine Learning," *arXiv preprint arXiv:2607.13670*, 2026.

Esponda, Ignacio and Demian Pouzo, "Berk-Nash equilibrium: A framework for modeling agents with misspecified models," *Econometrica*, 2016, *84* (3), 1093–1130.

Exabel, "Alternative Data Is Now Critical: Budgets, Leadership and Commitment Point to Long-Term Growth," February 2026.

Farboodi, Maryam, Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran, "Valuing financial data," Technical Report, National Bureau of Economic Research 2022.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson, "Learning about the long run," *Journal of Political Economy*, 2024, *132* (10), 3334–3377.

Fenichel, Neil, "Persistence and Smoothness of Invariant Manifolds for Flows," *Indiana University Mathematics Journal*, 1971, *21* (3), 193–226.

Gabaix, Xavier and Thomas Graeber, "The Complexity of Economic Decisions," Working Paper 33109, National Bureau of Economic Research 2024.

Gennotte, Gerard, "Optimal portfolio choice under incomplete information," *The Journal of Finance*, 1986, *41* (3), 733–746.

Ghorbani, Behrooz, Song Mei, Theodor Misiakiewicz, and Andrea Montanari, "Linearized two-layers neural networks in high dimension," *The Annals of Statistics*, 2021, *49* (2), 1029–1054.

Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri, "Economic predictions with big data: The illusion of sparsity," *Econometrica*, 2021, *89* (5), 2409–2437.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, "Test assets and weak factors," Technical Report, National Bureau of Economic Research 2021.

GLM-5 Team, "GLM-5: From Vibe Coding to Agentic Engineering," *arXiv preprint arXiv:2602.15763*, 2026.

Grossman, Sanford J and Joseph E Stiglitz, "On the impossibility of informationally efficient markets," *The American economic review*, 1980, *70* (3), 393–408.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 2020, *33* (5), 2223–2273.

Hedgeweek, "Hedge Funds Charging Billions in 'No Limit' Passthrough Fees," Hedgeweek February 2025.

Hirsch, Morris W., Charles C. Pugh, and Michael Shub, *Invariant Manifolds*, Vol. 583 of *Lecture Notes in Mathematics*, Berlin, Heidelberg: Springer, 1977.

Hobbhahn, Marius and Tamay Besiroglu, "Trends in GPU Price-Performance," Epoch AI 2022.

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre, "Training Compute-Optimal Large Language Models," *arXiv preprint arXiv:2203.15556*, 2022.

Jacot, Arthur, Franck Gabriel, and Clément Hongler, "Neural tangent kernel: Convergence and generalization in neural networks," *arXiv preprint arXiv:1806.07572*, 2018.

Jevons, William Stanley, *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*, London: Macmillan and Co., 1865.

Kadan, Ohad and Asaf Manela, "Estimating the value of information," *The Review of Financial Studies*, 2019, *32* (3), 951–991.

Kelly, Bryan and Dacheng Xiu, "Financial Machine Learning," *Foundations and Trends in Finance*, 2023, *13* (3–4), 205–363.

___, Boris Kuznetsov, Semyon Malamud, and Yuan Zhang, "Large and Deep Factor Models," *arXiv preprint arXiv:2402.06635*, 2026.

___, Semyon Malamud, and Kangying Zhou, "The virtue of complexity in return prediction," *The Journal of Finance*, 2024, *79* (1), 459–503.

Kelly, Bryan T and Semyon Malamud, "Understanding The Virtue of Complexity," *Available at SSRN 5346842*, 2025.

___, Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu, "Artificial intelligence asset pricing models," Technical Report, National Bureau of Economic Research 2025.

___, Semyon Malamud, and Kangying Zhou, "The virtue of complexity everywhere," *Available at SSRN*, 2025.

Kimi Team, "Kimi K3: Open Frontier Intelligence," *arXiv preprint arXiv:2607.24653*, 2026.

Ledoit, Olivier and Michael Wolf, "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 2017, *30* (12), 4349–4388.

Lee, Jaehoon, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington, "Wide neural networks of any depth evolve as linear models under gradient descent," *Advances in neural information processing systems*, 2019, *32*, 8572–8583.

Lewellen, Jonathan and Jay Shanken, "Learning, asset-pricing tests, and market efficiency," *The Journal of finance*, 2002, *57* (3), 1113–1145.

Liao, Yuan, Xin Tong, Wanjie Wang, and Dacheng Xiu, "Fixed-order PCA: Theory for Overestimated Factor Models," *arXiv preprint arXiv:2605.18448*, 2026.

Marčenko, Vladimir A and Leonid Andreevich Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, 1967, *1* (4), 457.

Marcet, Albert and Thomas J Sargent, "Convergence of least-squares learning in environments with hidden state variables and private information," *Journal of Political Economy*, 1989, *97* (6), 1306–1322.

Martin, Ian WR and Stefan Nagel, "Market efficiency in the age of big data," *Journal of Financial Economics*, 2022, *145* (1), 154–177.

Maslej, Nestor et al., "The AI Index 2025 Annual Report," Technical Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University 2025.

Misiakiewicz, Theodor, "Spectrum of Inner-Product Kernel Matrices in the Polynomial Regime and Multiple Descent Phenomenon in Kernel Ridge Regression," *arXiv preprint arXiv:2204.10425*, 2022.

Mistral AI, "Mistral Large 3," Model card December 2025. Accessed July 13, 2026.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin, "Model complexity, expectations, and asset prices," *Review of Economic Studies*, 2024, *91* (4), 2462–2507.

Moll, Benjamin, "The Trouble with Rational Expectations in Heterogeneous Agent Models: A Challenge for Macroeconomics," *The Economic Journal*, 2026, *136* (676), 1173–1216.

Neudata, "How Big Is the Alternative Data Market for Investment Managers?," Neudata research note 2026.

Olea, José Luis Montiel, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat, "Competing models," *The Quarterly Journal of Economics*, 2022, *137* (4), 2419–2457.

Pálvölgyi, Dömötör and Gyuri Venter, "Multiple equilibria in noisy rational expectations economies," *Available at SSRN 2524105*, 2015.

Pástor, Ľuboš and Pietro Veronesi, "Stock valuation and learning about profitability," *The Journal of Finance*, 2003, *58* (5), 1749–1789.

___ **and** ___ , "Learning in financial markets," *Annu. Rev. Financ. Econ.*, 2009, *1* (1), 361–381.

Puri, Indira and Laura Veldkamp, "Artificial Intelligence and Cognitive Inequality," *Journal of Monetary Economics*, 2026, *157*, 103884.

Ritov, Ya'acov, Peter J Bickel, Anthony C Gamst, and Bastiaan Jan Korneel Kleijn, "The Bayesian analysis of complex, high-dimensional models: Can it be CODA?," *Statistical Science*, 2014, *29* (4), 619–639.

Ross, Jillian, Eric So, Zoe De Simone, Charles Pozniak, and Andrew W. Lo, "Why Better Models Can Create Riskier Systems: Evidence from LLM Agents in Financial Markets," *arXiv preprint arXiv:2609.04373*, 2026.

Routledge, Bryan R, "Adaptive learning in financial markets," *The Review of Financial Studies*, 1999, *12* (5), 1165–1202.

Rubio, Francisco and Xavier Mestre, "Spectral convergence for a general class of random matrices," *Statistics & probability letters*, 2011, *81* (5), 592–602.

Shen, Zhouyu and Dacheng Xiu, "Can Machines Learn Weak Signals?," Working Paper 2024-29, University of Chicago, Becker Friedman Institute for Economics 2024.

Silverstein, Jack W and ZD Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, 1995, *54* (2), 175–192.

Timmermann, Allan, "How learning in financial markets generates excess volatility and predictability in stock prices," *The Quarterly Journal of Economics*, 1993, *108* (4), 1135–1145.

___ **and Luka Vulicevic**, "Compute, Complexity, and the Scaling Laws of Return Predictability," 2026. Working paper, University of California San Diego.

Torbey, Henri, Ju-Hon Kwek, Farhan Banani, and Victoria Nguyen, "Asset Management 2025: The Great Convergence," Technical Report, McKinsey & Company September 2025.

Veldkamp, Laura L, *Information choice in macroeconomics and finance*, Princeton University Press, 2011.

Vives, Xavier, *Information and learning in markets: the impact of market microstructure*, Princeton University Press, 2008.

Wang, Jiang, "A model of intertemporal asset prices under asymmetric information," *The Review of Economic Studies*, 1993, *60* (2), 249–282.

Weitzman, Martin L, "Subjective expectations and asset-return puzzles," *American Economic Review*, 2007, *97* (4), 1102–1130.

Welch, Ivo and Amit Goyal, "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 2008, *21* (4), 1455–1508.

Wellington Global Multi-Strategy Fund, "Prospectus Supplement, Form 424B3," U.S. Securities and Exchange Commission filing 2024.

Yin, Yong-Qua, Zhi-Dong Bai, and Pathak R Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," *Probability theory and related fields*, 1988, *78* (4), 509–521.

Zuckerman, Gregory, *The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution*, New York: Portfolio, 2019.

Organization of the appendix

The appendix contains the proofs of the main-text results, the required random-matrix tools, the calibration, supporting evidence on compute and data scale, and the finite-economy foundation of CREE. Appendix A proves Bayesian updating. Appendix B constructs the limiting economy and proves informed demand, and Appendix C proves the single-scale equilibrium and acquisition results. Appendix D constructs nested information and proves pre-price residual risk. Appendix E shows that excluding past prices from the history is exact when the scaling technology appears after the last history date, and negligible in the limit when scaled trading in the history was rare. Appendix F proves post-price residual risk and Corollary 4, recovers the scale distribution, and proves equilibrium existence, uniqueness, and characterization. Appendix G shows that nesting emerges under flexible signal acquisition when marginal source costs are ordered. Appendix H proves the acquisition-cost comparative statics. Appendix I collects the random-matrix tools used in these proofs. Appendix J presents the calibration, empirical targets, baseline outcomes, and motivation for quadratic research costs. Appendix K reports public disclosures of compute and data scale, adapted from Timmermann and Vulicevic (2026) and updated to November 2026. Appendix L constructs rational expectations equilibria of regularized finite economies that converge to CREE, including endogenous binary information acquisition; it requires $\sigma_\varepsilon^2 > 0$. The limiting economy throughout uses the common prior of Assumption 1.

A Bayesian updating: proof of Lemma 1

Proof of Lemma 1. *The current signals carry no information about β.* Under the prior, β, Ψ, and the strong signals are independent, the history and current signals are conditionally independent draws given Ψ, and the history payoffs depend on β only through $\mathbf{d} = \boldsymbol{\mathcal{S}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The joint density of $(\beta, \Psi, \boldsymbol{\mathcal{S}}, \mathbf{d}, s, S)$ therefore factors as

$$p(\beta)\, p(\mathbf{d} \mid \boldsymbol{\mathcal{S}}, \beta)\, p(\mathbf{s})\, p(s)\, p(\Psi)\, p(\mathbf{S} \mid \Psi)\, p(S \mid \Psi).$$

Integrating out Ψ and conditioning on $\mathcal{F} = \sigma(\boldsymbol{\mathcal{S}}, \mathbf{d})$, the density of (β, s, S) is proportional to

$$p(\beta)\, p(\mathbf{d} \mid \boldsymbol{\mathcal{S}}, \beta) \;\times\; p(s) \int p(\Psi)\, p(\mathbf{S} \mid \Psi)\, p(S \mid \Psi)\, d\Psi,$$

a function of β times a function of (s, S). Hence β and (s, S) are independent given $\mathcal{F}$, and $\beta \mid \mathcal{F}_I$ has the law of $\beta \mid \mathcal{F}$. It remains to compute the latter.

Posterior of β. Because the strong block (b, ψ) is known, the investor first removes the strong component and works with the residual $\mathbf{d}^\perp := \mathbf{d} - \mathbf{s}b$, which in vector form is

$$\mathbf{d}^\perp \;=\; \mathbf{S}\beta \;+\; \boldsymbol{\varepsilon},$$

where $\mathbf{d}^\perp = (d_\tau - s_\tau' b)_{\tau=1}^T$ and $\mathbf{S} = (S_\tau)_{\tau=1}^T \in \mathbb{R}^{T\times N}$. To lighten notation, we write $\mathbf{d}$ for $\mathbf{d}^\perp$ throughout the rest of the proof. Under the investor's prior, $\beta \in \mathbb{R}^N$ is drawn from $\mathcal{N}(0, \sigma_\beta^2 I/N)$, and the signals have covariance matrix Ψ. The Gaussian projection theorem gives

$$\begin{aligned} E[\beta \mid \mathbf{S}, \mathbf{d}] \;&=\; \mathrm{Cov}[\beta, \mathbf{d} \mid \mathbf{S}]\mathrm{Cov}[\mathbf{d} \mid \mathbf{S}]^{-1}\mathbf{d} \\ \mathrm{Var}[\beta \mid \mathbf{S}, \mathbf{d}] \;&=\; \mathrm{Var}[\beta] \;-\; \mathrm{Cov}[\beta, \mathbf{d} \mid \mathbf{S}]\mathrm{Cov}[\mathbf{d} \mid \mathbf{S}]^{-1}\mathrm{Cov}[\beta, \mathbf{d} \mid \mathbf{S}]' \end{aligned}$$

We have

$$\begin{aligned} \mathrm{Cov}[\mathbf{d} \mid \mathbf{S}] \;&=\; (\sigma_\varepsilon^2 I_T + c_N^{-1}\sigma_\beta^2 \check{\Psi}) \\ \mathrm{Cov}[\beta, \mathbf{d} \mid \mathbf{S}] \;&=\; E[\beta \mathbf{d}' \mid \mathbf{S}] \;=\; \sigma_\beta^2 \mathbf{S}'/N \end{aligned}$$

where $c_N = N/T$ and $\check{\Psi} := \mathbf{S}\mathbf{S}'/T \in \mathbb{R}^{T\times T}$ is the *Gram dual* of the sample covariance $\hat{\Psi} = \mathbf{S}'\mathbf{S}/T \in \mathbb{R}^{N\times N}$ of (6). The two share their nonzero eigenvalues and are exchanged by the push-through identity

$$\mathbf{S}'\big(\ell I_T + \check{\Psi}\big)^{-1} \;=\; \big(\ell I_N + \hat{\Psi}\big)^{-1}\mathbf{S}', \tag{29}$$

which is the only matrix fact this proof needs: it moves every $T \times T$ inverse to the $N \times N$ side. Hence

$$\mathrm{Var}[\beta \mid \mathbf{S}, \mathbf{d}] \;=\; \sigma_\beta^2 I/N \;-\; \sigma_\beta^4 \mathbf{S}'(\sigma_\varepsilon^2 I_T + c_N^{-1}\sigma_\beta^2 \check{\Psi})^{-1}\mathbf{S}/N^2.$$

Now we plug everything in; by (29), we have

$$\hat{\beta} \;=\; \mathbf{S}'/N(\sigma_\beta^{-2}\sigma_\varepsilon^2 I_T + c_N^{-1}\check{\Psi})^{-1}\mathbf{d} \;=\; (z c_N I_N + \hat{\Psi})^{-1}\mathbf{S}'\mathbf{d}/T$$

where $z = \sigma_\varepsilon^2 \sigma_\beta^{-2}$. Applying (29) once more to the posterior covariance,

$$\begin{aligned}\hat{\Sigma}_\beta &= \sigma_\beta^2 I/N \;-\; c_N \sigma_\beta^2 \mathbf{S}'(z c_N I_T + \check{\Psi})^{-1}\mathbf{S}/N^2 \\ &= \sigma_\beta^2 I/N \;-\; \sigma_\beta^2 (z c_N I + \hat{\Psi})^{-1}\hat{\Psi}/N \\ &= \sigma_\varepsilon^2 T^{-1}(z c_N I + \hat{\Psi})^{-1}.\end{aligned}$$

This proves part (i).

Posterior of the payoff. Given $\mathcal{F}_I$, the strong component $b's$ and the current weak signals S are known, so

$$d \;=\; b's \;+\; \beta' S \;+\; \varepsilon,$$

where $\beta \mid \mathcal{F}_I \sim \mathcal{N}(\hat{\beta}, \hat{\Sigma}_\beta)$ by part (i) and $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is independent of β and of $\mathcal{F}_I$. Hence $\beta' S \mid \mathcal{F}_I \sim \mathcal{N}(\hat{\beta}' S,\; S' \hat{\Sigma}_\beta S)$, and adding the independent ε gives part (ii). □

Inverse-Wishart convention. For a positive-definite scale matrix Λ and $\nu_\Psi > N - 1$, the distribution $\mathcal{W}_N^{-1}(\Lambda, \nu_\Psi)$ in Assumption 1 has density

$$p(\Psi) = \frac{|\Lambda|^{\nu_\Psi/2}}{2^{\nu_\Psi N/2}\Gamma_N(\nu_\Psi/2)} |\Psi|^{-(\nu_\Psi + N + 1)/2} \exp\!\left\{-\tfrac{1}{2}\operatorname{tr}(\Lambda \Psi^{-1})\right\},$$

where $\Gamma_N(a) = \pi^{N(N-1)/4} \prod_{i=1}^{N} \Gamma(a + (1-i)/2)$. For $\nu_\Psi > N + 1$, its mean is $\Lambda/(\nu_\Psi - N - 1)$. The case $V_0 = 0$ is understood as the degenerate weak block $\Psi = 0$.

B The limiting economy and informed demand

We construct the common probability space used in Definition 1 and prove Lemma 2 directly under the common prior. The only asymptotic inputs are the concentration and deterministic equivalents in Appendix I.

B.1 Proof of Lemma 2

Proof of Lemma 2. Let $c_N = N/T_N \to c$ and write $O_N = (\boldsymbol{\mathcal{S}}_N, \mathbf{d}_N, s_N, S_N)$ for the informed observations. Lemma 1 gives the exact posterior

$$d_N \mid O_N \sim \mathcal{N}(M_N, R_N), \qquad M_N = b's_N + \hat{\beta}_N' S_N, \qquad R_N = \sigma_\varepsilon^2 + S_N' \hat{\Sigma}_{\beta,N} S_N. \tag{30}$$

Concentration of residual risk. Conditional on Ψ_N and the history, $S_N \sim \mathcal{N}(0, \Psi_N)$ is independent of the history, and $N\hat{\Sigma}_{\beta,N} \preceq \sigma_\beta^2 I_N$. Lemma I.1, localized using Lemma I.6, therefore gives

$$R_N - \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{T_N} \operatorname{tr}\Big[\Psi_N (z c_N I_N + \hat{\Psi}_N)^{-1}\Big] \xrightarrow{p} 0.$$

Proposition I.4, evaluated at the prior spectrum H_0, implies

$$R_N \xrightarrow{p} r := \sigma_\varepsilon^2[1 + \xi(zc; c)] = \sigma_\varepsilon^2 \frac{Z_*}{zc}, \qquad Z_* := zc[1 + \xi(zc; c)].$$

The change from zc_N to zc is harmless by the resolvent identity and the uniform lower bound on these positive ridge parameters.

The effective-ridge quadratic. Apply Theorem I.5 to the normalized trace with $A_N = \Psi_N$. Its fixed-point identity is

$$\int \frac{x}{Z_* + x}\, dH_0(x) = \frac{1}{c} - \frac{z}{Z_*}. \tag{31}$$

By Lemma I.6, H_0 is the image of MP_y under $x = a/t$, where $y = (1+\kappa)^{-1}$ and $a = V_0/(1+\kappa)$. For $V_0 > 0$, set $u = a/Z_*$ and let m_{MP} be the Stieltjes transform of MP_y. The left-hand side of (31) is $u m_{\mathrm{MP}}(-u)$, and

$$ywm_{\mathrm{MP}}(w)^2 - (1 - y - w)m_{\mathrm{MP}}(w) + 1 = 0.$$

Substituting $m_{\mathrm{MP}}(-u) = (c^{-1} - z/Z_*)/u$ yields

$$(1 + c\kappa)Z_*^2 + c[V_0(1 - c) - z(2 + c\kappa)]Z_* - c^2 z(V_0 - z) = 0.$$

The polynomial is negative at $Z_* = zc$ when $V_0 > 0$ and has positive leading coefficient, so its root above zc is the larger root. This proves the formula for $\sigma_I^2(c) = r$ in the lemma. When $V_0 = 0$, the weak block vanishes and $r = \sigma_\varepsilon^2$, with $Z_* = zc$ by continuity.

The joint Gaussian limit. Conditional on (β_N, Ψ_N), the payoff is centered Gaussian with variance $b'\psi b + \beta_N' \Psi_N \beta_N + \sigma_\varepsilon^2$. Independence of β_N and Ψ_N, quadratic-form concentration, and the prior spectral mean give

$$\beta_N' \Psi_N \beta_N \xrightarrow{p} \sigma_\beta^2 \frac{V_0}{\kappa}.$$

Consequently $d_N \Rightarrow \mathcal{N}(0, \sigma_d^2)$. For every real t, the exact posterior implies

$$E[e^{itd_N}] = E[e^{itM_N - t^2 R_N/2}] = e^{-t^2 r/2} E[e^{itM_N}] + o(1).$$

The error vanishes by bounded convergence in probability, since $R_N \to r$. Thus $M_N \Rightarrow \mathcal{N}(0, \sigma_d^2 - r)$. More generally, for real t, u,

$$E[e^{itM_N + iu(d_N - M_N)}] = E[e^{itM_N - u^2 R_N/2}] \longrightarrow \exp\left\{-\tfrac{1}{2}t^2(\sigma_d^2 - r) - \tfrac{1}{2}u^2 r\right\}.$$

This proves that the limiting forecast and residual are independent Gaussians. The supply shock is independent of both at every N and in the limit.

Coupling. The steps above give $(M_N, R_N) \Rightarrow (\tilde{S}, r)$ with $\tilde{S} \sim \mathcal{N}(0, \sigma_d^2 - r)$. The convergence $R_N \to r$ in probability does not depend on how the economies for different N are coupled, because r is deterministic. Weak convergence on $\mathbb{R}^2$ has an almost-sure representation: there are copies of the pairs (M_N, R_N) on one probability space with $(M_N, R_N) \to (\tilde{S}, r)$ almost surely. Appendix B.2 extends this space to the limiting economy. That construction uses only the weak limit above, and it gives the informed posterior $\mathcal{N}(\tilde{S}, r)$ in (32).

Demand convergence and optimality. Because $R_N, r \geq \sigma_\varepsilon^2$, for every compact price set $C \subset \mathbb{R}$, almost surely on the coupled space,

$$\sup_{p \in C} \left| \frac{M_N - p}{\gamma R_N} - \frac{\tilde{S} - p}{\gamma r} \right| \leq \frac{|M_N - \tilde{S}|}{\gamma \sigma_\varepsilon^2} + \sup_{p \in C} |\tilde{S} - p| \frac{|R_N - r|}{\gamma \sigma_\varepsilon^4} \longrightarrow 0.$$

Under the posterior (32), conditional expected utility is

$$-\exp\left\{-\gamma x(\tilde{S} - p) + \tfrac{1}{2}\gamma^2 x^2 r\right\}.$$

Its unique maximizer is $(\tilde{S} - p)/(\gamma r)$, proving (10) in every CREE. The coupled copies have the original finite-economy laws, so the demands also converge in distribution. □

B.2 Construction and verification of the limiting economy

The proof of Lemma 2 establishes $(M_N, R_N) \Rightarrow (M^\infty, r)$, with $M^\infty \sim \mathcal{N}(0, \sigma_d^2 - r)$ and $r = \sigma_I^2(c)$. Its coupling step realizes copies of these pairs on one probability space with almost-sure convergence. We extend that space to the whole economy. For each N, draw O_N conditional on (M_N, R_N) according to the original finite common-prior law, using ran-

domness independent of the payoff residual introduced below. These regular conditional distributions exist because the observations and statistics take values in Euclidean spaces. The observations have their original marginal laws and reproduce their prescribed posterior statistics.

Take independent $Z, \eta \sim \mathcal{N}(0,1)$, independent also of the entire observation array, and define

$$d_N = M_N + \sqrt{R_N}\, Z, \qquad d^\infty = M^\infty + \sqrt{r}\, Z.$$

By (30), every (O_N, d_N, η) has its original finite-economy joint law. The coupling chooses only dependence across different economy sizes, and gives $d_N \to d^\infty$ almost surely.

Formally, take Ω to be the coordinate space of arrays $((O_N, d_N)_{N\geq 1}, \eta)$ on which M_N and d_N have finite limits and $R_N \to r$. Equip it with its Borel sigma-field $\mathcal{A}$ and the law $\mathbb{P}_\infty$ induced by this construction. Set

$$\mathcal{F}_I^\infty = \sigma(O_1, O_2, \ldots), \qquad \tilde{S} = M^\infty = \lim_N M_N, \qquad d^\infty = \lim_N d_N.$$

Independence of Z from the whole observed array gives

$$d^\infty \mid \mathcal{F}_I^\infty \sim \mathcal{N}(\tilde{S}, r). \tag{32}$$

It also ensures that conditioning on the full observation array leaves each finite posterior $\mathcal{N}(M_N, R_N)$ unchanged. The limiting posterior is therefore conditional on the full limiting information, not just the forecast statistic.

At every information state in Ω, including null states, select the kernel $\mathcal{N}(\lim_N M_N, r)$. It is the weak limit of the finite Bayesian posteriors along that observed array and a version of (32). This supplies the statewise convention in Definition 1. For any candidate price P, the uninformed investors condition on $\sigma(P)$ under the same $\mathbb{P}_\infty$. Appendix C.1 computes this conditional law in a CREE and specifies its continuous Gaussian version at every price.

B.3 Learning the strong-signal block

The following auxiliary result concerns the known strong-signal parameters in Assumption 1. Informed investors observe the history and can estimate these parameters. Proposition B.1 shows that the estimation does not change the limit of their payoff posterior. Uninformed

investors do not observe the history. That they know (b, ψ) is an assumption. Work on the nondegenerate strong-signal coordinates, so that ψ is positive definite.

For this auxiliary calculation, suppose the informed investor estimates the strong block under the following priors:

$$b \sim \mathcal{N}(0, \sigma_b^2 I_n), \qquad \psi \sim \mathcal{W}^{-1}(V_0^n,\, n + k_\psi),$$

independent of the weak-block priors in (2); signals are conditionally Gaussian given their covariances. Write $z_b := \sigma_\varepsilon^2/\sigma_b^2$, $\hat{\psi} := T^{-1}\sum_{\tau \le T} s_\tau s_\tau'$, and let $\boldsymbol{\mathcal{S}} = [\mathbf{s}, \mathbf{S}] \in \mathbb{R}^{T \times (n+N)}$ be the augmented signal matrix.

Applying the Gaussian updating calculation of Lemma 1 to the augmented coefficient vector gives the following posterior. Writing $\boldsymbol{\beta} = (b, \beta)$ for the joint coefficient, the posterior is Gaussian, $\boldsymbol{\beta} \mid \mathcal{F} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}, \hat{\Sigma}_{\boldsymbol{\beta}})$ with $\hat{\boldsymbol{\beta}} = (D + \boldsymbol{\mathcal{S}}'\boldsymbol{\mathcal{S}})^{-1}\boldsymbol{\mathcal{S}}'\mathbf{d}$, $\hat{\Sigma}_{\boldsymbol{\beta}} = \sigma_\varepsilon^2(D + \boldsymbol{\mathcal{S}}'\boldsymbol{\mathcal{S}})^{-1}$, and $D = \operatorname{diag}(z_b I_n,\ zN\, I_N)$; the strong covariance updates by conjugacy, $\psi \mid \mathcal{F} \sim \mathcal{W}^{-1}(T\hat{\psi} + V_0^n,\ T + n + k_\psi)$, so $\bar{\psi} := E[\psi \mid \mathcal{F}] = (T\hat{\psi} + V_0^n)/(T + k_\psi - 1)$. Let $Q_S := I_T - \mathbf{S}(zN\, I_N + \mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ be the residualization matrix that partials the weak block out of the strong regression, so that

$$\hat{b} = (z_b I_n + \mathbf{s}' Q_S \mathbf{s})^{-1}\mathbf{s}' Q_S\, \mathbf{d}, \qquad \hat{\Sigma}_{bb} = \sigma_\varepsilon^2 (z_b I_n + \mathbf{s}' Q_S \mathbf{s})^{-1}.$$

Let $B_S := (zN\, I_N + \mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{s} \in \mathbb{R}^{N \times n}$, and let $\hat{\beta}$ and $\hat{\Sigma}_\beta$ be the known-block posterior moments of Lemma 1. Block inversion of $D + \boldsymbol{\mathcal{S}}'\boldsymbol{\mathcal{S}}$ gives the other blocks of $\hat{\Sigma}_{\boldsymbol{\beta}}$ and the weak block $\hat{\beta}^{\text{est}}$ of $\hat{\boldsymbol{\beta}}$:

$$\hat{\Sigma}_{\beta b} = -B_S \hat{\Sigma}_{bb}, \qquad \hat{\Sigma}_{\beta\beta} = \hat{\Sigma}_\beta + B_S \hat{\Sigma}_{bb} B_S', \qquad \hat{\beta}^{\text{est}} = \hat{\beta} - B_S(\hat{b} - b). \tag{33}$$

The current signals carry no information about $\boldsymbol{\beta}$, as in the proof of Lemma 1. With an estimated strong block, the informed payoff posterior is therefore $\mathcal{N}(M_N^{\text{est}}, R_N^{\text{est}})$, where

$$M_N^{\text{est}} = \hat{b}'s + (\hat{\beta}^{\text{est}})'S, \qquad R_N^{\text{est}} = \sigma_\varepsilon^2 + \mathcal{S}'\hat{\Sigma}_{\boldsymbol{\beta}}\,\mathcal{S}.$$

Its known-block counterpart is $\mathcal{N}(M_N, R_N)$ in (30).

Proposition B.1 (The strong block is learned perfectly) *Under these priors, in the limit $T, N \to \infty$ with $N/T \to c > 0$ and n fixed, and with $m(\cdot; c)$ as in Theorem I.3:*

(i) $\dfrac{1}{T}\operatorname{tr}(Q_S) \xrightarrow{p} \dfrac{\sigma_\varepsilon^2}{\sigma_I^2(c)} = 1 - c\big(1 - zc\, m(-zc; c)\big)$;

(ii) the strong-coefficient posterior covariance vanishes at the parametric rate, $T\,\hat{\Sigma}_{bb} \xrightarrow{p} \sigma_I^2(c)\,\psi^{-1}$*, and* $\hat{b} \xrightarrow{p} b$ *with* $\hat{b} - b = O_p(T^{-1/2})$*;*

(iii) the strong-covariance posterior mean is consistent, $\bar{\psi} \xrightarrow{p} \psi$*;*

(iv) the payoff posterior has the known-block limit, $M_N^{\text{est}} - M_N = O_p(T^{-1/2})$ *and* $0 \le R_N^{\text{est}} - R_N = O_p(T^{-1})$*.*

By item (iv) and $R_N \ge \sigma_\varepsilon^2$, the informed demand (9) with an estimated strong block has the limit in Lemma 2. For informed investors, taking (b, ψ) as known in the main text is therefore without loss.

Proof of Proposition B.1. *(i) Trace of* Q_S*.* By the cyclic property of the trace and $\mathbf{S}'\mathbf{S} = T\hat{\Psi}$ (so $zN\,I_N + \mathbf{S}'\mathbf{S} = T(zc_N I_N + \hat{\Psi})$ with $c_N = N/T$),

$$\tfrac{1}{T}\operatorname{tr}(Q_S) = 1 - \tfrac{1}{T}\operatorname{tr}\big((zN\,I_N + \mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{S}\big) = 1 - \tfrac{N}{T} + zc_N\,\tfrac{N}{T}\,\hat{m}(zc_N),$$

where $\hat{m}(zc_N) := \frac{1}{N}\operatorname{tr}\big((zc_N I_N + \hat{\Psi})^{-1}\big)$. Since $c_N \to c$ and $\hat{m}(zc_N) \to m(-zc; c)$ (Theorem I.3), the right-hand side tends to $1 - c + zc^2\,m(-zc;c) = 1 - c(1 - zc\,m(-zc;c))$; by Proposition I.4 and Lemma 2, $\sigma_I^2(c) = \sigma_\varepsilon^2/(1 - c(1 - zc\,m(-zc;c)))$, which is (i).

(ii) Posterior covariance and consistency. Q_S is symmetric with eigenvalues in $(0, 1]$ and independent of $\mathbf{s}$, whose rows are i.i.d. $\mathcal{N}(0, \psi)$. Conditional on Q_S, the Gaussian quadratic form $\frac{1}{T}\mathbf{s}'Q_S\mathbf{s}$ has mean $\frac{1}{T}\operatorname{tr}(Q_S)\,\psi$ and variance $O\big(\operatorname{tr}(Q_S^2)/T^2\big) = O(1/T)$ (using $\operatorname{tr}(Q_S^2) \le \operatorname{tr}(Q_S) \le T$); hence $\frac{1}{T}\mathbf{s}'Q_S\mathbf{s} \xrightarrow{p} \frac{1}{T}\operatorname{tr}(Q_S)\,\psi \to \frac{\sigma_\varepsilon^2}{\sigma_I^2(c)}\psi$, the concentration now running over the T time periods (with n fixed) rather than over the N signals. As $z_b/T \to 0$,

$$T\,\hat{\Sigma}_{bb} = \sigma_\varepsilon^2\big(\tfrac{z_b}{T} I_n + \tfrac{1}{T}\mathbf{s}'Q_S\mathbf{s}\big)^{-1} \xrightarrow{p} \sigma_\varepsilon^2\big(\tfrac{\sigma_\varepsilon^2}{\sigma_I^2(c)}\psi\big)^{-1} = \sigma_I^2(c)\,\psi^{-1}.$$

By the partitioned-regression formula, with $\mathbf{d} = \mathbf{s}\,b + u$ and $u = \mathbf{S}\beta + \boldsymbol{\varepsilon}$,

$$\hat{b} - b = (z_b I_n + \mathbf{s}'Q_S\mathbf{s})^{-1}\big(-z_b\,b + \mathbf{s}'Q_S u\big).$$

The inverse is $O_p(1/T)$ by the display above. Since $\mathbf{s}$ is independent of $(\mathbf{S}, \beta, \boldsymbol{\varepsilon})$ with mean-zero rows, $E[\mathbf{s}'Q_S u \mid \mathbf{S}, \beta, \boldsymbol{\varepsilon}] = 0$ and $\operatorname{Var}[\mathbf{s}'Q_S u \mid \mathbf{S}, \beta, \boldsymbol{\varepsilon}] = (u'Q_S^2 u)\,\psi$; as $\|Q_S\| \le 1$ and $u'u = \sum_\tau u_\tau^2 = O_p(T)$, we have $u'Q_S^2 u \le u'u = O_p(T)$, so $\mathbf{s}'Q_S u = O_p(\sqrt{T})$. Combining, $\hat{b} - b = O_p(1/T) \cdot O_p(\sqrt{T}) = O_p(T^{-1/2}) \to 0$.

(iii) Covariance consistency. With n, k_ψ fixed, $\hat{\psi} = T^{-1}\sum_{\tau \le T} s_\tau s_\tau' \xrightarrow{p} \psi$ by the law of large numbers, and $\bar{\psi} = (T\hat{\psi} + V_0^n)/(T + k_\psi - 1) \to \psi$ since $V_0^n/T \to 0$ and $T/(T + k_\psi - 1) \to 1$.

(iv) Payoff posterior. Substituting (33) into the definitions of M_N^{est} and R_N^{est} and writing $v := s - B_S' S \in \mathbb{R}^n$,

$$M_N^{\text{est}} - M_N = (\hat{b} - b)'v, \qquad R_N^{\text{est}} - R_N = v'\hat{\Sigma}_{bb}\, v \;\geq\; 0.$$

It remains to show $v = O_p(1)$. Since $t/(zN + t^2) \leq (2\sqrt{zN})^{-1}$ for $t \geq 0$, the singular values of $(zN\, I_N + \mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ are at most $(2\sqrt{zN})^{-1}$. For fixed n, $\|\mathbf{s}\|^2 \leq \operatorname{tr}(\mathbf{s}'\mathbf{s}) = O_p(T)$. Hence $\|B_S\| \leq \|\mathbf{s}\|/(2\sqrt{zN}) = O_p(1)$. Given Ψ and the history, $S \sim \mathcal{N}(0, \Psi)$, so

$$E\big[\|B_S' S\|^2 \mid \Psi, \boldsymbol{\mathcal{S}}\big] = \operatorname{tr}\big(B_S' \Psi B_S\big) \leq n\, \|\Psi\|\, \|B_S\|^2,$$

which is $O_p(1)$ by Lemma I.6. Thus $v = O_p(1)$. Item (ii) gives $\hat{b} - b = O_p(T^{-1/2})$ and $\|\hat{\Sigma}_{bb}\| = O_p(T^{-1})$, which proves (iv). □

C Equilibrium and information acquisition

The proofs follow Lemma 3, Theorems 1 and 2, and Proposition 3 in the main text. As in the main text, this appendix suppresses the superscript ∞ on limiting random variables and information sets and writes E, Var, and Cov for moments under $\mathbb{P}_\infty$.

C.1 Proof of Lemma 3

Proof of Lemma 3. Write $\mathsf{A} := \gamma(1-\lambda)/\lambda \geq 0$ and suppress the argument c. Market clearing with the informed demand (10) gives (13), $s_P = P - \mathsf{A}\sigma_I^2\, x_U(P)$. By the restriction following Definition 1, x_U is continuous and nonincreasing; since $\mathsf{A}\sigma_I^2 \geq 0$, the map $p \mapsto p - \mathsf{A}\sigma_I^2\, x_U(p)$ is continuous, strictly increasing, and tends to $\pm\infty$ as $p \to \pm\infty$, so it is a bijection of $\mathbb{R}$. Hence $\sigma(P) = \sigma(s_P)$, and conditioning on $P = p$ is conditioning on $s_P = s_P(p) := p - \mathsf{A}\sigma_I^2\, x_U(p)$. By Lemma 2, $\tilde{S} \sim \mathcal{N}(0, \tau_S^{-1})$, $\eta \sim \mathcal{N}(0,1)$, and $e \sim \mathcal{N}(0, \sigma_I^2)$ are independent under $\mathbb{P}_\infty$. Hence $(d, s_P) = (\tilde{S} + e,\ \tilde{S} - \nu\eta)$ is jointly Gaussian, with $\operatorname{Cov}[d, s_P] = \tau_S^{-1}$, $\operatorname{Var}[s_P] = \tau_S^{-1} + \tau_P^{-1}$, and $\operatorname{Var}[d] = \tau_S^{-1} + \sigma_I^2$. Gaussian projection gives $d \mid s_P \sim \mathcal{N}(\alpha\, s_P,\ \sigma_U^2)$, with $\alpha = \tau_S^{-1}/(\tau_S^{-1} + \tau_P^{-1}) = \tau_P/(\tau_S + \tau_P)$ and

$$\sigma_U^2 \;=\; \tau_S^{-1} + \sigma_I^2 - \frac{\tau_S^{-2}}{\tau_S^{-1} + \tau_P^{-1}} \;=\; \sigma_I^2 + \frac{1}{\tau_S + \tau_P},$$

as claimed. We use this Gaussian version at every $p \in \mathbb{R}$. □

C.2 Proof of Theorem 1

Proof of Theorem 1. Write $\mathsf{A} := \gamma(1-\lambda)/\lambda \geq 0$ and suppress the argument c.

Informed demand. By Lemma 2, the informed demand of every CREE is $x_I(p) = (\tilde{S} - p)/(\gamma\sigma_I^2)$.

Uninformed demand. Take a CREE; by the restriction following Definition 1, x_U is continuous and nonincreasing. By Lemma 3, $d \mid P = p \sim \mathcal{N}(\alpha\, s_P(p), \sigma_U^2)$, with $s_P(p) = p - \mathsf{A}\sigma_I^2\, x_U(p)$. As in the proof of Lemma 2, the CARA problem in Definition 1(ii) under this posterior has the unique solution

$$x_U(p) \;=\; \frac{\alpha\, s_P(p) - p}{\gamma\sigma_U^2} \;=\; \frac{\alpha\big(p - \mathsf{A}\sigma_I^2\, x_U(p)\big) - p}{\gamma\sigma_U^2}\,.$$

Collecting $x_U(p)$ gives $x_U(p)\big(\gamma\sigma_U^2 + \alpha\mathsf{A}\sigma_I^2\big) = -(1-\alpha)\,p$, so $x_U(p) = -\vartheta p$ with

$$\vartheta \;=\; \frac{1-\alpha}{\gamma\sigma_U^2 + \alpha\mathsf{A}\sigma_I^2} \;=\; \frac{\lambda(1-\alpha)}{\gamma\big[\lambda\sigma_U^2 + (1-\lambda)\alpha\sigma_I^2\big]}\,,$$

which is (14). Since $\sigma_\eta > 0$, $\tau_P < \infty$ and $\alpha < 1$, so $\vartheta > 0$.

Price. With these schedules, market clearing reads $\lambda(\tilde{S} - P)/(\gamma\sigma_I^2) - (1-\lambda)\vartheta P = \sigma_\eta\eta$. The left-hand side is strictly decreasing in P, so the equation has a unique solution. By (13), it is $P = s_P/(1 + \mathsf{A}\sigma_I^2\vartheta) = a\,(\tilde{S} - \nu\eta)$ with $a = 1/(1 + \mathsf{A}\sigma_I^2\vartheta)$. Substituting ϑ,

$$\mathsf{A}\sigma_I^2\vartheta \;=\; \frac{(1-\lambda)(1-\alpha)\sigma_I^2}{\lambda\sigma_U^2 + (1-\lambda)\alpha\sigma_I^2}\,, \qquad a \;=\; \frac{\lambda\sigma_U^2 + (1-\lambda)\alpha\sigma_I^2}{\lambda\sigma_U^2 + (1-\lambda)\sigma_I^2}\,,$$

which is (15). This proves uniqueness.

Existence. Let x_I, x_U, and P be given by (14) and (15). The price is a function of $(\tilde{S}, \eta)$, hence $\mathcal{F}_I^\infty \vee \sigma(\eta)$-measurable, and $x_U(p)$ is deterministic. Informed optimality holds by Lemma 2, and market clearing holds by construction. The schedule x_U is continuous and decreasing, so Lemma 3 and the demand computation above give uninformed optimality. □

C.3 Proof of Theorem 2

Proof of Theorem 2. Suppress the argument c. By the definition of R_I^2, $\sigma_I^2 = (1 - R_I^2)\sigma_d^2$ and $B := \sigma_d^2 - \sigma_I^2 = R_I^2\sigma_d^2$. Recall $\zeta = 1 - e^{-2\gamma\chi} \in (0,1)$, so $e^{-2\gamma\chi} = 1 - \zeta$.

Ex-ante utilities. Fix $\lambda \in (0,1]$ and the CREE of Theorem 1. Let m_j and σ_j^2 be the posterior mean and variance of investor $j \in \{I, U\}$, so $m_I = \tilde{S}$, $m_U = \alpha s_P$, and $x_j(P) =$

$(m_j - P)/(\gamma\sigma_j^2)$. Conditional on the investor's information, the CARA utility of this demand is $-\exp\{-(m_j - P)^2/(2\sigma_j^2)\}$. Write $d - P = (m_j - P) + (d - m_j)$. The error $d - m_I = e$ is independent of $(\tilde{S}, \eta)$, hence of $m_I - P$. The error $d - m_U$ is independent of s_P, hence of P and of $m_U - P$. The variable $m_j - P$ is therefore Gaussian with mean zero and variance $\mathrm{Var}[d - P] - \sigma_j^2$. For $X \sim \mathcal{N}(0, s^2)$, $E[-\exp\{-X^2/(2\sigma^2)\}] = -(1 + s^2/\sigma^2)^{-1/2}$. Hence

$$V_I \;=\; -e^{\gamma\chi}\,\frac{\sigma_I}{\sqrt{\mathrm{Var}[d-P]}}\,, \qquad V_U \;=\; -\frac{\sigma_U}{\sqrt{\mathrm{Var}[d-P]}}\,,$$

and $V_I \gtreqless V_U$ if and only if $(1-\zeta)\sigma_U^2 \gtreqless \sigma_I^2$.

The zero-mass economy. Let $\lambda = 0$ and take a CREE. Market clearing reads $x_U(P) = \sigma_\eta\eta$ in every state. Since $\sigma_\eta > 0$, $\eta = x_U(P)/\sigma_\eta$ is a function of P. For the converse, x_U is continuous and nonincreasing, so its level sets $\{p : x_U(p) = y\}$ are disjoint intervals. At most countably many of them contain more than one point; let D be the countable set of the corresponding values y. For $y \notin D$ in the range of x_U, let $g(y)$ be the unique solution of $x_U(p) = y$; the map g is decreasing, hence Borel. The variable $\sigma_\eta\eta$ has a continuous distribution, so $\sigma_\eta\eta \notin D$ and $P = g(\sigma_\eta\eta)$ almost surely. Hence $\sigma(P) = \sigma(\eta)$ up to null sets. The supply shock η is independent of the payoff d, so $d \mid P \sim \mathcal{N}(0, \sigma_d^2)$, and we use this version at every $p \in \mathbb{R}$. CARA optimization then gives $x_U(p) = -p/(\gamma\sigma_d^2)$. This schedule is strictly decreasing, so market clearing gives $P = -\gamma\sigma_\eta\sigma_d^2\eta$. The informed demand is that of Lemma 2. Conversely, these schedules and this price satisfy Definition 1, and x_U is continuous and decreasing. They are therefore the unique CREE at $\lambda = 0$.

In this CREE, $m_U = 0$ and $\sigma_U^2 = \sigma_d^2$, and $x_j(P) = (m_j - P)/(\gamma\sigma_j^2)$ for $j \in \{I, U\}$. The error $d - m_I = e$ is independent of $(\tilde{S}, \eta)$, and $d - m_U = d$ is independent of P. The ex-ante computation above therefore applies at $\lambda = 0$. Hence $V_I \le V_U$ if and only if $(1-\zeta)\sigma_d^2 \le (1 - R_I^2)\sigma_d^2$, that is, $R_I^2 \le \zeta$.

The acquisition margin. For $\lambda \in (0, 1]$, Lemma 3 gives

$$\sigma_U^2(\lambda) \;=\; \sigma_I^2 + \frac{B\,\nu^2}{B + \nu^2}\,, \qquad \nu \;=\; \frac{\gamma\sigma_\eta\sigma_I^2}{\lambda}\,.$$

It is continuous and strictly decreasing in λ, and it tends to $\sigma_I^2 + B = \sigma_d^2$ as $\lambda \downarrow 0$. The net gain $(1-\zeta)\sigma_U^2(\lambda) - \sigma_I^2$ is therefore strictly decreasing on $(0, 1]$, with limit $\sigma_d^2(R_I^2 - \zeta)$ at $\lambda = 0$.

If $R_I^2 \le \zeta$, the gain is negative for every $\lambda \in (0, 1]$, so no positive fraction satisfies Definition 2, while $\lambda = 0$ does by the previous step.

If $R_I^2 > \zeta$, $\lambda = 0$ is not an overall equilibrium by the previous step. A positive fraction is an interior equilibrium if and only if the gain is zero, that is,

$$\frac{B\,\nu^2}{B+\nu^2} \;=\; \frac{\zeta}{1-\zeta}\,\sigma_I^2 \quad\Longleftrightarrow\quad \nu^2 \;=\; \frac{\zeta\,(1-R_I^2)\,R_I^2}{R_I^2-\zeta}\,\sigma_d^2\,,$$

which, with $\nu = \gamma\sigma_\eta\sigma_I^2/\lambda$ and $\sigma_I^2 = (1-R_I^2)\sigma_d^2$, gives

$$\lambda \;=\; \gamma\sigma_\eta\sigma_d\sqrt{\frac{(1-R_I^2)(R_I^2-\zeta)}{\zeta R_I^2}}\,.$$

By monotonicity this is the unique zero of the gain. If it lies in $(0,1)$, it is the unique overall equilibrium, since at $\lambda = 1$ the gain is negative. If it is at least one, the gain is nonnegative on $(0,1]$, so $\lambda = 1$ is the unique overall equilibrium. □

C.4 Proof of Proposition 3

Proof of Proposition 3. We first establish the limiting information shares, then derive every row of Table 1 from the acquisition condition. All moments below are under the common prior, as in Section 4.7.

Residual risk and the two complexity limits. Substituting $Z_* = c\sigma_I^2/\sigma_\beta^2$ and $z = \sigma_\varepsilon^2/\sigma_\beta^2$ into the quadratic of Lemma 2 shows that $\sigma_I^2(c)$ is the larger root of

$$\begin{aligned} 0 = (1+c\kappa)(\sigma_I^2)^2 &+ \left[\sigma_\beta^2 V_0(1-c) - \sigma_\varepsilon^2(2+c\kappa)\right]\sigma_I^2 \\ &- \sigma_\varepsilon^2(\sigma_\beta^2 V_0 - \sigma_\varepsilon^2). \end{aligned} \tag{34}$$

At $c = 0$, the polynomial factors as

$$(\sigma_I^2 - \sigma_\varepsilon^2)(\sigma_I^2 - \sigma_\varepsilon^2 + \sigma_\beta^2 V_0),$$

so its larger root is σ_ε^2 and is simple. For $c > 0$, the polynomial is negative at $\sigma_I^2 = \sigma_\varepsilon^2$, where it equals $-c\sigma_\varepsilon^2\sigma_\beta^2 V_0$. At $\sigma_I^2 = \sigma_\varepsilon^2 + \sigma_\beta^2 V_0/\kappa$, it equals $(1+\kappa)(\sigma_\beta^2 V_0/\kappa)^2 > 0$. Consequently,

$$\sigma_\varepsilon^2 < \sigma_I^2(c) < \sigma_\varepsilon^2 + \frac{\sigma_\beta^2 V_0}{\kappa}, \qquad c > 0,$$

and the polynomial has a strictly positive derivative at its larger root. Its partial derivative

with respect to c is

$$\kappa\sigma_I^2\left(\sigma_I^2-\sigma_\varepsilon^2-\frac{\sigma_\beta^2 V_0}{\kappa}\right)<0.$$

Implicit differentiation therefore gives $\partial_c\sigma_I^2>0$, including the right derivative $\partial_c\sigma_I^2(0)=\sigma_\varepsilon^2$. Dividing (34) by c and using the preceding bounds gives

$$\sigma_I^2(c)\longrightarrow\sigma_\varepsilon^2+\frac{\sigma_\beta^2 V_0}{\kappa}.$$

Since σ_d^2 is independent of c, $R_I^2(c)$ strictly decreases and the limits in (17) follow.

Acquisition costs and supply noise. Theorem 2 gives (16) and the equivalent identity

$$\lambda^2=\gamma^2\sigma_\eta^2\left[\frac{1-\zeta}{\zeta}\sigma_I^2-\frac{(\sigma_I^2)^2}{\sigma_d^2-\sigma_I^2}\right]. \tag{35}$$

The learning quantities σ_I^2 and σ_d^2 are independent of χ and σ_η. Thus $\partial_{\sigma_\eta}\lambda=\lambda/\sigma_\eta>0$. Since $\partial_\chi\zeta=2\gamma(1-\zeta)$, differentiation yields

$$\frac{\partial\lambda^2}{\partial\chi}=-\frac{2\gamma^3\sigma_\eta^2\sigma_I^2(1-\zeta)}{\zeta^2}<0.$$

These are the σ_η and χ claims in part (i).

Weak explainable variance at $c=0$. Using (17) in (35),

$$\lambda(0)^2=\gamma^2\sigma_\eta^2\sigma_\varepsilon^2\left[\frac{1-\zeta}{\zeta}-\frac{\sigma_\varepsilon^2}{A_s+A_w}\right],\qquad A_s+A_w=b'\psi b+\frac{\sigma_\beta^2 V_0}{\kappa}.$$

Since $\sigma_I^2(0)=\sigma_\varepsilon^2$ is independent of A_w, the right-hand side strictly increases with A_s+A_w, hence with A_w. This is the low-complexity A_w entry.

Strong explainable variance. At every complexity level, σ_I^2 is independent of A_s, while $\partial_{A_s}\sigma_d^2=1$. Differentiating (35) therefore gives

$$\frac{\partial\lambda^2}{\partial A_s}=\frac{\gamma^2\sigma_\eta^2(\sigma_I^2)^2}{(\sigma_d^2-\sigma_I^2)^2}>0.$$

This establishes the strong-signal claim in part (i).

Weak explainable variance at $c \to \infty$. In the other limit, (35) becomes

$$\lambda(\infty)^2 = \gamma^2 \sigma_\eta^2 \sigma_I^2(\infty) \left[\frac{1-\zeta}{\zeta} - \frac{\sigma_I^2(\infty)}{b'\psi b} \right], \qquad \sigma_I^2(\infty) = \sigma_\varepsilon^2 + \frac{\sigma_\beta^2 V_0}{\kappa}.$$

Let $A_s = b'\psi b > 0$. The margin is interior only if the bracket is positive, that is, $\sigma_I^2(\infty) < \frac{1-\zeta}{\zeta} A_s$, or $A_w < \frac{1-\zeta}{\zeta} A_s - \sigma_\varepsilon^2$. Holding A_s, γ, χ, and σ_η fixed, differentiation with respect to the limiting residual $\sigma_I^2(\infty)$ gives

$$\frac{\partial \lambda(\infty)^2}{\partial \sigma_I^2(\infty)} = \gamma^2 \sigma_\eta^2 \left[\frac{1-\zeta}{\zeta} - \frac{2\sigma_I^2(\infty)}{A_s} \right].$$

This is positive for $\sigma_I^2(\infty) < \frac{1-\zeta}{2\zeta} A_s$ and negative for $\sigma_I^2(\infty) > \frac{1-\zeta}{2\zeta} A_s$. Since $\sigma_I^2(\infty) = \sigma_\varepsilon^2 + A_w$, $\lambda(\infty)$ increases in A_w for $A_w < A_w^*$ and decreases for $A_w > A_w^*$, with $A_w^* = \frac{1-\zeta}{2\zeta} A_s - \sigma_\varepsilon^2$. At the peak, $R_I^2(\infty) = A_s/(A_s + \sigma_I^2(\infty)) = 2\zeta/(1+\zeta)$. This is the high-complexity A_w entry of Table 1.

Complexity. For the information share, holding σ_d^2 and ζ fixed, (16) gives

$$\frac{\partial \lambda^2}{\partial R_I^2} = \frac{\gamma^2 \sigma_\eta^2 \sigma_d^2}{\zeta} \left[\frac{\zeta}{(R_I^2)^2} - 1 \right].$$

Combining this derivative with $\partial_c R_I^2 < 0$ proves $\operatorname{sgn}(\partial_c \lambda) = \operatorname{sgn}(R_I^2 - \sqrt{\zeta})$, including the right derivative at zero. Since $R_I^2(c)$ decreases continuously from $R_I^2(0)$ to $R_I^2(\infty)$, participation is hump-shaped precisely when $R_I^2(\infty) < \sqrt{\zeta} < R_I^2(0)$. If $R_I^2(\infty) \geq \sqrt{\zeta}$, it increases throughout; if $R_I^2(0) \leq \sqrt{\zeta}$, it decreases throughout, on interior acquisition margins. This completes the verification of every row of Table 1. □

The proof also gives two facts that Proposition 3 and Table 1 do not state.

Additional observation: risk aversion. At fixed ζ, (35) gives $\partial_\gamma \lambda = \lambda/\gamma > 0$, since σ_I^2 and σ_d^2 are independent of γ. At fixed χ, however, $\partial_\gamma \zeta = 2\chi(1-\zeta)$ and

$$\frac{\partial \lambda^2}{\partial \gamma} = 2\gamma \sigma_\eta^2 \sigma_I^2 \left[\frac{1-\zeta}{\zeta} - \frac{\sigma_I^2}{\sigma_d^2 - \sigma_I^2} - \frac{\gamma\chi(1-\zeta)}{\zeta^2} \right].$$

This expression can have either sign at any fixed complexity level with informative signals. As $\gamma \downarrow 0$, (35) gives $\lambda^2 \sim \gamma \sigma_\eta^2 \sigma_I^2/(2\chi)$, so the response is positive for sufficiently small γ. At the entry boundary $\gamma = \log(\sigma_d^2/\sigma_I^2)/(2\chi)$, the first two terms in brackets cancel and the

third is strictly negative. The response is therefore negative just below that boundary. Both neighborhoods have interior acquisition, even if the population cap binds between them.

Additional observation: finite complexity. The A_w entries of Table 1 hold at the two complexity limits. At finite $c > 0$, by (34), $\sigma_I^2(c)$ depends on $\sigma_\beta^2 V_0$ and on κ separately, so λ is not a function of A_w alone. We therefore vary $\sigma_\beta^2 V_0$ at fixed κ, so that $dA_w = d(\sigma_\beta^2 V_0)/\kappa$. The root $\sigma_I^2(0)$ of (34) is simple, so the implicit function theorem makes $\sigma_I^2(c)$ jointly smooth in $(\sigma_\beta^2 V_0, \kappa, c)$ near $c = 0$. For the other endpoint, divide (34) by c and regard it as a polynomial in σ_I^2, $\sigma_\beta^2 V_0$, κ, and $1/c$. At $1/c = 0$ its positive root is $\sigma_I^2(\infty)$, and its derivative with respect to σ_I^2 is $\kappa\sigma_I^2(\infty) > 0$. The implicit function theorem gives joint smoothness in $(\sigma_\beta^2 V_0, \kappa, 1/c)$ near $1/c = 0$. By (35), the derivative of λ^2 with respect to $\sigma_\beta^2 V_0$ at fixed κ therefore converges, as $c \downarrow 0$ and as $c \to \infty$, to $1/\kappa$ times the limiting derivative with respect to A_w computed above. Hence, on interior acquisition margins and for $A_w \neq A_w^*$, the derivatives of λ with respect to σ_β^2 and to V_0 at fixed κ have the sign of the high-complexity A_w entry at all sufficiently large finite c, and the sign of the low-complexity entry at all sufficiently small $c > 0$. The same argument applies to κ at fixed $\sigma_\beta^2 V_0$, with the opposite sign: the limits depend on κ only through $A_w = \sigma_\beta^2 V_0/\kappa$, which decreases in κ. Finally, $\zeta < 2\zeta/(1+\zeta) < \sqrt{\zeta} < 1$ for $0 < \zeta < 1$, so at the high-complexity peak $R_I^2(\infty) < \sqrt{\zeta}$, and λ decreases in c at all sufficiently large finite c.

D Limiting information across scales

We define the limiting information sets $\mathcal{F}_q^\infty$ used in Section 5. The objective distribution and investors' common beliefs coincide under $\mathbb{P}_\infty$. Rather than define them directly as the limit of the information generated by the acquired signals, we define them through the forecasts: a larger model contains every smaller one, so an investor at scale q knows the forecast $M(u)$ that each model of scale $u \leq q$ would produce, and $\mathcal{F}_q^\infty$ is the information in this forecast path, together with the strong signals every informed investor observes. The finite-dimensional laws of the finite-economy forecasts converge to those of this forecast path (Lemma D.3). The forecasts are also easier to work with: the forecast path $q \mapsto M(q)$ is a diffusion in the scale coordinate (Appendix D.2). We first build this forecast path from the nested posteriors, then let $\mathcal{F}_q^\infty$ be the information it carries.

D.1 Proof of Lemma 4: pre-price residual risk

D.1.1 Observed and omitted signals

Partition $S = (S_1, S_2)$ and $\beta = (\beta_1, \beta_2)$ into N_1 observed and $N_2 = N - N_1$ omitted coordinates, and partition the history $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2]$ in the same way. Write $q_N := N_1/N$ and let $q_N \to q \in (0, 1]$. The information acquired at scale q in the finite economy is

$$\mathcal{F}_q^N := \sigma(\mathbf{s}, \mathbf{S}_1, \mathbf{d}, s, S_1). \tag{36}$$

We take the degrees of freedom $\nu_N := (1+\kappa)N$ to be an integer. Otherwise replace ν_N by its integer part: the exact formulas below then hold with κN replaced by $\nu_N - N$, and no limit changes, because $\nu_N/N \to 1 + \kappa$ in both cases. The following two auxiliary results supply the parameter transformation used in Lemma 4.

Lemma D.1 (Inverse-Wishart sub-block) *Partition $\Psi \sim \mathcal{W}_N^{-1}(V_0 N\, I_N,\, \nu_N)$ with leading $N_1 \times N_1$ block Ψ_1, and put $A_1 := \Psi_1^{-1}\Psi_{12}$ and $\Psi_{2\cdot 1} := \Psi_2 - \Psi_{21}\Psi_1^{-1}\Psi_{12}$. Then:*

(i) the marginal law of the block is inverse-Wishart, with degrees of freedom reduced by the codimension,

$$\Psi_1 \;\sim\; \mathcal{W}_{N_1}^{-1}\big(V_0 N\, I_{N_1},\, \nu_N - N_2\big), \qquad \nu_N - N_2 = N_1 + \kappa N;$$

equivalently, Ψ_1 has the prior law of Assumption 1 in dimension N_1 with (V_0, κ) replaced by $(V_0/q_N, \kappa/q_N)$;

(ii) $E[\Psi_1] = \frac{V_0 N}{\kappa N - 1} I_{N_1}$;

(iii) as $N \to \infty$, the empirical spectral distribution of Ψ_1 converges weakly, in probability, to the law $H_0^{V_0/q, \kappa/q}$ of Lemma I.6, and $\|\Psi_1\| \le \|\Psi\|$;

(iv) Ψ_1 is independent of $(A_1, \Psi_{2\cdot 1})$; conditional on $\Psi_{2\cdot 1}$, the rows of A_1 are independent $\mathcal{N}\big(0, \Psi_{2\cdot 1}/(V_0 N)\big)$ vectors; and $\Psi_{2\cdot 1} \sim \mathcal{W}_{N_2}^{-1}(V_0 N\, I_{N_2}, \nu_N)$.

Proof of Lemma D.1. The inverse-Wishart definition gives

$$\Omega_\Psi := \Psi^{-1} \sim \mathrm{Wishart}_N\big((V_0 N)^{-1} I_N, \nu_N\big),$$

where, for an integer $\nu_\Psi \ge N$, $\mathrm{Wishart}_N(\Sigma, \nu_\Psi)$ denotes the law of $\sum_{i=1}^{\nu_\Psi} g_i g_i'$ for i.i.d. $g_i \sim \mathcal{N}(0, \Sigma)$. Realize $\Omega_\Psi = G'G$ with $G \in \mathbb{R}^{\nu_N \times N}$ having i.i.d. $\mathcal{N}(0, (V_0 N)^{-1})$ entries, and split

its columns as $G = [\, G_1 \mid G_2 \,]$ with $G_1 \in \mathbb{R}^{\nu_N \times N_1}$ and $G_2 \in \mathbb{R}^{\nu_N \times N_2}$. The blocks of Ω_Ψ are $\Omega_{\Psi,1} = G_1'G_1$, $\Omega_{\Psi,12} = G_1'G_2$, and $\Omega_{\Psi,2} = G_2'G_2$. Let $P_2 := G_2(G_2'G_2)^{-1}G_2'$ be the orthogonal projection onto $\mathrm{col}(G_2)$, which has rank N_2 almost surely.

(i) Marginal law. By block inversion, the leading block of $\Psi = \Omega_\Psi^{-1}$ is

$$\Psi_1 = \big(\Omega_{\Psi,1\cdot 2}\big)^{-1}, \qquad \Omega_{\Psi,1\cdot 2} := \Omega_{\Psi,1} - \Omega_{\Psi,12}\Omega_{\Psi,2}^{-1}\Omega_{\Psi,21} = G_1'\big(I_{\nu_N} - P_2\big)G_1.$$

Choose $U \in \mathbb{R}^{\nu_N \times (\nu_N - N_2)}$, a measurable function of G_2, with orthonormal columns spanning $\mathrm{col}(G_2)^\perp$, so that $I_{\nu_N} - P_2 = UU'$. The matrices G_1 and G_2 are independent, and the columns of G_1 are independent $\mathcal{N}(0, (V_0N)^{-1}I_{\nu_N})$ vectors. Since $U'U$ is the identity, $H := U'G_1$ has i.i.d. $\mathcal{N}(0, (V_0N)^{-1})$ entries conditional on G_2. Thus, conditional on G_2, $\Omega_{\Psi,1\cdot 2} = H'H \sim \mathrm{Wishart}_{N_1}\big((V_0N)^{-1}I_{N_1},\, \nu_N - N_2\big)$. This law does not depend on G_2. It is therefore the unconditional law, and Ψ_1 is independent of G_2. Inversion gives (i). For the equivalent form, $V_0N = (V_0/q_N)N_1$ and $N_1 + \kappa N = (1 + \kappa/q_N)N_1$.

(ii) Mean. For $W \sim \mathrm{Wishart}_{N_1}(\Sigma, \nu_\Psi)$ with $\nu_\Psi > N_1 + 1$, $E[W^{-1}] = \Sigma^{-1}/(\nu_\Psi - N_1 - 1)$. With $\Sigma = (V_0N)^{-1}I_{N_1}$ and $\nu_\Psi = N_1 + \kappa N$,

$$E[\Psi_1] = E\big[\Omega_{\Psi,1\cdot 2}^{-1}\big] = \frac{V_0N\, I_{N_1}}{(N_1 + \kappa N) - N_1 - 1} = \frac{V_0N}{\kappa N - 1} I_{N_1}.$$

(iii) Spectrum. By (i), $\Psi_1 = V_0N\, W^{-1}$ with $W \sim \mathrm{Wishart}_{N_1}(I_{N_1}, \nu_N - N_2)$, and

$$\frac{N_1}{\nu_N - N_2} = \frac{q_N}{q_N + \kappa} \;\longrightarrow\; y_1 := \frac{q}{q + \kappa} \in (0,1), \qquad \frac{V_0N}{\nu_N - N_2} \;\longrightarrow\; \frac{V_0}{q + \kappa} = \frac{V_0/q}{1 + \kappa/q}.$$

The proof of Lemma I.6 uses the dimensions only through the limits of these two ratios. It therefore applies although q_N varies with N, and gives the limit law $H_0^{V_0/q, \kappa/q}$. We claim convergence only in probability, because no coupling across N is fixed here. Finally, Ψ_1 is a principal block of Ψ, so $\|\Psi_1\| \le \|\Psi\|$.

(iv) Independence. Block inversion also gives

$$A_1 = -\Omega_{\Psi,12}\Omega_{\Psi,2}^{-1} = -(P_2G_1)'G_2(G_2'G_2)^{-1}, \qquad \Psi_{2\cdot 1} = \Omega_{\Psi,2}^{-1} = (G_2'G_2)^{-1}.$$

Conditional on G_2, the matrices P_2G_1 and $(I_{\nu_N} - P_2)G_1$ are jointly Gaussian and uncorrelated, hence independent. So A_1 and Ψ_1 are independent conditional on G_2. The conditional law of Ψ_1 does not depend on G_2, so Ψ_1 is independent of (A_1, G_2) and therefore of $(A_1, \Psi_{2\cdot 1})$. Row i of A_1 is $-(G_2'G_2)^{-1}G_2'g_i$, where g_i is column i of G_1. Conditional on G_2, these rows

are independent centered Gaussian vectors with covariance

$$(V_0N)^{-1}(G_2'G_2)^{-1}G_2'G_2(G_2'G_2)^{-1} = \frac{\Psi_{2\cdot1}}{V_0N}.$$

This conditional law depends on G_2 only through $\Psi_{2\cdot1}$, so it is also the law conditional on $\Psi_{2\cdot1}$. Finally, $\Omega_{\Psi,2} = G_2'G_2 \sim \text{Wishart}_{N_2}\big((V_0N)^{-1}I_{N_2}, \nu_N\big)$, which gives the law of $\Psi_{2\cdot1}$. □

Lemma D.2 (Folding omitted signals into residual risk) *Let $V_0 > 0$ and $\Psi_{2\cdot1} = \Psi_2 - \Psi_{21}\Psi_1^{-1}\Psi_{12}$. The payoff has the exact conditional regression representation*

$$d = b's + \tilde{\beta}_1'S_1 + \varepsilon_1, \qquad \tilde{\beta}_1 = \beta_1 + \Psi_1^{-1}\Psi_{12}\beta_2, \tag{37}$$

where ε_1 is Gaussian and independent of S_1 conditional on the parameters, with variance $\sigma_\varepsilon^2 + R_{\text{om}}$ and $R_{\text{om}} = \beta_2'\Psi_{2\cdot1}\beta_2$. Let r_q be the value of σ_I^2 in Lemma 2 at the parameters

$$\begin{aligned} c_q &= qc, & \kappa_q &= \frac{\kappa}{q}, & V_{0,q} &= \frac{V_0}{q}, \\ \sigma_{\beta,q}^2 &= \sigma_\beta^2\frac{q(1+\kappa)}{\kappa+q}, & \sigma_{\varepsilon,q}^2 &= \sigma_\varepsilon^2 + \sigma_\beta^2 V_0\frac{1-q}{\kappa+q}. \end{aligned} \tag{38}$$

Then the posterior given the information $\mathcal{F}_q^N$ of (36), with mean $M_N(q) := E[d \mid \mathcal{F}_q^N]$, satisfies

$$\text{Var}[d \mid \mathcal{F}_q^N] \xrightarrow{p} r_q, \qquad E\Big|E\big[e^{itd} \mid \mathcal{F}_q^N\big] - e^{itM_N(q) - t^2r_q/2}\Big| \longrightarrow 0 \quad \text{for every } t \in \mathbb{R}. \tag{39}$$

The parameters (38) leave the total payoff variance σ_d^2 unchanged.

Proof of Lemma D.2. *Conditional regression.* Gaussian projection gives $S_2 = \Psi_{21}\Psi_1^{-1}S_1 + e_2$, where e_2 and S_1 are independent conditional on Ψ and $\text{Var}[e_2 \mid \Psi] = \Psi_{2\cdot1}$. Substitution proves (37), with $\varepsilon_1 = \beta_2'e_2 + \varepsilon$. The same decomposition applies independently to every historical observation conditional on the parameters.

The effective coefficient prior. With $A_1 = \Psi_1^{-1}\Psi_{12}$, the effective coefficient is $\tilde{\beta}_1 = \beta_1 + A_1\beta_2$. By Lemma D.1(iv), Ψ_1 is independent of $(A_1, \Psi_{2\cdot1})$, and conditional on $\Psi_{2\cdot1}$ the rows of A_1 are independent centered Gaussian vectors with covariance $\Psi_{2\cdot1}/(V_0N)$. Consequently, conditional on $(\beta_2, \Psi_{2\cdot1})$,

$$A_1\beta_2 \sim \mathcal{N}\left(0, \frac{R_{\text{om}}}{V_0N}I_{N_1}\right), \qquad \tilde{\beta}_1 \sim \mathcal{N}\left(0, \frac{\sigma_\beta^2 + R_{\text{om}}/V_0}{N}I_{N_1}\right).$$

These coefficients are independent of Ψ_1. After the regression, the law of the residual history depends on the omitted parameters only through R_{om} and is independent of the effective coefficient and observed signals conditional on R_{om}. Thus, conditional on this one scalar, the observed problem is exactly a Gaussian regression of the baseline form. Its coefficient prior is $\mathcal{N}\big(0, N^{-1}(\sigma_\beta^2 + R_{\mathrm{om}}/V_0) I_{N_1}\big)$, its noise variance is $\sigma_\varepsilon^2 + R_{\mathrm{om}}$, and its covariance prior is the law in Lemma D.1(i).

Concentration of omitted risk. Since $\Psi_{2\cdot 1} \preceq \Psi_2$ and Ψ_2 is a principal block of Ψ, we have $\|\Psi_{2\cdot 1}\| \le \|\Psi\|$. Let $\mathcal{E}_N := \{\|\Psi\| \le \Lambda + 1\}$, with Λ the edge bound of Lemma I.6; then $\mathbb{P}(\mathcal{E}_N) \to 1$. By Lemma D.1(iv), $\Psi_{2\cdot 1} \sim \mathcal{W}_{N_2}^{-1}(V_0 N\, I_{N_2}, \nu_N)$, with dimension ratio $N_2/\nu_N \to y_2 := (1-q)/(1+\kappa) < 1$ and $V_0 N/\nu_N \to V_0/(1+\kappa)$. Let $q < 1$. The proof of Lemma I.6, applied with these two ratios, shows that the empirical spectral distribution of $\Psi_{2\cdot 1}$ converges weakly, in probability, to the law of $\frac{V_0}{1+\kappa}\ell^{-1}$ with $\ell \sim \mathrm{MP}_{y_2}$. On $\mathcal{E}_N$ the spectrum of $\Psi_{2\cdot 1}$ lies in $[0, \Lambda + 1]$, so the first moments converge:

$$\frac{1}{N_2}\operatorname{tr}(\Psi_{2\cdot 1}) \xrightarrow{p} \frac{V_0}{1+\kappa} \cdot \frac{1}{1-y_2} = \frac{V_0}{\kappa + q}, \qquad \frac{1}{N}\operatorname{tr}(\Psi_{2\cdot 1}) \xrightarrow{p} \frac{(1-q)V_0}{\kappa + q}.$$

For $q = 1$ the second limit also holds, because $N^{-1}\operatorname{tr}(\Psi_{2\cdot 1}) \le (N_2/N)(\Lambda + 1)$ on $\mathcal{E}_N$. The vector $\sqrt{N}\beta_2 \sim \mathcal{N}(0, \sigma_\beta^2 I_{N_2})$ is independent of Ψ. Apply the L^2 bound of Lemma I.1 in dimension N_2 to the matrix $\Psi_{2\cdot 1}\mathbf{1}_{\mathcal{E}_N}$ with $C = \Lambda + 1$, and multiply it by $(N_2/N)^2$:

$$E\left[\left(R_{\mathrm{om}} - \frac{\sigma_\beta^2}{N}\operatorname{tr}(\Psi_{2\cdot 1})\right)^2 \mathbf{1}_{\mathcal{E}_N}\right] \le \frac{2(\Lambda+1)^2 \sigma_\beta^4 N_2}{N^2} \longrightarrow 0.$$

Therefore

$$R_{\mathrm{om}} \xrightarrow{p} r_{\mathrm{om}} := \sigma_\beta^2 V_0 \frac{1-q}{\kappa + q}.$$

Multiplying the prior variance of each effective coefficient by N_1 and replacing R_{om} by r_{om} gives $\sigma_{\beta,q}^2$ in the limit; adding r_{om} to the residual variance gives $\sigma_{\varepsilon,q}^2$. The covariance parameters in (38) follow from Lemma D.1(i) and (iii).

The conditional Gaussian kernel. The scalar R_{om} is latent to an investor who observes $\mathcal{F}_q^N$. Conditional on $(\mathcal{F}_q^N, R_{\mathrm{om}})$, Lemma 1 applies to the reduced regression and gives $d \sim$

$\mathcal{N}\big(m_N(R_{\mathrm{om}}), v_N(R_{\mathrm{om}})\big)$. Here, for $\rho \geq 0$,

$$m_N(\rho) = b's + S_1'\big(c_N\zeta(\rho)I_{N_1} + \hat{\Psi}_1\big)^{-1}\frac{\mathbf{S}_1'(\mathbf{d} - \mathbf{s}b)}{T},$$
$$v_N(\rho) = (\sigma_\varepsilon^2 + \rho)\Big[1 + \frac{1}{T}S_1'\big(c_N\zeta(\rho)I_{N_1} + \hat{\Psi}_1\big)^{-1}S_1\Big],$$

with $\hat{\Psi}_1 = \mathbf{S}_1'\mathbf{S}_1/T$, $c_N = N/T$, and $\zeta(\rho) = (\sigma_\varepsilon^2+\rho)/(\sigma_\beta^2+\rho/V_0)$. The function ζ is continuous and $\zeta(\rho) \geq \underline{\zeta} := \min\{z, V_0\} > 0$.

The statistic $v_N(r_{\mathrm{om}})$ depends on the data only through $(\mathbf{S}_1, S_1)$, whose law is the baseline law with covariance Ψ_1. Its ridge and dimension ratio satisfy $c_N\zeta(r_{\mathrm{om}}) \to c_q\sigma_{\varepsilon,q}^2/\sigma_{\beta,q}^2$ and $N_1/T \to c_q$. Lemma D.1(iii) supplies the spectral hypotheses of Appendix I with $H = H_0^{V_{0,q},\kappa_q}$, in probability and with $\|\Psi_1\| \leq \Lambda + 1$ on $\mathcal{E}_N$. This is sufficient: every subsequence has a further subsequence along which these hypotheses hold almost surely, and a sequence converges in probability if every subsequence has a further subsequence that does. The reduced parameters $(V_0/q_N, \kappa/q_N, q_Nc_N)$ vary with N, but the deterministic equivalents of Appendix I use the sequence only through the limits of the dimension ratio, the ridge, and the spectral distribution. The risk-concentration step in the proof of Lemma 2 therefore gives $v_N(r_{\mathrm{om}}) \xrightarrow{p} r_q$.

Next, the resolvent identity gives, for $\rho, \rho' \geq 0$,

$$\Big\|\big(c_N\zeta(\rho)I_{N_1} + \hat{\Psi}_1\big)^{-1} - \big(c_N\zeta(\rho')I_{N_1} + \hat{\Psi}_1\big)^{-1}\Big\| \leq \frac{|\zeta(\rho) - \zeta(\rho')|}{c_N\underline{\zeta}^2}.$$

The norm $\|\Psi_1\| \leq \|\Psi\|$ is bounded on $\mathcal{E}_N$. Also $\|\hat{\Psi}_1\| \leq \|\Psi_1\|\,\|X\|^2/T$ for a $T \times N_1$ matrix X of independent standard Gaussians, and $\|X\|^2/T$ is bounded in probability by the extreme-eigenvalue limit used in Lemma I.6 (Bai and Silverstein, 2010). The variables $\|S_1\|^2/T$ and $\|\mathbf{d}-\mathbf{s}b\|^2/T$ have bounded means on $\mathcal{E}_N$. Hence $v_N(R_{\mathrm{om}})-v_N(r_{\mathrm{om}}) \xrightarrow{p} 0$, because $R_{\mathrm{om}} \xrightarrow{p} r_{\mathrm{om}}$. For the mean, $m_N(R_{\mathrm{om}}) - m_N(r_{\mathrm{om}}) = S_1'w_N$, where w_N is a function of the history and the parameters with

$$\|w_N\| \leq \frac{|\zeta(R_{\mathrm{om}}) - \zeta(r_{\mathrm{om}})|}{c_N\underline{\zeta}^2}\,\|\hat{\Psi}_1\|^{1/2}\,\frac{\|\mathbf{d} - \mathbf{s}b\|}{\sqrt{T}} \xrightarrow{p} 0.$$

Conditional on the history and the parameters, $S_1 \sim \mathcal{N}(0, \Psi_1)$, so $S_1'w_N$ has conditional variance at most $\|\Psi_1\|\|w_N\|^2$. Therefore

$$v_N(R_{\mathrm{om}}) \xrightarrow{p} r_q, \qquad m_N(R_{\mathrm{om}}) - m_N(r_{\mathrm{om}}) \xrightarrow{p} 0. \tag{40}$$

Uniform integrability. The kernel variance is unbounded in R_{om}, so (40) alone does not control the event on which R_{om} is far from r_{om}. A fourth-moment bound does. Conditional on (β, Ψ), d is centered Gaussian with variance $b'\psi b + \beta'\Psi\beta + \sigma_\varepsilon^2$, and

$$E\big[(\beta'\Psi\beta)^2 \mid \Psi\big] = \frac{\sigma_\beta^4}{N^2}\big[(\operatorname{tr}\Psi)^2 + 2\operatorname{tr}(\Psi^2)\big] \le 3\sigma_\beta^4\Big(\frac{\operatorname{tr}\Psi}{N}\Big)^2.$$

By Lemma D.1(i) with a block of size one, $V_0 N/\Psi_{ii}$ has the chi-square law with $\kappa N + 1$ degrees of freedom, so $E[\Psi_{ii}^2] = (V_0 N)^2/[(\kappa N - 1)(\kappa N - 3)]$ for $\kappa N > 3$. Convexity gives $E[(\operatorname{tr}\Psi/N)^2] \le \max_i E[\Psi_{ii}^2]$. Hence $\sup_N E[d^4] < \infty$, where the supremum is over all large N. Since $m_N(R_{\mathrm{om}})$ and $v_N(R_{\mathrm{om}})$ are the conditional mean and variance of d given $(\mathcal{F}_q^N, R_{\mathrm{om}})$, convexity of $x \mapsto x^4$ and $x \mapsto x^2$ under conditional expectations also gives

$$E\big[m_N(R_{\mathrm{om}})^4\big] \le E[d^4], \qquad E\big[v_N(R_{\mathrm{om}})^2\big] \le E\big[(d - m_N(R_{\mathrm{om}}))^4\big] \le 16E[d^4].$$

Passing to the observed posterior. The observed posterior is a mixture of the kernels over R_{om}, so

$$\operatorname{Var}[d \mid \mathcal{F}_q^N] = E\big[v_N(R_{\mathrm{om}}) \mid \mathcal{F}_q^N\big] + \operatorname{Var}\big[m_N(R_{\mathrm{om}}) \mid \mathcal{F}_q^N\big], \qquad M_N(q) = E\big[m_N(R_{\mathrm{om}}) \mid \mathcal{F}_q^N\big].$$

By (40) and the second-moment bound, $v_N(R_{\mathrm{om}}) \to r_q$ in L^1. Conditional expectation is an L^1 contraction, so the first term converges to r_q in L^1. For the second term, fix $K > 0$ and let $\phi_K(x) = \max\{-K, \min\{x, K\}\}$. The variable $\phi_K(m_N(r_{\mathrm{om}}))$ is $\mathcal{F}_q^N$-measurable, and the conditional variance is the smallest mean squared distance to an $\mathcal{F}_q^N$-measurable variable. Since ϕ_K is a contraction with range $[-K, K]$,

$$\begin{aligned} E\big[\operatorname{Var}[m_N(R_{\mathrm{om}}) \mid \mathcal{F}_q^N]\big] &\le E\big[\big(m_N(R_{\mathrm{om}}) - \phi_K(m_N(r_{\mathrm{om}}))\big)^2\big] \\ &\le \frac{2E[d^4]}{K^2} + 2E\big[\min\big\{(m_N(R_{\mathrm{om}}) - m_N(r_{\mathrm{om}}))^2, 4K^2\big\}\big]. \end{aligned}$$

The last term tends to zero as $N \to \infty$ by (40) and bounded convergence. Letting $N \to \infty$ and then $K \to \infty$ gives

$$E\big[(m_N(R_{\mathrm{om}}) - M_N(q))^2\big] = E\big[\operatorname{Var}[m_N(R_{\mathrm{om}}) \mid \mathcal{F}_q^N]\big] \longrightarrow 0.$$

This proves the first claim in (39). For the second claim,

$$E\big[e^{itd} \mid \mathcal{F}_q^N\big] = E\big[e^{itm_N(R_{\mathrm{om}})-t^2 v_N(R_{\mathrm{om}})/2} \mid \mathcal{F}_q^N\big].$$

We have shown $m_N(R_{\mathrm{om}}) - M_N(q) \to 0$ in L^2 and $v_N(R_{\mathrm{om}}) \xrightarrow{p} r_q$. The integrand therefore differs from the $\mathcal{F}_q^N$-measurable variable $e^{itM_N(q)-t^2 r_q/2}$ by a term that is bounded by 2 and tends to zero in probability, hence in L^1. The contraction property gives the claim.

Finally,

$$\sigma_{\beta,q}^2 \frac{V_{0,q}}{\kappa_q} + \sigma_{\varepsilon,q}^2 = \sigma_\beta^2 \frac{V_0}{\kappa} + \sigma_\varepsilon^2.$$

Adding $b'\psi b$ proves conservation of total payoff variance. □

D.1.2 The residual-risk curve

We prove Lemma 4 and record the explicit root and derivatives used in the equilibrium characterization. The limit $R(q)$ of $\mathrm{Var}[d \mid \mathcal{F}_q^N]$ incorporates both omitted signals and estimation of the predictive loadings and signal covariance. Use the abbreviations

$$A_w = \frac{\sigma_\beta^2 V_0}{\kappa}, \qquad A_s = b'\psi b, \qquad a_R = 1 + c\kappa,$$
$$t_R = \kappa A_w - \sigma_\varepsilon^2, \quad \upsilon(q) = \frac{\kappa(1-q)}{\kappa+q}.$$

In particular, $\sigma_d^2 = A_s + A_w + \sigma_\varepsilon^2$. Define

$$b_R(q) = A_w[\kappa(1-c) - \upsilon(q)] - \sigma_\varepsilon^2(2 + c\kappa),$$
$$d_R(q) = -[\sigma_\varepsilon^2 + A_w \upsilon(q)]t_R, \qquad \Delta_R(q) = \sqrt{b_R(q)^2 - 4a_R d_R(q)}.$$

Applying the parameter transformation (38) to Lemma 2 gives the quadratic (22), with the larger root

$$R(q) = \frac{-b_R(q) + \Delta_R(q)}{2a_R}. \tag{41}$$

Thus the curve depends on $(c, \kappa, A_w, \sigma_\varepsilon^2)$; the known strong block affects total payoff variance but contributes no residual risk. Acquisition costs, risk aversion, supply noise, and the scale distribution do not enter (41).

Proof of Lemma 4. Let $q \in (0,1]$. By (39) in Lemma D.2, $\mathrm{Var}[d \mid \mathcal{F}_q^N] \xrightarrow{p} r_q$, where r_q is the limiting risk of Lemma 2 at the parameters (38). Substitute these parameters into the residual-risk quadratic (34) implied by Lemma 2. The quadratic coefficient is $1 + c_q\kappa_q = a_R$, and the identity $\sigma^2_{\beta,q}V_{0,q} - \sigma^2_{\varepsilon,q} = \kappa A_w - \sigma^2_\varepsilon = t_R$ gives the constant coefficient $d_R(q)$. The linear coefficient simplifies to $b_R(q)$, so r_q solves (22). Lemma 2 selects the larger root, so $r_q = R(q)$ in (41). At $q = 1$ the transformation is the identity, so $R(1) = \sigma^2_I(c)$.

Let $\mathcal{P}_q(x) = a_R x^2 + b_R(q)x + d_R(q)$ and $n(q) = \sigma^2_\varepsilon + A_w\upsilon(q)$. At zero its two roots are $\sigma^2_\varepsilon + A_w$ and $(\sigma^2_\varepsilon - \kappa A_w)/(1 + c\kappa)$, with the former strictly larger. For $q > 0$,

$$\mathcal{P}_q(n(q)) = -c\kappa A_w n(q)[1 - \upsilon(q)] < 0.$$

Hence the roots are distinct and the larger one satisfies $R(q) > n(q)$. The discriminant remains strictly positive on $[0,1]$, proving smoothness, including at zero. Viewing the root as a function of υ gives

$$\begin{aligned} R_\upsilon &= \frac{A_w(R + t_R)}{\Delta_R}, & R_{\upsilon\upsilon} &= \frac{2R_\upsilon(A_w - a_R R_\upsilon)}{\Delta_R}, \\ \upsilon' &= -\frac{\kappa(1+\kappa)}{(\kappa+q)^2}, & \upsilon'' &= \frac{2\kappa(1+\kappa)}{(\kappa+q)^3}, \\ R' &= R_\upsilon \upsilon', & R'' &= R_{\upsilon\upsilon}(\upsilon')^2 + R_\upsilon \upsilon''. \end{aligned} \tag{42}$$

Since $R + t_R \geq A_w(\upsilon + \kappa) > 0$, we have $R_\upsilon > 0$ and therefore $R' < 0$. The displayed formulas give every derivative used in Theorem 5 and its explicit distribution formulas. □

Smoothness, positive residual risk, and decreasing risk are therefore consequences of the primitives. The formula for R'' need not have a constant sign, so no concavity of $-\log R$ or of the scale objective is assumed. Holding $(\kappa, A_w, \sigma^2_\varepsilon)$ fixed, implicit differentiation also gives

$$\partial_c R(q) = \frac{\kappa R(q)[R(0) - R(q)]}{\Delta_R(q)} > 0 \qquad (q > 0).$$

Thus greater complexity raises residual risk at each positive scale.

D.2 The posterior-mean process

Use the curve R in (41), with $A_s = \sigma^2_d - R(0)$ and $j = -R'$. The index $q = 0$ records the strong-signal information common to subscribers; it is not the uninformed option. The

strong-signal vector $s \sim \mathcal{N}(0, \psi)$ has fixed dimension n and the same distribution for all N and T, so we retain the notation s in the limiting economy.

We first show that the finite-economy forecasts at different scales have asymptotically independent Gaussian increments. Fix scales $0 = q_0 < q_1 < \cdots < q_k \leq 1$ and integers $N_1(q_1) \leq \cdots \leq N_1(q_k) \leq N$ with $N_1(q_\ell)/N \to q_\ell$. The information sets $\mathcal{F}^N_{q_\ell}$ of (36) are then nested, with forecasts $M_N(q_\ell) = E[d \mid \mathcal{F}^N_{q_\ell}]$. Put $\mathcal{F}^N_0 := \sigma(s) \subseteq \mathcal{F}^N_{q_1}$ and $M_N(0) := b's$.

Lemma D.3 (Joint convergence of forecasts across scales) *Let* $\Delta_{N,\ell} = M_N(q_\ell) - M_N(q_{\ell-1})$ *for* $1 \leq \ell \leq k$ *and* $\Delta_{N,k+1} = d - M_N(q_k)$. *In the limit,* $(s, \Delta_{N,1}, \ldots, \Delta_{N,k+1})$ *converges in distribution to a vector with independent components* $s \sim \mathcal{N}(0, \psi)$, $\Delta_\ell \sim \mathcal{N}\big(0, R(q_{\ell-1}) - R(q_\ell)\big)$ *for* $1 \leq \ell \leq k$, *and* $\Delta_{k+1} \sim \mathcal{N}\big(0, R(q_k)\big)$.

Proof of Lemma D.3. *Single scales.* For $u \in \{q_0, \ldots, q_k\}$ and $t \in \mathbb{R}$, define

$$\varphi_{N,u}(t) := E\big[e^{itd} \mid \mathcal{F}^N_u\big] - e^{itM_N(u) - t^2 R(u)/2}.$$

We claim that $E[|\varphi_{N,u}(t)|] \to 0$. For $u > 0$ this is (39), because $r_u = R(u)$ by the proof of Lemma 4. For $u = 0$, the variable $d - b's = \beta' S + \varepsilon$ is independent of s and, conditional on (β, Ψ), centered Gaussian with variance $\beta'\Psi\beta + \sigma^2_\varepsilon$. Hence

$$\varphi_{N,0}(t) = e^{itb's}\Big(E\big[e^{-t^2(\beta'\Psi\beta + \sigma^2_\varepsilon)/2}\big] - e^{-t^2 R(0)/2}\Big).$$

The bracket tends to zero, because $\beta'\Psi\beta \xrightarrow{p} A_w$ by the proof of Lemma 2 and $R(0) = \sigma^2_\varepsilon + A_w$.

Increments. Let $u < u'$ be two consecutive scales. Since $\mathcal{F}^N_u \subseteq \mathcal{F}^N_{u'}$, iterated expectations give

$$e^{itM_N(u) - t^2 R(u)/2} + \varphi_{N,u}(t) = e^{-t^2 R(u')/2} E\big[e^{itM_N(u')} \mid \mathcal{F}^N_u\big] + E\big[\varphi_{N,u'}(t) \mid \mathcal{F}^N_u\big].$$

Multiply by $e^{-itM_N(u) + t^2 R(u')/2}$, which is $\mathcal{F}^N_u$-measurable and has modulus $e^{t^2 R(u')/2}$:

$$E\big[e^{it(M_N(u') - M_N(u))} \mid \mathcal{F}^N_u\big] = e^{-t^2(R(u) - R(u'))/2} + \tilde{\varphi}_{N,u}(t), \tag{43}$$

where

$$E[|\tilde{\varphi}_{N,u}(t)|] \leq e^{t^2 R(u')/2}\big(E[|\varphi_{N,u}(t)|] + E[|\varphi_{N,u'}(t)|]\big) \longrightarrow 0.$$

At the largest scale, the definition of φ_{N,q_k} gives directly

$$E\big[e^{it(d-M_N(q_k))} \mid \mathcal{F}^N_{q_k}\big] = e^{-t^2R(q_k)/2} + e^{-itM_N(q_k)}\varphi_{N,q_k}(t).$$

Induction. Fix $a \in \mathbb{R}^n$ and $t_1, \dots, t_{k+1} \in \mathbb{R}$. Let $\delta_\ell = R(q_{\ell-1}) - R(q_\ell)$ for $\ell \le k$ and $\delta_{k+1} = R(q_k)$, and put

$$\Phi_{N,\ell} := \exp\Big\{i\Big(a's + \sum_{m\le\ell} t_m \Delta_{N,m}\Big)\Big\}, \qquad 0 \le \ell \le k+1.$$

For $\ell \ge 1$, the variable $\Phi_{N,\ell-1}$ has modulus one and is $\mathcal{F}^N_{q_{\ell-1}}$-measurable. Conditioning on $\mathcal{F}^N_{q_{\ell-1}}$ and using (43) and the display after it gives

$$E[\Phi_{N,\ell}] = E\Big[\Phi_{N,\ell-1}E\big[e^{it_\ell\Delta_{N,\ell}} \mid \mathcal{F}^N_{q_{\ell-1}}\big]\Big] = e^{-t_\ell^2\delta_\ell/2}E[\Phi_{N,\ell-1}] + o(1).$$

Iterating from $\ell = k+1$ down to $\ell = 1$, and using $E[\Phi_{N,0}] = e^{-a'\psi a/2}$, we obtain

$$E[\Phi_{N,k+1}] \longrightarrow e^{-a'\psi a/2}\prod_{\ell=1}^{k+1} e^{-t_\ell^2\delta_\ell/2}.$$

The right-hand side is the characteristic function of the stated Gaussian vector. Pointwise convergence of characteristic functions proves convergence in distribution. □

The limiting economy imposes this increment structure at all scales simultaneously. Its initial forecast is $M(0) = G := b's \sim \mathcal{N}(0, A_s)$. Consider nested information sets $(\mathcal{G}_q)_{0\le q\le 1}$ with $\sigma(s) \subseteq \mathcal{G}_0$ and the conditional payoff laws

$$d^\infty \mid \mathcal{G}_q \sim \mathcal{N}(M(q), R(q)), \qquad M(q) = E^\infty[d^\infty \mid \mathcal{G}_q]. \tag{44}$$

These requirements determine how the forecasts at different scales are related.

Proposition D.4 (Forecast innovations) *Suppose the nested information sets satisfy* (44)*, with $M(0) = G$ and $V_0 > 0$. For $q < q'$, the innovation $M(q') - M(q)$ is independent of $\mathcal{G}_q$ and has law $\mathcal{N}(0, R(q) - R(q'))$. The process has a continuous version satisfying*

$$M(q) = G + \int_0^q \sqrt{j(u)}\, dW_u, \qquad d^\infty = M(1) + \sqrt{R(1)}\, Z, \tag{45}$$

where W is a standard Brownian motion, independent of $\mathcal{G}_0$, and Z is standard normal, independent of $\mathcal{G}_1$. In particular, the joint law of the payoff and forecast process is determined by A_s and R.

Proof of Proposition D.4. For $t \in \mathbb{R}$, (44) gives $E^\infty[e^{itd^\infty} \mid \mathcal{G}_q] = e^{itM(q)-t^2R(q)/2}$. Nesting and iterated expectations imply

$$e^{itM(q)-t^2R(q)/2} = e^{-t^2R(q')/2}E^\infty[e^{itM(q')} \mid \mathcal{G}_q].$$

Multiplying by $e^{-itM(q)+t^2R(q')/2}$ yields

$$E^\infty[e^{it(M(q')-M(q))} \mid \mathcal{G}_q] = e^{-t^2(R(q)-R(q'))/2}.$$

This deterministic characteristic function proves normality and independence from $\mathcal{G}_q$; successive forecast increments are therefore independent. Since j is bounded,

$$E^\infty[|M(q') - M(q)|^4] = 3(R(q) - R(q'))^2 \leq 3\|j\|_\infty^2|q' - q|^2.$$

The continuity criterion gives a continuous version, with quadratic variation $[M]_q = \int_0^q j(u)\,du$. The process

$$W_q := \int_0^q j(u)^{-1/2}\,dM(u)$$

has independent centered Gaussian increments of variance equal to the increment in q, so it is Brownian and independent of $\mathcal{G}_0$. Finally, the posterior at $q = 1$ implies that $Z = (d^\infty - M(1))/\sqrt{R(1)}$ is standard normal and independent of $\mathcal{G}_1$. □

The forecast process is a diffusion in the scale coordinate, $dM(q) = \sqrt{j(q)}\,dW_q$, similar to the continuous and heterogeneous information (CHI) structure of Avdis and Glebkin (2026). Here $j(q)$ is the rate at which acquiring additional signals reduces posterior variance. The derivation uses Gaussian posteriors conditional on the nested information sets; it does not require a separate assumption of jointly Gaussian forecasts.

D.3 Information generated by the forecasts

For the nested information sets above, define the information generated by the observed forecast path as

$$\mathcal{G}_q^M := \sigma\big(s, M(u) : 0 \le u \le q\big) \subseteq \mathcal{G}_q.$$

Take the supply shock $\eta \sim \mathcal{N}(0,1)$ independent of $(d^\infty, \mathcal{G}_1)$. The forecast path retains all information in $\mathcal{G}_q$ about the payoff, the other forecasts, and the supply shock: writing M for the entire continuous path,

$$\mathcal{L}(d^\infty, s, M, \eta \mid \mathcal{G}_q) = \mathcal{L}(d^\infty, s, M, \eta \mid \mathcal{G}_q^M). \tag{46}$$

To see this, the strong signals and the past forecast path are already $\mathcal{G}_q^M$-measurable. Proposition D.4 implies that the entire process of future increments $(M(u) - M(q))_{q \le u \le 1}$ is independent of $\mathcal{G}_q$. This follows first for finite collections of increments and then for the continuous path. The terminal residual $\sqrt{R(1)}Z$ is independent of $\mathcal{G}_1$, and η is independent of both the acquired information and the payoff. Thus the conditional law on the left of (46) depends on $\mathcal{G}_q$ only through the observed strong signals and forecast path.

In particular, for any candidate price P measurable with respect to $\mathcal{G}_1^M \vee \sigma(\eta)$, disintegration of (46) gives

$$\mathcal{L}(d^\infty \mid \mathcal{G}_q \vee \sigma(P)) = \mathcal{L}(d^\infty \mid \mathcal{G}_q^M \vee \sigma(P)).$$

The two information sets therefore produce the same conditional trading problem for every such price. Informed demands then depend on $\mathcal{G}_q$ only through $\mathcal{G}_q^M$ and P, and market clearing aggregates these demands and the supply shock. Restricting the price to be measurable with respect to the forecast paths of the acquired scales and η is therefore self-consistent, and we impose it. The entire past path matters here: the current forecast $M(q)$ alone need not reveal the smaller-scale forecasts entering prices. This is the sense in which acquired information is equivalent to the information generated by the forecasts.

We now construct this representation and use it to define the limiting economy. On $\Omega = \mathbb{R}^n \times C_0([0,1]) \times \mathbb{R}^2$, take independent coordinates

$$s \sim \mathcal{N}(0, \psi), \qquad W = (W_q)_{0 \le q \le 1} \text{ a standard Brownian motion}, \qquad Z, \eta \sim \mathcal{N}(0,1).$$

Here $C_0([0,1])$ is the space of continuous paths starting at zero; $\mathbb{P}_\infty$ is the product of the indicated Gaussian laws and Wiener measure, and $\mathcal{A}$ is its completed Borel sigma-field. Set $G = b's$, define M and d^∞ by (45), and define the acquired information by

$$\mathcal{F}_q^\infty := \sigma\big(s, M(u) : 0 \le u \le q\big), \qquad 0 \le q \le 1, \tag{47}$$

completed under $\mathbb{P}_\infty$. Thus knowing $\mathcal{F}_q^\infty$ means knowing the strong signals and all forecasts up to scale q. When $j > 0$, it is equivalently $\sigma(s, W_u : 0 \le u \le q)$, since the deterministic transformation in (45) can be inverted. The information sets are nested and continuous from below at every positive scale, and the supply shock is independent of the entire acquired information family and the payoff.

The construction satisfies the required posterior laws. Indeed,

$$d^\infty - M(q) = \int_q^1 \sqrt{j(u)}\, dW_u + \sqrt{R(1)}\, Z$$

is independent of $\mathcal{F}_q^\infty$ and Gaussian with variance $\int_q^1 j(u)\, du + R(1) = R(q)$. Consequently,

$$d^\infty \mid \mathcal{F}_q^\infty \sim \mathcal{N}(M(q), R(q)), \qquad M(q) = E^\infty[d^\infty \mid \mathcal{F}_q^\infty]. \tag{48}$$

Moreover,

$$\mathrm{Var}^\infty[d^\infty] = \sigma_d^2, \qquad \mathrm{Cov}^\infty[M(q), M(q')] = \sigma_d^2 - R(\min\{q, q'\}).$$

This supplies a common payoff and a jointly specified family of information sets for every feasible scale. By Lemma D.3 and the continuous mapping theorem, the finite-economy vector $(s, M_N(q_1), \ldots, M_N(q_k), d)$ converges in distribution to $(s, M(q_1), \ldots, M(q_k), d^\infty)$ for all scales $0 < q_1 < \cdots < q_k \le 1$. At $q = 1$, it has the payoff–forecast decomposition of Appendix B. Equation (47) is the definition of acquired information in the limiting economy; the equilibrium conditions use this information together with the observed price.

D.4 Equilibrium conventions

The scale distribution F is a probability distribution on $(0,1]$, extended to a right-continuous CDF on $[0,1]$ by $F(0) = 0$ and $F(1) = 1$. It may have atoms and has no atom at zero. The informed fraction λ is separate; the uninformed have mass $1 - \lambda$ and acquire no strong or weak signals. Integrals against dF are Lebesgue–Stieltjes integrals. When $\lambda = 0$, F is

an arbitrary unused conditional distribution and imposes no restrictions on the values of unchosen scales.

For $\lambda > 0$, let

$$q_H := \inf\{q \in [0,1] : F(q) = 1\}.$$

The information collectively acquired by positive population shares is

$$\bigvee_{\substack{0<q\le 1\\ 1-F(q-)>0}} \mathcal{F}_q^\infty = \mathcal{F}_{q_H}^\infty,$$

where the equality follows from nesting and continuity from below. Each $\mathcal{F}_q^\infty$ is generated by the strong signals and the forecast path, which is the self-consistent restriction imposed in Appendix D.3. The price restriction in Definition 3 is therefore

$$P \text{ is } \begin{cases} \mathcal{F}_{q_H}^\infty \vee \sigma(\eta)\text{-measurable}, & \lambda > 0, \\ \sigma(\eta)\text{-measurable}, & \lambda = 0. \end{cases}$$

The schedules $(q, p, \omega) \mapsto x_q(p, \omega)$ are jointly measurable and, when $\lambda > 0$,

$$\int_{(0,1]} |x_q(P)|\, dF(q) < \infty \qquad \mathbb{P}_\infty\text{-almost surely.}$$

The uninformed schedule x_U is Borel measurable. When $\lambda = 0$, the informed population integral is defined to be zero. Market clearing holds $\mathbb{P}_\infty$-almost surely.

All conditional beliefs are regular conditional distributions under the same $\mathbb{P}_\infty$ and the actual candidate price function P. For each fixed scale, informed optimality holds almost surely in the observed information and price; uninformed optimality holds almost surely in the observed price. Joint measurability then makes these conditions valid for population aggregation, while imposing them at every feasible scale defines all deviation values. Before conditioning on the price, we select the Gaussian kernel in (48) as the posterior version. Conditioning further on an arbitrary candidate price is defined almost surely; requiring optimality also at null information–price pairs would require an additional choice of versions there.

The acquisition cost of Assumption 2 is finite, continuous, and nonnegative on $(0, 1]$. Since holding zero shares is feasible, the optimized conditional trading utility lies in $[-1, 0]$; hence

each ex-ante value in Definition 4 is well defined and finite. These values are deterministic, with the equilibrium price rule implicit in their expectations. Formally, Definition 4 requires a common optimal value $\bar{V}$ such that

$$\begin{aligned} V(q) &\le \bar{V} && \text{for every } q \in (0,1], \\ \lambda\,[\bar{V} - V(q)] &= 0 && \text{for } dF\text{-almost every } q, \\ V_U \le \bar{V}, &\qquad (1-\lambda)(\bar{V} - V_U) = 0. \end{aligned}$$

Thus F assigns probability one to optimal scales whenever $\lambda > 0$; when $\lambda = 0$, the conditions reduce to $V(q) \le V_U$ for every q. If the acquisition menu contains only one scale, the definitions reduce to Definitions 1 and 2.

E Past prices and the birth of scaling

The history (3) records signals and payoffs and contains no prices. This appendix shows when the omission is without loss. In the single-scale economy of Sections 3 and 4 it always is: an informed investor holds every signal that entered a past price, so a past price can add only the supply shock of its date, which is independent of everything else. With heterogeneous scales the question has content. Definition 3 lets an investor at scale $q < q_H$ learn from the current price because the price reflects the forecasts of larger models. A price formed at a past date with scaled trading reflects the larger models' forecasts at that date, so an investor at scale q could use it as a noisy observation of the contribution of the omitted signals to that date's payoff. Such observations are information about the qN-dimensional effective coefficient $\tilde{\beta}_1$ of Lemma D.2, and T of them are as many as the history itself. Under $N/T \to c$ their effect on residual risk can be first order, and a model in which investors learn from them must solve a sequence of past equilibria with a time-varying scale distribution.

We instead date the scaling technology. Investors could scale only after the last history date. Before it, prices reflected the strong signals, which every informed investor observes, and noise trading, so excluding them from the history is exact (Appendix E.1). Appendix E.2 relaxes the timeline and shows that past prices are negligible in the limit whenever scaled trading in the history was rare, because few dates had it or because few investors took part. The reason is that the price history carries a bounded amount of information about a coefficient with qN directions, of which the forecast of d uses one. A past price is a signal plus an independent supply shock, so what it can teach an investor at scale q about $\tilde{\beta}_1$ is bounded by its signal-to-noise ratio, and that ratio is governed by the loadings on

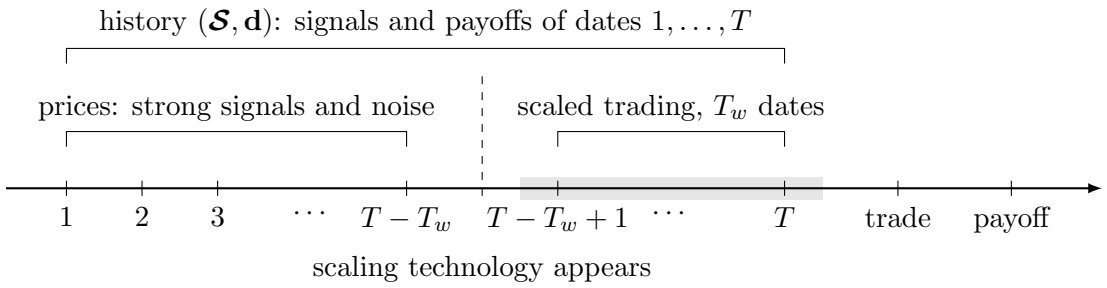

Figure 7: Timeline of the data and of prices. The axis shows the history dates $1, \ldots, T$, the trading date, and the payoff date. The history available at the trading date consists of the signals and payoffs of dates $1, \ldots, T$ and contains no prices. Before the scaling technology appears, prices reflect the strong signals and noise trading only. At the T_w shaded dates after it, investors trade on scaled models and prices reflect their forecasts. Assumption E.1 is the case $T_w = 0$, under which excluding past prices from the history is exact (Proposition E.2). Proposition E.6 shows that the prices of the shaded dates are negligible in the limit when $T_w/T \to 0$, or when few investors had scaled at those dates.

larger models' forecasts, which vanish with the mass of investors that had scaled beyond q at that date, by (58). The current signal vector is drawn independently of the past, so the investor's forecast of d loads on a fresh direction of the coefficient, and past prices with summed signal-to-noise ratios of order k move it by $O(k/N)$ in mean square. Past prices are therefore negligible when their summed signal-to-noise ratios are $o(N)$, for the same reason that a history of $o(N)$ dates teaches nothing about the N weak coefficients in the limit, and can be first order when they are of order T, as in the repeated economy above. Figure 7 shows the timeline.

E.1 The scaling technology is born after the history

Consider a past date $\tau \in \{1, \ldots, T\}$ at which the economy traded, with the signals of dates $1, \ldots, \tau$ and the payoffs of dates $1, \ldots, \tau - 1$ then available, and let P_τ be the price that cleared the market. We do not model these past equilibria. We only use that a price is a function of the information collectively acquired at its date and of that date's supply shock, as in Definitions 1 and 3, and we state what that information was. Write $\mathcal{P} := \sigma(P_1, \ldots, P_T)$ for the price history and $\eta_1, \ldots, \eta_T$ for the past supply shocks, which are i.i.d. $\mathcal{N}(0, 1)$ and independent of the parameters, of the signals and residuals of all dates, and of the current supply shock η.

Assumption E.1 (Birth of scaling) *For every $\tau = 1, \ldots, T$, the price P_τ is $\sigma(s_\tau, \eta_\tau)$-measurable.*

This is the price of a date at which every informed investor traded on the strong signals alone, with (b, ψ) and the residual risk of the strong-signal forecast known, so that demands and the price depended on the data only through s_τ, and the uninformed learned from the price. It also covers dates without informed trading, at which P_τ is a function of η_τ alone. A date at which investors instead estimated that residual risk from past payoffs is covered by Assumption E.3 below, with a zero contribution to the budget of Lemma E.5, because its price signal is a function of strong signals and payoffs. Under the assumption the scaling technology appears after date T, and the trading date is the first on which models of scale $q > 0$ trade. The data record is unaffected: signals and payoffs were recorded all along.

Proposition E.2 (Past prices are uninformative) *Let the current price P depend only on the information collectively acquired and the supply shock η, as in Definition 3. Under Assumption E.1, at every finite (N, T) and every scale $q \in (0, 1]$, the conditional law of d given $\mathcal{F}_q^N \vee \sigma(P) \vee \mathcal{P}$ equals its conditional law given $\mathcal{F}_q^N \vee \sigma(P)$, and the conditional law of d given $(P, \mathcal{P})$ equals its conditional law given P.*

Proof of Proposition E.2. Every acquired information set (36) contains the strong-signal history $\mathbf{s}$, so $\mathcal{P} \subseteq \mathcal{F}_q^N \vee \sigma(\eta_1, \ldots, \eta_T)$ for every $q > 0$. The past supply shocks are independent of $(d, \mathcal{F}_q^N, P)$, which proves the first claim. For the uninformed, $\mathcal{P} \subseteq \sigma(\mathbf{s}, \eta_1, \ldots, \eta_T)$. The current price clears the market, so it is a function of η and of the variables entering the demands, and every demand depends on the history only through the posterior of (β, Ψ), which is a function of $(\mathbf{S}, \mathbf{d} - \mathbf{s}b)$ because b is known (Lemma 1). Hence (d, P) is a function of $(\beta, \Psi, \mathbf{S}, \mathbf{d} - \mathbf{s}b, s, S, \varepsilon, \eta)$, where $\mathbf{d} - \mathbf{s}b = \mathbf{S}\beta + \boldsymbol{\varepsilon}$. This vector is independent of $(\mathbf{s}, \eta_1, \ldots, \eta_T)$, because the strong signals are Gaussian, uncorrelated with the weak signals, and independent of the residuals, and the past supply shocks are independent of everything else. So $\mathcal{P}$ is independent of (d, P), which proves the second claim. □

Under this timeline, excluding prices from the history is without loss at every scale and for the uninformed, exactly and not only in the limit. The one-period model is the complete description of the first trading date after the technology appears.

E.2 Scaled trading in the history

Assumption E.1 makes the trading date the first with scaled trading. This subsection allows scaled trading in the history and shows that its prices are negligible in the limit when, in

total, they carry little about the omitted signals. For an investor at scale q, the price history re-weights the posterior of the effective coefficient $\tilde{\beta}_1$ by the likelihood of the observed prices. We show three things. Lemma E.4 expresses the conditional law of d, with and without the price history, through the two posteriors of $\tilde{\beta}_1$ and the current observed signals, which are independent of the past. Lemma E.5 bounds the expected log-likelihood ratio between the two posteriors by the summed signal-to-noise ratios of the past prices. Lemma I.2 in Appendix I bounds what a history can teach about a prediction: a history that carries information $o(N)$ about a coefficient with N_1 independent directions of variance $O(1/N)$ carries information $o(1)$ about a noisy prediction on an independent direction. Thus the two conditional laws of d converge to each other, the Gaussian limit of Lemma D.2 holds with the price history included, and the forecast increments of Lemma D.3 are asymptotically independent of the price history (Proposition E.6).

Assumption E.3 (Price history) *For every $\tau = 1, \dots, T$,*

$$P_\tau = a_\tau (X_\tau - \eta_\tau),$$

where the price signal X_τ is square integrable and measurable with respect to the signals of dates $u \le \tau$, the payoffs of dates $u < \tau$, and the prices of dates $u < \tau$, and the loading $a_\tau \neq 0$ is measurable with respect to the strong signals of dates $u \le \tau$, the payoffs of dates $u < \tau$, and the prices of dates $u < \tau$.

The form is that of the normalized price (56) at each past date, with the coefficient of the supply shock normalized to one. The loading a_τ is otherwise free; it may depend on the strong signals and payoffs of earlier dates, which every acquired information set contains, as when past investors scaled their demands by an estimate of residual risk. It contains Assumption E.1 for prices affine in the supply shock, with X_τ a function of s_τ. The price signal may be any function of the data available at date τ and of earlier prices, so the scale distribution may differ across dates, and nothing about past equilibria is used beyond measurability and the second moment. For an investor at scale q, the size of a past price signal is $\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]$, the variance of the part the investor cannot reproduce from its own data. When the date-τ price has the linear form of Appendix F with loadings k_τ, the part that an investor at scale q could not reproduce at date τ is $\int_q^1 k_\tau(s)\, dM_\tau(s)$, with variance $v_\tau(q) - 1$ in (57). The investor at the trading date holds more data, so by the law of total variance $E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]$ is at most $v_\tau(q) - 1$, which vanishes with the mass of investors at scales above q, by (58).

Fix $q \in (0,1]$ and write $\mathcal{F}_q^{N,\mathcal{P}} := \mathcal{F}_q^N \vee \mathcal{P}$ for the information of the investor at scale q augmented with the price history, S_1 for the current observed signals, Ψ_1 for their covariance, and $v_q := \sigma_\varepsilon^2 + R_{\mathrm{om}}$. Conditional on $(\mathcal{F}_q^N, R_{\mathrm{om}})$, the effective coefficient has the Gaussian posterior of Lemma 1 applied to the reduced regression in the proof of Lemma D.2,

$$\tilde{\beta}_1 \mid (\mathcal{F}_q^N, R_{\mathrm{om}}) \sim \mathcal{N}(\hat{\beta}_1, \hat{\Sigma}_1), \qquad \hat{\Sigma}_1^{-1} = \frac{N}{\tilde{\sigma}_q^2} I_{N_1} + \frac{\mathbf{S}_1'\mathbf{S}_1}{v_q}, \qquad \hat{\beta}_1 = \hat{\Sigma}_1 \frac{\mathbf{S}_1'(\mathbf{d} - \mathbf{s}b)}{v_q}, \tag{49}$$

with $\tilde{\sigma}_q^2 := \sigma_\beta^2 + R_{\mathrm{om}}/V_0$. Every eigenvalue of $\hat{\Sigma}_1$ is at most $\tilde{\sigma}_q^2/N$, because $\mathbf{S}_1'\mathbf{S}_1$ is nonnegative definite. Write $\pi := \mathcal{N}(\hat{\beta}_1, \hat{\Sigma}_1)$, and let π' be the conditional law of $\tilde{\beta}_1$ given $(\mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}})$.

Lemma E.4 (Reduction to the observed block) *Let Assumption E.3 hold and fix $q \in (0,1]$. Let $\mathcal{H}$ be the sigma-field generated by the history $(\mathbf{s}, \mathbf{S}_1, \mathbf{d})$, the current strong signals s, the omitted risk R_{om}, and the block Ψ_1.*

(i) Conditional on $\mathcal{H}$, the current observed signals satisfy $S_1 \sim \mathcal{N}(0, \Psi_1)$ and are independent of $(\tilde{\beta}_1, \mathcal{P})$; the current residual ε_1 of (37) is $\mathcal{N}(0, v_q)$ and independent of $(\tilde{\beta}_1, \mathcal{P}, S_1)$; and the conditional law of $\tilde{\beta}_1$ is π given $\mathcal{H}$ and π' given $\mathcal{H} \vee \mathcal{P}$.

(ii) The conditional law of d given $(\mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}})$ is the law of $b's + \theta' S_1 + \varepsilon_1$ with $\theta \sim \pi'$ and $\varepsilon_1 \sim \mathcal{N}(0, v_q)$ independent, and the conditional law of d given $(\mathcal{F}_q^N, R_{\mathrm{om}})$ is the same with π in place of π'. These are also the conditional laws of d given $(\mathcal{H}, S_1, \mathcal{P})$ and given $(\mathcal{H}, S_1)$.

Proof of Lemma E.4. Let $\mathcal{Y}$ collect the parameters β, $A_1 = \Psi_1^{-1}\Psi_{12}$, and $\Psi_{2\cdot1}$, the residuals ε_u and $e_u := S_{2,u} - A_1' S_{1,u}$ of the omitted signals on the observed ones at all dates u, including the current one, and the supply shocks η_u of the past dates. The vector $(\Psi_1, \mathbf{S}_1, S_1)$ is independent of $(\mathcal{Y}, \mathbf{s}, s)$: by Lemma D.1(iv) and the prior, Ψ_1 is independent of $(\beta, A_1, \Psi_{2\cdot1})$; conditional on the parameters, the residuals e_u are $\mathcal{N}(0, \Psi_{2\cdot1})$ and independent of the observed signals, which are i.i.d. $\mathcal{N}(0, \Psi_1)$ given Ψ_1; and the strong signals, the payoff residuals, and the supply shocks are independent of everything else. The payoffs, the prices, and R_{om} are functions of $(\mathbf{S}_1, \mathbf{s}, s)$ and $\mathcal{Y}$: the payoffs by (37), since $\tilde{\beta}_1 = \beta_1 + A_1\beta_2$ and the residual $\varepsilon_{1,u} = \beta_2' e_u + \varepsilon_u$ are functions of $\mathcal{Y}$; the prices by Assumption E.3, since the omitted signals are $A_1' S_{1,u} + e_u$ and the loadings are functions of strong signals, payoffs, and earlier prices; and $R_{\mathrm{om}} = \beta_2' \Psi_{2\cdot1} \beta_2$. Let $\mathcal{H}_0$ be the sigma-field generated by $(\mathbf{S}_1, \mathbf{s}, s)$, the payoffs,

and R_{om}, and let $\mathcal{H}_0^{\mathcal{P}} := \mathcal{H}_0 \vee \mathcal{P}$, so that

$$\mathcal{H} = \mathcal{H}_0 \vee \sigma(\Psi_1), \qquad \sigma(\mathcal{F}_q^N, R_{\text{om}}) = \mathcal{H}_0 \vee \sigma(S_1),$$
$$\mathcal{H} \vee \mathcal{P} = \mathcal{H}_0^{\mathcal{P}} \vee \sigma(\Psi_1), \qquad \sigma(\mathcal{F}_q^{N,\mathcal{P}}, R_{\text{om}}) = \mathcal{H}_0^{\mathcal{P}} \vee \sigma(S_1).$$

Conditional independence survives conditioning on functions of the variables involved: if U is independent of V given Z, then U is independent of V given $(Z, h(Z,V))$ for every measurable h, because for bounded f both $E[f(U) \mid Z, h(Z,V), V] = E[f(U) \mid Z, V]$ and its conditional expectation given $(Z, h(Z,V))$ equal $E[f(U) \mid Z]$. With $U = (\Psi_1, S_1)$, $V = \mathcal{Y}$, and $Z = (\mathbf{S}_1, \mathbf{s}, s)$, the pair (Ψ_1, S_1) is independent of $\mathcal{Y}$ given $\mathcal{H}_0$ and given $\mathcal{H}_0^{\mathcal{P}}$, so S_1 is independent of $\mathcal{Y}$ given $\mathcal{H}$ and given $\mathcal{H} \vee \mathcal{P}$. Since $\tilde{\beta}_1$ is a function of $\mathcal{Y}$, its conditional law is the same given $\mathcal{H}_0$, given $\mathcal{H}$, and given $\mathcal{H}_0 \vee \sigma(S_1)$, where it is π by (49); likewise it is the same given $\mathcal{H}_0^{\mathcal{P}}$, given $\mathcal{H} \vee \mathcal{P}$, and given $\mathcal{H}_0^{\mathcal{P}} \vee \sigma(S_1)$, where it is π' by definition. Since $(\tilde{\beta}_1, \mathcal{P})$ is a function of $\mathcal{Y}$ and of the generators of $\mathcal{H}_0$, the variable S_1 is independent of $(\tilde{\beta}_1, \mathcal{P})$ given $\mathcal{H}$. Also, S_1 is $\mathcal{N}(0, \Psi_1)$ and independent of $(\mathbf{S}_1, \mathbf{s}, s, \mathcal{Y})$ given Ψ_1, and $\mathcal{H} \subseteq \sigma(\Psi_1, \mathbf{S}_1, \mathbf{s}, s, \mathcal{Y})$, so the conditional law of S_1 given $\mathcal{H}$ is $\mathcal{N}(0, \Psi_1)$. Finally, the current residual $\varepsilon_1 = \beta_2' e + \varepsilon$ is $\mathcal{N}(0, v_q)$ conditional on the parameters, on the signals, residuals, and supply shocks of the past dates, and on the current signals (s, S_1), by Lemma D.2; $\mathcal{H}$, $\tilde{\beta}_1$, $\mathcal{P}$, and S_1 are generated by functions of these variables and v_q is $\mathcal{H}$-measurable, so ε_1 is $\mathcal{N}(0, v_q)$ and independent of $(\tilde{\beta}_1, \mathcal{P}, S_1)$ given $\mathcal{H}$. This proves (i).

For (ii), write $d = b's + \tilde{\beta}_1' S_1 + \varepsilon_1$ as in (37). Conditional on $(\mathcal{H}, S_1, \mathcal{P})$, the term $b's$ is known, $\tilde{\beta}_1 \sim \pi'$ because by (i) S_1 is independent of $\tilde{\beta}_1$ given $\mathcal{H} \vee \mathcal{P}$, and $\varepsilon_1 \sim \mathcal{N}(0, v_q)$ is independent of $\tilde{\beta}_1$. This conditional law is a function of (s, S_1, π', v_q), hence measurable with respect to $\sigma(\mathcal{F}_q^{N,\mathcal{P}}, R_{\text{om}}) \subseteq \sigma(\mathcal{H}, S_1, \mathcal{P})$, so it is also the conditional law of d given $(\mathcal{F}_q^{N,\mathcal{P}}, R_{\text{om}})$. The same argument without $\mathcal{P}$ gives the second claim, with π in place of π'. □

Lemma E.5 (Information in the price history) *Let Assumption E.3 hold and fix $q \in (0,1]$. Then $\pi' \ll \pi$, and the divergence of the posterior with the price history from the posterior without it,*

$$D := \mathrm{KL}(\pi' \,\|\, \pi) = \int \log \frac{d\pi'}{d\pi}\, d\pi',$$

satisfies

$$E[D] \le \frac{1}{2}\sum_{\tau=1}^{T} \log\bigl(1 + E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]\bigr) \le \frac{1}{2}\sum_{\tau=1}^{T} E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]. \tag{50}$$

Proof of Lemma E.5. Write $\theta := \tilde{\beta}_1$, $\mathcal{C}_0 := \sigma(\mathcal{F}_q^N, R_{\mathrm{om}})$, and $\mathcal{C}_\tau := \mathcal{C}_0 \vee \sigma(P_1, \dots, P_\tau)$, so that $\mathcal{C}_T = \sigma(\mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}})$. The loading a_τ is $\mathcal{C}_{\tau-1}$-measurable, because $\mathcal{F}_q^N$ contains the strong signals and payoffs of all dates and $\mathcal{C}_{\tau-1}$ the earlier prices. Conditional on θ, on $\mathcal{C}_{\tau-1}$, and on any further variables dated before τ, the price $P_\tau = a_\tau(X_\tau - \eta_\tau)$ is therefore an affine function, with a known nonzero coefficient, of the Gaussian shock η_τ, which is independent of all of them, so the prices have conditional densities $p(\cdot \mid \cdot)$ with respect to Lebesgue measure under every conditioning used below. By Bayes' rule, π' is π re-weighted by the likelihood ratio of the price history,

$$\frac{d\pi'}{d\pi}(\theta) = \frac{p(P_1, \dots, P_T \mid \theta, \mathcal{C}_0)}{p(P_1, \dots, P_T \mid \mathcal{C}_0)},$$

which gives $\pi' \ll \pi$, and D is the conditional expectation given $\mathcal{C}_T$ of this log-likelihood ratio at the true coefficient. Factorizing both densities over dates,

$$E[D] = E\Bigl[\log \frac{p(P_1, \dots, P_T \mid \theta, \mathcal{C}_0)}{p(P_1, \dots, P_T \mid \mathcal{C}_0)}\Bigr] = \sum_{\tau=1}^{T} E[\ell_\tau], \qquad \ell_\tau := \log \frac{p(P_\tau \mid \theta, \mathcal{C}_{\tau-1})}{p(P_\tau \mid \mathcal{C}_{\tau-1})}.$$

Each term is bounded in three steps, each using Gibbs' inequality: the expected log-likelihood ratio of the law a variable has against any other law is nonnegative.[29] First,

$$E[\ell_\tau] \le E\Bigl[\log \frac{p(P_\tau \mid \theta, X_\tau, \mathcal{C}_{\tau-1})}{p(P_\tau \mid \mathcal{C}_{\tau-1})}\Bigr],$$

because the difference of the two sides is the expected log-likelihood ratio of the law of P_τ given $(\theta, X_\tau, \mathcal{C}_{\tau-1})$ against its law given $(\theta, \mathcal{C}_{\tau-1})$. Given $(\theta, X_\tau, \mathcal{C}_{\tau-1})$, the price is $\mathcal{N}(a_\tau X_\tau, a_\tau^2)$, because η_τ is independent of $(\theta, X_\tau, \mathcal{C}_{\tau-1})$, which are functions of the parameters, the signals and residuals, and the shocks of earlier dates. Second, let $m_\tau := E[P_\tau \mid \mathcal{C}_{\tau-1}]$, $V_\tau := \mathrm{Var}[P_\tau \mid \mathcal{C}_{\tau-1}] = a_\tau^2\bigl(1 + \mathrm{Var}[X_\tau \mid \mathcal{C}_{\tau-1}]\bigr)$, and let g_τ be the $\mathcal{N}(m_\tau, V_\tau)$ density. Replacing $p(P_\tau \mid \mathcal{C}_{\tau-1})$ by $g_\tau(P_\tau)$ raises the expectation, because the difference is the expected log-

[29] For laws $\nu \ll \mu$, Jensen's inequality for the concave logarithm gives $\int \log \frac{d\mu}{d\nu}\, d\nu \le \log \int \frac{d\mu}{d\nu}\, d\nu \le 0$, so $\int \log \frac{d\nu}{d\mu}\, d\nu \ge 0$.

likelihood ratio of the law of P_τ given $\mathcal{C}_{\tau-1}$ against $\mathcal{N}(m_\tau, V_\tau)$.[30] Third, the divergence between two Gaussians gives, conditional on $(X_\tau, \mathcal{C}_{\tau-1})$,

$$E\Big[\log \frac{p(P_\tau \mid \theta, X_\tau, \mathcal{C}_{\tau-1})}{g_\tau(P_\tau)} \;\Big|\; X_\tau, \mathcal{C}_{\tau-1}\Big] = \frac{1}{2}\log\frac{V_\tau}{a_\tau^2} + \frac{a_\tau^2 + (a_\tau X_\tau - m_\tau)^2}{2V_\tau} - \frac{1}{2},$$

and averaging over X_τ given $\mathcal{C}_{\tau-1}$, with $E[(a_\tau X_\tau - m_\tau)^2 \mid \mathcal{C}_{\tau-1}] = V_\tau - a_\tau^2$, leaves $\frac{1}{2}\log\big(1 + \mathrm{Var}[X_\tau \mid \mathcal{C}_{\tau-1}]\big)$. Hence $E[\ell_\tau] \le \frac{1}{2}E\log\big(1 + \mathrm{Var}[X_\tau \mid \mathcal{C}_{\tau-1}]\big)$. Jensen's inequality for the concave logarithm, and $E\mathrm{Var}[X_\tau \mid \mathcal{C}_{\tau-1}] \le E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]$ because $\mathcal{F}_q^N \subseteq \mathcal{C}_{\tau-1}$, give the first inequality in (50); $\log(1+x) \le x$ gives the second. □

A unit of D is worth little to the investor. By (49) the posterior π has N_1 independent directions of variance at most $\tilde\sigma_q^2/N$, and by Lemma E.4(i) the forecast $E[d \mid \mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}}] - E[d \mid \mathcal{F}_q^N, R_{\mathrm{om}}] = (m' - \hat\beta_1)'S_1$, with m' the mean of π', loads on a direction independent of the past; its conditional mean square given $\mathcal{H}\vee\mathcal{P}$ is $(m'-\hat\beta_1)'\Psi_1(m'-\hat\beta_1) \le \|\Psi_1\|\,|m'-\hat\beta_1|^2 \le 2\|\Psi_1\|\tilde\sigma_q^2 D/N$. The last inequality holds because the Gaussian with the moments of π' has the smallest divergence from π among laws with those moments, so its divergence is at most D, while the divergence between two Gaussians is at least $\frac{1}{2}(m' - \hat\beta_1)'\hat\Sigma_1^{-1}(m' - \hat\beta_1) \ge \frac{1}{2}|m' - \hat\beta_1|^2 N/\tilde\sigma_q^2$, every eigenvalue of $\hat\Sigma_1$ being at most $\tilde\sigma_q^2/N$. Lemma I.2 extends this bound from the mean to the whole conditional law. Imagine r investors with independent current signals who all read the price history. Together they cannot learn more about $\tilde\beta_1$ than the history contains, and their prediction problems overlap only through the coefficient, with cross-covariances $S_i'\hat\Sigma_1 S_j$ of order $N^{-1/2}$, so each learns at most $D/r + O(r/N)$ about its own payoff. The best r makes this $O(\sqrt{D/N})$, which is the lemma.

Proposition E.6 (Negligible price history) *Let Assumption E.3 hold and let $q \in [0,1]$. If $q > 0$, suppose that*

$$\sum_{\tau=1}^{T} E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N] = o(N). \tag{51}$$

[30] The second and third steps say that, for a given conditional variance, a Gaussian price signal is the one from which the price reveals the most, and the per-date bound $\frac{1}{2}\log(1 + \mathrm{Var}[X_\tau \mid \mathcal{C}_{\tau-1}])$ is what a Gaussian signal with that variance reveals when observed in standard Gaussian noise. The derivation here uses only the Gaussian divergence formula and the fact that, averaged over the signal, the divergence of the price's conditional law from the mixture is at most its divergence from the Gaussian with the mixture's mean and variance.

Then for every $t \in \mathbb{R}$,

$$E\Big|E\big[e^{itd} \mid \mathcal{F}_q^{N,\mathcal{P}}\big] - E\big[e^{itd} \mid \mathcal{F}_q^{N}\big]\Big| \longrightarrow 0, \tag{52}$$

so the characteristic-function convergence in (39) holds with $\mathcal{F}_q^{N,\mathcal{P}}$ in place of $\mathcal{F}_q^N$, with the same $M_N(q)$ and r_q. If (51) holds at the scales $q_1 < \cdots < q_k$ of Lemma D.3, that lemma holds for the information sets $\mathcal{F}_{q_\ell}^{N,\mathcal{P}}$ with the same increments $\Delta_{N,\ell}$, and the increments are asymptotically independent of the price history: with $\Phi_{N,k+1} = \exp\{i(a's + \sum_{m\le k+1} t_m \Delta_{N,m})\}$ for $a \in \mathbb{R}^n$ and $t_1, \dots, t_{k+1} \in \mathbb{R}$, as in that proof, for every sequence of $\mathcal{P}$-measurable ξ_N with $|\xi_N| \le 1$,

$$E[\Phi_{N,k+1}\,\xi_N] - E[\Phi_{N,k+1}]\,E[\xi_N] \longrightarrow 0. \tag{53}$$

By Lévy's continuity theorem, (53) says that $(s, \Delta_{N,1}, \dots, \Delta_{N,k+1})$ converges to the Gaussian vector of Lemma D.3 also conditional on any events $A_N \in \mathcal{P}$ with $\inf_N P(A_N) > 0$. Two cases of (51) are of interest. If scaled trading occurred only at the T_w dates of a window W, with X_τ a function of s_τ outside it and $\sup_{\tau\in W} E[X_\tau^2] \le C_X$ inside it, then $\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N] = 0$ outside the window, the sum in (51) is at most $T_w C_X$, and the condition holds at every scale whenever $T_w/T \to 0$. If instead every past date had scaled trading, the condition requires the summed sizes $\sum_\tau E\mathrm{Var}[X_\tau \mid \mathcal{F}_q^N]$ of the price signals to grow more slowly than N, which holds when the mass of investors at scales above q was small enough at the past dates for the variances $v_\tau(q) - 1$ to be small on average. With T dates and loadings of order one the sum is of order T, hence of order N, and the proposition does not apply; this is the repeated economy described at the start of this appendix.

Proof of Proposition E.6. *Conditional on the omitted risk.* Let $q > 0$, fix t, and take $\mathcal{H}$, π, π' from Lemma E.4 and D from Lemma E.5. Let Λ be the spectral edge of Lemma I.6 and put

$$\mathcal{E}_N := \{\|\Psi_1\| \le \Lambda + 1\} \cap \{R_{\mathrm{om}} \le r_{\mathrm{om}} + 1\}, \qquad C_1 := (\Lambda + 1)\Big(\sigma_\beta^2 + \frac{r_{\mathrm{om}} + 1}{V_0}\Big).$$

Then $P(\mathcal{E}_N) \to 1$, because $\|\Psi_1\| \le \|\Psi\|$, $\limsup_N \|\Psi\| \le \Lambda$ almost surely, and $R_{\mathrm{om}} \to r_{\mathrm{om}}$ in probability by the proof of Lemma D.2; and $\mathcal{E}_N \in \mathcal{H}$. On $\mathcal{E}_N$ every eigenvalue of $\hat{\Sigma}_1$ is at

most $\tilde{\sigma}_q^2/N \le C_1/((\Lambda+1)N)$, so, with $B := \Psi_1^{1/2}\hat{\Sigma}_1\Psi_1^{1/2}$,

$$V_N := \operatorname{tr}(\hat{\Sigma}_1\Psi_1\hat{\Sigma}_1\Psi_1) = \operatorname{tr}(B^2) \le \|B\|\operatorname{tr}(B) \le \|\Psi_1\|\,\|\hat{\Sigma}_1\|\cdot\|\Psi_1\|\operatorname{tr}(\hat{\Sigma}_1) \le \frac{C_1}{N}\cdot\frac{C_1 N_1}{N} \le \frac{C_1^2}{N}.$$

Conditional on $\mathcal{H}$, Lemma E.4(i) places $d - b's - \hat{\beta}_1' S_1 = (\tilde{\beta}_1 - \hat{\beta}_1)' S_1 + \varepsilon_1$ in the setting of Lemma I.2, with $(\theta, m, \Sigma, \Psi, Z, S, v) = (\tilde{\beta}_1, \hat{\beta}_1, \hat{\Sigma}_1, \Psi_1, \mathcal{P}, S_1, v_q)$, dimension N_1, and $V = V_N$. Conditional on $\mathcal{H}$, the information in the price history about the coefficient is $E[D \mid \mathcal{H}]$, because the conditional laws of $\tilde{\beta}_1$ without and with the history are π and π'. Let J_N be the information, conditional on $\mathcal{H}$, in the price history about the payoff given the current signals, and let Γ' and Γ be the conditional laws of d given $(\mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}})$ and given $(\mathcal{F}_q^N, R_{\mathrm{om}})$. By Lemma E.4(ii), $J_N = E[\mathrm{KL}(\Gamma'\|\Gamma) \mid \mathcal{H}]$; the known shift $b's + \hat{\beta}_1' S_1$ changes no divergence and no distance. Since $v_q \ge \sigma_\varepsilon^2$, (96) gives, on $\mathcal{E}_N$,

$$J_N \le \frac{\sqrt{E[D \mid \mathcal{H}]\, V_N}}{v_q} \le \frac{C_1}{\sigma_\varepsilon^2}\sqrt{\frac{E[D \mid \mathcal{H}]}{N}},$$

and (97) gives $E[\mathrm{TV}(\Gamma', \Gamma)^2 \mid \mathcal{H}] \le \frac{1}{2} J_N$; off $\mathcal{E}_N$ the distance is at most one. Taking expectations, with Jensen's inequality for the square root and (50),

$$E\big[\mathrm{TV}(\Gamma', \Gamma)^2\big] \le \frac{C_1}{2\sigma_\varepsilon^2}\sqrt{\frac{E[D]}{N}} + P(\mathcal{E}_N^c) \longrightarrow 0,$$

because (51) makes $E[D] = o(N)$. Since $|E[e^{itd} \mid \mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}}] - E[e^{itd} \mid \mathcal{F}_q^N, R_{\mathrm{om}}]| \le 2\,\mathrm{TV}(\Gamma', \Gamma)$ and $E[\mathrm{TV}(\Gamma', \Gamma)] \le (E[\mathrm{TV}(\Gamma', \Gamma)^2])^{1/2}$,

$$E\Big|E\big[e^{itd} \mid \mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}}\big] - E\big[e^{itd} \mid \mathcal{F}_q^N, R_{\mathrm{om}}\big]\Big| \longrightarrow 0. \tag{54}$$

Removing the omitted risk. In the notation of the proof of Lemma D.2, let

$$\Xi := E\big[e^{itd} \mid \mathcal{F}_q^N, R_{\mathrm{om}}\big] = \exp\{itm_N(R_{\mathrm{om}}) - t^2 v_N(R_{\mathrm{om}})/2\},$$
$$\bar{\Xi} := \exp\{itm_N(r_{\mathrm{om}}) - t^2 v_N(r_{\mathrm{om}})/2\},$$

where $\bar{\Xi}$ is $\mathcal{F}_q^N$-measurable. By (40), $\Xi - \bar{\Xi} \to 0$ in probability and, both having modulus at

most one, in L^1. Since $\mathcal{F}_q^N \subseteq \mathcal{F}_q^{N,\mathcal{P}}$,

$$E\big[e^{itd} \mid \mathcal{F}_q^{N,\mathcal{P}}\big] - E\big[e^{itd} \mid \mathcal{F}_q^N\big] = E\Big[E\big[e^{itd} \mid \mathcal{F}_q^{N,\mathcal{P}}, R_{\mathrm{om}}\big] - \Xi \;\Big|\; \mathcal{F}_q^{N,\mathcal{P}}\Big] + E\big[\Xi - \bar{\Xi} \mid \mathcal{F}_q^{N,\mathcal{P}}\big] - E\big[\Xi - \bar{\Xi} \mid \mathcal{F}_q^N\big].$$

Conditional expectation contracts L^1 norms, so the three terms have L^1 norms at most the left side of (54), $E|\Xi - \bar{\Xi}|$, and $E|\Xi - \bar{\Xi}|$, which proves (52) for $q > 0$. For $q = 0$, $\mathcal{F}_0^N = \sigma(s)$ and $\mathcal{F}_0^{N,\mathcal{P}} = \sigma(s, \mathcal{P})$. Conditional on (β, Ψ), the variable $d - b's = \beta' S + \varepsilon$ is centered Gaussian with variance $\beta'\Psi\beta + \sigma_\varepsilon^2$ and independent of $(s, \mathcal{P})$, so both conditional characteristic functions equal $e^{itb's}$ times a conditional expectation of $e^{-t^2(\beta'\Psi\beta+\sigma_\varepsilon^2)/2}$, and each is within $E|e^{-t^2(\beta'\Psi\beta+\sigma_\varepsilon^2)/2} - e^{-t^2R(0)/2}| \to 0$ of $e^{itb's - t^2R(0)/2}$ in L^1, as in the proof of Lemma D.3. With (39), the triangle inequality gives the characteristic-function convergence for $\mathcal{F}_q^{N,\mathcal{P}}$.

Joint convergence and asymptotic independence. The proof of Lemma D.3 uses only the characteristic-function convergence at each scale and the nesting of the information sets, and both hold for the sets $\mathcal{F}_{q_\ell}^{N,\mathcal{P}}$ with the forecasts $M_N(q_\ell)$ unchanged. Its induction step multiplies an $\mathcal{F}_{q_{\ell-1}}^{N,\mathcal{P}}$-measurable variable of modulus at most one by the conditional characteristic function of the next increment, so it applies to $\Phi_{N,\ell-1}\xi_N$ for any $\mathcal{P}$-measurable ξ_N with $|\xi_N| \le 1$, which is $\mathcal{F}_0^{N,\mathcal{P}}$-measurable, with the same error terms, and yields

$$E[\Phi_{N,k+1}\xi_N] - e^{-a'\psi a/2} \prod_{\ell=1}^{k+1} e^{-t_\ell^2 \delta_\ell/2}\, E[\xi_N] \longrightarrow 0,$$

because the current strong signals s are independent of $\mathcal{P}$. The product is the limit of $E[\Phi_{N,k+1}]$, which gives (53). □

Proposition E.6 places the price history outside the limiting economy. Every object of Appendix D.3 is a functional of the independent coordinates (s, W, Z, η): the forecasts $M(q)$, the payoff d^∞, the information sets (47), the prices, the posteriors, and the demands and equilibria of Appendix F. The finite-economy vector of strong signals and forecast increments converges to its Gaussian limit jointly with any bounded function of the price history, with a product limit law, and the current supply shock is independent of the price history by assumption. The limiting economy with a price history is therefore the economy of Appendix D.3 with one additional coordinate, independent of (s, W, Z, η), and this description is justified by Proposition E.6 in the same sense in which Lemma D.3

justifies the construction without it. Conditioning on an independent coordinate changes no conditional law of d^∞, at any scale or for the uninformed, and hence no demand, no residual risk, and no equilibrium object of Definition 3 and Theorem 5. The proposition takes the past prices as given and says nothing about the equilibria that produced them; it shows that disclosing them leaves the limiting economy, in which Definition 3 is stated, unchanged.

F Equilibrium with endogenous scales

We prove Lemmas 5 and 6, Corollary 4, and Theorem 5 in the Gaussian economy constructed in Appendix D. All expectations and variances below use $\mathbb{P}_\infty$, which is both the objective law and investors' common beliefs. A linear price is an affine, square-integrable function of the strong signals, the acquired forecast innovations, and η, with deterministic coefficients. Uniqueness throughout concerns this class and the almost-sure conventions of Appendix D.4. Write $j = -R' > 0$, $K = \gamma\sigma_\eta > 0$, and $A_s = \sigma_d^2 - R(0) \geq 0$. For general acquisition costs, define

$$\mathcal{J}_0(q) = -\frac{\log R(q)}{2\gamma} - \chi(q), \qquad q_H = \min \operatorname*{argmax}_{0<q\leq 1} \mathcal{J}_0(q).$$

The maximum is attained at a positive scale because $\mathcal{J}_0'(0) > 0$. Set $R_H = R(q_H)$, $B_H = \sigma_d^2 - R_H$, and $\overline{R} = R_H e^{2\gamma\chi(q_H)}$. For the proof only, abbreviate the absolute population measure by $\mu(dq) = \lambda\, dF(q)$, so $\mu((0,1]) = \lambda \leq 1$. We first prove the two lemmas and the corollary in their main-text order: indifference determines residual risk, and market clearing recovers the distribution from that risk. The proof of Theorem 5 then checks admissibility and participation, excludes other support shapes under the additional restrictions, and establishes general existence and uniqueness.

F.1 Proofs of Lemma 5 and Corollary 4: post-price residual risk

Proof of Lemma 5. Write $R_P(q) = \mathrm{Var}^\infty[d^\infty \mid \mathcal{F}_q^\infty, P]$ as in (21), $R_U = \mathrm{Var}^\infty[d^\infty \mid P]$, and $S_P = \mathrm{Var}^\infty[d^\infty - P]$. Linear equilibrium prices are centered, as verified in Appendix F.3.4. For any investor with information $\mathcal{I}$ including the price and deterministic posterior variance r, Gaussian CARA optimization gives

$$\max_x E^\infty[-e^{-\gamma x(d^\infty - P)} \mid \mathcal{I}] = -\exp\left\{-\frac{(E^\infty[d^\infty \mid \mathcal{I}] - P)^2}{2r}\right\}.$$

The conditional mean of $d^\infty - P$ is centered Gaussian with variance $S_P - r$. Integrating the optimized utility therefore yields

$$V(q) = -e^{\gamma\chi(q)}\sqrt{\frac{R_P(q)}{S_P}}, \qquad V_U = -\sqrt{\frac{R_U}{S_P}}.$$

The denominator $\sqrt{S_P}$ is common to all acquisition choices. Thus $-e^{\gamma\chi(q)}\sqrt{R_P(q)}$ is constant on the support of any positive-entry equilibrium. At its highest acquired scale q_H, the investor already knows every forecast entering the price; the remaining price variation reveals only the independent supply shock. The last paragraph of this proof makes this precise and gives $R_P(q_H) = R(q_H)$, so indifference gives

$$e^{\gamma\chi(q)}\sqrt{R_P(q)} = e^{\gamma\chi(q_H)}\sqrt{R(q_H)},$$
$$R_P(q) = R(q_H)e^{2\gamma[\chi_1(q_H)-\chi_1(q)]}$$

at every occupied scale, with equality at support endpoints by continuity. The fixed cost cancels from this comparison.

For (23), use the Gaussian posterior (57) of Appendix F.2: $R_P(q) = R(q) - H(q)^2/v(q)$ with $v(q) \geq 1$ and $H(q) = \int_q^1 k(s)j(s)\,ds$. By (58) and the bound $L \geq K$ of Appendix F.3.4, $k(s) = \Pi(s)/L(s)$, where $\Pi(s) = \int_{[s,1]} \mu(du)/R_P(u)$ is positive if and only if $\mu([s,1]) > 0$. Since q_H is the largest point of the support, $k > 0$ on $(0, q_H)$ and $k = 0$ on $(q_H, 1]$. With $j > 0$, this gives $H(q) > 0$ for $q < q_H$ and $H(q) = 0$ for $q \geq q_H$, which proves (23). □

The ex-ante certainty equivalent of a value $V < 0$ is $-\log(-V)/\gamma$. By the value formulas in the proof of Lemma 5, differences in ex-ante certainty equivalents therefore equal differences in

$$\mathcal{J}(q) = -\frac{\log R_P(q)}{2\gamma} - \chi(q), \qquad \mathcal{J}_U = -\frac{\log R_U}{2\gamma}. \tag{55}$$

The common term $\log S_P/(2\gamma)$ cancels; the indices are evaluated before observing the price.

Proof of Corollary 4. Fix q and write $r = R_P(q)$ and $a = S_P - r$. By the proof of Lemma 5, $m = E^\infty[d^\infty \mid \mathcal{F}_q^\infty, P] - P$ is centered Gaussian with variance a, and the optimal demand is $x_q(P) = m/(\gamma r)$. Here $a > 0$: if $a = 0$, then $R_P(q) = S_P$ and $V(q) = -e^{\gamma\chi(q)} < -1 \leq V_U$, so q could not be in the support of F. Given $(\mathcal{F}_q^\infty, P)$, the payoff d^∞ is Gaussian with mean $m + P$ and deterministic variance r. The forecast error $e = d^\infty - P - m$ is therefore Gaussian

with mean zero and variance r, and it is independent of m. Hence $\pi_q = m(m+e)/(\gamma r)$, and

$$E^\infty[\pi_q] = \frac{a}{\gamma r}, \qquad \mathrm{Var}^\infty[\pi_q] = \frac{3a^2 + ar - a^2}{(\gamma r)^2} = \frac{a(2S_P - r)}{(\gamma r)^2}.$$

Their ratio gives (25). The map $r \mapsto (S_P - r)/(2S_P - r)$ has derivative $-S_P/(2S_P - r)^2 < 0$. By (24) and $\chi_1' > 0$ in Assumption 2, R_P is strictly decreasing on the support of F, so the ratio is strictly increasing there.

With mean supply, as in Appendix J, let $\delta = E^\infty[d^\infty - P]$. Then m has mean δ and variance $a = S_P - r$, the expected profit $E^\infty[\pi_q] = (a + \delta^2)/(\gamma r)$ is positive, and the same calculation gives

$$\frac{E^\infty[\pi_q]^2}{\mathrm{Var}^\infty[\pi_q]} = \frac{(a+\delta^2)^2}{a^2 + a(3\delta^2 + S_P) + \delta^2 S_P}.$$

Its derivative with respect to a has the sign of $a(\delta^2 + S_P) + \delta^2(S_P - 3\delta^2)$, which is positive whenever $S_P \geq 3\delta^2$. A Gaussian integral gives $E^\infty[e^{-m^2/(2r)}] = \sqrt{r/S_P}\, e^{-\delta^2/(2S_P)}$, so

$$V(q) = -e^{\gamma\chi(q)} \sqrt{\frac{R_P(q)}{S_P}}\, e^{-\delta^2/(2S_P)}.$$

The last factor is common to all scales. Indifference therefore still gives (24), so R_P is strictly decreasing on the support of F and $a = S_P - R_P$ increases with scale. The ratio then increases with scale whenever $S_P \geq 3\delta^2$. □

F.2 Proof of Lemma 6: recovering the distribution

We recover the price information and population tail from R and the actual post-price risk R_P in (21) on an occupied interval. The calculation uses the following identities for a linear CREE, whose coefficient matching and existence at any population distribution are verified in Appendix F.3.4. Normalize the price's supply-noise coefficient and write

$$P = a_P \mathcal{X}, \qquad \mathcal{X} = k_0 M(0) + \int_0^1 k(s)\, dM(s) - \eta. \tag{56}$$

Here $M(0) = b's$, and $k(s)$ is the loading on the forecast innovation $dM(s)$. An investor at scale q learns from the price through its residual $Y_q = \mathcal{X} - E^\infty[\mathcal{X} \mid \mathcal{F}_q^\infty]$. Define its covariance

with the unknown payoff and its variance by

$$H(q) = \mathrm{Cov}^{\infty}[d^{\infty} - M(q), Y_q \mid \mathcal{F}_q^{\infty}] = \int_q^1 k(s)j(s)\,ds,$$
$$v(q) = \mathrm{Var}^{\infty}[Y_q \mid \mathcal{F}_q^{\infty}] = 1 + \int_q^1 k(s)^2 j(s)\,ds, \qquad g(q) = \frac{H(q)}{v(q)}.$$

Gaussian conditioning gives

$$\begin{aligned} R_P(q) &= R(q) - \frac{H(q)^2}{v(q)}, \\ E^{\infty}[d^{\infty} \mid \mathcal{F}_q^{\infty}, P] &= M(q) + g(q)\big(\mathcal{X} - E^{\infty}[\mathcal{X} \mid \mathcal{F}_q^{\infty}]\big). \end{aligned} \tag{57}$$

Let

$$\Pi(q) = \int_{[q,1]} \frac{\mu(ds)}{R_P(s)}, \qquad L(q) = K + \int_{[q,1]} \frac{g(s)\mu(ds)}{R_P(s)}.$$

Here Π is a precision-weighted tail; F continues to denote the conditional CDF of scales. Matching the coefficient of each forecast increment in market clearing gives

$$k(q) = \frac{\Pi(q)}{L(q)}, \qquad h(q) := 1 - k(q)g(q) > 0, \qquad L(q)h(q)v(q) = K. \tag{58}$$

The population tail is $m(q) = \mu([q,1])$, and the same market-clearing identities give

$$\begin{aligned} &m = R_P\Pi - Kg, \quad k = \frac{mv + KH}{KR + mH}, \\ &H' = -kj, \qquad\qquad v' = -k^2 j, \qquad H(1) = 0, \quad v(1) = 1. \end{aligned} \tag{59}$$

These identities follow from the precision tails and hold across atoms because R_P and g are continuous.

In particular, $h = KR_P/(KR + mH) > 0$. Differentiating the Gaussian posterior with $R' = -j$, $H' = -kj$, and $v' = -k^2 j$ gives

$$R_P' = -j + \frac{2kHj}{v} - \frac{k^2H^2j}{v^2} = -j\left(1 - \frac{kH}{v}\right)^2.$$

Hence, away from atoms,

$$R_P' = -jh^2, \qquad \mathcal{J}' = \frac{jh^2}{2\gamma R_P} - \chi'. \tag{60}$$

Proof of Lemma 6. On $[q_L, q_H)$, positive population above q implies $H(q) > 0$, so $R(q) > R_P(q)$. Set $u = H/\sqrt{v} = \sqrt{R - R_P}$. The identities above give

$$h = \sqrt{\frac{R_P'}{R'}}, \qquad \frac{k}{\sqrt{v}} = \frac{1-h}{u}, \qquad m = \frac{K}{u\sqrt{v}}\left(\frac{R_P}{h} - R\right).$$

The last equality follows by solving $h = KR_P/(KR + mH)$ for m. Because no investor acquires above q_H, $k = 0$ there and $v(q_H) = 1$. Moreover,

$$\frac{v'}{v} = -j\left(\frac{k}{\sqrt{v}}\right)^2 = -\frac{\left[\sqrt{-R'} - \sqrt{-R_P'}\right]^2}{R - R_P}.$$

Integrating backward from q_H determines $v^{-1/2}$ as the exponential factor in (26). The integrand equals $k^2 j/v$ and extends continuously to q_H, since $k(q_H-) = \mu(\{q_H\})/(KR_H)$ and $v(q_H) = 1$. Substituting u, h, and $K = \gamma\sigma_\eta$ into the formula for m and using $F(q) = 1 - m(q)/\lambda$ for $q < q_H$ proves (26).

To evaluate the tail at the top, set $\lambda_H = \mu(\{q_H\})$ and $\varepsilon := q_H - q$. The terminal values and the left limit of k give

$$H(q) = \frac{\lambda_H j(q_H)}{KR_H}\,\varepsilon + o(\varepsilon), \qquad R(q) - R_P(q) = \left[\frac{\lambda_H j(q_H)}{KR_H}\right]^2 \varepsilon^2 + o(\varepsilon^2).$$

By Lemma 5 and $\chi_1 \in C^2$, R_P has a continuous second derivative up to q_H from the occupied interval. Since $g(q_H) = H(q_H)/v(q_H) = 0$, (58) gives $h(q_H-) = 1$, and (60) gives $R_P'(q_H-) = R'(q_H)$. With $R_P(q_H) = R(q_H)$, the zeroth- and first-order terms of $R - R_P$ vanish at q_H. Comparing the last display with the second-order Taylor expansion of $R - R_P$ at q_H therefore yields

$$\frac{R''(q_H) - R_P''(q_H-)}{2} = \left[\frac{\lambda_H j(q_H)}{KR_H}\right]^2.$$

Solving for the nonnegative mass λ_H and using $1 - F(q_H-) = \lambda_H/\lambda$ proves (27). Finally, the inclusive tail $m(q) = \mu([q, q_H])$ is continuous on $[q_L, q_H]$ because there are no atoms below

q_H. It satisfies $m(q) = \lambda[1 - F(q)]$ for $q < q_H$, $m' = -\lambda F'$ on (q_L, q_H), $m(q_L) = \lambda$, and $m(q_H) = \lambda_H$. These identities prove the normalization and atom formulas. □

An occupied interval that reaches q_H therefore requires $-R'(q_H) = 2\gamma\chi_1'(q_H)R(q_H)$: differentiate (24) and use $R_P'(q_H-) = R'(q_H)$.

F.3 Proof of Theorem 5

Appendices F.3.1–F.3.6 together prove Theorem 5. They first construct the candidate distribution, then verify its optimality and establish that it is the unique equilibrium.

F.3.1 A candidate density and a top mass

Consider a positive-entry candidate whose largest acquired scale is initially denoted by $\bar{q}$. For the mixed candidate, suppose

$$\mu(dq) = \rho(q)\mathbf{1}_{(q_L,\bar{q})}\,dq + \lambda_H\delta_{\bar{q}}(dq), \qquad 0 < q_L < \bar{q} \le 1,$$
$$\lambda = \lambda_H + \int_{q_L}^{\bar{q}} \rho(q)\,dq \le 1,$$

where ρ is continuous and strictly positive on the occupied interval and $\lambda_H > 0$ is an absolute population share. The conditional CDF among informed investors is $F(q) = \mu((0,q])/\lambda$. We also allow a single-scale candidate $\mu = \lambda\delta_{\bar{q}}$ and consider no acquisition separately. At this stage the proposed mixed form is an assumption used to derive a candidate; Appendix F.3.3 proves its necessity under (74).

In any equilibrium with $\lambda > 0$, let $\bar{q} = \sup\operatorname{supp}\mu$. The preceding top-scale argument gives $R_P(\bar{q}) = R(\bar{q})$ without requiring an atom there. Continuity and optimality of occupied scales imply that the common informed index is $\mathcal{J}_0(\bar{q})$, even if $\bar{q}$ has no atom. Since $R_P(q) \le R(q)$, every deviation has $\mathcal{J}(q) \ge \mathcal{J}_0(q)$, so $\bar{q}$ must maximize $\mathcal{J}_0$. If a smaller q tied that maximum, the positive population above it would make the price strictly informative beyond its acquired information, giving $\mathcal{J}(q) > \mathcal{J}_0(q) = \mathcal{J}_0(\bar{q})$, a contradiction. Thus $\bar{q} = q_H$, the smallest global maximizer. For the construction and deviation checks, write $R_*(q) = R_H e^{2\gamma[\chi_1(q_H)-\chi_1(q)]}$ for the extension of this indifference curve to all scales. Throughout this appendix, R_P denotes the post-price risk (21) under the actual candidate price. The two curves coincide on the equilibrium support by Lemma 5, but can differ at unoccupied scales or for a candidate that is not an overall equilibrium. Comparing (55) to $\mathcal{J}_0(q_H)$ shows that a positive-mass linear

CREE is an overall equilibrium if and only if

$$\begin{aligned} R_P(q) &\geq R_*(q) && \text{for every } q \in (0,1], \\ R_P(q) &= R_*(q) && \text{for } dF\text{-almost every } q, \\ R_U &\geq \overline{R}, \qquad (1-\lambda)(R_U - \overline{R}) = 0. \end{aligned} \tag{61}$$

In particular,

$$R(q) > R_*(q) \quad \text{for } 0 \leq q < q_H.$$

For the mixed construction, take $q_H < 1$ with $\mathcal{J}_0''(q_H) < 0$; Appendix F.3.3 derives these properties from the additional primitive restrictions. On the candidate interval, indifference fixes $R_P = R_*$ and $R_*' = -2\gamma\chi_1' R_*$. The risk reduction and its derivative therefore identify $u = H/\sqrt{v}$ and h; the identity $1-h = kH/v$ then identifies $\ell = k/\sqrt{v}$. Since $v' = -j\ell^2 v$ and $v(q_H) = 1$, the remaining price variance is obtained by one integral. For $0 < q < q_H$, define

$$\begin{aligned} u_*(q) &= \sqrt{R(q) - R_*(q)}, & h_*(q) &= \sqrt{\frac{2\gamma\chi_1'(q)R_*(q)}{j(q)}}, \\ \ell_*(q) &= \frac{1-h_*(q)}{u_*(q)}, & e_*(q) &= \frac{1}{2}\left[\frac{\chi_1''(q)}{\chi_1'(q)} - \frac{j'(q)}{j(q)} - 2\gamma\chi_1'(q)\right], \\ z_\rho(q) &= e_*(q) - j(q)\ell_*(q)^2, & v_*(q) &= \exp\left\{\int_q^{q_H} j(s)\ell_*(s)^2\, ds\right\}. \end{aligned} \tag{62}$$

The root and derivative formulas (41)–(42) make every function in (62) explicit in $(c, \kappa, A_w, \sigma_\varepsilon^2, \gamma, \chi_1)$. In particular, they do not depend on the fixed cost, strong-signal predictability, supply noise, or an equilibrium price. Logarithmic differentiation gives $h_*'/h_* = e_*$, so $z_\rho = h_*'/h_* - j\ell_*^2$. The candidate density and absolute population tail are

$$\begin{aligned} \rho_*(q) &= \frac{KR_*(q)z_\rho(q)}{h_*(q)u_*(q)\sqrt{v_*(q)}}, \\ m_*(q) &= \frac{K}{\sqrt{v_*(q)}}\left[\frac{R_*(q)\ell_*(q)}{h_*(q)} - u_*(q)\right]. \end{aligned} \tag{63}$$

On an equilibrium occupied interval, $m_* = m$ and $\rho_* = -m'$.

The tail identities imply

$$k' = -\frac{\rho h}{R_P L}, \qquad g' = -\frac{kjh}{v}, \qquad \frac{h'}{h} = \frac{\rho g}{R_P L} + \frac{k^2 j}{v}, \qquad Lhv = K.$$

Substitution of $g = u_*/\sqrt{v_*}$ and $L = K/(h_* v_*)$ therefore yields the density in (63). Furthermore,

$$k = \ell_* \sqrt{v_*}, \qquad g = \frac{u_*}{\sqrt{v_*}}, \qquad \Pi = \frac{K\ell_*}{h_* \sqrt{v_*}}, \qquad L = \frac{K}{h_* v_*}.$$

Substituting these expressions into $m = R_P \Pi - Kg$ proves the tail formula and $m'_* = -\rho_*$.

The same identities imply $(u_* \sqrt{v_*})' = -kj$ with zero terminal value, verifying the price covariance H and thus the candidate residual risk throughout the interval.

For the interior nondegenerate top in the theorem, define the curvature and associated mass threshold by

$$\begin{aligned} C_H &= -\gamma R_H \mathcal{J}_0''(q_H) = \frac{1}{2}\left[R''(q_H) - \frac{j(q_H)^2}{R_H} + 2\gamma R_H \chi_1''(q_H)\right] > 0, \\ \Lambda_H &= \frac{K R_H \sqrt{C_H}}{j(q_H)}. \end{aligned} \tag{64}$$

By indifference, $C_H = [R''(q_H) - R_P''(q_H)]/2$ in a mixed equilibrium, with the derivative taken from the occupied interval. For $\varepsilon := q_H - q$,

$$R(q) - R_*(q) = C_H \varepsilon^2 + o(\varepsilon^2).$$

The mixed candidate has occupied scales arbitrarily close to the top. Since $v(q_H) = 1$, $H(q_H) = 0$, and $k(q_H -) = \lambda_H/(K R_H)$, its price implies

$$R(q) - R_P(q) = \left[\frac{\lambda_H j(q_H)}{K R_H}\right]^2 \varepsilon^2 + o(\varepsilon^2).$$

Matching these coefficients with C_H in (64) pins down the top mass:

$$\lambda_H = K R_H \frac{\sqrt{C_H}}{j(q_H)} = \Lambda_H. \tag{65}$$

The corresponding endpoint limits are

$$u_*(q_H) = 0, \qquad h_*(q_H) = v_*(q_H) = 1, \qquad \ell_*(q_H) = \frac{\sqrt{C_H}}{j(q_H)}, \qquad m_*(q_H) = \Lambda_H.$$

These limits and $m'_* = -\rho_*$ determine the entire candidate upper tail independently of its as-yet unknown lower endpoint.

F.3.2 Admissibility, participation, and the equilibrium cases

A mixed candidate.

The density is admissible on a proposed interval only if

$$0 < h_*(q) < 1, \qquad z_\rho(q) > 0 \qquad (q_L < q < q_H). \tag{66}$$

The quantity u_* is already strictly positive below q_H. Appendix F.3.3 will show that these inequalities hold throughout the mixed equilibrium interval under (74). Without those restrictions, a candidate interval containing negative z_ρ is inadmissible.

Write $B(q) = \sigma_d^2 - R(q)$. If q is the lowest acquired scale, the uninformed residual risk is the following explicit function:

$$U_*(q) = \sigma_d^2 - \frac{[\ell_*(q)B(q) + u_*(q)]^2}{1 + \ell_*(q)^2 B(q)}. \tag{67}$$

The endpoint must solve

$$m_*(q_L) \le 1, \qquad U_*(q_L) \ge \overline{R}, \qquad [1 - m_*(q_L)]\,[U_*(q_L) - \overline{R}] = 0. \tag{68}$$

If uninformed investors remain, solve $U_*(q_L) = \overline{R}$ subject to $m_*(q_L) < 1$. If everyone acquires, solve $m_*(q_L) = 1$ and check $U_*(q_L) \ge \overline{R}$. Both functions are strictly monotone on an admissible interval: m_* decreases and U_* increases with q.

For general primitives, a candidate must also pass a global lower-scale deviation check. For every $0 < q < q_L$, let $y(q) = R(q) - R(q_L)$ and require

$$R(q) - \frac{[u_*(q_L) + \ell_*(q_L)y(q)]^2}{1 + \ell_*(q_L)^2 y(q)} \ge R_*(q). \tag{69}$$

This rules out buying a smaller model and using its information together with the price.

Deviations above q_H are already ruled out by the global definition of q_H. Under (74), Appendix F.3.3 proves that the lower-scale inequality follows automatically from admissibility.

A connected equilibrium with a continuous, strictly positive density on (q_L, q_H) exists if and only if some $0 < q_L < q_H$ satisfies (66), (68), and (69). Its population distribution is

$$\lambda\, dF(q) = \rho_*(q)\mathbf{1}_{(q_L,q_H)}\, dq + \Lambda_H \delta_{q_H}(dq), \qquad \lambda = m_*(q_L). \tag{70}$$

In particular $\Lambda_H < 1$ and $m_*(q) = \Lambda_H + \int_q^{q_H} \rho_*(s)\, ds$. The inclusive population tail $m(q) = \lambda[1 - F(q-)]$ equals $m_*(q)$ and is continuous on $[q_L, q_H]$. The CDF is

$$F(q) = \begin{cases} 0, & q < q_L, \\ 1 - m_*(q)/\lambda, & q_L \le q < q_H, \\ 1, & q \ge q_H. \end{cases} \tag{71}$$

The conditional density is ρ_*/λ and the jump in F at q_H is Λ_H/λ. The formula for m_* avoids a second integral when evaluating the CDF.

Below q_L there are no additional informed investors, so k is constant. Continuing the price moments through the unobserved forecast variance gives (69); continuing through all forecast variance, including the strong block, gives (67). For monotonicity, write $C_{\mathcal{X}} = kB + H$ and $S_{\mathcal{X}} = k^2 B + v$ at a proposed endpoint. The identities $B' = j$, $H' = -kj$, and $v' = -k^2 j$ give $C'_{\mathcal{X}} = k'B$ and $S'_{\mathcal{X}} = 2kk'B$, and hence

$$U'_* = -\frac{2k'BC_{\mathcal{X}}}{S_{\mathcal{X}}}\left(1 - \frac{kC_{\mathcal{X}}}{S_{\mathcal{X}}}\right) > 0, \qquad 1 - \frac{kC_{\mathcal{X}}}{S_{\mathcal{X}}} = \frac{vh}{S_{\mathcal{X}}} > 0.$$

Here $k' < 0$ because the density is positive. Together with $m'_* < 0$, this shows that at most one admissible endpoint satisfies the participation and population conditions.

Finally, these conditions are sufficient as well as necessary. Given such an endpoint, take $k = \ell_*\sqrt{v_*}$ on (q_L, q_H), extend it constantly below q_L, set $k_0 = k(q_L)$, and set $k = 0$ above q_H. The identities just proved recover the precision tails and Gaussian posteriors for the population in (70). Equations (77) and (78) then give the price and optimal demands and clear the market. Indifference on the occupied interval, the endpoint conditions, and all lower-scale deviation inequalities give (61) at every feasible scale. Thus every candidate satisfying these checks is an overall equilibrium. General uniqueness is established in Appendix F.3.6.

No acquisition. With no acquisition, $R_U = \sigma_d^2$ and a deviator faces $R_P(q) = R(q)$. Such an

equilibrium exists exactly when $\overline{R} \geq \sigma_d^2$. Substituting $\overline{R} = R_H e^{2\gamma[\chi_0+\chi_1(q_H)]}$ gives the entry threshold (28). Every positive informed population makes the price informative about the payoff, so $R_U < \sigma_d^2$. Since positive-entry equilibrium requires $R_U \geq \overline{R}$, it is impossible when $\overline{R} \geq \sigma_d^2$.

A single acquired scale. Suppose $\overline{R} < \sigma_d^2$. Any equilibrium with all informed investors at one scale must use q_H. Its price reveals $M(q_H) - \nu_{\text{one}}\eta$, where $\nu_{\text{one}} = KR_H/\lambda_{\text{one}}$, and uninformed risk is $R_H + \nu_{\text{one}}^2 B_H/(\nu_{\text{one}}^2 + B_H)$. Solving $R_U = \overline{R}$, subject to the cap $\lambda \leq 1$, gives

$$\lambda_{\text{one}} = \min\left\{ K\sqrt{\frac{R_H}{e^{2\gamma\chi(q_H)} - 1} - \frac{R_H^2}{B_H}},\, 1 \right\}. \tag{72}$$

The denominator in (72) is positive because $\chi_1(q_H) > 0$. The radicand is positive precisely because $\overline{R} < \sigma_d^2$. This candidate is an equilibrium if and only if

$$R_H + \frac{\nu_{\text{one}}^2[R(q) - R_H]}{\nu_{\text{one}}^2 + R(q) - R_H} \geq R_*(q) \qquad \text{for every } 0 < q < q_H. \tag{73}$$

The left side is the candidate's residual risk at a lower scale. All higher-scale deviations are already ruled out by the definition of q_H, and participation is built into λ_{one}. For $q < q_H$, strict increase of χ_1 gives $R_*(q) > R_H$, and minimality of the benchmark maximizer gives $R(q) > R_*(q)$. Rearranging (73) therefore gives

$$\nu_{\text{one}}^2 \geq \frac{[R(q) - R_H][R_*(q) - R_H]}{R(q) - R_*(q)}.$$

Thus, without the additional connected-support restrictions, the exact single-scale test is

$$\nu_{\text{one}}^2 \geq \Gamma_H, \qquad \Gamma_H := \sup_{0 \leq q < q_H} \frac{[R(q) - R_H][R_*(q) - R_H]}{R(q) - R_*(q)}.$$

Including $q = 0$ in that supremum is equivalent by continuity to testing all strictly positive scales. The supremum may be infinite at a degenerate top. Under (74), the curvature argument in Appendix F.3.3 reduces this global test to $\lambda_{\text{one}} \leq \Lambda_H$. If it holds, the equilibrium has a single acquired scale. If it fails, general existence and uniqueness together with the support restrictions proved below imply that the equilibrium is the mixed candidate above, with a unique admissible lower endpoint. Without the additional primitive restrictions, the exact test remains (73), and a mixed equilibrium may have gaps.

F.3.3 Why the candidate forms are exhaustive

We now drop the candidate support assumption and consider an arbitrary positive-entry equilibrium. The indifference and pricing identities used above apply to it as well.

No interior atoms or singular continuous component.

Suppose there is an atom of mass $a > 0$ at $q < q_H$. At this point R_P and g are continuous, $g > 0$, and the tails on the two sides satisfy

$$\Pi_- = \Pi_+ + \frac{a}{R_P}, \qquad L_- = L_+ + \frac{ag}{R_P}, \qquad h_- = h_+ \frac{L_+}{L_+ + ag/R_P} < h_+.$$

Equation (60) then gives $\mathcal{J}'_+ > \mathcal{J}'_-$. A local maximum would require $\mathcal{J}'_- \geq 0 \geq \mathcal{J}'_+$, which is impossible. At q_H, however, $g = 0$, so this argument permits an atom there.

There is no singular continuous component below q_H either. Once atoms have been excluded, k and h are continuous on $(0, q_H)$ and R_P is C^1 there. At every point of the support, the first-order condition is

$$h(q) = \sqrt{\frac{2\gamma\chi'(q)R_P(q)}{j(q)}}.$$

The right side is C^1 on each compact subinterval of $(0, q_H)$. On complementary gaps, the tails Π, L, and thus k, are constant, and $h' = hk^2 j/v$ is uniformly bounded locally. These bounds and agreement with the displayed C^1 function on the support make h locally Lipschitz: decompose an interval at its first and last support points, bounding the two exterior gap pieces and the difference between those support points. But, in the sense of measures,

$$dh = \frac{gh}{R_P L}\mu(dq) + \frac{hk^2 j}{v}\, dq.$$

The coefficient on μ is strictly positive and bounded away from zero on compact subintervals below q_H. Local absolute continuity of h therefore implies local absolute continuity of μ.

The limiting acquisition action $q = 0$ cannot be optimal, even in the enlarged action space used for compactness. Its right marginal value is $j(0)h(0+)^2/(2\gamma R_P(0)) > 0$, since $\chi'_1(0) = 0$ by Assumption 2. By continuity, each equilibrium has a neighborhood of zero containing no occupied scales. Consequently $\mu(dq) = \rho(q)dq + \lambda_H \delta_{q_H}(dq)$ with an integrable density on $(0, q_H)$.

The top mass.

Suppose $q_H < 1$ and $\mathcal{J}_0''(q_H) < 0$. The expansion defining C_H above applies whether or not the support is connected. Without a top mass, $k(q) \to 0$ and $H(q) = o(q_H - q)$ as $q \uparrow q_H$. Then $H(q)^2/v(q) = o((q_H - q)^2)$ cannot equal $R(q) - R_*(q)$ along occupied scales approaching q_H. Since q_H is the supremum of the support, this contradiction proves $\lambda_H > 0$. If occupied lower scales accumulate at q_H, coefficient matching gives exactly (65). If there is a gap below the top, the no-deviation inequality $R_P \geq R_*$ instead gives the bound $\lambda_H \leq \Lambda_H$.

Primitive restrictions that rule out gaps.

For reference, the additional restrictions in Theorem 5 are $\chi_1 \in C^3([0,1])$ and

$$\sigma_\varepsilon^2 \leq \sigma_\beta^2 V_0, \qquad \chi_1''(q) > 0, \qquad \chi_1'''(q) < 2\sqrt{2\gamma}\,[\chi_1''(q)]^{3/2} \quad (0 < q < 1), \\ \chi_1'(1) > \frac{j(1)}{2\gamma R(1)}. \tag{74}$$

The two curvature inequalities hold for every $\chi_1(q) = \bar{\chi} q^2$ with $\bar{\chi} > 0$; the endpoint inequality then requires $\bar{\chi} > j(1)/(4\gamma R(1))$. We first reduce the required information curvature to the residual-risk quadratic. Use a_R, t_R, A_w, υ from Appendix D.1, and write $b_1 = b_R(1)$ and $\Delta_1 = \Delta_R(1)$. Solving that quadratic for υ gives

$$\upsilon = \frac{a_R}{A_w} R + \frac{b_1 - a_R t_R}{A_w} + \frac{c\kappa(1+\kappa)t_R}{R + t_R}, \qquad q = \frac{\kappa(1-\upsilon)}{\kappa + \upsilon}.$$

Differentiating this inverse relation yields the curvature identity

$$\mathcal{S}_R(q) := \frac{R'''(q)}{R'(q)} - \frac{3}{2}\left[\frac{R''(q)}{R'(q)}\right]^2 = \frac{6 a_R c \kappa^3 (1+\kappa)^3 A_w^3 t_R}{(\kappa + q)^4 \Delta_R(q)^4} \geq 0. \tag{75}$$

The sign follows from $t_R = \sigma_\beta^2 V_0 - \sigma_\varepsilon^2 \geq 0$; all other factors are positive. Thus the needed curvature is a consequence of an inequality on the original information parameters. The endpoint restriction in (74) is also entirely primitive:

$$\chi_1'(1) > \frac{\kappa A_w [R(1) + t_R]}{2\gamma(\kappa+1)\Delta_1 R(1)}, \qquad \begin{aligned} b_1 &= \kappa A_w (1-c) - \sigma_\varepsilon^2 (2 + c\kappa), \\ \Delta_1 &= \sqrt{b_1^2 + 4 a_R \sigma_\varepsilon^2 t_R}. \end{aligned}$$

Indeed, the right side equals $j(1)/(2\gamma R(1))$ by (42) and the quadratic at one.

Consider an interval with no acquired scales. The price loading k is constant there.

Starting from price moments H_b, v_b at its upper endpoint b, the residual risk is

$$r_f(q) = R(q) - \frac{[H_b + ky(q)]^2}{v_b + k^2 y(q)}, \qquad y(q) = R(q) - R(b).$$

Cancellation of the quadratic terms makes r_f a fractional-linear function of R, with $r_f > 0$ and $r_f' < 0$ on its feasible domain. The differential expression on the left of (75) is invariant under such a transformation, so $\mathcal{S}_{r_f} = \mathcal{S}_R$. Writing marginal information benefit and net marginal value as

$$\mathcal{B}_f = -\frac{1}{2\gamma}(\log r_f)' > 0, \qquad g_f = \mathcal{B}_f - \chi_1',$$

and differentiating gives

$$\mathcal{B}_f'' = \mathcal{S}_R \mathcal{B}_f + \frac{3(\mathcal{B}_f')^2}{2\mathcal{B}_f} + 2\gamma^2 \mathcal{B}_f^3.$$

At a stationary point of g_f, we have $\mathcal{B}_f' = \chi_1'' > 0$. Using (75) and minimizing over $\mathcal{B}_f > 0$ therefore gives

$$\begin{aligned} g_f'' &\geq \frac{3(\chi_1'')^2}{2\mathcal{B}_f} + 2\gamma^2 \mathcal{B}_f^3 - \chi_1''' \\ &\geq 2\sqrt{2\gamma}\,(\chi_1'')^{3/2} - \chi_1''' > 0. \end{aligned} \tag{76}$$

The second inequality is sharp, with equality in its minimization at $\mathcal{B}_f = \sqrt{\chi_1''/(2\gamma)}$. Thus net marginal value on any constant-loading interval has no interior local maximum.

This argument also applies to $r_f = R$, corresponding to zero loading. The derivative $\mathcal{J}_0'$ is positive at zero and negative at one by the endpoint restriction. Since every stationary point of $\mathcal{J}_0'$ is a strict local minimum, it crosses zero exactly once, with strictly negative derivative there. Consequently $q_H \in (0,1)$ is the unique maximizer and $\mathcal{J}_0''(q_H) < 0$.

Suppose an equilibrium has a gap (a,b) between two occupied scales, allowing $b = q_H$. The acquisition index takes its common maximum at both endpoints, and its derivative vanishes there because both endpoints are interior feasible scales. This also holds at q_H: $H(q_H) = 0$ makes $h(q_H-) = 1$, even in the presence of a top mass, so the left derivative equals $\mathcal{J}_0'(q_H) = 0$. On the gap, $\mathcal{J}' = g_f$, hence $g_f(a) = g_f(b) = 0$ and $\int_a^b g_f(q)\,dq = 0$. An identically zero g_f contradicts (76); otherwise the zero integral requires a positive interior maximum, which contradicts the same inequality. There can therefore be no gap, including

immediately below the top mass. When more than one scale is acquired, the support is a single interval $[q_L, q_H]$ with $q_L > 0$.

The density is strictly positive throughout the interior of this interval. To verify this stronger claim, indifference gives $R_P = R_*$ and $h = \sqrt{2\gamma\chi_1' R_*/j}$. Thus h is C^2, $u = H/\sqrt{v} = \sqrt{R - R_*} > 0$ is C^3, and $k/\sqrt{v} = (1-h)/u$ is C^2 locally. Using $v' = -k^2 j$ shows that k is C^2, and $k' = -\rho h/(R_P L)$ then gives a C^1, nonnegative version of the density. If $\rho(q_0) = 0$ at an interior point, nonnegativity implies $\rho'(q_0) = 0$, and hence $k'(q_0) = k''(q_0) = 0$. Freeze k at its value at q_0 and start the constant-loading posterior from the same H, v there. The equations for H, v show that its posterior-risk derivatives through order three agree with those of R_P at q_0. Its marginal benefit therefore satisfies $\mathcal{B}_f = \chi_1'$, $\mathcal{B}_f' = \chi_1''$, and $\mathcal{B}_f'' = \chi_1'''$ at that point, contradicting (76). This proves strict positivity. The general top-mass argument in Appendix F.3.3 now gives $\lambda_H = \Lambda_H > 0$.

It remains to simplify the single-scale test. For the candidate with mass λ_{one}, the left derivative of net marginal value at the top is

$$g_f'(q_H-) = \mathcal{J}''(q_H-) = \frac{C_H}{\gamma R_H}\left[\left(\frac{\lambda_{\text{one}}}{\Lambda_H}\right)^2 - 1\right].$$

If $\lambda_{\text{one}} > \Lambda_H$, slightly smaller scales strictly improve acquisition value, so the candidate cannot be an equilibrium. If $\lambda_{\text{one}} \le \Lambda_H$, then $g_f(q_H) = 0$ and $g_f'(q_H-) \le 0$. The function g_f is strictly positive immediately below q_H: this follows from the negative derivative when the inequality is strict, and from (76) when that derivative is zero. If g_f were nonpositive at any smaller scale, it would have a positive interior maximum before returning to zero at q_H, which is impossible. Hence acquisition value is strictly increasing on $(0, q_H)$ for this candidate, ruling out every lower-scale deviation. Higher-scale deviations are excluded by the definition of q_H, and the uninformed participation condition is built into λ_{one}. The candidate is therefore an equilibrium exactly when $\lambda_{\text{one}} \le \Lambda_H$. If this test fails, the existence and uniqueness results in Appendices F.3.5–F.3.6 imply that the equilibrium has the connected, strictly positive density and top mass just established.

Lower-scale deviations from the mixed candidate.

Under (74), no separate search over lower-scale deviations is needed. For any admissible branch truncated at q_L, let g_f be its net marginal value after extending k constantly below

q_L. At the endpoint $g_f(q_L) = 0$ and

$$g_f'(q_L-) = -\frac{2\mathcal{B}_f(q_L)\rho_*(q_L)g(q_L)}{R_*(q_L)L(q_L)} \leq 0.$$

This follows by comparing $h'/h = \rho g/(R_P L) + k^2 j/v$ on the indifferent branch with the same identity without the density term below q_L. If this derivative is negative, g_f is positive just below q_L; if it is zero, (76) gives the same conclusion. That inequality then prevents g_f from becoming nonpositive at a smaller scale, as in the single-scale argument. Thus acquisition value increases strictly toward q_L from below, proving (69). General existence and the connected-support proof ensure that the mixed regime has an admissible endpoint; strict monotonicity of m_*, U_* makes it unique. This proves the claimed classification, conditional on the general existence and uniqueness results established next.

F.3.4 The linear CREE at any population distribution

We verify the pricing identities used in the construction and prove that every population measure μ has a unique linear CREE. This result also supplies the continuity needed to establish general existence.

Fix any population measure μ. A linear equilibrium price must have a nonzero coefficient on η: otherwise all demands would be measurable with respect to the acquired information and could not clear an independent, nondegenerate supply shock. Its intercept is zero, since taking expectations in market clearing gives zero expected aggregate demand, whereas a nonzero intercept makes every type's expected demand have the opposite sign.

Use the normalized price (56) and its Gaussian posterior (57). The noise coefficient implies that the coefficient multiplying $\mathcal{X}$ in aggregate demand, after multiplication by γ, is $-K$. The coefficient on $dM(q)$ is then $\Pi(q) - k(q)L(q)$. The same calculation for the vector s gives $k_0 = \Pi(0)/L(0)$ and excludes any additional strong-signal loading, up to null directions. For the last identity in (58), differentiate $Lv - \Pi H$ in the sense of measures, using $H' = -kj$, $v' = -k^2 j$, $d\Pi = -\mu/R_P$, and $dL = -g\mu/R_P$. Every term cancels, and the terminal value is K. Positivity follows by continuation from $H(1) = 0$, $v(1) = 1$, and $L(1) = K$; a possible atom at one changes $\Pi(1)$ but not $L(1)$.

The uninformed posterior and the price level are

$$\begin{aligned} g_U &= \frac{k_0 A_s + H(0)}{k_0^2 A_s + v(0)}, & R_U &= \sigma_d^2 - \frac{[k_0 A_s + H(0)]^2}{k_0^2 A_s + v(0)}, \\ a_P &= \frac{L(0) + (1-\lambda) g_U / R_U}{\Pi(0) + (1-\lambda)/R_U} > 0. \end{aligned} \tag{77}$$

Consequently, the demand schedules, including those at unoccupied scales, are

$$\begin{aligned} x_q(p) &= \frac{M(q) + g(q)\{p/a_P - E^\infty[\mathcal{X} \mid \mathcal{F}_q^\infty]\} - p}{\gamma R_P(q)}, \\ x_U(p) &= \frac{g_U p/a_P - p}{\gamma R_U}. \end{aligned} \tag{78}$$

These formulas match all coefficients in market clearing.

For existence and uniqueness, use the tail system (59). The identity $m = R_P \Pi - Kg$ follows by differentiation and the terminal condition, including its jumps; eliminating Π, L gives the stated expression for k. The resulting ODE is a terminal-value system with a bounded measurable coefficient m and a locally Lipschitz vector field in (H, v) wherever its denominator is positive. In the descending variance coordinate $t = R(q) - R(1)$, its local solution satisfies

$$\frac{dR_P}{dt} = (1 - kg)^2, \qquad R_P \geq R(1) > 0, \qquad h = \frac{KR_P}{KR + mH} > 0.$$

Reconstructing $\Pi = (m + Kg)/R_P$ and $L = K/(hv)$ recovers $d\Pi = -\mu/R_P$ and $dL = -g\mu/R_P$, including their jumps. Thus $\Pi \leq 1/R(1)$, $L \geq K$, and

$$0 \leq k \leq \frac{1}{KR(1)}. \tag{79}$$

These bounds keep the solution finite on $[0, 1]$, prove global existence and uniqueness, and recover the original tails in (58). Bound (79) also applies to k_0. All residual risks are bounded below by $R(1)$, so the demands in (78) are jointly measurable and have an almost-surely finite population integral. For example, $M(q)$ and $E^\infty[\mathcal{X} \mid \mathcal{F}_q^\infty]$ have continuous paths, and their suprema on $[0, 1]$ are finite almost surely. Above the largest acquired scale $m = H = k = 0$, so the price uses only collectively acquired information and η. When $\lambda = 0$, these formulas give $P = -K\sigma_d^2 \eta$.

We also need continuity in the population distribution. Temporarily allow the limiting

acquisition action $q = 0$, distinct from remaining uninformed; it reveals the strong signals and costs χ_0. Let $\mu_n \to \mu$ weakly on $[0,1]$, with tails m_n and m and solutions (H_n, v_n) and (H, v) of (59). Weak convergence implies almost-everywhere and L^1 convergence of the tails and convergence of the total masses. Write $k(m, H, v) = (mv + KH)/(KR + mH)$ for the loading in (59). By (79), every solution stays in the region

$$0 \le H \le \bar{H} := \frac{R(0) - R(1)}{KR(1)}, \qquad 1 \le v \le \bar{v} := 1 + \frac{R(0) - R(1)}{[KR(1)]^2}.$$

On this region and for $0 \le m \le 1$, the denominator satisfies $KR + mH \ge KR(1)$. Hence k and k^2 are Lipschitz in (H, v) with a constant independent of m and q, and

$$\frac{\partial k}{\partial m} = \frac{K(vR - H^2)}{(KR + mH)^2}, \qquad |k(m_n, H, v) - k(m, H, v)| \le C_m |m_n - m|, \qquad C_m = \frac{\bar{v} R(0)}{KR(1)^2}.$$

The integral form of (59) and Gronwall's inequality then bound $\sup_q(|H_n - H| + |v_n - v|)$ by a constant multiple of $\int_0^1 |m_n - m|\, dq$, which tends to zero. This gives uniform convergence of H, v, and $R_P = R - H^2/v$. The formula for k_0 uses the total mass $m(0)$, so it remains continuous even if an atom appears at zero. Equation (77) therefore also gives continuity of R_U and a_P.

F.3.5 General existence

First restrict the acquisition menu to a finite grid of positive scales. Let q_H be the smallest grid maximizer of $\mathcal{J}_0$, and define $\overline{R}$ from it as at the start of Appendix F and R_* as in Appendix F.3.1; in this subsection these are the grid quantities. If $\overline{R} \ge \sigma_d^2$, the all-uninformed allocation is an equilibrium. Otherwise start at $s = q_H$ with no population, $H = \Pi = 0$, $v = 1$, and $L = K$. At each stage H, v, Π, L record the price moments and precision tails generated by previously assigned, higher-scale investors. The current scale s is indifferent to the top, with $R_s := R(s) - H^2/v = R_*(s)$. Adding mass $z \ge 0$ at s changes the loading below it to

$$k(z) = \frac{\Pi + z/R_s}{L + zH/(vR_s)}.$$

It changes neither the residual risk at s nor any risk at a higher scale. Before assigning investors below s, a lower grid scale q faces

$$R_P(q;z) = R(q) - \frac{[H + k(z)y]^2}{v + k(z)^2 y}, \qquad y = R(q) - R(s).$$

The uninformed risk has the same expression with $R(q)$ replaced by σ_d^2 and $y = \sigma_d^2 - R(s)$. The loading increases with z, and

$$\frac{\partial}{\partial k}\frac{(H+ky)^2}{v+k^2y} = \frac{2y(H+ky)(v-kH)}{(v+k^2y)^2} > 0$$

for $y > 0$ and positive loadings, since $v - kH > 0$. At the initial zero loading the derivative is zero, but the information gain increases strictly with every positive addition. Thus lower-scale and uninformed values rise monotonically with z.

Add mass until the first of three events: total informed mass reaches one; uninformed risk reaches $\overline{R}$; or a lower grid scale reaches $R_P(q;z) = R_*(q)$. In the first case stop with everyone informed; in the second stop with the remaining population uninformed. In the third case stop adding mass at s and move to the smallest lower scale attaining equality. A stopping event takes precedence over a simultaneous activation. Selecting the smallest tied scale ensures that subsequent additions leave the other tied scales' values unchanged. If z was added before moving to q, update

$$\begin{aligned} \Pi_{\text{new}} &= \Pi + z/R_s, & L_{\text{new}} &= L + zH/(vR_s), \\ H_{\text{new}} &= H + k(z)[R(q) - R(s)], & v_{\text{new}} &= v + k(z)^2[R(q) - R(s)]. \end{aligned}$$

There are at most as many activation steps as grid scales. At the smallest scale the procedure must stop through uninformed participation or the unit-population constraint. Every occupied scale remains indifferent to the top, and every unused grid scale is weakly worse because the next binding constraint was always the first event. The uninformed option is optimal if it has positive mass and weakly worse otherwise. The price and demand formulas (77) and (78) therefore complete a linear equilibrium on every finite grid.

Now take grids whose mesh, including the distances to zero and one, tends to zero. Adjoin the limiting acquisition action zero and keep the uninformed action as a separate isolated point. This enlarged action space is compact, so the grid population distributions have weakly convergent subsequences. Index the grids by n, and let G_n, μ_n, λ_n, $\mathcal{J}^n$, and $\mathcal{J}_U^n$ be the nth grid and its equilibrium population measure, informed mass, and indices (55).

Along such a subsequence, with limit μ of mass λ and indices $\mathcal{J}$ and $\mathcal{J}_U$, the continuity established in Appendix F.3.4 gives $\mathcal{J}^n \to \mathcal{J}$ uniformly on $[0,1]$ and $\mathcal{J}_U^n \to \mathcal{J}_U$. The grid top and the grid thresholds $\overline{R}$ and R_* vary with n, so we pass to the limit with the common index

$$\bar{\mathcal{J}}_n = \max\Big\{\max_{q \in G_n} \mathcal{J}^n(q),\ \mathcal{J}_U^n\Big\}.$$

Grid optimality is

$$\int_{[0,1]} \big[\bar{\mathcal{J}}_n - \mathcal{J}^n(q)\big]\, \mu_n(dq) + (1-\lambda_n)\big[\bar{\mathcal{J}}_n - \mathcal{J}_U^n\big] = 0,$$

with a nonnegative integrand and a nonnegative second term. Because $\mathcal{J}$ is continuous and the mesh tends to zero, $\bar{\mathcal{J}}_n \to \bar{\mathcal{J}} := \max\{\max_{[0,1]} \mathcal{J}, \mathcal{J}_U\}$. Uniform convergence of the indices and weak convergence of μ_n preserve the equality. The limiting distribution therefore assigns mass only to maximizers of $\mathcal{J}$ over all scales, including the uninformed option whenever its mass is positive. If $\lambda_n \to 0$, then $\mu = 0$ and the equality reduces to $\mathcal{J}(q) \le \mathcal{J}_U$ for every q, which is the no-acquisition equilibrium. The argument at zero in Appendix F.3.3 excludes mass on the added acquisition action. The limit is consequently an equilibrium of the original economy. This proves existence for the general costs of Assumption 2. The uniqueness argument below then implies that every subsequence has the same limit, so the entire sequence of grid equilibria converges weakly to it. No step requires a quadratic variable cost.

F.3.6 General uniqueness

Every positive-entry equilibrium has the same q_H, common informed index $\mathcal{J}_0(q_H)$, and target risk R_*. Run from q_H toward zero using the descending coordinate

$$t = R(q) - R_H, \qquad u = \frac{H}{\sqrt{v}}, \qquad \ell = \frac{k}{\sqrt{v}}, \qquad u_*(t) = \sqrt{R(q(t)) - R_*(q(t))}.$$

The no-deviation condition is $u \le u_*$; population can be added only at contact points $u = u_*$. The dynamics, with dots denoting differentiation in t, are

$$\dot{u} = \ell - \tfrac{1}{2}\ell^2 u, \qquad \dot{\ell} = a - \tfrac{1}{2}\ell^3, \qquad a := \frac{\dot{k}}{\sqrt{v}} \ge 0. \tag{80}$$

The controls are locally integrable, and $a = 0$ off the occupied set. The top mass is encoded in $u(0) = 0$, $v(0) = 1$, and $\ell(0) = \lambda_H/(KR_H)$. Throughout, $1 - \ell u = h > 0$.

Compare two candidate equilibria and write $w = u_1 - u_2$. On any interval where $w > 0$, feasibility gives $u_2 < u_1 \le u_*$, so the second distribution adds no population and $a_2 = 0$. At the left endpoint of a positive component, $w = 0$ and its right derivative is nonnegative. Because $\ell \mapsto \ell - \ell^2 u/2$ is strictly increasing on $\ell u < 1$, this implies $\ell_1 \ge \ell_2$ there. The second equation in (80), with $a_1 \ge 0$ and $a_2 = 0$, preserves this order. Moreover,

$$\dot{w} + \tfrac{1}{2}\ell_1^2 w = (\ell_1 - \ell_2)\left[1 - \tfrac{1}{2}u_2(\ell_1 + \ell_2)\right] \ge 0,$$

where the bracket is positive because $u_2 \le u_1$ and $\ell_i u_i < 1$. Thus a positive component of w cannot end at a finite later time. Interchanging the indices proves the analogous assertion for a negative component. The paths therefore either coincide everywhere or coincide initially and are strictly ordered at every smaller scale after they separate.

Let $t_0 \ge 0$ be the separation time and $T = R(0) - R_H$, and suppose $w > 0$ on $(t_0, T]$. Write m_i for the tail of μ_i. Since $a_2 = 0$ on $(t_0, T]$, μ_2 has no mass below $q(t_0)$, so $\lambda_2 = m_2(q(t_0))$. At t_0, $u_1 = u_2$ and $\ell_1 \ge \ell_2$. If $\ell_1(t_0) = \ell_2(t_0)$ and $a_1 = 0$ almost everywhere on $(t_0, T]$, both paths solve (80) with $a = 0$ from the same data at t_0. They then coincide on $(t_0, T]$, which contradicts separation. Hence either $\ell_1(t_0) > \ell_2(t_0)$, or $a_1 > 0$ on a set of positive measure in $(t_0, T]$. In the second case μ_1 has positive mass below $q(t_0)$, because $a = \dot{k}/\sqrt{v}$ and $k' = -\rho h/(R_P L)$.

Suppose first that $t_0 > 0$. On $[0, t_0]$ the paths $u_1 = u_2$ determine equal ℓ from the first equation in (80), because $\ell \mapsto \ell - \ell^2 u/2$ is strictly increasing on $\ell u < 1$ and ℓ is continuous below the top. They then determine equal v from $\dot{v} = \ell^2 v$, and hence equal H, k, m from (59). Thus $\ell_1(t_0) = \ell_2(t_0)$ and $m_1(q(t_0)) = m_2(q(t_0))$, and μ_1 has positive mass below $q(t_0)$. Suppose next that $t_0 = 0$, where $\ell_i(0) = \lambda_{H,i}/(KR_H)$ and $m_i(q_H) = \lambda_{H,i}$ for the top masses $\lambda_{H,i}$. Then either $\lambda_{H,1} > \lambda_{H,2}$, or $\lambda_{H,1} = \lambda_{H,2}$ and μ_1 has positive mass below q_H. In both cases,

$$\lambda_1 = m_1(q(t_0)) + \mu_1\big((0, q(t_0))\big) > m_2(q(t_0)) = \lambda_2.$$

Continue the free dynamics, with no further population, through the strong-block variance A_s. In this extension $k = k_0$ is constant, and the terminal residual risk is exactly R_U in (77). The same comparison preserves strict ordering of u and gives $R_{U,1} < R_{U,2}$. Since $\lambda_2 < \lambda_1 \le 1$, equilibrium 2 has uninformed investors and hence $R_{U,2} = \overline{R}$. Equilibrium 1

requires $R_{U,1} \geq \overline{R}$, a contradiction. Thus both distributions coincide; if the information paths coincide everywhere, the reconstruction above already gives this conclusion. The price and demands are then unique by Appendix F.3.4. The entry argument also excludes coexistence with an all-uninformed equilibrium.

General existence and uniqueness, the entry threshold, and the support classification together prove Theorem 5. In the mixed regime under (74), the unique equilibrium supplies an admissible endpoint for the construction, and the recovered density and top mass give (71). All population shares are absolute until normalization by λ; F remains the conditional scale distribution among informed investors. □

G Endogenous nesting with flexible information acquisition

We allow investors to acquire arbitrary subsets of signals in the limiting Gaussian economy, keeping the original scale index q. The scale economy restricts an investor to a prefix $[0, q]$; here the investor may also choose a bundle with gaps. Acquisition costs are common across investors and increase along the signal index. A block of signals of size dq contributes $j(q)\,dq$ to explained payoff variance, so its cost per unit of information is $\chi_1'(q)/j(q)$. We maintain the assumption that this ratio is nondecreasing: earlier signals are cheaper when compared with later signals providing the same amount of information. This ordering suggests buying the cheapest signals first, then the next cheapest, and continuing toward more expensive signals. An investor's scale specifies how far it proceeds in the ordering; larger bundles then contain every signal in smaller ones. Price feedback makes this ordering a substantive equilibrium result, since signals can differ in how much the price already reveals about them and in how they help interpret it. Under the maintained ordering, every linear equilibrium consists of prefixes, and the equilibrium obtained by choosing only scale remains an equilibrium under flexible acquisition.

Heuristic proof. Consider first a nested equilibrium in which investors choose cutoffs in the cost ordering, and a trader contemplating a bundle that skips some cheaper signals and acquires more expensive ones. Evaluate this bundle by adding its selected signals in increasing cost order. For each selected signal in an interval of occupied equilibrium cutoffs, compare the trader with an equilibrium investor whose bundle ends at that signal. For that investor, who has acquired all preceding signals, the marginal benefit of this signal equals its marginal acquisition cost, by the indifference condition in (60). The deviating trader has less preceding information. Signals are complements at the equilibrium price, so this

missing information lowers the marginal benefit of the same signal. The gaps make it harder to extract information from price and leave greater residual payoff risk, while the signal's acquisition cost is unchanged. Adding the later signal to the incomplete bundle therefore yields a weakly negative net marginal benefit. Lemma G.1 establishes the price properties behind this complementarity, and (91) gives its precise form.

Signals beyond the largest equilibrium bundle, whose scale is q_H, cannot overturn the comparison. No investor acquires them, so they are absent from price. For any given amount of this additional information, all selections have the same gross value, and the cheapest selection is the block immediately beyond q_H. Adding that block to the largest equilibrium bundle cannot improve its net value, because q_H is an optimal acquisition choice at the equilibrium price. Complementarity makes the block's gross benefit weakly smaller when it is added to an incomplete bundle of lower-cost signals. Thus acquiring unused signals cannot make a bundle with gaps profitable.

The same reasoning accommodates gaps in the distribution of equilibrium cutoffs. Within such a gap, the population acquiring each signal is unchanged, so every signal has the same price loading. A trader can therefore move any selected signals in the gap to its cheapest end, preserving their total explained payoff variance and hence their gross information value. For a gap starting at an occupied cutoff, acquiring this cheapest block cannot improve on the equilibrium bundle ending at that cutoff. Signals below the smallest occupied cutoff can have positive net benefits, but their combined contribution, after the entry fee, is bounded by the value of an optimal equilibrium bundle. The ordered marginal bound (92) combines these comparisons into a bound on the entire deviating bundle. Proposition G.4 gives the full argument, including gaps, atoms, and the case of no information acquisition.

This explains why a nested equilibrium survives flexible acquisition. To show that every flexible-acquisition equilibrium is nested, we also use strict complementarity to rule out crossing optimal bundles: two investors cannot optimally choose bundles that each contain signals missing from the other (Lemma G.2). Market clearing then implies that replacing an acquired expensive signal by a skipped cheaper one weakly raises gross value and weakly lowers cost. If the replacement leaves the investor indifferent, it creates crossing optimal bundles, which is itself impossible. Theorem G.3 formalizes this exchange argument, and Theorem G.5 combines the two directions to obtain equilibrium equivalence and uniqueness.

The formal construction below expresses signals in the innovation representation of Appendix D.3, which preserves the information relevant for trading at every scale. Appendix D.2 derives the underlying diffusion from the information primitives. The extension allows its

elementary signals to be acquired separately, charges acquisition costs in the original q coordinate, and maintains the ordering in (82). A change of index is used only in the proof to compare equally informative blocks.

G.1 From the information primitives to flexible acquisition

Retain the information parameters of Assumption 1, the cost function of Assumption 2, and the maintained conditions $V_0 > 0$, $\sigma_\varepsilon^2 > 0$, and $\sigma_\eta > 0$. Appendix D.1 derives the pre-price risk curve $R(q)$ from these parameters. Lemma D.3 derives joint Gaussian forecast increments from the finite-economy posteriors, and Proposition D.4 gives

$$M(q) = G + \int_0^q \sqrt{j(u)}\, dW_u, \qquad d^\infty = M(1) + \sqrt{R(1)}\, Z, \qquad j = -R' > 0,$$

where $G = b's$ has variance A_s, and G, W, Z, η are independent. Thus the diffusion and its variance rate are determined by the original information primitives. Because $j > 0$, the transformation between the forecast path and its innovations is invertible:

$$\mathcal{F}_q^\infty = \sigma\big(s, M(u) : 0 \le u \le q\big) = \sigma\big(s, W_u : 0 \le u \le q\big).$$

Appendix D.3 establishes the corresponding equivalence of conditional trading problems.

An investor chooses a measurable set $B \subseteq [0,1]$ of elementary signals $dM(q)$ before observing their realizations or price. A nonempty bundle also gives access to the fixed strong-signal vector s. The investor observes each acquired signal. To express this in the continuum, record the cumulative sum of acquired signal increments up to every cutoff x:

$$\mathcal{I}_B := \sigma\left(s, \int_{B\cap[0,x]} dM(u) : 0 \le x \le 1\right), \qquad |B| > 0.$$

Here x runs through the signal ordering. The integral adds only the signals in B whose indices are at most x; differences between two cutoffs reveal the acquired increments between them. Observing these sums at every cutoff therefore records the whole collection of acquired signals. The empty set denotes nonacquisition and conveys no private signals. Buying the same signal reveals the same realization to every investor. Disjoint sets of elementary signals reveal independent weak payoff components, and the component revealed by B has variance $\int_B j(q)\, dq$. In particular, a prefix $B = [0, q]$ with $q > 0$ gives precisely the information $\mathcal{F}_q^\infty$.

All set comparisons are modulo Lebesgue-null sets, and acquisition allocations are jointly measurable in investor and signal index.

The acquisition cost is stated directly per unit of the original signal index:

$$C(B) = \chi_0 \mathbb{1}_{\{|B|>0\}} + \int_B \chi_1'(q)\, dq. \tag{81}$$

Thus a nonempty prefix $[0, q]$ costs $\chi_0 + \chi_1(q) = \chi(q)$. The variable charge on a block dq is $\chi_1'(q)\, dq$. Throughout this appendix, we maintain the additional assumption

$$\frac{d}{dq}\left(\frac{\chi_1'(q)}{j(q)}\right) = \frac{\chi_1''(q)j(q) + \chi_1'(q)R''(q)}{j(q)^2} \geq 0, \qquad 0 < q < 1. \tag{82}$$

This is a restriction on costs relative to the information supplied by each signal block. Since $j(q) = -R'(q)$ is determined before observing price, the restriction involves only acquisition costs and the information primitives. It ensures that exchanging a later block for an equally informative earlier block does not raise acquisition cost. Increasing marginal cost per unit of q alone need not imply this ordering.

Sufficient conditions. Under the maintained information assumptions, including the Gaussian coefficient prior and inverse-Wishart covariance prior in Assumption 1, either of the following conditions guarantees (82):

1. *Quadratic costs and an endpoint restriction.* If $\chi_1(q) = \bar{\chi} q^2$, with $\bar{\chi} > 0$, it suffices that

$$j(1) - j'(1) = -R'(1) + R''(1) \geq 0. \tag{83}$$

   Under these maintained assumptions, the endpoint restriction is also necessary for (82) on the entire interval when costs are quadratic. The baseline calibration satisfies (83).[31]

2. *General increasing, convex costs.* For any χ_1 satisfying Assumption 2, it suffices that

$$\sigma_\varepsilon^2 \geq \kappa A_w = \sigma_\beta^2 V_0. \tag{84}$$

The proof below verifies both conditions from the derivatives of R. Neither is imposed separately when (82) holds directly.

[31] Evaluating the risk derivatives at the parameters in Table 2, including its zero-noise specification $\sigma_\varepsilon^2 = 0$, gives $j(1) - j'(1) \approx 6.87 \times 10^{-5} > 0$

The flexible economy retains the unit mass of identical investors, common risk aversion γ, common beliefs, and supply $\sigma_\eta\eta$. Write $K = \gamma\sigma_\eta$ as in Appendix F. Equilibrium requires optimal acquisition over all measurable B, optimal trading given $(\mathcal{I}_B, P)$, and market clearing. Prices are linear in the collectively acquired signals and the supply shock. All expectations are under the common limiting law $\mathbb{P}_\infty$.

G.2 Proof preparation and market clearing

Verification of the sufficient conditions. Use $a_R, t_R, \Delta_R, \upsilon$ from Appendix D.1. Differentiating the risk quadratic gives

$$R_{\upsilon\upsilon} = -\frac{2c\kappa(1+\kappa)A_w^3 t_R}{\Delta_R^3}, \qquad t_R = \kappa A_w - \sigma_\varepsilon^2.$$

If $t_R \le 0$, then $R_{\upsilon\upsilon} \ge 0$; since $R_\upsilon > 0$ and $\upsilon'' > 0$, (42) implies $R'' > 0$. Together with $\chi_1' \ge 0$ and $\chi_1'' \ge 0$, this proves (82) for every cost function in Assumption 2. If $t_R > 0$, the curvature identity (75) and $R' < 0$ imply $R''' < 0$. Consequently,

$$\frac{d}{dq}\big[j(q) - qj'(q)\big] = qR'''(q) \le 0,$$

so $j(q) - qj'(q)$ attains its minimum at $q = 1$. Under quadratic costs, the derivative in (82) is $2\bar{\chi}[j(q) - qj'(q)]/j(q)^2$. Combining the cases $t_R \le 0$ and $t_R > 0$ proves the equivalence with (83).

A change of index for the proof. The proof exchanges blocks with the same contribution to explained payoff variance. To express such exchanges as exchanges of equal-length intervals, define

$$t(q) := \sigma_d^2 - R(q) = A_s + \int_0^q j(u)\,du, \qquad \bar{t} := t(1). \tag{85}$$

Here $t(q) = \mathrm{Var}^\infty[M(q)]$ is the unconditional variance of the forecast based on private signals, with no conditioning on price, and $dt = j(q)\,dq$. This deterministic change of index is used only for the proof; the acquisition problem and its costs remain those in (81). Independent forecast increments have variance equal to their increment in t, so we may write

$$M(q) = \widetilde{W}_{t(q)}, \qquad d^\infty = \widetilde{W}_{\bar{t}} + \sqrt{R(1)}\,Z,$$

for a standard Brownian motion $\widetilde{W}$. For $t \geq A_s$, this is an invertible relabelling of the forecast path. When $A_s > 0$, complete $G = \widetilde{W}_{A_s}$ to a path on $[0, A_s]$ using an independent Brownian bridge; this initial interval represents the strong block. The bridge is auxiliary and will be shown to convey no additional information about payoff or price at equilibrium. For $A_s = 0$ it is unnecessary.

We use A for sets in the proof coordinate, reserving B for the original signal sets. A nonempty bundle B corresponds to

$$A(B) = [0, A_s] \cup t(B), \qquad t(B) := \{t(q) : q \in B\},$$

and nonacquisition corresponds to $A = \varnothing$. As in the original coordinate, the information generated by a proof-coordinate set records the cumulative sum of selected increments up to every cutoff:

$$\widetilde{\mathcal{I}}_A := \sigma\left(\int_{A\cap[0,x]} d\widetilde{W}_t : 0 \leq x \leq \bar{t}\right).$$

Each interval of length δ reveals a payoff component with variance and payoff covariance δ. The corresponding cost is

$$\widetilde{C}(A) = \chi_0 \mathbb{1}_{\{|A|>0\}} + \int_A \mathfrak{c}(t)\, dt, \qquad \mathfrak{c}(t) = \begin{cases} 0, & 0 \leq t < A_s, \\ \dfrac{\chi_1'(q)}{j(q)}, & t = t(q), \quad 0 \leq q \leq 1. \end{cases} \tag{86}$$

Thus $\widetilde{C}(A(B)) = C(B)$, because $\mathfrak{c}(t(q))\, dt = \chi_1'(q)\, dq$. Condition (82) makes $\mathfrak{c}$ nondecreasing, including across the free strong block since $\chi_1'(0) = 0$. For the proof we temporarily allow all measurable $A \subseteq [0, \bar{t}]$, including partial selections from the strong block. After establishing positive marginal information values, we show that this enlargement leaves equilibrium choices unchanged.

Consider any linear trading equilibrium at an arbitrary acquisition allocation. Its supply-shock loading is nonzero, since otherwise aggregate demand could not clear independent random supply. Its intercept is zero by the expected market-clearing equation, as in Appendix F.3.4. Normalize the price as

$$P = a_P \mathcal{X}, \qquad \mathcal{X} = \int_0^{\bar{t}} \widetilde{k}(t)\, d\widetilde{W}_t - \eta, \qquad \widetilde{k} \in L^2([0, \bar{t}]). \tag{87}$$

At this stage no sign restriction is imposed on a_P or $\widetilde{k}$. The tilde distinguishes the loading in the variance coordinate t from the scale-coordinate loading $k(q)$ in (56). For any acquired or counterfactual set A, define

$$H_A = \int_{A^c} \widetilde{k}(t)\,dt, \qquad v_A = 1 + \int_{A^c} \widetilde{k}(t)^2\,dt, \quad g_A = \frac{H_A}{v_A},$$
$$R_P(A) = R(1) + |A^c| - \frac{H_A^2}{v_A}.$$

The complement is relative to $[0, \bar{t}]$. Gaussian conditioning gives

$$E^\infty[d^\infty \mid \widetilde{\mathcal{I}}_A, P] = \int_A d\widetilde{W}_t + g_A\left(\mathcal{X} - \int_A \widetilde{k}(t)\,d\widetilde{W}_t\right),$$
$$\mathrm{Var}^\infty[d^\infty \mid \widetilde{\mathcal{I}}_A, P] = R_P(A), \qquad R(1) \le R_P(A) \le \sigma_d^2.$$

Consequently the acquisition index, in the same certainty-equivalent units as (55), is

$$\mathcal{J}(A) = U(A) - \widetilde{C}(A), \qquad U(A) = -\frac{\log R_P(A)}{2\gamma}.$$

The omitted term $\log \mathrm{Var}^\infty[d^\infty - P]/(2\gamma)$ is common to all sets.

Index investors by $i \in [0,1]$, with acquired sets A_i, and define

$$d\nu(i) = \frac{di}{R_P(A_i)}, \qquad \Pi(t) = \int \mathbb{1}_{\{t\in A_i\}}\,d\nu(i), \qquad L(t) = K + \int \mathbb{1}_{\{t\in A_i\}} g_{A_i}\,d\nu(i).$$

The notation Π continues to denote precision-weighted acquisition, while F is reserved for the scale CDF. Comparing source and supply coefficients in aggregate CARA demand yields

$$\widetilde{k}(t)L(t) = \Pi(t), \qquad a_P = \frac{K + \int g_{A_i}\,d\nu(i)}{\int d\nu(i)}. \tag{88}$$

These equations use the actual acquisition sets and do not presume that they are prefixes.

Lemma G.1 (Positive residual source loadings) *At any linear trading equilibrium,*

$$0 \le \widetilde{k}(t) \le \frac{1}{KR(1)}, \qquad a_P > 0.$$

Let $g_U = g_\varnothing$ and $v_U = v_\varnothing$. Then

$$u(t) := 1 - g_U \widetilde{k}(t) > 0 \quad \text{for almost every } t, \tag{89}$$

and $1 - g_A \widetilde{k}(t) \ge u(t) > 0$ for every measurable set A.

Proof of Lemma G.1. Equation (88) implies that $\Pi(t) = 0$ if and only if $\widetilde{k}(t) = 0$: when $\Pi(t) = 0$, the acquiring population has zero ν-measure, so $L(t) = K$. On the active sources, where $\Pi(t) > 0$, put $z(t) = 1/\widetilde{k}(t)$. Then

$$z(t) = \frac{K}{\Pi(t)} + \frac{\int \mathbb{1}_{\{t \in A_i\}} g_{A_i}\, d\nu(i)}{\Pi(t)}.$$

Cauchy–Schwarz gives $|g_A| \le \sqrt{\bar{t}}/2$, so the essential infimum m of z over active sources is bounded below. If $m \le 0$, then for every A,

$$g_A = \frac{\int_{A^c} \widetilde{k}(t)^2 z(t)\, dt}{v_A} \ge m \frac{v_A - 1}{v_A} \ge m \left(1 - \frac{1}{v_U}\right),$$

where the integrand on inactive sources is defined as zero. Since $\Pi(t) \le 1/R(1)$, taking essential infima in the expression for z gives

$$m \ge KR(1) + m \left(1 - \frac{1}{v_U}\right),$$

contradicting $m \le 0$. Thus $\widetilde{k}$ is positive on active sources and zero elsewhere. All filters g_A are now nonnegative; (88) gives the loading bound and $a_P > 0$. If there are no active sources, these conclusions and (89) hold immediately.

For the remaining claim, observe that

$$g_U - g_A = \frac{\int_A \widetilde{k}(s) u(s)\, ds}{v_A}.$$

Substitution into market clearing gives

$$\Pi(t)u(t) + \widetilde{k}(t) \int_{i:t\in A_i} \frac{\int_{A_i} \widetilde{k}(s)u(s)\,ds}{v_{A_i}}\,d\nu(i) = K\widetilde{k}(t).$$

Multiplying by $\widetilde{k}(t)$ and integrating over sources shows that

$$\int \frac{\int_{A_i} \widetilde{k}(s)u(s)\,ds}{v_{A_i}}\,d\nu(i) = \frac{K(v_U - 1)}{v_U}.$$

Subtracting $\widetilde{k}(t)$ times this identity from the previous equation yields

$$\Pi(t)u(t) - \widetilde{k}(t) \int_{i:t\notin A_i} \frac{\int_{A_i} \widetilde{k}(s)u(s)\,ds}{v_{A_i}}\,d\nu(i) = \frac{K\widetilde{k}(t)}{v_U}. \tag{90}$$

Suppose $N = \{t : u(t) < 0\}$ has positive measure, and write $w = -u > 0$ there. All sources in N are active. Since $u \geq 0$ outside N, (90) implies, on N,

$$\Pi(t)w(t) - \widetilde{k}(t) \int_{i:t\notin A_i} \frac{\int_{A_i\cap N} \widetilde{k}(s)w(s)\,ds}{v_{A_i}}\,d\nu(i) < 0.$$

Multiply by $\widetilde{k}(t)$ and integrate over N. Reversing the order of integration makes the left side

$$\int \left[1 - \frac{\int_{N\setminus A_i} \widetilde{k}(t)^2\,dt}{v_{A_i}}\right] \left[\int_{A_i\cap N} \widetilde{k}(s)w(s)\,ds\right] d\nu(i) \geq 0,$$

because the first bracket is at least $1/v_{A_i} > 0$. This contradicts strict negativity. Thus $u \geq 0$ almost everywhere. If $u(t) = 0$ at an active source, the left side of (90) is nonpositive and the right side is positive. At an inactive source, $u(t) = 1$. Hence $u > 0$ almost everywhere. Finally, $g_U - g_A = \int_A \widetilde{k}u/v_A \geq 0$, which proves the assertion for every A. Moreover, (90) gives $u(t) \geq K/[v_U L(t)]$ on active sources. Since $L(t) \leq K + \sqrt{t}/[2R(1)]$, this is a uniform positive lower bound; on inactive sources $u(t) = 1$. □

Equivalence of the proof representation. The representation preserves trading equilibria for any original acquisition allocation: set $A_i = A(B_i)$ for entrants and $A_i = \varnothing$ for nonentrants. Everyone who observes the strong block observes it in full, so market clearing makes $\widetilde{k}$ constant on $[0, A_s]$. The price depends on this interval only through $\widetilde{W}_{A_s} = G$. The

auxiliary bridge is independent of payoff and price; the other components of s are likewise irrelevant conditional on G. Thus conditional trading problems and costs coincide with those in the original coordinate.

The enlargement to arbitrary A introduces no additional equilibrium choices. By Lemma G.1, the marginal reduction in residual risk from acquiring a source is $[1 - \widetilde{k}(t)g_A]^2 > 0$. Completing any nonempty set with the free strong block therefore weakly improves its value, strictly if a positive-measure part of that block is missing. The strong block alone is also suboptimal: the uniform bound just established and $j(0) > 0$ give a positive lower bound on the marginal gross value of weak signals near $q = 0$, while $\chi_1'(0) = 0$. Hence every nonempty optimal A consists of the full strong block and a positive-measure set of weak signals, and corresponds to an original bundle B. These comparisons also apply at a trading equilibrium generated by an original allocation, since Lemma G.1 uses only trading optimality and market clearing. Thus unrestricted acquisition optimality is equivalent in the two representations. For prefix-restricted acquisition, the same argument excludes every additional prefix ending at or below A_s. The remaining prefixes are exactly $A = [0, t(q)]$, corresponding to $B = [0, q]$ with cost $\chi(q)$.

At any such allocation, the relation to the original price notation is

$$\widetilde{k}(t) = k_0 \quad (0 \le t < A_s), \qquad \widetilde{k}(t(q)) = k(q) \quad (0 < q \le 1).$$

The normalized price (87) is therefore (56), and $R_P([0, t(q)]) = R_P(q)$. The acquired forecast path at a prefix gives the same trading information as $\mathcal{F}_q^\infty$ by Appendix D.3.

G.3 Complementarity and endogenous nesting

For an algebraic interpolation $a : [0, \bar{t}] \to [0, 1]$ of an acquired-set indicator, define $H_a, v_a, g_a, R_P(a)$ by replacing $\mathbb{1}_{A^c}$ with $1 - a$ in the preceding formulas. The identity

$$g_U - g_a = \frac{\int_0^{\bar{t}} a(t)\widetilde{k}(t)u(t)\,dt}{v_a} \ge 0$$

extends Lemma G.1 to these interpolations. Write $h_a(t) = 1 - \widetilde{k}(t)g_a > 0$. The marginal gross value of source t is

$$\mathcal{M}(t; a) = \frac{h_a(t)^2}{2\gamma R_P(a)}.$$

Differentiating the posterior moments gives

$$\frac{\delta R_P(a)}{\delta a(t)} = -h_a(t)^2, \qquad \frac{\delta g_a}{\delta a(t)} = -\frac{\widetilde{k}(t)h_a(t)}{v_a},$$

and hence

$$\frac{\delta \mathcal{M}(t;a)}{\delta a(s)} = \frac{1}{2\gamma}\left[\frac{2\widetilde{k}(t)\widetilde{k}(s)h_a(t)h_a(s)}{v_a R_P(a)} + \frac{h_a(t)^2 h_a(s)^2}{R_P(a)^2}\right] > 0. \tag{91}$$

The second term is strictly positive even for sources absent from price. This interpolation compares feasible source sets; it does not add a fractional acquisition technology.

Lemma G.2 (Optimal sets form a chain) *At any equilibrium price, two optimal acquisition sets are comparable by inclusion, up to null sets.*

Proof of Lemma G.2. If $|A \setminus D| > 0$ and $|D \setminus A| > 0$, integrate (91) over the two set differences to obtain

$$U(A \cup D) + U(A \cap D) > U(A) + U(D).$$

Additive variable costs satisfy equality in the corresponding set identity. The nonnegative entry fee satisfies $\widetilde{C}(A) + \widetilde{C}(D) \geq \widetilde{C}(A \cup D) + \widetilde{C}(A \cap D)$. Consequently,

$$\mathcal{J}(A \cup D) + \mathcal{J}(A \cap D) > \mathcal{J}(A) + \mathcal{J}(D).$$

If A and D both attain the common optimal value $\mathcal{J}^*$, the right side is $2\mathcal{J}^*$, while each term on the left is at most $\mathcal{J}^*$. This is impossible. □

Identical preferences, beliefs, and acquisition costs are essential here: all investors maximize the same net acquisition index. Complementarity alone would not order the choices of investors with different objectives. Lemma G.2 does not use (82): nesting by inclusion follows from complementarity and common additive costs. The cost ordering is needed to identify the nested bundles with prefixes in the original q coordinate.

Theorem G.3 (Every flexible-acquisition equilibrium is nested) *Under (82), every linear equilibrium of the flexible-acquisition economy has, for almost every investor i,*

$$B_i = \varnothing \quad \text{or} \quad B_i = [0, q_i], \qquad q_i \in (0, 1],$$

up to null sets. Thus acquisitions are nested in the original signal order. Weak cost ordering suffices; strict cost differences are unnecessary.

Proof of Theorem G.3. Work in the equivalent proof representation. By Lemma G.2, the optimal sets chosen by investors are pairwise comparable. Fix such a set A, an included source $s \in A$, and an excluded source $t \notin A$. Every other chosen set containing t must contain A, hence also s. Joint measurability makes this population inclusion valid for almost every such pair (s, t). Using (88),

$$\begin{aligned}[\widetilde{k}(s) - \widetilde{k}(t)]L(s) &= \Pi(s) - \Pi(t) - \widetilde{k}(t)[L(s) - L(t)] \\ &= \int_{i:s\in A_i,\ t\notin A_i} [1 - \widetilde{k}(t) g_{A_i}]\, d\nu(i) \geq 0.\end{aligned}$$

The last inequality follows from Lemma G.1, and $L(s) \geq K > 0$. Thus the price loading of an included source is at least that of an excluded source.

If A is not a prefix, there are equal-positive-measure sets $D_+ \subseteq A$ and $D_- \subseteq A^c$ such that every source in D_- precedes every source in D_+. For $t \in D_-$ and $s \in D_+$, we then have $\mathfrak{c}(t) \leq \mathfrak{c}(s)$ and $\widetilde{k}(t) \leq \widetilde{k}(s)$. Consider the exchange

$$a_\theta = \mathbb{1}_A - \theta \mathbb{1}_{D_+} + \theta \mathbb{1}_{D_-}, \qquad 0 \leq \theta \leq 1.$$

Its gross-value derivative is

$$\frac{dU(a_\theta)}{d\theta} = \frac{1}{2\gamma R_P(a_\theta)} \left[\int_{D_-} h_{a_\theta}(t)^2\, dt - \int_{D_+} h_{a_\theta}(s)^2\, ds \right] \geq 0.$$

Indeed, $g_{a_\theta} \geq 0$ and $h_{a_\theta} > 0$, so h_{a_θ} decreases with the price loading. Variable costs weakly fall and the entry fee is unchanged. The exchanged set $A' = (A \setminus D_+) \cup D_-$ therefore satisfies $\mathcal{J}(A') \geq \mathcal{J}(A) = \mathcal{J}^*$. A strict inequality contradicts optimality; equality makes A and A' incomparable optimal sets, contradicting Lemma G.2. Thus every chosen proof-coordinate set is a prefix. Every nonempty chosen prefix ends at $t(q_i)$ for some $q_i > 0$, by the equivalence established above; its original signal set is $B_i = [0, q_i]$. □

G.4 Equilibrium equivalence and uniqueness

To establish equivalence, we must also exclude arbitrary-subset deviations at the equilibrium of the prefix-restricted economy. The following argument includes gaps and atoms in the distribution of acquired scales.

Proposition G.4 (Preservation under flexible acquisition) *Under (82), every linear equilibrium of the scale economy, which restricts acquisition to prefixes $B = [0, q]$, is a linear equilibrium of the flexible-acquisition economy.*

Proof of Proposition G.4. Fix a scale-equilibrium price and use its equivalent proof representation, with prefixes $A = [0, t(q)]$. Lemma G.1 applies because its proof uses only trading optimality and market clearing. Thus (91) also holds at this price. Building any set A in increasing source order gives

$$U(A) - U(\varnothing) = \int_A \mathcal{M}(t; \mathbb{1}_{A\cap[0,t]})\, dt \leq \int_A \mathcal{M}(t; \mathbb{1}_{[0,t]})\, dt.$$

The inequality follows from complementarity and $A \cap [0, t] \subseteq [0, t]$. Bounded price loadings and $R_P(A) \geq R(1) > 0$ justify the absolutely continuous integration. Define the continuous prefix index, charging the entry fee also at its virtual empty endpoint, by

$$J(t) = U([0, t]) - \chi_0 - \int_0^t \mathfrak{c}(s)\, ds, \qquad 0 \leq t \leq \bar{t}.$$

The actual uninformed value is $U(\varnothing) = J(0) + \chi_0$. For every nonempty A, the preceding bound yields

$$\mathcal{J}(A) \leq J(0) + \int_A J'(t)\, dt. \tag{92}$$

Suppose first that entry is positive. Let E be the closed support of the population distribution of acquired prefix lengths, and let $\mathcal{J}^*$ be their common optimal value. Restricted optimality gives $J \leq \mathcal{J}^*$ everywhere and $J = \mathcal{J}^*$ on E, so $J' = 0$ almost everywhere on E. On each component of $[0, \bar{t}] \setminus E$, the acquiring population is constant; therefore (88) makes $\widetilde{k}$ constant there. Replace the portion of A in each such gap by the leftmost interval of the same length. This replacement preserves $|A|$ and the acquired integrals of $\widetilde{k}$ and $\widetilde{k}^2$, hence preserves $R_P(A)$. Since $\mathfrak{c}$ is nondecreasing, it weakly lowers cost. Call the resulting set $\widehat{A}$.

For a gap with left endpoint $a \in E$ and acquired length ℓ,

$$\int_a^{a+\ell} J'(t)\,dt = J(a+\ell) - J(a) \le 0.$$

This includes the terminal gap above the largest acquired prefix. An initial gap starting at zero contributes at most $\mathcal{J}^* - J(0)$; if there is no initial gap, $J(0) = \mathcal{J}^*$. Summing over the countably many gaps and using (92) gives

$$\mathcal{J}(A) \le \mathcal{J}(\widehat{A}) \le J(0) + \int_{\widehat{A}} J'(t)\,dt \le \mathcal{J}^*.$$

The uninformed action was already a feasible restricted deviation, so acquisition remains optimal. If all investors are uninformed, $\widetilde{k} = 0$ and gross value depends only on $|A|$. A prefix minimizes cost at any given measure, so the restricted no-entry conditions exclude every subset deviation. Trading optimality and market clearing are unchanged. □

Theorem G.5 (Equivalence and uniqueness) *Under the maintained information primitives, Assumption 2, and (82), the flexible-acquisition economy and the scale economy have the same linear equilibrium allocations, prices, and trading strategies, up to payoff-irrelevant information. The equilibrium is essentially unique: the informed fraction λ, the acquired-set distribution induced by $\lambda\,dF(q)$ through $B = [0, q]$, and the price and optimal demands are unique, with the same almost-sure conventions as Theorem 5. There are no additional nonnested linear equilibria.*

Proof of Theorem G.5. Theorem G.3 maps every flexible equilibrium to a prefix equilibrium. The preceding identification maps each nonempty equilibrium prefix to its original scale $q > 0$, with identical cost, posterior beliefs, and trading opportunities. Conversely, Proposition G.4 excludes every arbitrary-subset deviation from a restricted equilibrium. Existence and essential uniqueness now follow from Theorem 5, proved in Appendices F.3.5 and F.3.6. No separate uniqueness argument or connected-support restriction is required. □

The result concerns flexible acquisition of signals in the limiting Gaussian representation. Their diffusion law is derived from the primitives in Appendix D.2. The information equivalence above holds at each scale; it does not identify arbitrary subsets of the original finite-dimensional raw signals with subsets of the elementary signals in this representation. It also does not cover nonlinear-price equilibria, investor-specific acquisition costs or risk

aversions, adaptive acquisition, or the design of new linear-combination signals. Common source realizations, identical investors, additive ordered costs, and the positive lower bound $R(1)$ on residual risk enter the proof explicitly.

H Acquisition-cost comparative statics

We prove Propositions 6 and 7 using the distribution and pricing formulas in Appendix F. Throughout, $\chi(q) = \chi_0 + \bar{\chi}q^2$, $K = \gamma\sigma_\eta$, and R is fixed by the information primitives.

H.1 Proof of Proposition 6

Proof of Proposition 6. Quadratic costs satisfy the curvature restrictions in Theorem 5, and $\bar{\chi} > -R'(1)/(4\gamma R(1))$ is its interior-endpoint condition because $\chi_1'(1) = 2\bar{\chi}$. Together with $\sigma_\varepsilon^2 \le \sigma_\beta^2 V_0$, these assumptions give the connected-support characterization used below. The maximization defining q_H is unchanged by adding the constant $-\chi_0$ to $\mathcal{J}_0$. The support indifference curve

$$R_*(q) = R(q_H)e^{2\gamma\bar{\chi}(q_H^2 - q^2)}$$

is also independent of χ_0. Consequently the functions in (62), the density and tail in (63), and the top mass Λ_H in (64) do not depend on the fixed cost. Only the entry target

$$\overline{R} = R(q_H)e^{2\gamma(\chi_0 + \bar{\chi}q_H^2)}$$

changes.

On a mixed branch with uninformed investors, the lower endpoint solves $U_*(q_L) = \overline{R}$ and $\lambda = m_*(q_L) < 1$. Let $\chi_0^a < \chi_0^b$ be two costs on this branch, with endpoints q_L^a, q_L^b and targets $\overline{R}^a < \overline{R}^b$, and put $q_m = \min\{q_L^a, q_L^b\}$. By Theorem 5, the equilibrium with endpoint q_m has density $\rho_* > 0$ on (q_m, q_H). The monotonicity argument following (71) then gives $U_*' > 0$ and $m_*' = -\rho_* < 0$ on (q_m, q_H). Since $U_*(q_L^a) = \overline{R}^a < \overline{R}^b = U_*(q_L^b)$, it follows that $q_L^a < q_L^b < q_H$ and

$$\lambda^a - \lambda^b = \int_{q_L^a}^{q_L^b} \rho_*(q)\, dq > 0.$$

Thus a fall in the fixed cost extends the same population distribution toward smaller scales; its density at retained scales and its absolute top mass remain unchanged. Where $\rho_*(q_L) > 0$,

also $U_*'(q_L) > 0$, and implicit differentiation gives

$$\frac{dq_L}{d\chi_0} = \frac{2\gamma\overline{R}}{U_*'(q_L)} > 0, \qquad \frac{d\lambda}{d\chi_0} = -\rho_*(q_L)\frac{2\gamma\overline{R}}{U_*'(q_L)} < 0. \tag{93}$$

On the common support of the two equilibria,

$$F^a(q) = 1 - \frac{m_*(q)}{\lambda^a} \geq 1 - \frac{m_*(q)}{\lambda^b} = F^b(q).$$

Between the two lower endpoints, $F^b = 0 \leq F^a$; at and above q_H, both CDFs are one. This also shows that the conditional distribution shifts toward smaller scales in first-order stochastic dominance, and Λ_H/λ decreases as entry expands.

For completeness, the signs extend across changes of regime. Above the entry threshold (28), $\lambda = 0$. When a single scale is acquired, the interior mass is

$$\lambda_{\text{one}} = K\sqrt{\frac{R_H}{e^{2\gamma(\chi_0+\bar{\chi}q_H^2)}-1} - \frac{R_H^2}{\sigma_d^2 - R_H}}, \qquad R_H = R(q_H),$$

capped at one; it is strictly decreasing in χ_0 before the cap binds. Under the additional restrictions of Theorem 5, the single-scale equilibrium remains valid precisely while $\lambda_{\text{one}} \leq \Lambda_H$.

This threshold also explains the marginal-benefit mechanism in the main text. For a single-scale population of mass λ at q_H, put $\nu = KR_H/\lambda$. At a smaller scale, Gaussian conditioning gives

$$R_P(q) = R_H + \frac{\nu^2[R(q) - R_H]}{\nu^2 + R(q) - R_H}, \qquad q < q_H.$$

Differentiating at the top from the left and using (64) yields

$$\text{MB}'(q_H-) - \text{MC}'(q_H) = \frac{j(q_H)^2}{\gamma K^2 R_H^3}\left(\lambda^2 - \Lambda_H^2\right).$$

Because $\text{MB}(q_H) = \text{MC}(q_H)$, the net marginal benefit is positive just below q_H when $\lambda < \Lambda_H$, is tangent to zero there when $\lambda = \Lambda_H$, and is negative just below it when $\lambda > \Lambda_H$. In the last case a downward scale deviation is profitable. Above q_H, the price is redundant and marginal benefit remains below marginal cost; the failure concerns the smaller-scale side of

the candidate optimum. On the mixed support, differentiating R_* gives $\mathrm{MB}(q) = \mathrm{MC}(q)$ throughout $[q_L, q_H]$.

If $\Lambda_H < 1$ and the fixed cost falls enough to reach this threshold, the continuous component starts with $q_L = q_H$ and $\lambda = \Lambda_H$, so the two branches join continuously. If the population cap is reached first, the distribution remains a unit atom at q_H. On an all-informed mixed branch, $m_*(q_L) = 1$ fixes the lower endpoint independently of χ_0. More generally, once all investors acquire, lowering the common fixed cost preserves their ranking of all positive scales and makes nonacquisition less attractive, so the same distribution remains an equilibrium. Uniqueness then makes it the equilibrium distribution. Together these cases prove weak monotonicity of participation over the entire fixed-cost range. □

H.2 Proof of Proposition 7

Proof of Proposition 7. Write $J(q, b) = -\log R(q)/(2\gamma) - bq^2$, where $b = \bar{\chi}$. The top-scale argument in Appendix F.3.1 applies without the interior-endpoint restriction and identifies the largest acquired scale as the smallest maximizer of J. For $b_1 < b_2$, let q_i maximize $J(\cdot, b_i)$. Adding their two optimality inequalities gives

$$(b_2 - b_1)(q_1^2 - q_2^2) \geq 0.$$

Hence $q_1 \geq q_2$. At a differentiable interior strict maximum, implicit differentiation of $J_q(q_H, b) = 0$ gives

$$\frac{dq_H}{db} = \frac{2q_H}{J_{qq}(q_H, b)} < 0.$$

To establish full-scale acquisition at a positive cost coefficient, define

$$\bar{\chi}_{\text{full}} := \inf_{0 \leq q < 1} \frac{\log[R(q)/R(1)]}{2\gamma(1 - q^2)}. \tag{94}$$

For $q < 1$, full-scale acquisition beats scale q in the standalone problem exactly when

$$J(1, b) - J(q, b) = \frac{\log[R(q)/R(1)]}{2\gamma} - b(1 - q^2) > 0.$$

The ratio in (94) extends continuously to $q = 1$, with value $-R'(1)/(4\gamma R(1)) > 0$. It is strictly positive on the whole compact interval $[0, 1]$, because $R > 0$ and $R' < 0$. Its infimum is therefore positive, and $0 < b < \bar{\chi}_{\text{full}}$ makes $q = 1$ the unique standalone maximizer. The

definition at $q = 0$ is only a continuous extension for computing the threshold; nonacquisition is a separate option.

It remains to show that, for sufficiently small b, every informed investor chooses the full scale. Put $R_1 = R(1)$, $B_1 = \sigma_d^2 - R_1 > 0$, and suppose $\chi_0 < \log(\sigma_d^2/R_1)/(2\gamma)$. For all sufficiently small $b > 0$, $R_1 e^{2\gamma(\chi_0+b)} < \sigma_d^2$, so the candidate single-scale equilibrium has positive mass

$$\lambda_b = \min\left\{1,\ K\sqrt{\frac{R_1}{e^{2\gamma(\chi_0+b)}-1} - \frac{R_1^2}{B_1}}\right\}.$$

With the acquisition menu restricted to $q = 1$, the economy is the binary one of Section 4.6, and λ_b is the informed fraction in Theorem 2 at cost $\chi_0 + b$. That theorem gives optimal participation and a linear CREE for this candidate. Its normalized price signal is $M(1) - \nu_b\eta$, with $\nu_b = KR_1/\lambda_b \geq KR_1$. For a deviating investor at $q < 1$, put $y(q) = R(q) - R_1 > 0$. Gaussian conditioning gives

$$R_{P,b}(q) = R_1 + \frac{\nu_b^2 y(q)}{\nu_b^2 + y(q)}.$$

Writing $v_0 = (KR_1)^2 > 0$, monotonicity in ν_b^2 implies

$$\frac{\log[R_{P,b}(q)/R_1]}{2\gamma(1-q^2)} \geq \frac{1}{2\gamma(1-q^2)}\log\left(1 + \frac{v_0 y(q)}{R_1[v_0 + y(q)]}\right).$$

The right side is strictly positive and continuous on $[0, 1)$, with positive limit $-R'(1)/(4\gamma R_1)$ at one. It therefore has a strictly positive minimum, say b_0. For $0 < b < b_0$, the loss of trading value from every smaller scale exceeds its variable-cost saving $b(1 - q^2)$. No smaller-scale deviation is profitable. Together with participation, this verifies the single-scale candidate as an overall equilibrium; uniqueness identifies it as the equilibrium. Finally, the formula for λ_b gives $\lambda_b = 1$ whenever

$$\chi_0 + b < \frac{1}{2\gamma}\log\left(1 + \frac{K^2 R_1 B_1}{B_1 + K^2 R_1^2}\right).$$

The right side is strictly positive. For sufficiently small χ_0 and b, this participation condition and the bound $b < b_0$ both hold, so every investor acquires the full model. □

H.3 Participation under changes in variable cost

On a smooth mixed branch with an interior top and $0 < \lambda < 1$, write $U_*(q, b)$ and $m_*(q, b)$ for the functions in (67) and (63). The envelope theorem at $q_H(b)$ gives

$$\frac{\partial \log \overline{R}}{\partial b} = 2\gamma q_H^2, \qquad \frac{\partial \log R_*(q, b)}{\partial b} = 2\gamma(q_H^2 - q^2).$$

Differentiating the entry condition and population accounting yields

$$\frac{dq_L}{db} = \frac{2\gamma q_H^2 \overline{R} - \partial_b U_*(q_L, b)}{\partial_q U_*(q_L, b)}, \qquad \frac{d\lambda}{db} = \partial_b m_*(q_L, b) - \rho_*(q_L, b)\frac{dq_L}{db}.$$

The denominator in the first expression is positive. Its numerator compares the change in the entry target with the change in the uninformed residual risk at a fixed lower endpoint. The second expression combines the change in the population tail with the movement of that endpoint. The fixed-cost argument does not sign these effects, because R_*, the price loadings, and the population tail all change with b. Panel (b) of Figure 4 evaluates the equilibrium along the variable-cost path and illustrates nonmonotone participation. The full-model panel of Figure 2 is verified with the single-scale posterior and deviation comparison in the preceding proof, rather than by extrapolating the interior-top density formula to a binding upper constraint.

H.4 Sharpe ratios and residual risk: Proposition H.1 and Corollary 8

This subsection states and proves the exact bounds behind Corollary 8, proves the corollary, and evaluates both at the calibrated values. The setting is that of Corollary 4 and its proof in Appendix F.1, with mean supply as in Appendix J.

Proposition H.1 (Sharpe ratios and residual risk) *Under Assumption 2, in a linear overall equilibrium with $\lambda > 0$ and mean supply as in Appendix J, let $\delta = E[d - P]$ and $S_P = \mathrm{Var}[d - P]$. For a fund at scale q with $R_P(q) < S_P$, let $\pi_q = x_q(P)(d - P)$ be its trading profit, $\mathrm{SR}(q) = E[\pi_q]/\sqrt{\mathrm{Var}[\pi_q]}$ its gross Sharpe ratio, and*

$$t(q) = \frac{S_P - R_P(q)}{R_P(q)}, \qquad a(q) = \frac{\delta^2}{R_P(q)}, \qquad u(q) = t(q) + a(q),$$

the payoff variance the fund forecasts and the squared premium, each relative to the residual variance $R_P(q)$.

(i) One plus the squared Sharpe ratio is inversely proportional to residual variance up to a bounded factor:

$$\frac{1+u(q)}{1+4t(q)} \le 1+\mathrm{SR}(q)^2 \le 1+u(q) = \frac{S_P+\delta^2}{R_P(q)}.$$

(ii) Holding S_P and δ fixed, the elasticity of the Sharpe ratio with respect to residual variance is

$$-\frac{\partial \log \mathrm{SR}(q)}{\partial \log R_P(q)} = \frac{1}{2u(q)} - \kappa(q), \qquad \kappa(q) = \frac{3a(q)^2 + 2a(q)t(q) + t(q)^2}{2u(q)\big[u(q) + 2t(q)^2 + 4a(q)t(q)\big]} \in \big(0, \tfrac{3}{2}\big).$$

(iii) For any two scales $q < q'$ in the support of F,

$$\frac{e^{2\gamma[\chi_1(q')-\chi_1(q)]}}{1+4t(q')} \le \frac{1+\mathrm{SR}(q')^2}{1+\mathrm{SR}(q)^2} \le \big(1+4t(q)\big)\, e^{2\gamma[\chi_1(q')-\chi_1(q)]}.$$

Proof of Proposition H.1. The proof of Corollary 4 computes the two moments of π_q in closed form. Expressed in the ratios t, a, and u, those moments make $\mathrm{SR}(q)^2$ equal to u divided by one plus a correction that is at most $4t$; this gives (i). Differentiating the same expression gives (ii), and combining (i) with the indifference formula (24) gives (iii).

Fix q, write $r = R_P(q)$, and drop the argument q from t, a, u, and κ. By the proof of Corollary 4 with mean supply, the conditional mean $m = E[d - P \mid \mathcal{F}_q, P]$ is Gaussian with mean δ and variance $S_P - r$, the optimal demand is $x_q(P) = m/(\gamma r)$, and the forecast error $d - P - m$ is independent of m with variance r. The moments of $\pi_q = m(d-P)/(\gamma r)$ computed there are

$$E[\pi_q] = \frac{S_P - r + \delta^2}{\gamma r}, \qquad \mathrm{Var}[\pi_q] = \frac{(S_P-r)^2 + (S_P-r)(3\delta^2 + S_P) + \delta^2 S_P}{(\gamma r)^2}.$$

Dividing numerators and denominators by r and r^2 and using $S_P/r = 1+t$ and $\delta^2/r = a$,

$$E[\pi_q] = \frac{u}{\gamma}, \qquad \mathrm{Var}[\pi_q] = \frac{t^2 + t(1+t+3a) + a(1+t)}{\gamma^2} = \frac{u + 2t(t+2a)}{\gamma^2},$$

so that

$$\mathrm{SR}(q)^2 = \frac{u^2}{u + 2t(t+2a)} = \frac{u}{1+c}, \qquad c = \frac{2t(t+2a)}{u}. \tag{95}$$

The correction c is nonnegative, and since $t + 2a \le 2(t + a) = 2u$ it is at most $4t$. Hence $u/(1 + 4t) \le \mathrm{SR}(q)^2 \le u$. The upper bound in (i) is the second inequality, and the lower bound follows from the first,

$$1 + \mathrm{SR}(q)^2 \ge 1 + \frac{u}{1 + 4t} = \frac{1 + 4t + u}{1 + 4t} \ge \frac{1 + u}{1 + 4t},$$

together with $1 + u = (S_P + \delta^2)/r$.

For (ii), the three ratios respond to r through their common denominator: $\partial t/\partial \log r = -(1 + t)$, $\partial a/\partial \log r = -a$, and $\partial u/\partial \log r = -(1 + u)$. Write the denominator of (95) as $D = u + 2t^2 + 4at$, so that $\mathrm{SR}(q)^2 = u^2/D$ and

$$-\frac{\partial \log \mathrm{SR}(q)}{\partial \log r} = -\frac{\partial \log u}{\partial \log r} + \frac{1}{2}\frac{\partial \log D}{\partial \log r} = \frac{1 + u}{u} - \frac{(1 + t)(1 + 4t + 4a) + a(1 + 4t)}{2D}.$$

Bringing the right side over the common denominator $2uD$ and collecting terms, the numerator is

$$2(1 + u)D - u\big[(1 + t)(1 + 4t + 4a) + a(1 + 4t)\big] = D - (3a^2 + 2at + t^2),$$

which gives $1/(2u) - \kappa$ with κ as stated. The bounds on κ follow from $D \ge u$ and $u^2 = a^2 + 2at + t^2$: the numerator $3a^2 + 2at + t^2$ is positive, and it is less than $3u^2 \le 3uD$, so $0 < \kappa < 3/2$.

For (iii), take $q < q'$ in the support of F. Equation (24) gives $R_P(q)/R_P(q') = e^{2\gamma[\chi_1(q') - \chi_1(q)]}$. Applying the upper bound in (i) at q' and the lower bound at q,

$$\frac{1 + \mathrm{SR}(q')^2}{1 + \mathrm{SR}(q)^2} \le \frac{(S_P + \delta^2)/R_P(q')}{(S_P + \delta^2)/[R_P(q)(1 + 4t(q))]} = \big(1 + 4t(q)\big)\frac{R_P(q)}{R_P(q')},$$

and the reverse choice of bounds gives the lower inequality. □

Proof of Corollary 8. Two facts about the calibrated economy are used. First, per month, $t(q) \le 0.008$ and $a(q) \le 0.008$ for every scale q on the support, so every fund's squared monthly Sharpe ratio is below 0.016. Second, the price reveals nearly all of the smallest model's signals, so $t(q_L) = 5 \times 10^{-5}$ is below one percent of $s = 0.0072$. Each approximation in the corollary drops a factor that Proposition H.1 bounds; we record the exact form and the size of the dropped factor at the baseline.

For the first display, (95) gives $1 + \mathrm{SR}(q)^2 = (1 + u)\,\theta$ with $\theta = (1 + u + c)/[(1 + u)(1 + c)] \in [1/(1 + 4t), 1]$, so the relative error is below $4t(q) \le 0.032$; at q_H it is 0.0003.

For the elasticity, part (ii) of the proposition, multiplied by two to pass from $\log R_P$ to $\log\sqrt{R_P}$, gives the exact value $1/u - 2\kappa$, while (95) gives $(1+\mathrm{SR}^2)/\mathrm{SR}^2 = 1 + (1+c)/u$. The approximation exceeds the exact value by $1 + c/u + 2\kappa < 4 + 4t/u$, a fraction below $4(u+t) \le 0.093$ of the approximation on the support; at q_L and q_H the fraction is 0.03 and 0.06.

The volatility gap is exact by the definition of Δ. For the Sharpe gap, write $u_L = u(q_L)$, $u_H = u(q_H)$, and likewise t_L, t_H, c_L, c_H. Since $R_P(q_L) = (1+\Delta)R_P(q_H)$, the definitions give $t_H = (1+t_L)(1+\Delta) - 1$ and $a(q_H) = (1+\Delta)a(q_L)$, hence the exact identity

$$u_H = (1+\Delta)u_L + \Delta, \qquad \text{so} \qquad \frac{\mathrm{SR}(q_H)^2}{\mathrm{SR}(q_L)^2} = \frac{u_H}{u_L}\,\frac{1+c_L}{1+c_H} = \Big[1 + \Delta\Big(1 + \frac{1}{u_L}\Big)\Big]\frac{1+c_L}{1+c_H}$$

by (95). Taking half the logarithm,

$$\log\frac{\mathrm{SR}(q_H)}{\mathrm{SR}(q_L)} = \frac{1}{2}\log\Big[1 + \Delta + \frac{\Delta}{u_L}\Big] + \frac{1}{2}\log\frac{1+c_L}{1+c_H},$$

where the last term lies between $-\frac{1}{2}\log(1+4t_H)$ and $\frac{1}{2}\log(1+4t_L)$. The stated form replaces $u_L = s + t_L(1+s)$ by s and drops the Δ inside the bracket and the last term. At the baseline, $u_L = 1.006\,s$, so the bracket moves from 2.0901 to $1 + \Delta/s = 2.0892$, and the last term is -0.0115; together the error is 0.011 on a gap of 0.357, or 3%. □

Calibrated values. Per month at the baseline, $\gamma = 1.80$, $\chi_1(q_H) = 2.18\times10^{-3}$, $q_L/q_H = 0.035$, $\delta = 4.01\times10^{-3}$, and $S_P = 2.227\times10^{-3}$, so

$$\Delta = e^{2\times1.80\times2.18\times10^{-3}\times(1-0.035^2)} - 1 = 0.0079, \qquad s = \frac{(4.01\times10^{-3})^2}{2.227\times10^{-3}} = 0.0072.$$

At q_L and q_H, $t = 5\times10^{-5}$ and 0.0079, $a = 0.0072$ and 0.0073, and $u = 0.0073$ and 0.0152. In part (i), the factor $1 + 4t$ is 1.0002 and 1.03, and $1 + \mathrm{SR}^2$ equals $(S_P + \delta^2)/R_P$ to within 0.03%. In part (ii), the elasticity $1/(2u) - \kappa$ is $68.8 - 1.5 = 67.3$ and $32.9 - 0.7 = 32.2$. In part (iii), the ratio $(1+\mathrm{SR}(q_H)^2)/(1+\mathrm{SR}(q_L)^2)$ is 1.0075, against $1 + \Delta = 1.0079$. The residual-volatility gap is $\frac{1}{2}\log(1+\Delta) = 0.0039$, a ratio of 1.004: 16.35% against 16.28% annualized. The Sharpe gap is $\frac{1}{2}\log(1+\Delta/s) = \frac{1}{2}\log 2.09 = 0.37$ to first order, a ratio of 1.45; the exact formula (95) gives 0.36 and a ratio of 1.43: 0.42 against 0.295 annualized. An annual Sharpe ratio of 0.3 is a monthly ratio of 0.085, whose square is $s = 0.0072$; the per-period squared Sharpe ratio is small for any asset with an ordinary annual Sharpe ratio,

so a forecastable share of the order of s doubles a fund's squared Sharpe ratio while leaving its residual volatility almost unchanged.

I Random-matrix toolkit

Lemma I.1 (Concentration of Gaussian quadratic forms) *For each N, let $S \in \mathbb{R}^N$ satisfy*

$$S \sim \mathcal{N}(0, \Psi),$$

where $\Psi = \Psi_N$ is a deterministic nonnegative definite matrix. Let A_N be a random $N \times N$ matrix such that:

1. *A_N is independent of S;*
2. *A_N is symmetric almost surely;*
3. *there exists a deterministic constant $C < \infty$, independent of N, such that*

$$\|A_N\| \le C \qquad a.s.$$

4.

$$\frac{1}{N^2}\operatorname{tr}(\Psi^2) \to 0 \qquad as\ N \to \infty.$$

Then

$$\frac{1}{N} S' A_N S - \frac{1}{N}\operatorname{tr}(A_N \Psi) \xrightarrow{p} 0.$$

More precisely,

$$E\left[\left(\frac{1}{N} S' A_N S - \frac{1}{N}\operatorname{tr}(A_N \Psi)\right)^2\right] \le \frac{2C^2}{N^2}\operatorname{tr}(\Psi^2),$$

and therefore the left-hand side converges to zero in L^2, hence in probability.

Condition (4) is mild: if $\|\Psi\| \le C_\Psi < \infty$ uniformly in N, then $\operatorname{tr}(\Psi^2) \le C_\Psi \operatorname{tr}(\Psi)$, so

$$\frac{1}{N^2}\operatorname{tr}(\Psi^2) \;\le\; \frac{C_\Psi}{N}\,\frac{\operatorname{tr}(\Psi)}{N} \;=\; O(N^{-1})$$

whenever $\operatorname{tr}(\Psi)/N = O(1)$*, which is therefore a sufficient condition. In particular, if* $\|\Psi\|$ *is uniformly bounded, then*

$$\frac{1}{N} S' A_N S - \frac{1}{N} \operatorname{tr}(A_N \Psi) \xrightarrow{p} 0.$$

Proof of Lemma I.1. Conditional on A_N, Gaussian quadratic-form identities give

$$E[S' A_N S \mid A_N] = \operatorname{tr}(A_N \Psi), \qquad \operatorname{Var}[S' A_N S \mid A_N] = 2 \operatorname{tr}(A_N \Psi A_N \Psi).$$

Diagonalize Ψ with eigenvalues $\ell_i \geq 0$ and express the symmetric A_N in this basis as (a_{ij}). Since $2\ell_i \ell_j \leq \ell_i^2 + \ell_j^2$,

$$\operatorname{tr}(A_N \Psi A_N \Psi) = \sum_{i,j} a_{ij}^2 \ell_i \ell_j \leq \sum_i \ell_i^2 \sum_j a_{ij}^2 \leq C^2 \operatorname{tr}(\Psi^2).$$

Divide the conditional variance by N^2 and take expectations to obtain the stated L^2 bound. If $\|\Psi\|$ is uniformly bounded, then $\operatorname{tr}(\Psi^2) \leq N \|\Psi\|^2 = O(N)$, proving the final claim. □

The next lemma is used in Appendix E. For random elements U, V, and W, write

$$\mathrm{I}(U; V \mid W) := E\big[\mathrm{KL}\big(\mathcal{L}(U \mid V, W) \,\big\|\, \mathcal{L}(U \mid W)\big)\big]$$

for the information in V about U given W, the expected log-likelihood ratio of the conditional law of U with V against the one without it.[32] The lemma says that a history carrying information $o(N)$ about a coefficient with N independent directions of variance $O(1/N)$ carries information $o(1)$ about a noisy prediction on a fresh direction. Its proof shares the information among r fresh predictions: together they cannot learn more than the history contains, and they are nearly separate problems because they overlap only through the coefficient's small prior covariance.

Lemma I.2 (Information in a fresh prediction) *Let* $\theta \sim \mathcal{N}(m, \Sigma)$ *on* $\mathbb{R}^k$*, let* Z *be a random element with* $\mathrm{I}(\theta; Z) < \infty$*, let* S *be independent of* (θ, Z) *with* $E[SS'] = \Psi$*, and let* $\varepsilon \sim \mathcal{N}(0, v)$ *with* $v > 0$ *be independent of* (θ, Z, S)*. Put* $Y := S'(\theta - m) + \varepsilon$ *and*

[32] With densities, $\mathrm{I}(U; V \mid W) = E \log \frac{p(u,v|w)}{p(u|w)\, p(v|w)}$, which is symmetric in U and V. The chain rule $\mathrm{I}(U; V_1, V_2 \mid W) = \mathrm{I}(U; V_1 \mid W) + \mathrm{I}(U; V_2 \mid V_1, W)$ is the factorization $\frac{p(u|v_1,v_2,w)}{p(u|w)} = \frac{p(u|v_1,w)}{p(u|w)} \cdot \frac{p(u|v_1,v_2,w)}{p(u|v_1,w)}$. If V is independent of U given (θ, W), the chain rule applied to $\mathrm{I}(U; \theta, V \mid W)$ in both orders gives $\mathrm{I}(U; V \mid W) \leq \mathrm{I}(U; \theta \mid W)$, because $\mathrm{I}(U; V \mid \theta, W) = 0$ and $\mathrm{I}(U; \theta \mid V, W) \geq 0$.

$V := \operatorname{tr}(\Sigma\Psi\Sigma\Psi)$. *Then*

$$\mathrm{I}(Z; Y \mid S) \le \min\Big\{\mathrm{I}(\theta; Z),\ \frac{\sqrt{\mathrm{I}(\theta; Z)\, V}}{v}\Big\}, \tag{96}$$

and, with $\mathrm{TV}(\nu, \mu) := \sup_A |\nu(A) - \mu(A)|$ *the total variation distance, the conditional laws of* Y *with and without* Z *satisfy, for every* $t \in \mathbb{R}$,

$$\begin{aligned} E\big[\mathrm{TV}\big(\mathcal{L}(Y \mid Z, S), \mathcal{L}(Y \mid S)\big)^2\big] &\le \tfrac{1}{2}\,\mathrm{I}(Z; Y \mid S), \\ E\big|E[e^{itY} \mid Z, S] - E[e^{itY} \mid S]\big|^2 &\le 2\ \mathrm{I}(Z; Y \mid S). \end{aligned} \tag{97}$$

Proof of Lemma I.2. *Replicas.* Let $(S_1, \varepsilon_1), \dots, (S_r, \varepsilon_r)$ be independent copies of (S, ε), independent of (θ, Z), and put $Y_j := S_j'(\theta - m) + \varepsilon_j$ and $S_{1:r} := (S_1, \dots, S_r)$. The replicas exist only in the proof; nobody observes them. Each (S_j, Y_j) is distributed as (S, Y) jointly with (θ, Z), and the other replicas are independent of (Z, S_j, Y_j), so $\mathrm{I}(Z; Y_j \mid S_{1:r}) = \mathrm{I}(Z; Y \mid S) =: J$ for every j.

Sharing the information. Write $Y_{<j} := (Y_1, \dots, Y_{j-1})$. The chain rule applied to $\mathrm{I}(Z, Y_{<j}; Y_j \mid S_{1:r})$ in both orders gives

$$\begin{aligned} \mathrm{I}(Z; Y_j \mid S_{1:r}) &= \mathrm{I}(Z; Y_j \mid Y_{<j}, S_{1:r}) + \mathrm{I}(Y_j; Y_{<j} \mid S_{1:r}) - \mathrm{I}(Y_j; Y_{<j} \mid Z, S_{1:r}) \\ &\le \mathrm{I}(Z; Y_j \mid Y_{<j}, S_{1:r}) + \mathrm{I}(Y_j; Y_{<j} \mid S_{1:r}). \end{aligned}$$

Summed over j, the first terms add up to $\mathrm{I}(Z; Y_1, \dots, Y_r \mid S_{1:r})$ by the chain rule, which is at most $\mathrm{I}(Z; \theta \mid S_{1:r}) = \mathrm{I}(\theta; Z)$, because $(Y_1, \dots, Y_r)$ depends on Z only through θ given $S_{1:r}$, and $S_{1:r}$ is independent of (θ, Z). Hence

$$rJ \le \mathrm{I}(\theta; Z) + \sum_{j=1}^{r} \mathrm{I}(Y_j; Y_{<j} \mid S_{1:r}). \tag{98}$$

The replicas are nearly independent. Fix a realization of $S_{1:r}$. With $\theta \sim \mathcal{N}(m, \Sigma)$ not yet updated by Z, the vector $(Y_1, \dots, Y_r)$ is Gaussian with covariance $C_{ij} = S_i'\Sigma S_j + v\mathbf{1}_{\{i=j\}}$. Let $C_{<j}$ be its leading $(j-1)\times(j-1)$ block and $c_j := (S_i'\Sigma S_j)_{i<j}$. By the Gaussian formula, the information in $Y_{<j}$ about Y_j is

$$\frac{1}{2}\log\frac{\mathrm{Var}[Y_j]}{\mathrm{Var}[Y_j \mid Y_{<j}]} = \frac{1}{2}\log\Big(1 + \frac{c_j' C_{<j}^{-1} c_j}{\mathrm{Var}[Y_j \mid Y_{<j}]}\Big) \le \frac{|c_j|^2}{2v^2},$$

because $\log(1+x) \le x$, $\mathrm{Var}[Y_j \mid Y_{<j}] \ge v$ as ε_j is independent of the rest, and $C_{<j} \succeq vI$. For $i \ne j$, $E[(S_i'\Sigma S_j)^2] = E[S_i'\Sigma\, E[S_jS_j']\,\Sigma S_i] = \mathrm{tr}(\Sigma\Psi\Sigma\Psi) = V$, so averaging over $S_{1:r}$ gives $\mathrm{I}(Y_j; Y_{<j} \mid S_{1:r}) \le (j-1)V/(2v^2)$, and the sum in (98) is at most $r(r-1)V/(4v^2)$.

The number of replicas. By (98), $J \le \mathrm{I}(\theta;Z)/r + (r-1)V/(4v^2)$ for every $r \ge 1$. With $r = 1$, $J \le \mathrm{I}(\theta;Z)$. If $\mathrm{I}(\theta;Z)$ and V are positive, $r := \lceil 2v\sqrt{\mathrm{I}(\theta;Z)/V}\rceil$ satisfies $2v\sqrt{\mathrm{I}(\theta;Z)/V} \le r < 2v\sqrt{\mathrm{I}(\theta;Z)/V} + 1$, so both terms are at most $\sqrt{\mathrm{I}(\theta;Z)V}/(2v)$, which gives (96); if either vanishes, so does the right side of (96), by $r = 1$ or by letting $r \to \infty$.

From information to laws. By definition, $J = E[\mathrm{KL}(\mathcal{L}(Y \mid Z,S) \,\|\, \mathcal{L}(Y \mid S))]$, and Pinsker's inequality $\mathrm{TV}(\nu,\mu)^2 \le \frac{1}{2}\mathrm{KL}(\nu\|\mu)$ gives the first bound in (97).[33] The second follows from $|\int e^{ity}\,(\nu-\mu)(dy)| \le 2\ \mathrm{TV}(\nu,\mu)$. □

For the following deterministic equivalents, condition on a covariance sequence Ψ_N whose operator norms are bounded and whose empirical spectral distributions converge weakly to a deterministic law H. The history consists of T independent $\mathcal{N}(0,\Psi_N)$ observations, with $N/T \to c > 0$. Write $a_N \approx b_N$ when their difference converges to zero in probability, and let

$$m_\Psi(w) = \int \frac{dH(x)}{x-w}, \qquad w < 0,$$

be the Stieltjes transform of the population law H. Lemma I.6 verifies these hypotheses almost surely under the common prior; conditional bounds can then be localized to events on which $\|\Psi_N\|$ is bounded.

Theorem I.3 (Silverstein–Bai deterministic equivalent) *(Silverstein and Bai, 1995; Bai and Zhou, 2008; Bai and Silverstein, 2010) Let $z > 0$ and*

$$\hat m(z) \;=\; N^{-1}\,\mathrm{tr}((zI + \hat\Psi)^{-1})\,.$$

Then,

$$\hat m(z) \;\approx\; m(-z;c),$$

[33] For $\nu \ll \mu$ and an event A with $p := \nu(A)$ and $q := \mu(A)$, Jensen's inequality for the convex function $x\log x$, applied under μ to the conditional expectation of $d\nu/d\mu$ given $\mathbf{1}_A$, gives $\mathrm{KL}(\nu\|\mu) \ge p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$. The right side is at least $2(p-q)^2$, because it vanishes with its first derivative at $p = q$ and its second derivative in p is $1/(p(1-p)) \ge 4$. Take the supremum over A.

where, for $w < 0$ and with m_Ψ the Stieltjes transform of the limiting spectral law H of Ψ_N, $m(w;c)$ is the unique solution with $m(w;c) > 0$ and $1 - c - c\,w\,m(w;c) > 0$ of the nonlinear master equation

$$m(w;c) = \frac{1}{1 \;-\; c \;-\; c\,w\,m(w;c)}\, m_\Psi\left(\frac{w}{1 \;-\; c \;-\; c\,w\,m(w;c)}\right). \tag{99}$$

The positivity condition is needed when $c > 1$: positivity of $m(w;c)$ alone does not select one solution of (99). With $k := 1 - c - c\,w\,m(w;c)$, equation (99) is equivalent to $k + c\int \frac{xk}{xk-w}\,dH(x) = 1$. The left-hand side is strictly increasing in $k > 0$ and vanishes at $k = 0$, so there is exactly one solution with $k > 0$.

We will also need

Proposition I.4 *For $z > 0$,*

$$T^{-1}\operatorname{tr}((zI + \hat\Psi)^{-1}\Psi) \;\approx\; \xi(z;c) = \frac{1 - z m(-z;c)}{c^{-1} - 1 + z m(-z;c)}\,. \tag{100}$$

Define the effective ridge

$$Z_*(z;c) \;=\; z\,(1 + \xi(z;c))\,.$$

Proof of Proposition I.4. Write $Q = (zI + \hat\Psi)^{-1}$ and $\xi_N = T^{-1}\operatorname{tr}(\Psi Q)$. The proposition follows from evaluating the *same* trace $N^{-1}\operatorname{tr}(\hat\Psi Q)$ in two ways.

First evaluation: the resolvent identity. Exactly, $\hat\Psi Q = I - zQ$, so by Theorem I.3

$$N^{-1}\operatorname{tr}(\hat\Psi Q) \;=\; 1 - z\,\hat m(z) \;\approx\; 1 - z\,m(-z;c) \;=:\; M\,. \tag{101}$$

Second evaluation: leave one observation out. Since $\hat\Psi = T^{-1}\sum_\tau S_\tau S_\tau'$,

$$N^{-1}\operatorname{tr}(\hat\Psi Q) \;=\; \frac{1}{NT}\sum_{\tau=1}^{T} S_\tau' Q\, S_\tau\,.$$

Let $Q_{(\tau)} = (zI + \hat\Psi - T^{-1}S_\tau S_\tau')^{-1}$, which is independent of S_τ. The rank-one update identity gives

$$S_\tau' Q\, S_\tau \;=\; \frac{S_\tau' Q_{(\tau)} S_\tau}{1 + T^{-1} S_\tau' Q_{(\tau)} S_\tau}\,,$$

and, $S_\tau \sim \mathcal{N}(0, \Psi)$ being independent of $Q_{(\tau)}$ with $\|Q_{(\tau)}\| \le z^{-1}$ and $\|\Psi\|$ uniformly bounded (the covariance bound above), Lemma I.1 gives $T^{-1}[S_\tau' Q_{(\tau)} S_\tau - \operatorname{tr}(\Psi Q_{(\tau)})] \to 0$ in L^2, uniformly in τ. Replacing $Q_{(\tau)}$ by Q costs $O(1)$: the update is rank one, so

$$\left|\operatorname{tr}\big(\Psi(Q - Q_{(\tau)})\big)\right| \;=\; \frac{T^{-1}\left|S_\tau' Q_{(\tau)} \Psi Q_{(\tau)} S_\tau\right|}{1 + T^{-1} S_\tau' Q_{(\tau)} S_\tau} \;\le\; T^{-1}\|\Psi\| z^{-2} \|S_\tau\|^2 \;=\; O(1)\,,$$

negligible against $\operatorname{tr}(\Psi Q) = T\,\xi_N = O(T)$. After division by T, each summand differs from $\xi_N/(1+\xi_N)$ by a term tending to zero in L^2, and since the per-τ L^2 bound does not depend on τ, the triangle inequality in L^2 carries the approximation through the average, whence

$$N^{-1} \operatorname{tr}(\hat{\Psi} Q) \;\approx\; \frac{T}{N} \frac{\xi_N}{1+\xi_N} \;=\; \frac{1}{c} \frac{\xi_N}{1+\xi_N}\,. \tag{102}$$

Solving. Equating (101) and (102) gives $\xi_N \approx c\,M(1+\xi_N)$. The division by $1 - cM$ is legitimate, and delivers $\xi > 0$: the positivity condition of Theorem I.3, read at $w = -z$, states exactly $1 - cM = 1 - c + c\,z\,m(-z;c) > 0$. Hence $\xi_N \approx cM/(1-cM)$. Dividing numerator and denominator by c and substituting $M = 1 - z\,m(-z;c)$,

$$\xi_N \;\approx\; \frac{M}{c^{-1} - M} \;=\; \frac{1 - z\,m(-z;c)}{c^{-1} - 1 + z\,m(-z;c)}\,,$$

which is (100). Equivalently $1+\xi = (1-cM)^{-1}$, so the effective ridge is $Z_*(z;c) = z(1+\xi) = z/\big(1 - c + c\,z\,m(-z;c)\big)$. At $w = -z$, the argument of m_Ψ in the master equation (99) is therefore $-Z_*(z;c)$. □

Theorem I.5 (Rubio and Mestre (2011)) *Let A_N be $N \times N$ matrices with uniformly bounded operator norm, independent of the history conditional on Ψ_N. Then, for $z > 0$,*

$$\frac{1}{N} \operatorname{tr}\Big[A_N (zI + \hat{\Psi})^{-1}\Big] \approx \frac{1 + \xi(z;c)}{N} \operatorname{tr}\big[A_N (Z_*(z;c) I + \Psi)^{-1}\big]\,.$$

We use this identity with $A_N = \Psi_N$.

Lemma I.6 (Prior spectrum: inverse Marchenko–Pastur) *Assume $V_0 > 0$. Let MP_y denote the Marchenko–Pastur law of ratio y (Marčenko and Pastur, 1967). Let $\Psi \sim \mathcal{W}^{-1}(V_0 N I_N, (1+\kappa)N)$ with $\kappa > 0$, and put $y := \frac{N}{(1+\kappa)N} = \frac{1}{1+\kappa} \in (0,1)$. Then, almost surely, the empirical spectral distribution $H_{\Psi_N}(x) = \frac{1}{N}\#\{i : \ell_i(\Psi) \le x\}$ converges weakly to the deterministic law $H_0 = H_0^{V_0,\kappa}$ obtained as the pushforward of the Marchenko–Pastur law*

MP_y *(ratio y, unit scale) under $\ell \mapsto \frac{V_0}{1+\kappa}\ell^{-1}$; equivalently, H_0 is the law of $x = \frac{V_0}{1+\kappa}\ell^{-1}$ with $\ell \sim \mathrm{MP}_y$. Consequently H_0 scales linearly in V_0, is supported on the bounded interval*

$$\left[\ \tfrac{V_0}{1+\kappa}(1+\sqrt{y})^{-2},\quad \tfrac{V_0}{1+\kappa}(1-\sqrt{y})^{-2}\ \right] \subset (0,\infty),$$

and has mean $\int x\, dH_0(x) = \frac{V_0}{\kappa}$, matching the inverse-Wishart prior mean $E[\Psi] = \frac{V_0 N}{\kappa N - 1} I \to \frac{V_0}{\kappa} I$. Moreover, almost surely, $\limsup_N \|\Psi_N\| \le \Lambda := \frac{V_0}{1+\kappa}(1-\sqrt{y})^{-2} < \infty$ (the upper spectral edge); in particular the $\{\Psi_N\}$ are almost surely eventually uniformly bounded, with spectra eventually contained in any fixed neighbourhood of the support interval above.

Proof of Lemma I.6. *Wishart representation.* Write $\nu_N := (1+\kappa)N$ for the degrees of freedom in Assumption 1. We take ν_N to be an integer. If $(1+\kappa)N$ is not an integer, the prior uses $\nu_N = \lfloor (1+\kappa)N \rfloor$; then $N/\nu_N \to y$ still holds, and the proof is unchanged. Let $(g_{ij})_{i,j\ge 1}$ be one infinite array of i.i.d. $\mathcal{N}(0,1)$ variables, and let $G_N = (g_{ij})_{i\le N, j\le \nu_N} \in \mathbb{R}^{N\times \nu_N}$ be its corner. Then $W_N := G_N G_N' \sim \mathrm{Wishart}_N(I_N, \nu_N)$, and $\Psi_N := V_0 N\, W_N^{-1}$ has the law $\mathcal{W}^{-1}(V_0 N I_N, \nu_N)$, because $\Psi_N^{-1} \sim \mathrm{Wishart}_N\big((V_0 N)^{-1} I_N, \nu_N\big)$. All Ψ_N are now defined on one probability space, and the almost-sure statements refer to it. The convergence in probability used in the applications does not depend on this coupling. We write Ψ, W, and G for Ψ_N, W_N, and G_N.

Marchenko–Pastur limit for W. Let $S_N := \frac{1}{\nu_N} W = \frac{1}{\nu_N} GG'$, the sample covariance of ν_N i.i.d. standard Gaussian vectors in $\mathbb{R}^N$. Since $N/\nu_N \to y \in (0,1)$, the Marchenko–Pastur theorem (Bai and Silverstein, 2010) gives that the empirical spectral distribution of S_N converges almost surely, weakly, to MP_y, with density $\frac{\sqrt{(\ell_+ - \ell)(\ell - \ell_-)}}{2\pi y\, \ell}$ on $[\ell_-, \ell_+]$, $\ell_\pm = (1 \pm \sqrt{y})^2$. Because $y < 1$ we have $\ell_- > 0$: the limiting support is bounded away from the origin. Moreover the extreme eigenvalues of such corners converge to the support edges, $\ell_{\min}(S_N) \to \ell_-$ and $\ell_{\max}(S_N) \to \ell_+$ almost surely (Yin et al., 1988; Bai and Yin, 2008), so almost surely, for all large N, the spectrum of S_N lies in a fixed compact subset $[\ell_- - \varepsilon, \ell_+ + \varepsilon] \subset (0,\infty)$.

Transfer to Ψ. On that compact set the map $\ell \mapsto \ell^{-1}$ is bounded and continuous, so the empirical spectral distribution of S_N^{-1} converges a.s. weakly to the pushforward of MP_y under $\ell \mapsto \ell^{-1}$. Now $\Psi = V_0 N\, W^{-1} = V_0 \frac{N}{\nu_N}\, S_N^{-1}$, and the scalar $V_0 \frac{N}{\nu_N} \to V_0 y = \frac{V_0}{1+\kappa}$; since the eigenvalues $V_0 \frac{N}{\nu_N} \ell_i(S_N)^{-1}$ are (uniformly over the compact spectrum) within $o(1)$ of $\frac{V_0}{1+\kappa}\ell_i(S_N)^{-1}$, multiplying by the convergent scalar preserves the weak limit. Hence the empirical spectral distribution of Ψ converges a.s. weakly to the law H_0 of $x = \frac{V_0}{1+\kappa}\ell^{-1}$, $\ell \sim \mathrm{MP}_y$. The same edge limits bound the whole spectrum of Ψ_N: a.s. eventually $\ell_{\min}(S_N) \ge \ell_- -$

ε, so $\|\Psi_N\| = V_0 \frac{N}{\nu_N}\, \ell_{\min}(S_N)^{-1} \to \frac{V_0}{1+\kappa}\ell_-^{-1} = \Lambda$, giving $\limsup_N \|\Psi_N\| \le \Lambda = \frac{V_0}{1+\kappa}(1-\sqrt{y})^{-2}$ almost surely.

Support and mean. As $\ell \in [\ell_-, \ell_+]$, $x = \frac{V_0}{1+\kappa}\ell^{-1} \in [\frac{V_0}{1+\kappa}\ell_+^{-1}, \frac{V_0}{1+\kappa}\ell_-^{-1}]$, the stated interval. For the mean, $\int x\, dH_0 = \frac{V_0}{1+\kappa}\int \ell^{-1}\, d\mathrm{MP}_y(\ell) = \frac{V_0}{1+\kappa}\cdot\frac{1}{1-y}$, using the negative-first moment $\int \ell^{-1} d\mathrm{MP}_y = \frac{1}{1-y}$ for $y < 1$. With $y = \frac{1}{1+\kappa}$, $\frac{1}{1-y} = \frac{1+\kappa}{\kappa}$, so $\int x\, dH_0 = \frac{V_0}{1+\kappa}\cdot\frac{1+\kappa}{\kappa} = \frac{V_0}{\kappa}$. $\square$

J Calibration

We calibrate the nested-information economy to monthly equity returns, the performance of market-timing strategies built on random features of different sizes, and survey and disclosure evidence on information acquisition and research spending. The baseline jointly fits five parameters to eleven moments, holding the monetary normalization, the quadratic cost exponent, the complexity of the signal universe, and the shape parameter of the covariance prior fixed.

J.1 Parameters and units

The period is one month. Write acquired complexity as $X = c_q = qc$ (see (38)), the limit of N_1/T, where $c = N/T$ is the complexity of the full signal universe. The forecasting exercises use $T = 120$ monthly observations, so a model with N_1 features is assigned complexity $X = N_1/120$, and we fix $c = 1000$, a universe of $N = 120{,}000$ weak signals. Population mass is one, and each modeled fund has assets under management of \$1 billion. We allocate its entire research budget $\mathcal{B}$ to the modeled strategy, so the monthly cost in units of fund capital is $\mathcal{B}/(12 \times 10^9)$. The acquisition cost per month is

$$C(X) = \chi_0 + \bar{\chi}_X X^2, \qquad \chi(q) = C(cq) = \chi_0 + \bar{\chi}_X c^2 q^2.$$

Thus the coefficient $\bar{\chi}$ in the main text equals $c^2\bar{\chi}_X$; Table 2 reports both coefficients. Nonacquisition has zero cost. Premia, volatilities, Sharpe ratios, and costs are reported in annual units: premia and costs are multiplied by 12, volatilities and Sharpe ratios by $\sqrt{12}$.

We parameterize the information structure by

$$A_s = b'\psi b, \qquad A_w = \frac{\sigma_\beta^2 V_0}{\kappa}, \qquad h_c = c\kappa,$$

where A_s and A_w are the strong- and weak-signal contributions to payoff variance. The calibration fixes $\sigma_\varepsilon^2 = 0$ and $A_s = 0$, giving $\sigma_d^2 = A_w$; the strong block is dropped because the linear predictors have no out-of-sample explanatory power at the monthly horizon (Appendix J.2). The moments do not identify h_c, which we fix at $h_c = 10$, and we fit the residual risk at full complexity, $R_\infty = A_w h_c/(1 + h_c)$, in place of A_w (Appendix J.3). The fitted parameters are $(R_\infty, \sigma_\eta, \gamma, \chi_0, \bar{\chi}_X)$. The calibration identifies neither σ_β^2 nor V_0 separately: their product is recovered as κA_w, with $\kappa = h_c/c$.

Table 2: Baseline calibration parameters

Parameter	Meaning	Baseline
$R_\infty = A_w h_c/(1+h_c)$	Residual payoff variance at full complexity	0.002171
A_w	Weak-signal payoff variance (implied)	0.002388
$h_c = c\kappa$	Covariance-learning shape parameter (fixed)	10
c	Full-universe complexity (fixed)	1000
A_s	Strong-signal payoff variance (fixed)	0
σ_η	Standard deviation of the supply shock	0.2006
γ	CARA risk aversion	1.803
χ_0	Fixed acquisition cost	2.484×10^{-6} (0.298 bp)
$\bar{\chi}_X$	Quadratic cost coefficient in X	1.264×10^{-6}
$\bar{\chi} = c^2\bar{\chi}_X$	Implied quadratic cost coefficient in q	1.264

Variances and costs are per month; costs are fractions of the assumed \$1 billion fund capital. The fixed cost is also given in annual basis points (bp) in parentheses. The coefficients $\bar{\chi}_X$ and $\bar{\chi}$ multiply X^2 and q^2 and are not themselves expenditure rates; $\bar{\chi}$ is implied by c and $\bar{\chi}_X$ and is not a separate fitted parameter.

Unit mean supply gives market clearing right-hand side $1+\sigma_\eta\eta$ and an affine price with the centered pricing equations of Appendix F.3.4. Let $\delta = E^\infty[d^\infty - P]$ and $S_P = \mathrm{Var}^\infty[d^\infty - P]$. Expected market clearing gives

$$\delta = \frac{\gamma}{\lambda \int_{(0,1]} R_P(q)^{-1}\, dF(q) + (1-\lambda) R_U^{-1}}, \qquad R_U = \mathrm{Var}^\infty[d^\infty \mid P].$$

Gaussian integration gives

$$V(q) = -e^{\gamma\chi(q)} \sqrt{\frac{R_P(q)}{S_P}} \exp\left(-\frac{\delta^2}{2S_P}\right).$$

The last factor is common to all scales and to nonacquisition, so changing mean supply leaves acquisition incentives unchanged.

J.2 Empirical targets

Table 3 reports the eleven targets and their fitted values. The eight financial targets use 708 monthly observations, from January 1967 through December 2025. Market excess returns

come from the Fama–French market series, and the forecasting exercises use Goyal–Welch predictors observed at the end of the previous month.[34]

Table 3: Calibration targets and baseline fit

Moment	Target	Fit
Market excess-return mean, monthly	0.608%	0.401%
Market excess-return volatility, monthly	4.548%	4.720%
Information ratio: $N_1 = 12$	0.0718	0.0183
Information ratio: $N_1 = 60$	0.1348	0.0199
Information ratio: $N_1 = 120$	0.1197	0.0217
Information ratio: $N_1 = 300$	0.0922	0.0276
Information ratio: $N_1 = 600$	0.1412	0.0420
Information ratio: $N_1 = 1{,}200$	0.1684	0.0760
Information-acquisition participation	53.0%	53.3%
Median annual research budget	\$750,000	\$754,328
Research expense of the largest funds/capital	4.83%	2.61%

Market moments are monthly; information ratios and the expense ratio are annual. The participation and budget targets use alternative-data surveys as proxies for information acquisition and research expenditure; the expense target uses pass-through expense disclosures of multi-strategy funds. Budget moments are conditional on acquisition; the assumed fund AUM is \$1 billion.

The forecasting exercise follows Kelly et al. (2024): random Fourier features of the predictors, five fixed random seeds, ridge regressions estimated on rolling windows of 120 months with the penalty and bandwidth chosen by the forecast errors of the preceding 120 months, and positions proportional to the predicted excess return. A strategy's information ratio is the intercept of a regression of its monthly profit on the buy-and-hold excess return, divided by the residual standard deviation and annualized; each target averages the five seeds. The feature counts $N_1 = 12$, 60, 120, 300, 600, and 1,200 correspond to $X = 0.1$, 0.5, 1, 2.5, 5, and 10. These are candidate forecasting models, rather than observed scales chosen by funds. We target information ratios rather than Sharpe ratios because the model's uninformed fund times the market on the price and earns at least the market Sharpe ratio, whereas every empirical timing strategy earns less than buy-and-hold; the information ratio is zero without information and has no such floor. The out-of-sample R^2 of a linear forecast

[34]The sources are the Kenneth R. French Data Library and Amit Goyal's updated equity-premium prediction data. The inputs are revised historical data, rather than real-time vintages.

from the dividend-price ratio, the term spread, and the Treasury-bill rate is negative at the monthly horizon, so we drop the strong-signal block and set $A_s = 0$.

The participation target is the 53% share of alternative-data users reported by AIMA and SS&C (2020). The median-budget target is \$750,000, the midpoint of the \$0.5–1 million bracket containing the median in Exabel (2026). The expense target is the annual research expense of the largest multi-strategy funds as a share of their capital, 4.8% per year, within the 3.5–8.4% range of pass-through disclosures.[35] Its model counterpart is the annual acquisition cost of the largest model, $12\chi(q_H)$. These surveys and disclosures differ in year, population, and coverage. They provide expenditure proxies under the assumed fund size and allocation, without a direct conversion from complexity to GPU-hours or dollar compute costs.

J.3 Learning, equilibrium, and the fitting criterion

For each candidate parameter vector, the private-information risk curve is determined before solving equilibrium. In complexity units, let $r(X) = R(X/c)$ and

$$n(X) = \frac{h_c A_w - \kappa A_w X}{h_c + X}.$$

Specializing the quadratic in Lemma 4 to $\sigma_\varepsilon^2 = 0$, $r(X)$ is the larger root of

$$(1 + h_c)r(X)^2 + \big[\kappa A_w - h_c A_w - n(X)\big]r(X) - \kappa A_w n(X) = 0.$$

This is a boundary specialization of the positive-noise specification. For $c > 1$, residual risk stays strictly positive even at the full scale:

$$r(c) = \frac{h_c A_w (1 - 1/c)}{1 + h_c} = R_\infty\Big(1 - \frac{1}{c}\Big) > 0.$$

[35] Pass-through structures charge investors the fund's research, data, technology, and compensation costs. Hedgeweek (2025), relaying Bloomberg's analysis of Form ADV filings, report that Balyasny's flagship fund charged over \$768 million in pass-through fees in 2023, with a gross return of 15.2% and a net return of 2.8%, and that ExodusPoint charged 8.4% in pass-through fees against an 11% gross return. By a BNP Paribas survey reported there, multi-strategy funds retained 59% of every dollar they earned in 2023, which implies pass-through expenses of 5–7% for gross returns of 10–15% and a 20% performance fee. The prospectus of the Wellington Global Multi-Strategy Fund caps the reimbursable pass-through expense, the compensation of its trading teams, at 3.50% of average net assets, with total annual expenses of 4.95% for its Class A shares (Wellington Global Multi-Strategy Fund, 2024).

Omitted and unlearned signals therefore leave payoff risk despite zero primitive noise. With $A_s = 0$, $r(0+) = A_w$: a model without weak signals forecasts nothing, while nonacquisition remains a separate action.

Given this risk curve, costs, risk aversion, and supply noise, we solve for the informed mass λ, the conditional CDF F, and the market-clearing price. A model with complexity X uses its acquired information and the price, with residual risk $r_P(X) = R_P(X/c)$ and R_P in (21); above the largest occupied complexity, the price adds no information and $r_P(X) = r(X)$. The model information ratio at complexity X is the alpha of the fund's trading profit on the excess payoff $d - P$, divided by the residual standard deviation: with $a = S_P - r_P(X)$,

$$\mathrm{IR}(X) = \frac{a\,(1 - \delta^2/S_P)}{\sqrt{a^2 + a(3\delta^2 + S_P) + \delta^2 S_P - \delta^2\big(2S_P - r_P(X)\big)^2/S_P}},$$

annualized by $\sqrt{12}$. It is zero for a fund without information and does not depend on γ. In the baseline, the probes $X \leq 1$ lie below the smallest occupied complexity and the probes $X \geq 2.5$ inside the occupied support (Appendix J.4).

Let m_f denote the vector of eight model financial moments, m_f^{data} their empirical counterparts, and Σ_f their covariance matrix from a paired circular block bootstrap with 24-month blocks and 5,000 draws, which resamples months and holds the forecasts fixed. Writing $LL' = \Sigma_f$, the fitting criterion is

$$\mathcal{Q} = \left\| L^{-1}(m_f - m_f^{\mathrm{data}}) \right\|^2 + z_{\mathcal{B}}^2 + z_{\lambda}^2 + z_C^2,$$

where

$$z_{\mathcal{B}} = \frac{\log(\mathcal{B}_{50}/750{,}000)}{0.30}, \quad z_{\lambda} = \frac{\lambda - 0.53}{0.08},$$
$$z_C = \frac{\log(C_H/0.0483)}{0.50}.$$

Here $\mathcal{B}_{50}$ is the median annual dollar budget among acquirers and $C_H = 12\chi(q_H)$ is the annual research expense of the largest model as a fraction of its capital. The financial weighting accounts for correlations across targets; block lengths of 12 and 60 months change the bootstrap standard errors by less than 15%. The three survey denominators are weighting tolerances, rather than estimated standard errors. The five parameters vary jointly in bounded optimization over their logarithms, using a global search, multiple local starts, and subsequent equilibrium-grid refinement. The reported fit has $\mathcal{Q} = 5.24$. Eleven moments for

five parameters yield six nominal excess restrictions, but this count does not establish full identification or justify a formal overidentification test.

The moments do not identify the shape parameter h_c: they constrain the residual risk at full complexity and the moments of the excess payoff $d - P$, not the payoff variance $\sigma_d^2 = R_\infty(1+h_c)/h_c$ itself. We therefore fix h_c and fit R_∞. At $h_c = 10$ the payoff volatility is 16.9% per year, close to the 16.3% volatility of the excess return; smaller values leave the fit unchanged but imply a payoff variance that grows without bound and that the price explains almost entirely. The results are robust to this choice: refitting with h_c fixed at 0.01, 1, or 3 changes the objective by less than 0.8, leaves participation and the ratio of the largest to the smallest model's Sharpe ratio unchanged, and moves the largest occupied complexity by less than a fifth, while $h_c = 30$ fits worse. The expense target sets the performance gap: moving it over its evidence range, from 3% to 10% of capital, moves the ratio of the largest to the smallest model's Sharpe ratio from 1.3 to 2.2 and the fitted premium from 5.8% to 2.6%, and leaves participation and the largest occupied complexity at their baseline values.

J.4 Baseline equilibrium and numerical checks

The baseline has both informed and uninformed investors. The informed distribution has a continuous component and an atom at its upper endpoint, as in Theorem 5. Table 4 reports the resulting distribution, performance, and spending outcomes. Complexities are measured in X units; the corresponding scale endpoints are $q_L = X_L/c$ and $q_H = X_H/c$.

For gross trading profit $\pi_q = x_q(P)(d^\infty - P)$, the gross Sharpe ratio is $E^\infty[\pi_q]/\sqrt{\mathrm{Var}^\infty[\pi_q]}$, annualized by $\sqrt{12}$. A deterministic research charge replaces its numerator by $E^\infty[\pi_q] - \chi(q)$, leaving its denominator unchanged. We call this performance *after allocated research costs*; management and performance fees are not modeled. Because F is conditional on acquisition, aggregate research spending as a fraction of total capital is

$$\lambda \int_{(0,1]} C(cq)\, dF(q) = \lambda\chi_0 + \lambda\bar{\chi}_X c^2 \int_{(0,1]} q^2\, dF(q).$$

In cost comparisons, all other baseline primitives are held fixed and equilibrium is solved again for each cost pair; along the history path, the signal universe, the prior, and the cost of each scale are held fixed and only $c = N/T$ changes.

With unit mean supply, the gross Sharpe ratio increases with scale when $S_P \geq 3\delta^2$, by the proof of Corollary 4. The baseline has $\delta = 0.00401$ and $S_P = 0.00223$ per month, so $\delta^2/S_P = 0.0072 < 1/3$. The condition also holds for every cost pair and history length.

Table 4: Baseline equilibrium outcomes

Outcome	Baseline
Share acquiring information, λ	53.3%
Mean / median complexity X among acquirers	17.0 / 6.9
Support of acquired complexity, $[X_L, X_H]$	[1.46, 41.5]
Weak signals in the smallest / largest model	175 / 4,979
Top mass as a share of acquirers	29.0%
Gross Sharpe ratio: uninformed / smallest / largest model	0.295 / 0.295 / 0.422
Mean gross Sharpe ratio among acquirers	0.338
Equity premium / excess-return volatility	4.81% / 16.3%
Price informativeness, $\rho(d, P)^2$	6.7%
Research expense of the largest model/capital	2.61% per year
Total research spending/total capital	46.7 bp per year

Annual units. Distributional moments are conditional on acquisition unless stated otherwise. The top mass is 15.5% of the whole population. The mean Sharpe is an average of fund-level ratios, rather than the Sharpe of a pooled portfolio.

Every posterior variance is at most σ_d^2, so $\delta \leq \gamma\sigma_d^2$. The price adds no information to the full signal set, so $S_P \geq \mathrm{Var}^{\infty}[d^{\infty} \mid \mathcal{F}_1^{\infty}, P] = R(1)$. Hence $\delta^2/S_P \leq \gamma^2\sigma_d^4/R(1) = 0.0086 < 1/3$.

Grid refinements and independent Gaussian profit quadrature verify the numerical calculations, and the saved equilibrium satisfies the acquisition optimality checks to numerical precision. The economic fit is less exact: the fitted information ratios lie below their targets at every feature count, by at most one and a quarter bootstrap standard errors, the fitted premium of 4.8% per year is below its 7.3% target, and the fitted expense of the largest model, 2.6% of capital, is below the 4.8% target. Multiple starts do not establish a global optimum.

J.5 Why quadratic research costs?

Two training-compute arguments motivate curvature without adding calibration moments. First, Timmermann and Vulicevic (2026) (August 30 version, Appendix A.1–A.2) use the epoch schedule

$$\mathrm{ep}(n) = \max\{\lfloor 0.2n^{0.8} \rfloor, 1\} + 10,$$

where n is the number of trainable parameters, and charge approximately $6nn_{\text{obs}}\,\text{ep}(n)$ floating-point operations for n_{obs} training observations. At fixed observation count, leading compute grows as $n^{1.8}$ because both operations per pass and the number of passes increase. The schedule exhausts validation performance; it does not estimate dollar-cost elasticities, and its floor and additive epochs prevent an exact power law.

Second, the GPT-3 accounting in Brown et al. (2020), Appendix D, gives approximately $6nn_{\text{tok}}$ training operations for n parameters and n_{tok} processed tokens. The Chinchilla results of Hoffmann et al. (2022) prescribe approximately proportional growth in parameters and tokens along the compute-optimal frontier. Together these relations give

$$\text{training operations} \;\propto\; nn_{\text{tok}}, \qquad n_{\text{tok}} \propto n \quad\Longrightarrow\quad \text{training operations} \;\propto\; n^2.$$

At fixed token count, compute is linear; these tokens are distinct from the monthly window T.

The variable charge represents research and training, including predictive-prior development. At fixed T, $X = N_1/T$ is proportional to model scale; under this analogy, the compute arguments motivate a quadratic approximation if dollar prices per operation are constant. Survey moments discipline χ_0 and $\bar{\chi}_X$; individual budget components need not scale quadratically.

K Public disclosures of compute and data scale

Table 5: Public Statements Emphasizing Compute and Data Scale

Firm	Timmermann and Vulicevic (2026), August 2026	November 2026
Renaissance Technologies	"A research database that grows by more than **40 terabytes a day**"; "**52,000 computer cores**"; "**Redundant computational facilities**."	—
Hudson River Trading	"Researchers have access to **massive compute**, custom tooling . . ."; "**high compute-per-researcher ratio**"	—
Two Sigma	"Store **300+ petabytes** of data"; "Conduct more than **100,000 market data simulations daily**"; "processing power measured in **zettaFLOPS**."	→ **600+ PB** of storage capacity; **110,000+** simulations a day
Jane Street	"We work with petabytes of data, a computing cluster with **hundreds of thousands of cores**, and a growing GPU cluster containing **thousands of high-end GPUs**."	→ **tens of thousands** of high-end GPUs
Citadel Securities	"turn cutting-edge ideas and **petabyte-scale data** into . . . models with direct trading impact."; "Leverage large scale compute and data (**petabytes; large budgets**) to run ambitious experiments and **push the frontier**"	—
Jump Trading	"Real-time inference on **petabytes** of market and alternative data"; "Massive HPC grid with **thousands of GPUs** . . ."; "**5M+ Simulations Run Daily**"	—
XTX Markets	"serve **hundreds of petabytes** and more than **100,000 compute nodes**"; "**650 Petabytes** of usable storage"; "**25,000 GPUs** in our research cluster"	→ **1+ EB** of usable storage; **€1bn+** data-center campus planned in Finland
D. E. Shaw	". . . designs and maintains the firm's **largest compute infrastructure** . . . **high-performance computing**, networking, and storage for research and trading."	—

Notes: The middle column reproduces the verbatim excerpts in Timmermann and Vulicevic (2026), Table A1, and firm names link to the sources cited there. The right column reports changes in the firms' public statements as of November 2026, and a dash marks no material change. November 2026 sources: Two Sigma, Jane Street, XTX Markets, and XTX Markets' January 2025 data-center announcement.

L CREE as a limit of finite REE

The main text defines CREE directly in the limiting economy. Under the nondegeneracy conditions stated below, this appendix recovers the positive-mass equilibria of Theorems 1 and 2 as limits of rational expectations equilibria of regularized finite economies. It therefore provides a microfoundation for the limiting construction of the main text.

Overview. We consider a sequence of finite economies of Section 3, indexed by the number N of weak signals, with history length T_N and $N/T_N \to c > 0$. To pass to the limit, we regularize these economies in two ways.

(i) *Covariance cutoff.* We condition the prior on $\Psi \preceq \bar{L} I_N$, as in (103), to control CARA expected utilities and their derivatives uniformly in N. The cutoff does not change the posterior of β.

(ii) *Price-support conditioning.* The uninformed demand we construct is a perturbation of the limiting linear demand, and it is exactly optimal only on a bounded price interval $[-J, J]$. We therefore condition the joint law of primitives on the event that the equilibrium price lies in this interval, as in (116). The event is determined by the price, which the uninformed investor observes. It is also determined by the informed information and the supply shock, and the supply shock carries no payoff information beyond the informed information. Conditioning on the event therefore leaves both investors' payoff posteriors unchanged, and the candidate becomes an exact equilibrium.

Both conditioning events have prior probability converging to one, the second because $J_N \to \infty$. Hence the regularized laws converge to the original prior in total variation,[36] and the regularization disappears in the limit: equilibrium prices, portfolios, optimized utilities, and informed fractions of the regularized economies converge to those of the CREE in the main text. The regularization serves only to construct finite equilibria; the limiting economy is that of the main text.

Table 6 lists the probability laws used in the paper.

[36] For probability laws $\mathbb{P}$ and $\mathbb{P}'$ on the same space, the total-variation distance is $\|\mathbb{P}-\mathbb{P}'\|_{\mathrm{TV}} = \sup_{\mathcal{E}} |\mathbb{P}(\mathcal{E}) - \mathbb{P}'(\mathcal{E})|$, the supremum over measurable events $\mathcal{E}$. It bounds the difference in probability of every event and controls differences in expectations of bounded functions of the state. If $\mathbb{P}(\mathcal{E}) > 0$, then $\|\mathbb{P}(\cdot \mid \mathcal{E}) - \mathbb{P}\|_{\mathrm{TV}} = \mathbb{P}(\mathcal{E}^c)$, so conditioning on an event of probability close to one moves the law only slightly.

Table 6: Probability laws. The table lists every probability law used in the paper, its definition, and where it is introduced. The limiting law $\mathbb{P}_\infty$ is both the objective law and the common belief of investors in the limiting economy, and Definition 1 and the equilibrium results of the main text use it. It is constructed on the probability space of Appendix B.2 for the single-scale economy and on that of Appendix D for the economy with scales. The other laws belong to the finite economies, indexed by the number N of weak signals. The finite-sample posteriors of Sections 4.1–4.3 are computed under the common-prior law $\mathbb{P}^0_N$. The laws $\mathbb{P}_N$, $\widetilde{\mathbb{P}}_N$, and $\widetilde{\mathbb{P}}_{N,\lambda}$, for λ in a compact subset of $(0,1]$, are obtained from $\mathbb{P}^0_N$ by conditioning on events whose probability converges to one, so they converge to $\mathbb{P}^0_N$ in total variation. The law $\widetilde{\mathbb{P}}_{N,J}$ at a fixed price bound J is an intermediate object and does not converge to $\mathbb{P}^0_N$ in total variation.

Law	Definition	Introduced in
$\mathbb{P}_\infty$	Law of the limiting economy	Section 3.3; Appendices B.2 and D
$\mathbb{P}^0_N$	Joint law of all random variables of the finite economy under the common prior	Appendix L.1
$\mathbb{P}_N$	$\mathbb{P}^0_N(\cdot \mid \mathcal{A}_N)$, with the covariance cutoff $\mathcal{A}_N = \{\Psi \preceq \bar{L} I_N\}$	Equation (103)
$\widetilde{\mathbb{P}}_{N,J}$	$\mathbb{P}_N(\cdot \mid \mathcal{E}_{N,J})$, with the candidate price in $[-J, J]$	Equation (116)
$\widetilde{\mathbb{P}}_N$	$\widetilde{\mathbb{P}}_{N,J}$ at a fixed informed fraction and the price bound $J = J_N \to \infty$ of Theorem L.2	Appendix L.4
$\widetilde{\mathbb{P}}_{N,\lambda}$	$\mathbb{P}_N(\cdot \mid \mathcal{E}_{N,\lambda})$, at informed fraction λ and the common price bound J_N of Lemma L.7; Theorem L.6 uses $\widetilde{\mathbb{P}}_{N,\lambda_N}$	Appendix L.6

Roadmap. Appendix L.1 specifies the finite primitives and the limiting Gaussian equilibrium. Appendices L.2–L.4 define and construct finite REEs and prove their convergence (Theorem L.2). Appendix L.5 establishes convergence of optimized utilities. Appendix L.6 endogenizes the informed fraction at the unchanged cost χ (Theorem L.6).

L.1 Finite primitives and the limiting equilibrium

Fix an informed fraction $\lambda \in (0,1]$ until Appendix L.6, where acquisition determines it endogenously. Let $\mathbb{P}_N^0$ be the joint law of all random variables of the finite economy in Section 3, with the unknown parameters distributed according to the common prior. It covers the weak-signal parameters (β, Ψ), the signal and payoff history, the current signals, the payoff, and the supply shock. This is the common law of the primitives and the law under which all investors form their beliefs. Maintain $V_0, \kappa, \sigma_\beta, \sigma_\varepsilon, \sigma_\eta, \gamma > 0$ and the known strong block (b, ψ). In particular, the finite implementation proved here assumes $\sigma_\varepsilon^2 > 0$. The resulting bound $R_N \geq \sigma_\varepsilon^2$ gives the derivative bounds in the proof of Lemma L.3 and the continuity of the informed demand $(M_N - P_N)/(\gamma R_N)$ in (M_N, R_N, P_N). Lemma L.1 does not use it. The zero-noise calibration of Appendix J, which sets $\sigma_\varepsilon^2 = 0$, is not covered. Under the prior, β and Ψ are independent, with $\beta \sim \mathcal{N}(0, \sigma_\beta^2 I_N/N)$, $\Psi \sim \mathcal{W}^{-1}(V_0 N I_N, (1+\kappa)N)$.

By Lemma I.6, the spectrum of Ψ under the prior has the finite limiting upper edge $V_0/(\sqrt{1+\kappa}-1)^2$, and $\limsup_N \|\Psi\|$ is at most this edge almost surely. We can therefore fix a constant $\bar{L}$ with

$$\bar{L} > \frac{V_0}{(\sqrt{1+\kappa}-1)^2}$$

and set

$$\mathcal{A}_N = \{\Psi \preceq \bar{L} I_N\}, \qquad \mathbb{P}_N = \mathbb{P}_N^0(\cdot \mid \mathcal{A}_N). \tag{103}$$

Since $\bar{L}$ lies strictly above the edge, $\mathbb{P}_N^0(\mathcal{A}_N) \to 1$. The cutoff is public and leaves the Gaussian prior on β unchanged. It also leaves the conditional posterior of β given the observed regression data unchanged: conditional on the past signals, the likelihood for β factors from the likelihood for Ψ. In particular, Lemma 1(i) continues to apply exactly. The cutoff is used only in this appendix's finite implementing economies.

Write d_N for the payoff d of the N-th economy and $\mathcal{I}_N = \sigma(\boldsymbol{\mathcal{S}}, \mathbf{d}, s, S)$ for the informed

information. Define

$$\begin{aligned} M_N &= b's + \hat{\beta}'S, \\ R_N &= r_0 + S'\hat{\Sigma}_\beta S, \qquad r_0 = \sigma_\varepsilon^2, \\ \mathsf{A} &= \frac{\gamma(1-\lambda)}{\lambda}, \qquad K = \gamma\sigma_\eta. \end{aligned}$$

Under $\mathbb{P}_N$, $d_N \mid \mathcal{I}_N \sim \mathcal{N}(M_N, R_N)$, $R_N \geq r_0$, and the standard Gaussian supply shock η is independent of all other primitives. Thus the informed demand at a price that reveals only the supply shock in addition to $\mathcal{I}_N$ is $(M_N - p)/(\gamma R_N)$.

The cutoff gives the informed statistics exponential moments that are uniform in N. These bounds justify differentiating the finite CARA integrals on every fixed compact set of positions.

Lemma L.1 (Uniform exponential moments) *For each fixed $a_1, a_2 \geq 0$ there is $N_0(a_1, a_2)$ such that*

$$\sup_{N \geq N_0(a_1,a_2)} E_{\mathbb{P}_N}\left[e^{a_1|M_N| + a_2 R_N}\right] < \infty. \tag{104}$$

The same holds with any fixed polynomial factor in (M_N, R_N) inside the expectation.

The proof uses a Gaussian mixture representation of (d_N, M_N), which later arguments also use, and chi-square envelopes for its variances. Recall $A_s = b'\psi b$ and put

$$H_N = A_s + S'\left(\frac{\sigma_\beta^2}{N} I_N - \hat{\Sigma}_\beta\right) S.$$

Conditional on the past weak signals $\mathbf{S}$, Ψ, and the current weak signal S, integration over the Gaussian coefficient, historical payoff noise, and current strong signal gives

$$(d_N, M_N) \mid (H_N, R_N) \sim \mathcal{N}\left(0, \begin{pmatrix} H_N + R_N & H_N \\ H_N & H_N \end{pmatrix}\right). \tag{105}$$

Indeed, the posterior-mean coefficient has covariance $\sigma_\beta^2 I_N/N - \hat{\Sigma}_\beta$ conditional on $\mathbf{S}$, and its covariance with β is the same matrix. Write $S = \Psi^{1/2} Z_N$ with $Z_N \sim \mathcal{N}(0, I_N)$ independent

of Ψ. Since $0 \preceq \hat{\Sigma}_\beta \preceq \sigma_\beta^2 I_N/N$ and $\Psi \preceq \bar{L} I_N$ under $\mathbb{P}_N$,

$$\begin{aligned} 0 \le R_N - r_0 &\le \frac{\sigma_\beta^2 \bar{L}}{N} \|Z_N\|^2, \\ H_N + R_N = A_s + r_0 + \frac{\sigma_\beta^2}{N}\|S\|^2 &\le A_s + r_0 + \frac{\sigma_\beta^2 \bar{L}}{N}\|Z_N\|^2. \end{aligned} \tag{106}$$

Proof of Lemma L.1. By (105), M_N is centered Gaussian with variance H_N given (H_N, R_N), so $E_{\mathbb{P}_N}[e^{a_1|M_N|} \mid H_N, R_N] \le 2e^{a_1^2 H_N/2}$. With $u = a_1^2/2 + a_2$ and $H_N, R_N \ge 0$, this gives

$$E_{\mathbb{P}_N}\left[e^{a_1|M_N| + a_2 R_N}\right] \le 2E_{\mathbb{P}_N}\left[e^{u(H_N+R_N)}\right] \le 2e^{u(A_s+r_0)}\left(1 - \frac{2u\sigma_\beta^2 \bar{L}}{N}\right)^{-N/2},$$

by (106) and the chi-square moment-generating function $E_{\mathbb{P}_N}[e^{v\|Z_N\|^2/N}] = (1-2v/N)^{-N/2}$ for $N > 2v$. The right-hand side is finite for $N > 2u\sigma_\beta^2\bar{L}$ and converges to $2e^{u(A_s+r_0+\sigma_\beta^2\bar{L})}$, which proves (104). A fixed polynomial factor is dominated by a constant times $e^{|M_N|+R_N}$, so it only increases a_1 and a_2 by one. □

The limiting equilibrium. The finite equilibria converge to the CREE of Theorem 1 at the fixed informed fraction $\lambda \in (0, 1]$. In the notation of this appendix, this limiting equilibrium has linear uninformed demand $g^\infty(p) = -\vartheta p$ and price

$$P^\infty = a\,(M^\infty - (K/\lambda) r \eta), \qquad \eta \sim \mathcal{N}(0,1) \text{ independent of } (d^\infty, M^\infty),$$

where (d^∞, M^∞, r) is the limiting law in (107) and the constants $\vartheta > 0$ and a are given in (108). We now derive these objects.

Limiting law. Lemma 2 and its proof in Appendix B.1 give the following limit under $\mathbb{P}_N^0$. It carries over to $\mathbb{P}_N$ because $\|\mathbb{P}_N - \mathbb{P}_N^0\|_{\mathrm{TV}} = \mathbb{P}_N^0(\mathcal{A}_N^c) \to 0$:

$$(d_N, M_N, R_N) \Rightarrow (d^\infty, M^\infty, r), \qquad \begin{pmatrix} d^\infty \\ M^\infty \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} B + r & B \\ B & B \end{pmatrix}\right), \tag{107}$$

Here $r = \sigma_I^2(c)$, $B = \sigma_d^2 - r$, and M^∞ is the limiting forecast $\tilde{S}$ of the main text.

Market clearing. Multiplying the clearing condition $\lambda(M - P)/(\gamma R) + (1-\lambda)g(P) = \sigma_\eta \eta$

by $\gamma R/\lambda$ gives

$$W = P - \mathsf{A}R\, g(P), \qquad W = M - (K/\lambda)R\eta.$$

In the limit, $R = r$ and $W = M^\infty - (K/\lambda)r\eta$ is the price signal s_P of the main text, with noise loading $\nu = Kr/\lambda$. It is Gaussian with variance $B + (K/\lambda)^2 r^2$ and covariance B with d^∞, so

$$E_{\mathbb{P}_\infty}[d^\infty \mid W] = \alpha W, \qquad \mathrm{Var}_{\mathbb{P}_\infty}[d^\infty \mid W] = U,$$

with α and U as in (108).

Uninformed demand. Suppose the uninformed demand is $g^\infty(p) = -\vartheta p$ with $\vartheta > 0$. Clearing gives $W = (1 + \mathsf{A}\vartheta r)P$, so the price reveals W and $P = aW$ with $a = 1/(1 + \mathsf{A}\vartheta r)$. The uninformed CARA demand is therefore

$$\frac{\alpha W - P}{\gamma U} = \frac{\alpha(1 + \mathsf{A}\vartheta r) - 1}{\gamma U}\, P.$$

Equating it to $-\vartheta P$ gives $\vartheta(\gamma U + \alpha \mathsf{A} r) = 1 - \alpha$. Collecting the constants,

$$\begin{aligned} &S_W = B + (K/\lambda)^2 r^2, \qquad \alpha = \frac{B}{S_W}, \qquad U = r + \alpha(K/\lambda)^2 r^2, \\ &\varsigma = \gamma U + \alpha \mathsf{A} r, \qquad \vartheta = \frac{1-\alpha}{\varsigma}, \qquad a = \frac{1}{1 + \mathsf{A}\vartheta r}. \end{aligned} \tag{108}$$

Here $\alpha = \alpha(c)$, $U = \sigma_U^2(c)$, and a is the price loading in (15), evaluated at the fixed λ; ϑ is the uninformed demand slope in (14). Since $\alpha < 1$, $\vartheta > 0$, which confirms the conjectured sign.

Theorem L.2 (Trading implementation at a fixed informed fraction) *Under the preceding assumptions, there are price bounds $J_N \to \infty$, common joint laws $\widetilde{\mathbb{P}}_N$, and continuous nonincreasing demand schedules g_N such that, for every sufficiently large N:*

(i) The price P_N, the informed demand $(M_N - P_N)/(\gamma R_N)$, and the uninformed schedule g_N form a finite REE at λ under $\widetilde{\mathbb{P}}_N$ (Definition 5), with $|P_N| \le J_N$ almost surely.

(ii) $\|\widetilde{\mathbb{P}}_N - \mathbb{P}_N\|_{\mathrm{TV}} \to 0$ and $\|\widetilde{\mathbb{P}}_N - \mathbb{P}_N^0\|_{\mathrm{TV}} \to 0$.

(iii) Under $\widetilde{\mathbb{P}}_N$, the payoff, price, and both portfolios converge jointly in distribution to

$$\left(d^\infty, P^\infty, \frac{M^\infty - P^\infty}{\gamma r}, -\vartheta P^\infty\right).$$

On every fixed price interval, $g_N \to -\vartheta p$ in C^1.

Here $\widetilde{\mathbb{P}}_N = \mathbb{P}_N(\cdot \mid |P_N| \le J_N)$ is a conditioned law. It differs from the common-prior law $\mathbb{P}_N^0$ and depends on λ and g_N.

The proof occupies Appendices L.2–L.4 and has three steps. Appendix L.2 constructs, for each bounded price interval, an uninformed demand that is exactly optimal on that interval under its own price law. Appendix L.3 defines a finite REE and conditions the primitive law on the event that the price lies in the interval, which turns this candidate into a finite REE. Appendix L.4 lets the interval expand with N and proves the convergence claims.

L.2 Exact conditional optimality on compact price intervals

This subsection proves the following lemma.

Lemma L.3 (Compact-price continuation) *For every fixed $J < \infty$ and all sufficiently large N, there is a bounded, continuous, globally nonincreasing schedule $\bar{g}_{N,J}$ with $\bar{g}_{N,J}(0) = 0$ whose candidate price clears uniquely in every state. On $[-J, J]$, the schedule is C^1 and is the unrestricted optimal uninformed demand under its own candidate price law. Its deviation from the limiting demand $g^\infty(p) = -\vartheta p$, $\delta_{N,J}(p) = \bar{g}_{N,J}(p) + \vartheta p$, satisfies*

$$\|\delta_{N,J}\|_{C^1([-J,J])} = \sup_{|p|\le J} |\delta_{N,J}(p)| + \sup_{|p|\le J} |\delta'_{N,J}(p)| \to 0.$$

No optimality is asserted outside that interval.

Optimality of the uninformed demand reduces to a first-order differential equation for the demand schedule, (111) below. We first derive the price law of a candidate schedule and this equation. The proof then solves the equation near the limiting linear demand.

Price law of a candidate schedule. Let g be a bounded, continuous, nonincreasing candidate with $g(0) = 0$. Its clearing price is unique in every state and, where g is

differentiable, has conditional density

$$\ell_g(p \mid M, R) = \left(\frac{1}{R} - \mathsf{A}g'(p)\right) \frac{\lambda}{K} \varphi\left(\frac{\lambda(p - M)}{KR} - \frac{\lambda \mathsf{A} g(p)}{K}\right), \tag{109}$$

where φ denotes the standard normal density. To see this, write market clearing as in the limiting equilibrium,

$$F_R^g(p) = p - \mathsf{A}Rg(p), \qquad F_{R_N}^g(P) = W_N, \qquad W_N = M_N - (K/\lambda)R_N\eta. \tag{110}$$

For each $R \geq r_0$, F_R^g is strictly increasing and onto $\mathbb{R}$, so the clearing price is unique. The density (109) is the Gaussian likelihood of the residual supply W_N at $F_R^g(p)$, multiplied by the clearing Jacobian. Global monotonicity is essential: there is exactly one root, so no additional price-selection factor enters this density.

Optimality condition. Under this price law, the uninformed investor's conditional CARA loss is strictly convex in the position. Hence $g(p)$ is the unrestricted optimal position at price p whenever it lies in the interior of the moment domain and satisfies

$$C_N(p, g(p)) - \mathsf{A}D_N(p, g(p))g'(p) = 0, \tag{111}$$

with C_N and D_N defined in (113). For an unrestricted position x, the conditional CARA loss is

$$\mathcal{L}_{N,g}(x; p) = \frac{E_{\mathbb{P}_N}\left[e^{-\gamma x(M_N - p) + \gamma^2 x^2 R_N/2} \ell_g(p \mid M_N, R_N)\right]}{E_{\mathbb{P}_N}[\ell_g(p \mid M_N, R_N)]}. \tag{112}$$

The investor minimizes this expression over $x \in \mathbb{R}$, allowing the loss to be $+\infty$ outside its moment domain. Since $e^{-\gamma x(M_N - p) + \gamma^2 x^2 R_N/2} = E_{\mathbb{P}_N}[e^{-\gamma x(d_N - p)} \mid \mathcal{I}_N]$, the loss is $\mathcal{L}_{N,g}(x; p) = E_{\mathbb{P}_N}[e^{-\gamma x(d_N - p)} \mid P = p]$. Its logarithm is the cumulant generating function of $-\gamma(d_N - p)$ under the uninformed posterior at price p. It is therefore convex on its domain, and strictly convex because d_N is not almost surely constant given the price. An interior zero of the loss derivative is therefore the unique unrestricted global minimizer. To compute the derivative, put

$$\mathcal{K}(p, y; M, R) = \frac{\lambda}{K} \varphi\left(\frac{\lambda(p - M)}{KR} - \frac{\lambda \mathsf{A} y}{K}\right), \qquad \mathcal{J}(p, y; M, R) = e^{-\gamma y(M - p) + \gamma^2 y^2 R/2},$$

and define

$$
\begin{aligned}
C_N(p,y) &= E_{\mathbb{P}_N}\left[\frac{p - M_N + \gamma R_N y}{R_N}\,\mathcal{J}(p,y;M_N,R_N)\mathcal{K}(p,y;M_N,R_N)\right], \\
D_N(p,y) &= E_{\mathbb{P}_N}[(p - M_N + \gamma R_N y)\,\mathcal{J}(p,y;M_N,R_N)\mathcal{K}(p,y;M_N,R_N)]\,.
\end{aligned}
\tag{113}
$$

Differentiating (112) in x, holding the price law fixed, and evaluating at $x = g(p)$ gives (111). The equation must be kept in this undivided form because D_N can vanish.

The proof of Lemma L.3 uses one result on the persistence of normally hyperbolic invariant manifolds. Let $\mathbb{T}^m$ denote the m-dimensional torus.

Lemma L.4 (Persistence of a repelling torus of rest points) *Fix $m \geq 1$ and $\varepsilon_0 > 0$. Let X_∞ and X_N, $N \geq 1$, be C^1 vector fields on $\mathbb{T}^m \times \mathbb{R}$, with coordinates (θ, z), such that*

$$
X_\infty(\theta,z) = \big(\omega(\theta)z, \varpi(\theta)z\big), \qquad \inf_\theta \varpi(\theta) > 0,
$$

$X_N = X_\infty$ for $|z| \geq \varepsilon_0$, and $\|X_N - X_\infty\|_{C^1} \to 0$. Then, for all sufficiently large N, there is a unique continuous function $h_N : \mathbb{T}^m \to \mathbb{R}$ whose graph $\{z = h_N(\theta)\}$ is invariant under X_N. It is C^1, and $\|h_N\|_{C^1} \to 0$.

Proof of Lemma L.4. *Existence.* The torus $\mathcal{T}_0 = \mathbb{T}^m \times \{0\}$ consists of rest points of X_∞. At $(\theta, 0)$, the linearization of X_∞ has eigenvalue zero on the tangent space of $\mathcal{T}_0$ and eigenvalue $\varpi(\theta)$ on the transverse direction $(\omega(\theta), \varpi(\theta))$. Put $\varpi_0 = \inf_\theta \varpi(\theta) > 0$. The tangential rate is zero and the normal expansion rate is at least ϖ_0, so the normal rate dominates every multiple of the tangential rate. Hence $\mathcal{T}_0$ is a compact normally repelling invariant manifold of X_∞; see Eldering (2016, Example 4.3). The fields grow at most linearly in z, so their flows are complete. After time reversal, $\mathcal{T}_0$ is normally attracting, which is the case that the last of the following references treats. A compact normally hyperbolic invariant manifold persists under C^1-small perturbations of the vector field as a C^1-close invariant manifold; see Fenichel (1971), Hirsch et al. (1977, Theorem 4.1), and Eldering (2016, Theorems 4.4 and 5.2). A manifold that is C^1-close to $\mathcal{T}_0$ is the graph of a C^1 function h_N with $\|h_N\|_{C^1} \to 0$.

Uniqueness. Let Γ be a compact set that is invariant under X_N, and put $\epsilon_N = \|X_N - X_\infty\|_{C^1}$. Take a point of Γ and let $\zeta = z - h_N(\theta)$ along its orbit. Write X_N^θ and X_N^z for the components of X_N. The graph of h_N is invariant, so the scalar function $z \mapsto X_N^z(\theta, z) - Dh_N(\theta) X_N^\theta(\theta, z)$ vanishes at $z = h_N(\theta)$, and the mean value theorem gives

$$
\dot\zeta = \big(\partial_z X_N^z - Dh_N(\theta)\,\partial_z X_N^\theta\big)\zeta,
$$

with the bracket evaluated at an intermediate value of z. Since $\partial_z X^z_\infty = \varpi(\theta) \geq \varpi_0$ and $\partial_z X^\theta_\infty = \omega(\theta)$ is bounded on the torus, $\partial_z X^z_N \geq \varpi_0 - \epsilon_N$ and $\partial_z X^\theta_N$ is bounded uniformly in N; also $\|Dh_N\| \to 0$. For large N the coefficient is at least $\varpi_0/2$, so $|\zeta|$ grows without bound along the orbit unless $\zeta = 0$. Since the orbit stays in the compact set Γ, $\zeta = 0$, and Γ lies in the graph of h_N. The graph of a continuous function on $\mathbb{T}^m$ is compact, so an invariant graph is the graph of h_N. □

Proof of Lemma L.3. Equation (111) is a first-order differential equation for g, but it cannot be solved for g': D_N can vanish, and in the limit both C_∞ and D_∞ vanish on the target line $y = -\vartheta p$. We therefore treat it geometrically. The planar system

$$\dot{p} = \mathsf{A} D_N(p, y), \qquad \dot{y} = C_N(p, y) \tag{114}$$

assigns a direction of motion to each point (p, y). A curve $y = g(p)$ is invariant under this system, meaning that trajectories starting on it stay on it, exactly when this direction is tangent to the curve at each of its points, that is, when g solves (111). At $N = \infty$, every point of the line $y = -\vartheta p$ is at rest, and nearby points move away from it at a positive rate. Such a line survives small perturbations of the system, so for large N there is an invariant curve C^1-close to it, and this curve is the finite demand. We make this precise in five steps.

Convergence of the field. The functions C_N, D_N and their first derivatives converge uniformly on compact sets of (p, y) to their Gaussian counterparts. Differentiating the kernels produces polynomial factors in M_N, R_N, R_N^{-1} multiplying bounded Gaussian factors and the exponential factor $\mathcal{J}$. Since $R_N^{-1} \leq r_0^{-1}$, the moment bounds of Lemma L.1 control these derivatives. Weak convergence (107), uniform integrability, and equicontinuity then give compact C^1 convergence. The argument applies to every fixed finite number of derivatives.

Limiting field. Write $w = p - \mathsf{A} r y$. Gaussian integration in (113) gives

$$C_\infty(p, y) = \Lambda(p, y)(y + \vartheta p), \qquad D_\infty(p, y) = r C_\infty(p, y),$$
$$\Lambda(p, y) = \frac{\varsigma}{\sqrt{2\pi S_W}} \exp\left\{ -\frac{w^2}{2S_W} - \gamma y(\alpha w - p) + \frac{\gamma^2 y^2 U}{2} \right\} > 0.$$

Indeed, $d^\infty \mid W = w$ has mean αw and variance U. The first-order factor, price minus conditional payoff mean plus γy times conditional payoff variance, is therefore $p - \alpha w + \gamma U y = \varsigma(y + \vartheta p)$. In the system (114), divide the field by the strictly positive function Λ and change

coordinates to $z = y + \vartheta p$. The limiting field is

$$\dot{p} = \mathsf{A}rz, \qquad \dot{z} = (1 + \vartheta \mathsf{A}r)z. \tag{115}$$

Thus $z = 0$ is a line of rest points with tangential eigenvalue zero and a strictly positive normal eigenvalue $1+\vartheta \mathsf{A}r$. The normalization is harmless on compact neighborhoods, where Λ is bounded away from zero.

Persistence. We apply Lemma L.4 with $m = 1$. In the coordinates (p, z), the normalized finite field is $Y_N = \big(\mathsf{A}D_N,\ C_N + \vartheta \mathsf{A}D_N\big)/\Lambda$, evaluated at $y = z - \vartheta p$, and its limit Y_∞ is the field (115). Fix $\varepsilon_0 > 0$. The band $|z| < \varepsilon_0$ is not invariant under these fields, and the chart in p has a boundary. We therefore modify Y_N away from the region of interest. Place the p coordinate on a circle that contains the chart $|p| \le J + 2$. Fix smooth cutoffs $\mathsf{c}_1(p)$, equal to one on $|p| \le J + 1$ and to zero outside $|p| < J + 2$, and $\mathsf{c}_2(z)$, equal to one on $|z| \le \varepsilon_0/2$ and to zero outside $|z| < \varepsilon_0$. Set

$$X_\infty = Y_\infty, \qquad X_N = Y_\infty + \mathsf{c}_1(p)\,\mathsf{c}_2(z)\,(Y_N - Y_\infty).$$

The cutoffs act only on the difference from the limiting field, so X_N is a C^2 field on the whole cylinder and equals X_∞ for $|z| \ge \varepsilon_0$. The convergence step, applied with two derivatives, gives $\|Y_N - Y_\infty\|_{C^2} \to 0$ on the compact set $\{|p| \le J+2,\ |z| \le \varepsilon_0\}$. Hence $\sup_N \|X_N - X_\infty\|_{C^2} < \infty$ and $\|X_N - X_\infty\|_{C^1} \to 0$. The limiting field has the form required by the lemma, with $\omega = \mathsf{A}r$ and $\varpi = 1 + \vartheta \mathsf{A}r \ge 1$. Lemma L.4 gives an invariant graph $z = h_N(p)$ of X_N with $\|h_N\|_{C^1} \to 0$. For large N, $|h_N| < \varepsilon_0/2$, so $X_N = Y_N$ on the part of the graph over $|p| \le J + 1$. Returning to the original chart,

$$g_{N,J}(p) = -\vartheta p + h_N(p)$$

therefore solves (111) exactly on $[-J-1, J+1]$ and has negative derivative for large N. The graph can depend on the cutoffs; we fix them once, because the proof uses only the existence of one such graph. When $\lambda = 1$, $\mathsf{A} = 0$ and (111) reads $C_N(p, g(p)) = 0$. The claim then reduces to the implicit function theorem, because $\partial_y C_\infty = \Lambda > 0$ on $z = 0$. The auxiliary circle is only a device for constructing this local graph; the economic price space remains $\mathbb{R}$.

Zero demand at zero price. The graph passes through the origin, $g_{N,J}(0) = 0$. Given

(H, R) in (105), set

$$a_{H,R} = \frac{H}{H + (K/\lambda)^2 R^2}, \qquad s_{H,R} = R + \frac{H(K/\lambda)^2 R^2}{H + (K/\lambda)^2 R^2}.$$

At $p = 0$ and $y = g_{N,J}(0)$, the first-order factor of the component with given (H, R) is $y(\mathsf{A} a_{H,R} R + \gamma s_{H,R})$. Its coefficient is strictly positive, and the likelihood and Jacobian weights are positive because $g'_{N,J} < 0$. The mixture first-order condition therefore forces $y = 0$. The decreasing graph consequently opposes the sign of price.

Completion outside the interval. Restrict the graph to $[-J-1, J+1]$ and extend it by its endpoint constants. The completed schedule $\bar{g}_{N,J}$ is bounded, continuous, nonincreasing, and sign opposing. It clears uniquely for every R. On $[-J, J]$, its exact likelihood is still (109), using the same local values and derivatives as in (111); changing the tails does not change this likelihood for a globally monotone schedule. The selected positions on that interval remain in a fixed compact set, inside the moment domain for large N. Strict convexity of (112) turns their first-order conditions into unrestricted global optimality. □

L.3 Joint conditioning and finite REE

We first define the finite equilibrium concept. It takes the law of the primitives as given.

Definition 5 (Finite REE) *Fix $\lambda \in (0, 1]$ and a probability law $\mathbb{P}'$ on the primitives of the N-th economy under which d_N and η are independent conditional on $\mathcal{I}_N$. A* finite REE *at λ under $\mathbb{P}'$ is a $\sigma(\mathcal{I}_N, \eta)$-measurable price P, a $\sigma(\mathcal{I}_N, P)$-measurable informed demand x_I, and a measurable uninformed schedule $x_U : \mathbb{R} \to \mathbb{R}$ such that:*

(i) $x_I \in \operatorname{argmax}_{x \in \mathbb{R}} E_{\mathbb{P}'}\left[-e^{-\gamma x(d_N - P)} \mid \mathcal{I}_N, P\right]$, $\mathbb{P}'$*-almost surely;*

(ii) $x_U(p) \in \operatorname{argmax}_{x \in \mathbb{R}} E_{\mathbb{P}'}\left[-e^{-\gamma x(d_N - p)} \mid P = p\right]$ *for almost every p under the $\mathbb{P}'$-law of P;*

(iii) $\lambda x_I + (1 - \lambda) x_U(P) = \sigma_\eta \eta$, $\mathbb{P}'$*-almost surely.*

Under $\mathbb{P}' = \mathbb{P}^0_N$, this is a rational expectations equilibrium of the finite economy of Section 3. We apply the definition under the conditioned law $\widetilde{\mathbb{P}}_{N,J}$ in (116), which differs from $\mathbb{P}^0_N$ in three ways. First, η is no longer independent of $\mathcal{I}_N$; only the conditional independence of d_N and η given $\mathcal{I}_N$ survives. Second, the marginal law of $\mathcal{I}_N$ is reweighted, so (β, Ψ) no longer has the prior (2). Third, the law is indexed by the candidate schedule and hence by equilibrium objects. The paragraph on the conditioned law after Lemma L.5 gives the

explicit form. Condition (ii) also differs from Definition 1(ii): it restricts the schedule only at almost every price in the support of P.

Conditioning the primitive law on the event that the candidate price lies in $[-J, J]$ turns the candidate of Lemma L.3 into a finite REE. Fix such a candidate, denote its clearing price by $\bar{P}_{N,J}$, and set

$$\mathcal{E}_{N,J} = \{|\bar{P}_{N,J}| \leq J\}, \qquad \widetilde{\mathbb{P}}_{N,J} = \mathbb{P}_N(\cdot \mid \mathcal{E}_{N,J}). \tag{116}$$

For $J > 0$ this event has positive conditional probability at every informed information realization: the nonempty price interval corresponds to a nonempty interval of the Gaussian supply shock. Investors know the new common joint law $\widetilde{\mathbb{P}}_{N,J}$.

Lemma L.5 (Restriction to the region of exact optimality) *Under $\widetilde{\mathbb{P}}_{N,J}$, the price $\bar{P}_{N,J}$ and demands*

$$x_{I,N} = \frac{M_N - \bar{P}_{N,J}}{\gamma R_N}, \qquad x_{U,N} = \bar{g}_{N,J}(\bar{P}_{N,J})$$

form a finite REE at λ under $\widetilde{\mathbb{P}}_{N,J}$ (Definition 5).

Proof of Lemma L.5. Conditioning on $\mathcal{E}_{N,J}$ leaves both investors' payoff posteriors unchanged. Their optimal demands are therefore the same as under $\mathbb{P}_N$, and market clearing holds state by state.

Uninformed investor. For almost every p in the price support and every measurable payoff set D,

$$\widetilde{\mathbb{P}}_{N,J}(d_N \in D \mid \bar{P}_{N,J} = p) = \mathbb{P}_N(d_N \in D \mid \bar{P}_{N,J} = p). \tag{117}$$

Indeed, the event $\mathcal{E}_{N,J}$ is measurable with respect to the observed price. The joint price-payoff law and its price marginal acquire the same factor, $\mathbf{1}_{\{|p| \leq J\}}/\mathbb{P}_N(\mathcal{E}_{N,J})$, which cancels in Bayes' rule. The unrestricted conditional optimization problem is unchanged at every realized price, so Lemma L.3 supplies its unique optimum.

Informed investor. The price carries no payoff information beyond $\mathcal{I}_N$:

$$\begin{aligned} &\widetilde{\mathbb{P}}_{N,J}(d_N \in D \mid \mathcal{I}_N) = \mathbb{P}_N(d_N \in D \mid \mathcal{I}_N), \\ &\quad d_N \mid (\mathcal{I}_N, \bar{P}_{N,J}) \sim_{\widetilde{\mathbb{P}}_{N,J}} \mathcal{N}(M_N, R_N). \end{aligned}$$

Indeed, the candidate price is a measurable function of $(\mathcal{I}_N, \eta)$. Under $\mathbb{P}_N$, d_N and η are independent conditional on $\mathcal{I}_N$. Since the conditioning indicator depends only on $(\mathcal{I}_N, \eta)$, this conditional independence is preserved under $\widetilde{\mathbb{P}}_{N,J}$. The unrestricted CARA optimum therefore remains the stated linear demand. The same argument preserves the posterior of latent payoff parameters conditional on the full informed information.

Market clearing. Equation (110) holds state by state before conditioning and hence on $\mathcal{E}_{N,J}$. Rearranging it gives

$$\lambda \frac{M_N - \bar{P}_{N,J}}{\gamma R_N} + (1-\lambda)\bar{g}_{N,J}(\bar{P}_{N,J}) = \sigma_\eta \eta.$$

Thus market clearing is exact, and optimality holds almost surely under the common law. Beliefs and choices outside the price support do not affect these equilibrium conditions. □

The conditioned law in terms of primitives. The conditioned law $\widetilde{\mathbb{P}}_{N,J}$ is a law on primitives, fixed before trade. Under it, the supply shock given the informed information is a truncated standard normal, and the marginal law of the informed information is reweighted. To see this, note that, once the candidate schedule is constructed, strict monotonicity of its clearing transformation gives

$$\begin{aligned} \mathcal{E}_{N,J} = \big\{ -J - \mathsf{A} R_N \bar{g}_{N,J}(-J) \le M_N - (K/\lambda) R_N \eta \\ \le J - \mathsf{A} R_N \bar{g}_{N,J}(J) \big\}. \end{aligned} \tag{118}$$

Its only schedule-dependent terms are deterministic endpoint values. One conditions the primitive law on (118), fixes that law, and then verifies equilibrium by Lemma L.5. Explicitly, conditional on informed information, define

$$\begin{aligned} a_{N,J}(\mathcal{I}_N) &= \frac{M_N - J + \mathsf{A} R_N \bar{g}_{N,J}(J)}{(K/\lambda) R_N}, \\ b_{N,J}(\mathcal{I}_N) &= \frac{M_N + J + \mathsf{A} R_N \bar{g}_{N,J}(-J)}{(K/\lambda) R_N}, \\ Z_{N,J}(\mathcal{I}_N) &= \Phi(b_{N,J}(\mathcal{I}_N)) - \Phi(a_{N,J}(\mathcal{I}_N)), \end{aligned}$$

where Φ is the standard normal distribution function. Under $\widetilde{\mathbb{P}}_{N,J}$, the supply shock conditional on $\mathcal{I}_N$ is standard normal truncated to $[a_{N,J}, b_{N,J}]$. The marginal law of $\mathcal{I}_N$ is simultaneously reweighted by $Z_{N,J}(\mathcal{I}_N)/\mathbb{P}_N(\mathcal{E}_{N,J})$. This reweighting is necessary for (117). Truncating the supply shock conditional on information while keeping the original infor-

mation marginal would generally change the uninformed posterior. The modified finite law therefore permits dependence between the supply shock and informed information. It describes a static trading economy with a finite observed history; the conditioned history need not remain an independently sampled time series.

L.4 Expanding price support and convergence

Proof of Theorem L.2. We let the price interval expand slowly enough that, on it, the finite demand stays within $1/J_N$ of the linear demand. Claim (i) is then Lemma L.5. Claim (ii) holds because the price is dominated by the tight residual supply, so it leaves the expanding interval with vanishing probability. Claim (iii) holds because the price stays within a vanishing distance of its value under linear demand.

Diagonal choice of the interval. For each integer $j \geq 1$, Lemma L.3 gives a threshold n_j such that for every $N \geq n_j$ the candidate on $[-j, j]$ exists and

$$\|\delta_{N,j}\|_{C^1([-j,j])} \leq \min\{j^{-1}, \vartheta/2\}. \tag{119}$$

Take the thresholds strictly increasing and large enough that all positions used on this interval lie in the finite conditional moment domain. Choose $j(N) \to \infty$ sufficiently slowly with $N \geq n_{j(N)}$ and put

$$J_N = j(N), \qquad g_N = \bar{g}_{N,J_N}, \qquad P_N = \bar{P}_{N,J_N},$$
$$\mathcal{E}_N = \{|P_N| \leq J_N\}, \qquad \widetilde{\mathbb{P}}_N = \mathbb{P}_N(\,\cdot \mid \mathcal{E}_N).$$

Lemma L.5 proves (i). No particular growth rate for J_N is required.

Total variation. The price is dominated by the residual supply,

$$|P_N| \leq |W_N|, \qquad W_N = M_N - (K/\lambda) R_N \eta. \tag{120}$$

Indeed, because g_N is nonincreasing and $g_N(0) = 0$, it opposes the sign of price. The clearing transformation therefore has the sign of price and absolute value at least its absolute value. The convergence (107) and independence of η make W_N tight. Since $J_N \to \infty$,

$$\mathbb{P}_N(\mathcal{E}_N^c) \leq \mathbb{P}_N(|W_N| > J_N) \longrightarrow 0.$$

For any probability law $\mathbb{P}$ and event $\mathcal{E}$ of positive probability, $\|\mathbb{P}(\cdot \mid \mathcal{E}) - \mathbb{P}\|_{\mathrm{TV}} = \mathbb{P}(\mathcal{E}^c)$: the

density is $\mathbf{1}_{\mathcal{E}}/\mathbb{P}(\mathcal{E})$, so half its $L^1(\mathbb{P})$ distance from one is $\mathbb{P}(\mathcal{E}^c)$. Consequently

$$\begin{aligned}\|\widetilde{\mathbb{P}}_N - \mathbb{P}_N\|_{\mathrm{TV}} &= \mathbb{P}_N(\mathcal{E}_N^c) \to 0,\\ \|\widetilde{\mathbb{P}}_N - \mathbb{P}_N^0\|_{\mathrm{TV}} &\le \mathbb{P}_N(\mathcal{E}_N^c) + \mathbb{P}_N^0(\mathcal{A}_N^c) \to 0.\end{aligned} \tag{121}$$

These comparisons concern the full joint state laws in each finite dimension. They prove (ii).

Convergence of outcomes. Write $\delta_N = \delta_{N,J_N}$, so that $\delta_N(p) = g_N(p) + \vartheta p$. Under $\mathbb{P}_N$, clearing gives

$$P_N = \frac{M_N - (K/\lambda)R_N\eta}{1 + \mathsf{A}\vartheta R_N} + \frac{\mathsf{A}R_N}{1 + \mathsf{A}\vartheta R_N}\delta_N(P_N). \tag{122}$$

The first term is the price under linear demand; the second term vanishes in probability. Indeed, the diagonal construction makes δ_N and its derivative vanish uniformly on every fixed interval, and the prices are tight by (120), so $\delta_N(P_N) \to 0$ in probability. Since $\mathsf{A}R_N/(1 + \mathsf{A}\vartheta R_N) \le 1/\vartheta$, the second term tends to zero in probability. Equation (107) and continuous mapping now give, jointly,

$$P_N \Rightarrow P^\infty, \qquad g_N(P_N) \Rightarrow -\vartheta P^\infty, \qquad \frac{M_N - P_N}{\gamma R_N} \Rightarrow \frac{M^\infty - P^\infty}{\gamma r}.$$

The lower bound $R_N \ge r_0$ ensures continuity of the last expression. The payoff can be included in the same joint convergence. Total variation decreases under measurable maps, so (121) transfers this convergence to $\widetilde{\mathbb{P}}_N$ and proves (iii). The C^1 assertion follows directly from (119).

Error bound on the realized support. Under the modified law, the price is uniformly close to its value under linear demand:

$$\left|P_N - \frac{M_N - (K/\lambda)R_N\eta}{1 + \mathsf{A}\vartheta R_N}\right| \le \frac{1}{\vartheta J_N} \qquad \widetilde{\mathbb{P}}_N\text{-almost surely.}$$

Indeed, by (119), both the demand-level error $|g_N(p)+\vartheta p|$ and the derivative error $|g_N'(p)+\vartheta|$ are at most J_N^{-1} on $[-J_N, J_N]$, which contains the realized support. The bound follows from (122). □

Mode of convergence. Theorem L.2(iii) is convergence in distribution under laws that vary with N. Convergence in probability holds on the coupled probability space of Appendix B.2. That space carries (M_N, R_N) for every N and one supply shock η, with

$(M_N, R_N) \to (\tilde{S}, r)$ $\mathbb{P}_\infty$-almost surely; there $M^\infty = \tilde{S}$. Because g_N is a deterministic schedule, the candidate price $P_N = (F_{R_N}^{g_N})^{-1}(M_N - (K/\lambda)R_N\eta)$ is defined there state by state. Equation (122) is a clearing identity, so it holds in every state. Its first term converges almost surely to $a(\tilde{S} - (K/\lambda)r\eta) = P^\infty$. By (120), $|P_N| \leq |W_N|$ in every state, and W_N converges almost surely. The prices are therefore tight, and $\delta_N(P_N) \to 0$ in $\mathbb{P}_\infty$-probability. Hence $P_N \to P^\infty$ in $\mathbb{P}_\infty$-probability, and $\mathbb{P}_\infty(\mathcal{E}_N) \geq \mathbb{P}_\infty(|W_N| \leq J_N) \to 1$. The coupled space does not carry Ψ; if the coupling also draws Ψ from its conditional law given the observations of the N-th economy, then $\mathbb{P}_\infty(\mathcal{A}_N) = \mathbb{P}_N^0(\mathcal{A}_N) \to 1$, and so $\mathbb{P}_\infty(\mathcal{A}_N \cap \mathcal{E}_N) \to 1$. The two conditioning events thus have limiting probability one on the coupled space, where the N-th economy has the law $\mathbb{P}_N^0$.

L.5 Optimized trading values and scope

Optimized utilities. The ex-ante optimized utilities of both investors converge to their Gaussian values. For the uninformed investor, let $v_{U,N}(p)$ denote the optimized conditional utility under $\widetilde{\mathbb{P}}_N$. At a realized price it equals the baseline value by (117). On each fixed compact price interval, (112), the moment bounds of Lemma L.1, and compact C^1 convergence of g_N imply uniform convergence to

$$v_{U,\infty}(p) = -\exp\left\{-\frac{[\alpha(1+\mathsf{A}\vartheta r)p - p]^2}{2U}\right\} = -\exp\left\{-\frac{\gamma^2 U\vartheta^2 p^2}{2}\right\}.$$

The ratio in (112) converges uniformly because its denominator stays away from zero. The denominator $E_{\mathbb{P}_N}[\ell_{g_N}(p \mid M_N, R_N)]$ is the density of P_N under $\mathbb{P}_N$. It converges uniformly on compact price intervals to the $\mathcal{N}(0, a^2 S_W)$ density of P^∞, which is bounded below there. The zero position is feasible and utility is nonpositive, so $-1 \leq v_{U,N}(p) \leq 0$ on the realized support. Price tightness therefore upgrades compact convergence to convergence of ex-ante optimized utility. For the informed investor the exact conditional optimized value is

$$v_{I,N} = -\exp\left\{-\frac{(M_N - P_N)^2}{2R_N}\right\} \in [-1, 0],$$

so its ex-ante value converges directly by joint weak convergence and boundedness. These ex-ante values are the finite counterparts of the values in (4), with the informed value gross of the acquisition cost. Indeed, let $x_{I,N} = (M_N - P_N)/(\gamma R_N)$ and $x_{U,N} = g_N(P_N)$. Since

$-e^{-\gamma x(d_N - P_N)} \le 0$, the tower property gives

$$E_{\widetilde{\mathbb{P}}_N}\left[-e^{-\gamma x_{I,N}(d_N - P_N)}\right] = E_{\widetilde{\mathbb{P}}_N}[v_{I,N}], \qquad E_{\widetilde{\mathbb{P}}_N}\left[-e^{-\gamma x_{U,N}(d_N - P_N)}\right] = E_{\widetilde{\mathbb{P}}_N}[v_{U,N}(P_N)],$$

with both conditional values in $[-1, 0]$; no uniform integrability is needed. Convergence of unbounded price or portfolio moments requires separate uniform-integrability bounds; total variation alone is insufficient.

Scope. Theorem L.2 establishes a finite-economy foundation for the trading component of CREE. It does not assert exact equilibrium for every sufficiently large member of the original finite family with an independent Gaussian supply shock. The joint conditioning changes that finite law while preserving conditional optimality at all retained prices. It imposes no position limits, rationing, or outside market maker. An exogenous price cap without a primitive-law change would not suffice: conditional on informed information, the Gaussian supply shock has unbounded support, whereas $F_R^g([-J, J])$ is bounded.

L.6 Endogenous information acquisition

We now endogenize the informed fraction at a fixed cost $\chi > 0$. Acquisition occurs before investors observe the price or the information $\mathcal{I}_N$. An investor either pays χ to observe $\mathcal{I}_N$, or remains uninformed and subsequently trades using only the price. The continuum of price-taking investors is the same as in the main model; it is the signal dimension and history length that are finite. A unilateral acquisition deviation changes neither the aggregate informed fraction nor the price map. Both acquisition choices are evaluated under the same common joint law, which is held fixed when an investor deviates. The main result is that the regularized finite economies have finite REEs with optimal acquisition whose informed fractions converge to the CREE fraction $\lambda(c)$.

Theorem L.6 (Finite REE with endogenous information acquisition) *Maintain the assumptions of this appendix, let $\chi > 0$, and suppose $v = B/r > \bar{C} = e^{2\gamma\chi} - 1$, with $r = \sigma_I^2(c)$ and $B = \sigma_d^2 - r$ as in (107). Define*

$$\lambda^* = \gamma\sigma_\eta\sqrt{\frac{r(v - \bar{C})}{v\bar{C}}}, \qquad \lambda(c) = \min\{\lambda^*, 1\}. \tag{123}$$

Here $\lambda(c)$ is the informed fraction of Theorem 2. There are informed fractions $\lambda_N \in (0, 1]$,

price bounds $J_N \to \infty$, prices P_N with price-support events $\mathcal{E}_{N,\lambda_N} = \{|P_N| \le J_N\}$, and common joint laws $\widetilde{\mathbb{P}}_{N,\lambda_N} = \mathbb{P}_N(\,\cdot \mid \mathcal{E}_{N,\lambda_N})$, defined for all sufficiently large N, such that:

(i) The price P_N and the unrestricted optimal portfolios form a finite REE at λ_N under $\widetilde{\mathbb{P}}_{N,\lambda_N}$ (Definition 5). At the unchanged acquisition cost χ and under the same law, both acquisition choices are indifferent if $\lambda_N < 1$, and acquiring information is weakly preferred if $\lambda_N = 1$.

(ii) $\lambda_N \to \lambda(c)$ and $\|\widetilde{\mathbb{P}}_{N,\lambda_N} - \mathbb{P}^0_N\|_{\mathrm{TV}} \to 0$. The conditioned law $\widetilde{\mathbb{P}}_{N,\lambda_N}$ differs from the common-prior law $\mathbb{P}^0_N$. The joint outcome limits of Theorem L.2 hold with all coefficients evaluated at $\lambda(c)$.

(iii) If $\lambda^* < 1$, then $\lambda_N < 1$ for all sufficiently large N. If $\lambda^* > 1$, one can choose $\lambda_N = 1$ for all sufficiently large N. If $\lambda^* = 1$, each finite equilibrium may be interior or all-informed, and $\lambda_N \to 1$ in either case.

The proof has three steps. First, we construct the fixed-λ equilibria of Theorem L.2 continuously in λ (Lemma L.7). Second, we show that the net gain from acquiring information converges uniformly in λ to its Gaussian limit (Lemma L.8). Third, we select λ_N as a zero of the finite gain and show that it converges to $\lambda(c)$.

L.6.1 A continuous family of finite trading equilibria

The fixed-λ construction can be made continuous in λ, uniformly on intervals bounded away from zero. Restore the dependence on λ in $\mathsf{A}(\lambda)$, K/λ, $\vartheta(\lambda)$, and the other constants of (108). The baseline primitive law $\mathbb{P}_N$ and the statistics (d_N, M_N, R_N) do not depend on λ. Only the candidate price map and the subsequent joint conditioning do.

Lemma L.7 (Uniform continuation in the informed fraction) *Let $\mathcal{L} = [\lambda_-, \lambda_+]$ with $0 < \lambda_- < \lambda_+ \le 1$. For each fixed $J < \infty$, the candidates of Lemma L.3 can be chosen jointly continuously in (p, λ), with continuous price derivatives on $[-J, J] \times \mathcal{L}$, and*

$$\sup_{\lambda \in \mathcal{L}} \|\delta_{N,J}(\,\cdot\,;\lambda)\|_{C^1([-J,J])} \longrightarrow 0, \qquad \delta_{N,J}(p;\lambda) = \bar{g}_{N,J}(p;\lambda) + \vartheta(\lambda)p. \tag{124}$$

At $\lambda = 1$, the schedule specifies the optimal demand of a potential uninformed deviator, although its aggregate weight is zero. A common diagonal choice $J_N \to \infty$ yields finite REEs with schedule $g_N(\,\cdot\,;\lambda)$ and price $P_N(\lambda)$ under laws $\widetilde{\mathbb{P}}_{N,\lambda}$ (Definition 5) for every $\lambda \in \mathcal{L}$, with

$$\sup_{\lambda\in\mathcal{L}} \|\widetilde{\mathbb{P}}_{N,\lambda} - \mathbb{P}^0_N\|_{\mathrm{TV}} \longrightarrow 0. \tag{125}$$

For each N, the candidate price is continuous in λ state by state, and the conditioned law is continuous in λ in total variation.

Proof of Lemma L.7. The proof repeats that of Lemma L.3 and Theorem L.2 with λ added as a coordinate that does not move.

Uniform convergence of the field. Because λ is bounded away from zero, A, K/λ, and their derivatives are bounded on $\mathcal{L}$, and K/λ is bounded away from zero. The envelopes in (106) therefore give convergence of C_N, D_N and their derivatives uniformly on compact sets of (p, y, λ). The constants $\vartheta(\lambda)$ and $\varsigma(\lambda)$ are positive and have positive minima on $\mathcal{L}$.

Invariant graphs. Add the equation $\dot{\lambda} = 0$ to the normalized vector field in the proof of Lemma L.3. With $z = y + \vartheta(\lambda)p$, its limit is

$$\dot{p} = \mathsf{A}(\lambda)rz, \qquad \dot{z} = [1 + \vartheta(\lambda)\mathsf{A}(\lambda)r]z, \qquad \dot{\lambda} = 0.$$

The normal expansion rate is at least one throughout $\mathcal{L}$, including $\lambda = 1$, where $\mathsf{A}(1) = 0$. As in the compact-price proof, extend the p chart to a circle. Extend the λ chart to a second circle, keeping the normal expansion rate strictly positive and retaining $\dot{\lambda} = 0$. This can be done smoothly on a neighborhood of the physical parameter interval; at its upper endpoint the mathematical coefficients extend smoothly past $\lambda = 1$. Use fixed cutoffs, as in the compact-price proof, to extend the finite perturbation by zero outside these charts and outside $|z| < \varepsilon_0$. The uniform convergence step, applied with two derivatives in (p, z, λ), gives C^2 convergence of the modified fields, which is stronger than the C^1 convergence that Lemma L.4 requires. No economic interpretation is assigned to the auxiliary parameter values. The torus $z = 0$ is then a compact normally repelling invariant manifold, with two zero tangential eigenvalues and one uniformly positive normal eigenvalue. Lemma L.4 with $m = 2$ and $\theta = (p, \lambda)$ gives a graph $z = h_N(p, \lambda)$ converging to zero in C^1. In particular, h_N and $\partial_p h_N$ are jointly continuous in (p, λ). Since $\dot{\lambda} = 0$, each parameter slice is invariant. Consequently $g_{N,J}(p; \lambda) = -\vartheta(\lambda)p + h_N(p, \lambda)$ solves (111) at that λ. For large N, its price derivative is uniformly negative. The argument at $p = 0$ in Lemma L.3 gives $g_{N,J}(0; \lambda) = 0$ uniformly, also at $\lambda = 1$, where the strictly positive component coefficient is $\gamma s_{H,R}$. Completing every slice by its endpoint constants at the same prices $\pm(J+1)$ preserves

joint continuity and unique clearing. This proves (124) and exact local optimality, including the uninformed deviation when $\lambda = 1$.

Common price bound. Choose the thresholds in (119) uniformly over $\mathcal{L}$, using $\min_{\lambda\in\mathcal{L}} \vartheta(\lambda) > 0$. For the resulting common J_N, set

$$\mathcal{E}_{N,\lambda} = \{|P_N(\lambda)| \le J_N\}, \qquad \widetilde{\mathbb{P}}_{N,\lambda} = \mathbb{P}_N(\cdot \mid \mathcal{E}_{N,\lambda}).$$

The conditioning proof is unchanged at $\lambda = 1$: clearing is simply $P_N(1) = M_N - KR_N\eta$. The uniform pathwise estimate

$$\sup_{\lambda\in\mathcal{L}} |P_N(\lambda)| \le |M_N| + \left(\sup_{\lambda\in\mathcal{L}} \frac{K}{\lambda}\right) R_N|\eta|$$

implies $\sup_{\lambda\in\mathcal{L}} \mathbb{P}_N(\mathcal{E}^c_{N,\lambda}) \to 0$ by tightness. Conditioning on an event moves a law by the probability of the complement, so $\|\widetilde{\mathbb{P}}_{N,\lambda} - \mathbb{P}_N\|_{\mathrm{TV}} = \mathbb{P}_N(\mathcal{E}^c_{N,\lambda})$ and $\|\mathbb{P}_N - \mathbb{P}^0_N\|_{\mathrm{TV}} = \mathbb{P}^0_N(\mathcal{A}^c_N)$. The triangle inequality then proves (125).

Continuity in λ. For fixed N and a fixed state, the clearing functions depend continuously on λ and are strictly increasing in p, with increments at least those of p itself. Their roots are therefore continuous in λ. At each fixed λ, the candidate price has a continuous distribution because the Gaussian supply shock has a density and the clearing map is strictly increasing. Thus $\mathbb{P}_N(|P_N(\lambda)| = J_N) = 0$. If $\lambda_m \to \lambda$, the indicators of $\mathcal{E}_{N,\lambda_m}$ converge almost surely to that of $\mathcal{E}_{N,\lambda}$. Dominated convergence gives convergence of the event probabilities and of their symmetric-difference probabilities. As every conditioning probability is positive, the densities $\mathbf{1}_{\mathcal{E}_{N,\lambda}}/\mathbb{P}_N(\mathcal{E}_{N,\lambda})$ are continuous in $L^1(\mathbb{P}_N)$. This is the asserted total-variation continuity. □

The outcome limits of Theorem L.2 also hold along any sequence $\lambda_N \to \lambda \in \mathcal{L}$, with all coefficients evaluated at λ. Indeed, (122) now holds with the corresponding parameter-dependent coefficients, and its error vanishes uniformly over $\mathcal{L}$ in probability. Joint convergence of the baseline statistics and continuous mapping then give the limiting price and portfolios evaluated at λ. Equation (125) transfers these limits to $\widetilde{\mathbb{P}}_{N,\lambda_N}$.

L.6.2 Finite acquisition values and their limit

The net gain from acquiring information is continuous in λ and converges uniformly to its Gaussian limit. For a member of the family of Lemma L.7, define the optimized ex-ante

trading values, before paying acquisition costs, by

$$V_{I,N}^{\mathrm{gr}}(\lambda) = E_{\widetilde{\mathbb{P}}_{N,\lambda}}\left[-\exp\left\{-\frac{(M_N - P_N(\lambda))^2}{2R_N}\right\}\right],$$
$$V_{U,N}(\lambda) = E_{\widetilde{\mathbb{P}}_{N,\lambda}}[v_{U,N}(P_N(\lambda);\lambda)]. \tag{126}$$

Both values lie in $[-1,0]$. Under CARA preferences, paying the deterministic cost χ multiplies the informed utility by $e^{\gamma\chi}$. The net acquisition gain is therefore

$$\Delta_N(\lambda) = e^{\gamma\chi}V_{I,N}^{\mathrm{gr}}(\lambda) - V_{U,N}(\lambda).$$

An interior informed fraction requires $\Delta_N(\lambda) = 0$; the all-informed corner requires $\Delta_N(1) \geq 0$. The latter inequality compares acquiring information with the optimal uninformed deviation under the same equilibrium price law.

Lemma L.8 (Continuous acquisition values and uniform convergence) *On every interval $\mathcal{L}$ of Lemma L.7, $V_{I,N}^{\mathrm{gr}}$ and $V_{U,N}$ are continuous, and*

$$\sup_{\lambda\in\mathcal{L}}|V_{I,N}^{\mathrm{gr}}(\lambda) - V_{I,\infty}^{\mathrm{gr}}(\lambda)| + \sup_{\lambda\in\mathcal{L}}|V_{U,N}(\lambda) - V_{U,\infty}(\lambda)| \longrightarrow 0, \tag{127}$$

where

$$S_P(\lambda) = \mathrm{Var}_{\mathbb{P}_\infty}[d^\infty - P^\infty(\lambda)] = B + r + a(\lambda)^2 S_W(\lambda) - 2a(\lambda)B,$$
$$V_{I,\infty}^{\mathrm{gr}}(\lambda) = -\sqrt{\frac{r}{S_P(\lambda)}}, \qquad V_{U,\infty}(\lambda) = -\sqrt{\frac{U(\lambda)}{S_P(\lambda)}}. \tag{128}$$

In particular, $S_P(\lambda) \geq U(\lambda) \geq r > 0$ and

$$\Delta_N \longrightarrow \Delta_\infty \quad \textit{uniformly on } \mathcal{L}, \qquad \Delta_\infty(\lambda) = \frac{\sqrt{U(\lambda)} - e^{\gamma\chi}\sqrt{r}}{\sqrt{S_P(\lambda)}}. \tag{129}$$

Proof of Lemma L.8. Continuity follows from dominated convergence, uniform convergence from sequential convergence and compactness of $\mathcal{L}$, and the limiting values from a Gaussian computation.

Continuity. For fixed N, the conditional loss (112) is continuous in (p,λ,x) on the retained compact price and position sets. Its denominator is positive, and the moment bounds of Lemma L.1 justify this continuity and differentiation under the integral. The

family g_N and its price derivative are continuous there, so $v_{U,N}(p;\lambda) = -\mathcal{L}_{N,g_N}(g_N(p;\lambda);p)$ is continuous on the retained support. The informed conditional value is continuous in the price as well. Write the expectations in (126) under $\mathbb{P}_N$ using the indicator of $\mathcal{E}_{N,\lambda}$ and its probability. Lemma L.7, the zero-probability boundary of the event, and the bound $[-1,0]$ on both conditional optimized values give continuity by dominated convergence. Only values on the retained support are used in this bound.

Uniform convergence. The compact posterior estimates are uniform in λ, by the same moment bounds and the uniform C^1 convergence in (124). The denominator of (112) converges uniformly on compact sets of (p,λ) to the $\mathcal{N}(0, a(\lambda)^2 S_W(\lambda))$ density of $P^\infty(\lambda)$. This density is bounded below on $[-J,J]\times\mathcal{L}$ for every J, because $a(\lambda)^2 S_W(\lambda)$ is continuous and positive on $\mathcal{L}$. The estimates therefore imply uniform convergence of $v_{U,N}(p;\lambda)$ to its Gaussian counterpart on every fixed compact set of (p,λ). For any sequence $\lambda_N \to \lambda \in \mathcal{L}$, the joint outcome convergence along λ_N, uniform price tightness, and the bound $[-1,0]$ imply convergence of both ex-ante values to their Gaussian values at λ. Compactness of $\mathcal{L}$ upgrades this sequential statement to (127): otherwise a sequence of parameters with an error bounded away from zero would have a convergent subsequence, contradicting that statement.

Gaussian values. Let m_j be the conditional payoff mean and s_j the conditional variance, with $s_I = r$ and $s_U = U(\lambda)$. The optimized conditional utility is $-\exp[-(m_j - P^\infty)^2/(2s_j)]$. The Gaussian variable $m_j - P^\infty$ is centered and has variance $S_P(\lambda) - s_j$ by total variance. Its Gaussian integral equals $-\sqrt{s_j/S_P(\lambda)}$, proving (128). Taking the difference gives (129). □

L.6.3 Existence and convergence of the acquisition fraction

The limiting gain Δ_∞ crosses zero strictly, because the limiting uninformed residual is strictly decreasing in λ when $B > 0$. Writing $\nu_0 = \gamma\sigma_\eta r$,

$$U(\lambda) = r + \frac{B\nu_0^2}{B\lambda^2 + \nu_0^2}, \qquad U'(\lambda) = -\frac{2B^2\nu_0^2\lambda}{(B\lambda^2+\nu_0^2)^2} < 0. \tag{130}$$

Thus an interior zero of Δ_∞ is a strict crossing, not an additional regularity assumption. At such a zero λ^*,

$$\Delta'_\infty(\lambda^*) = \frac{U'(\lambda^*)}{2\sqrt{U(\lambda^*)S_P(\lambda^*)}} < 0.$$

Proof of Theorem L.6. We compute the limiting fraction, select finite zeros of the gain by the intermediate value theorem, show that they converge, and verify that the selected economies are equilibria.

Limiting fraction. As $\lambda \downarrow 0$, $U(\lambda) \to B + r = r(1+v) > e^{2\gamma\chi} r$. Choose $\lambda_- > 0$ small enough that $\Delta_\infty(\lambda_-) > 0$, and construct the continuous family on $\mathcal{L} = [\lambda_-, 1]$. The limiting indifference equation is

$$U(\lambda) = e^{2\gamma\chi} r.$$

Substituting (130) and solving gives (123). Because $\sigma_d^2 = B + r = r(1+v)$, this formula agrees with Theorem 2. Here $V_{I,\infty}^{\mathrm{gr}}$ is gross of the acquisition cost, whereas V_I in (4) includes the factor $e^{\gamma\chi}$, so the indifference condition $e^{\gamma\chi} V_{I,\infty}^{\mathrm{gr}} = V_{U,\infty}$ is $V_I = V_U$.

Selection. By Lemma L.8, $\Delta_N(\lambda_-) > 0$ for all sufficiently large N. If $\Delta_N(1) \geq 0$, set $\lambda_N = 1$. If $\Delta_N(1) < 0$, continuity and the intermediate value theorem give a zero $\lambda_N \in (\lambda_-, 1)$. This construction enforces the exact finite acquisition condition, without requiring monotonicity or uniqueness of the finite gap.

Convergence of the fraction. When $\lambda^* < 1$, the limiting gap is positive below λ^* and negative above it. Uniform convergence gives $\Delta_N(1) < 0$ eventually, and any selected finite zero must converge to λ^*. When $\lambda^* > 1$, $\Delta_\infty(1) > 0$, so $\lambda_N = 1$ eventually. At the boundary $\lambda^* = 1$, $\Delta_\infty(\lambda) > 0$ for every $\lambda < 1$. For each $\varepsilon > 0$, its minimum on $[\lambda_-, 1-\varepsilon]$ is strictly positive whenever that interval is nonempty. Uniform convergence excludes finite zeros on that interval for large N. The selected fractions therefore converge to one, including when $\Delta_N(1)$ changes sign along the sequence. This proves the fraction assertions.

Equilibrium. Fix the selected economy's primitive law $\widetilde{\mathbb{P}}_{N,\lambda_N}$ and its price and portfolio maps. Within this economy, a single investor's decision to acquire information leaves $\widetilde{\mathbb{P}}_{N,\lambda_N}$, λ_N, and the price map unchanged. Its informed and uninformed opportunities are exactly those whose optimized values appear in (126). For an interior λ_N, allocate a fraction λ_N of the atomless investors to acquisition: all investors are indifferent and neither type has a profitable deviation. At $\lambda_N = 1$, the acquisition inequality makes acquiring optimal for everyone, including comparison with the hypothetical uninformed portfolio. By the parameterized version of Lemma L.5, the price and portfolios form a finite REE at λ_N under $\widetilde{\mathbb{P}}_{N,\lambda_N}$. Hence these are equilibria of the complete binary acquisition-and-trading problem.

Outcome limits. Equation (125) applies at λ_N. The moving-parameter convergence following Lemma L.7 and $\lambda_N \to \lambda(c) > 0$ give the stated joint outcome limits. □

Scope. Once an implementing law is selected, it remains fixed before acquisition and when an investor deviates. Under this appendix's assumptions, the construction covers every positive-mass overall equilibrium of Theorem 2, including the weak all-informed boundary. The zero-informed case is excluded. Endogenous scales require a separate deviation analysis because conditioning need not preserve payoff posteriors at intermediate information sets.