Compute, Complexity, and the Scaling Laws of Return Predictability

Allan Timmermann
UC San Diego

Luka Vulicevic
*UC San Diego**

March 27, 2026

Abstract

Investors' computing power ("compute") governs how much signal they can extract from increasingly high-dimensional datasets. Forecast performance follows scaling laws, defined as stable power-law relationships between accuracy and training compute, that pin down both an irreducible bound on return predictability and the rate at which additional compute improves accuracy. In firm-level cross-sectional return prediction, scaling laws explain over 80% of portfolio performance variation across models. Treating compute as a primitive yields economic implications: scaling laws quantify predictability limits and the value of new data, distinguish problems with strong versus weak returns to scale, and deliver a market-efficiency metric as the elasticity of risk-adjusted returns with respect to compute. Measured this way, computational superiority earns rents: a 25% increase in compute would have raised an investor's Sharpe ratio by about 14% over the last 30 years, indicating substantial returns to computational sophistication.

Keywords: Machine learning, Asset Pricing, Model Complexity, Return Predictability, Market Efficiency
JEL Classification Numbers: C3, C58, C61, G11, G12, G14

*Contact: atimmermann@ucsd.edu & lvulicevic@ucsd.edu. We thank William Mullins, Joseph Engelberg, Mo Pourmohammadi, and seminar participants at NYU and Virginia Tech for valuable comments. The Rady Research Fund and the Brandes Center provided critical funding for this project.

1. Introduction

Processing power ("compute") plays a key role in determining the magnitude of the gains in predictive accuracy investors can achieve from their information sets and, in equilibrium, how much of that information is reflected in prices. Leading quantitative firms increasingly emphasize compute and data as strategic capital. Renaissance Technologies, for example, advertises a "research database that grows by more than *40 terabytes a day*" and an infrastructure of "*52,000 computer cores*" with "*redundant computational facilities*" on their website. Hudson River Trading makes the same point differently, emphasizing a deliberately "**high compute-per-researcher ratio**" as core to its mission.[1]

With vast amounts of raw information widely and often publicly available, competitive advantage shifts to the rate and effectiveness with which investors can process high-dimensional signals, making compute a natural primitive. In this paper, we study the limits of return predictability as a function of compute and connect these limits to recent advances in forecasting-intensive domains—such as computer vision, meteorology, autonomous driving, and natural language processing—where predictive accuracy improves systematically with model scale and complexity (Kaplan et al. (2020), Henighan et al. (2020)).

A leading explanation for these improvements in predictive performance is the "double descent" phenomenon, in which sufficiently large models continue to improve beyond the classical bias–variance turning point (Belkin et al. (2019)). The insight that increasing model scale raises out-of-sample forecasting performance has reshaped modern machine learning and is often summarized by *scaling laws*: stable, approximately power-law relationships between forecast error and computational scale that characterize both the rate at which forecasting performance improves with compute and its asymptotic limit. Crucially, while any single model provides *an* estimate of predictability, performance patterns across a family

[1] A sort of arms race can be seen in Table A1, which has more quotes and sources from industry leading hedge funds. They openly advertise their computational resources to attract researchers and often describe their compute as an 'edge'.

of models discipline inference about the underlying data-generating process.

Figure 1: Modern Machine Learning

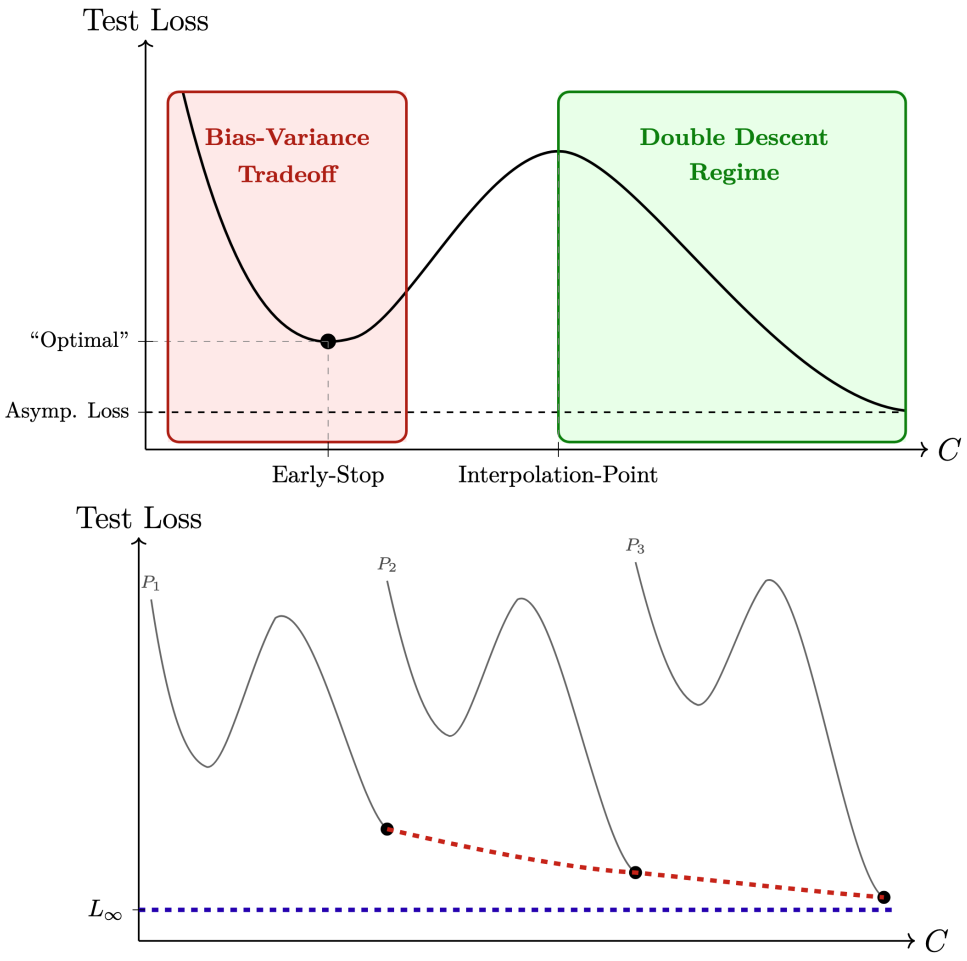


Figure 1 provides a stylized illustration of the double descent phenomenon and its relation to scaling laws, plotting test loss against model compute C.[2] The top panel depicts the classical U-shaped bias–variance tradeoff, where test loss initially falls as models grow before rising due to overfitting, followed by the double-descent regime in which test loss decreases again beyond the interpolation threshold and approaches an asymptotic lower bound. Training models with increasing parameter count ($P_1 < P_2 < P_3$) traces out their asymp-

[2]As Nakkiran et al. (2021) show, double descent arises in a wide variety of models and model primitives including parameters, data, and training compute. In this paper we focus on compute measured as the number of floating point operations (FLOPs) required to train a model. A floating point operation is a basic arithmetic operation on real-valued numbers; for example, $1.23 + 4.56 = 5.79$ constitutes one such operation.

totic losses once sufficient compute pushes each model into double descent. The scaling law (bottom panel) approaches its own asymptote L_∞—the irreducible loss—representing the theoretical lower bound no model can surpass regardless of size or compute, and therefore the limit to predictability for a given forecast problem.[3]

These developments stand in contrast to prevailing views in economics and finance. Conventional econometric practice cautions against complex models on grounds of overfitting and weak out-of-sample performance, favoring low-dimensional specifications and penalizing complexity through information criteria or regularization such as LASSO and ridge regression. This classical paradigm, grounded in a U-shaped bias–variance tradeoff, has recently been challenged by Kelly et al. (2024b), who show that more complex models can perform better in forecasting risk premia.

A key distinction between their paper and ours is that compute and model complexity, as measured by parameter count, are fundamentally different objects. Parameters are the conduit through which compute is expended, not a resource in their own right: one could set up a trillion-parameter model, but without sufficient compute to push it into double descent, the parameter count is economically meaningless. Compute, on the other hand, is a capital resource—the binding constraint on what a model can learn and thus on what insights an investor can extract from a given dataset.

Treating compute as a primitive yields several new insights for asset pricing. Scaling laws introduce a new set of econometric tools for quantifying the value of data, the marginal return to additional compute, and the performance of investors facing a compute budget constraint. Estimating these laws also uncovers new properties of large models: they exhibit substantially lower and more stable training risk, and the rate of decay along the power-law curve varies with the choice of input features.

More formally, scaling laws posit a power relation between test loss L as a function of

[3]At the interpolation threshold, the model has just enough parameters to perfectly fit the training data, forcing a typically jagged and fragile solution. Moving into the overparameterized regime, many interpolating solutions become feasible, and optimization procedures such as stochastic gradient descent implicitly select smoother, more regularized ones.

compute c of the form

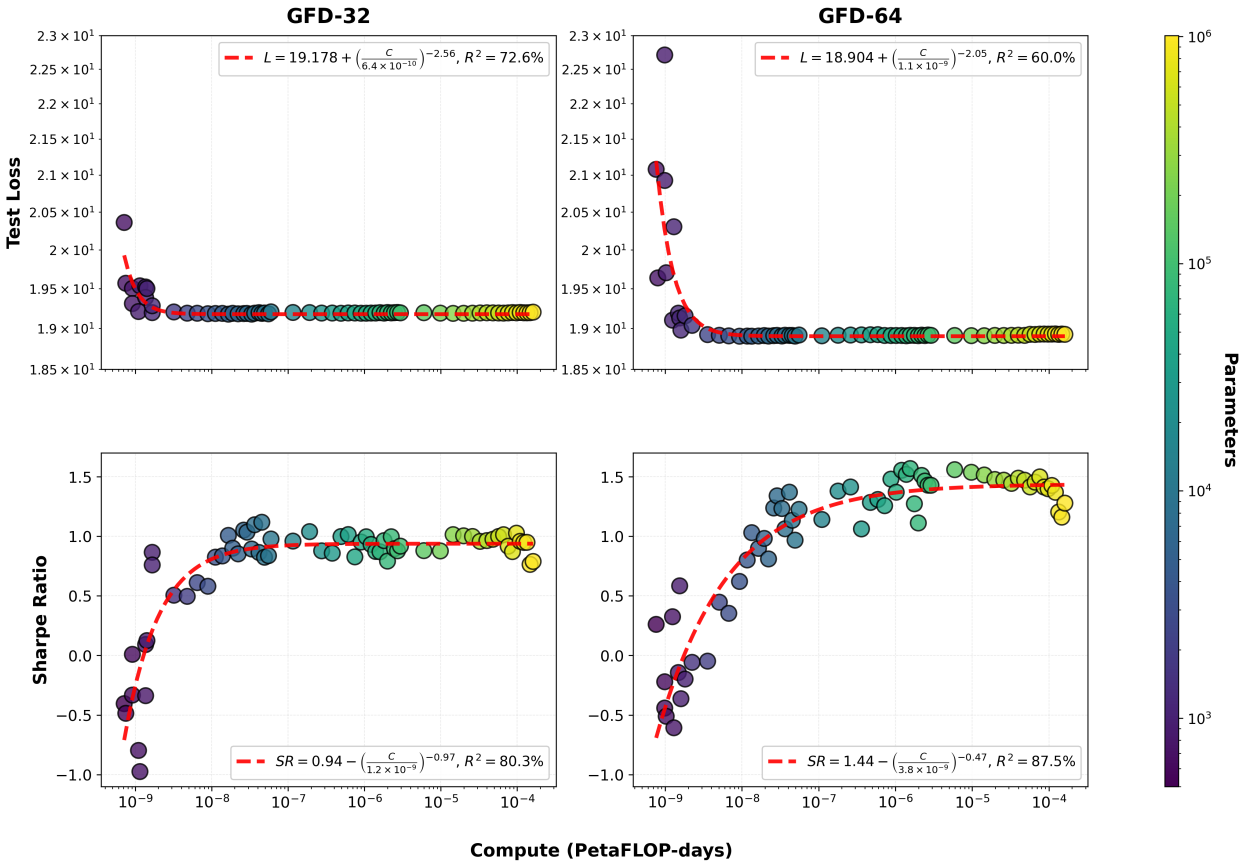
$$L(c) = L_\infty + \left(\frac{c_0}{c}\right)^\alpha, \tag{1}$$

where $L_\infty > 0$ is the irreducible loss, $c_0 > 0$ is a scale parameter, and $\alpha > 0$ is the scaling exponent. The power-law term $(c_0/c)^\alpha$ captures reducible error that decreases with compute; as $c \to \infty$, loss approaches L_∞ from above. The exponent α governs the rate of improvement: larger values indicate faster gains with additional compute, while smaller values imply slower convergence. The asymptote L_∞ represents the fundamental uncertainty that no model can eliminate given a particular dataset, and therefore bounds the limits to predictability.

Although scaling laws such as Equation (1) have been established across a variety of fields, their existence has not been demonstrated for financial return prediction. To examine whether scaling laws apply to financial data, we estimate neural networks ranging from 100 parameters to more than a million and exploit the double-descent regime to identify each model's asymptotic loss. We then estimate scaling laws relating return predictability to computational scale, showing that these relationships emerge robustly across training, validation, and test samples, as well as across a range of trading strategies.

Figure 2 illustrates our empirical results. Using the Jensen et al. (2021) characteristics data, the figure plots root mean squared error (top panels) and sample Sharpe ratio (bottom panels) over a thirty-year test sample (1994–2024), with the left panels using 32 firm characteristics and the right panels using 64. A clear scaling law emerges in each case: small models at low compute, shown on the left of each panel, produce high and volatile test loss alongside low—even negative—Sharpe ratios, while performance converges to a well-defined asymptote as compute increases. This limit—a Sharpe ratio of 1.44 in the bottom right panel—bounds the economically valuable information extractable from the underlying dataset.

Figure 2: Firm-Level Return Scaling Laws (1994 - 2024)



Note. The top panel plots out-of-sample forecast loss (RMSE) and the bottom panel plots the out-of-sample Sharpe ratio of a trading strategy constructed from the same forecasts, both as functions of cumulative training compute (in PetaFLOP-days, log scale). Each marker corresponds to a feedforward neural network trained to predict next-month firm-level excess returns using the Jensen et al. (2021) characteristics data, evaluated over a test sample spanning 1994–2024. The left panel depicts such models trained with a feature set of 32 characteristics while the right adds an additional 32 other features for a total of 64. The dashed red curve in each panel is the fitted scaling law estimated across the family of trained networks; the plotted equation reports the fitted asymptote and scaling exponent, and the associated R^2. In the bottom panel, the Sharpe ratio is computed from a zero-investment long–short portfolio formed each month by sorting firms on predicted returns using a median breakpoint and taking equal-weight positions long (above-median) and short (below-median); Sharpe ratios are computed from monthly portfolio returns and annualized.

The power laws in Figure 2 point to three underappreciated benefits of large models in finance. First, large models are less risky to deploy. Under double descent, sufficiently over-parameterized models are implicitly regularized and therefore produce more stable training outcomes than smaller models, which are prone to converging to poor local minima (Belkin et al. (2019)). Our estimated models exhibit pronounced heteroskedasticity at low compute:

some small models achieve performance comparable to models orders of magnitude larger, while others converge to unfavorable minima and perform markedly worse.

Second, low-compute models are more likely to exhibit test-sample performance driven by chance rather than genuine predictive content. At low compute, performance is less stable and less representative: models trained with slightly different compute can exhibit meaningfully different results, and outcomes are more sensitive to hyperparameter choices, complicating replication and interpretation. Both concerns diminish as compute increases and performance stabilizes near the power-law asymptote. Much of the machine learning literature in finance—largely framed within the bias–variance paradigm—has operated in this low-compute region, potentially understating both the reliability and economic value of these methods.

Third, comparing scaling laws across different input feature sets allows us to assess the marginal value of additional data. Doubling the number of input features from 32 to 64 yields, in the limit, a 53% higher Sharpe ratio, a difference our hypothesis tests confirm is statistically significant at the 1% level. These results offer a potential explanation for why institutional investors have increasingly sought both larger datasets and the computational infrastructure required to process them, as our findings suggest both translate directly into superior portfolio performance.

Scaling laws carry two implications for asset pricing and market efficiency. The first is that computing power is central to how information gets embedded into prices: firms with greater computational resources more effectively extract and forecast from public and private information, impounding more of it into prices. The second is that scaling laws explain why leading hedge funds are investing heavily in compute. Just as execution speed—enabled by fiber, microwave, and laser technology—defined high-frequency trading around 2010, processing power now drives investment firms' capital expenditure. The concavity of the scaling relationship prevents a winner-takes-all outcome in which a single firm dominates by virtue of superior resources alone, but our results suggest the arms race will push firms

toward unique datasets that can only be fully exploited at scale.

An additional advantage of the scaling-laws approach is its application to testing the informativeness of different datasets. We develop a formal statistical test for the economic value of augmenting an existing dataset with new data sources, asking whether doing so shifts the scaling law. This reflects a key point: all power-law estimates are conditional on a given information set. If the information set changes in a substantively informative way, the power-law relationship should change as well, and comparing estimates across datasets while holding the forecasting sample fixed identifies the economic and statistical value of any additional information.[4]

The remainder of the paper is organized as follows. Section 2 discusses related literature. Section 3 develops the econometric framework and estimation of scaling laws. Section 4 describes the data and empirical scaling-law results. Section 5 presents a simple asset-pricing model deriving economic implications of heterogeneity in compute. Section 6 concludes. Appendix A details neural network architecture and training, Appendix B contains additional empirical results, and Appendix C provides proofs for the asset-pricing model.

2. Related Literature

Modern financial markets are increasingly characterized by heterogeneity in computational capacity. While classic models such as Grossman and Stiglitz (1980), Hellwig (1980), Kyle (1985), Admati (1985), Glosten and Milgrom (1985) emphasize differentials in information sets, today's markets increasingly feature algorithmic traders, high-frequency firms, retail traders, and institutional investors who all differ drastically in their ability to process their information sets.[5] For instance, sophisticated algorithms powered by machine

[4]Both the irreducible loss L_∞ and the scaling exponent α may change when augmenting a dataset, as new data may make estimation of the conditional mean function more complex but also more accurate.

[5]See also the extensive literature on strategic information acquisition (Van Nieuwerburgh and Veldkamp (2009), Vives (2010)), information diffusion and learning (Veldkamp (2006), Banerjee et al. (2009)), rational inattention (Sims (2003), Kacperczyk et al. (2016), Mondria (2010)), and the more recent focus on machine learning and alternative data (Dugast and Foucault (2018), Farboodi and Veldkamp (2020, 2021)).

learning can extract signals from data that human traders or simpler algorithms cannot, even when both entities observe identical information sets such as earnings announcements. In parallel, access to high-dimensional public information about individual firms and the macro-economy has expanded through EDGAR, CRSP, Bloomberg, and the St. Louis Fed's FRED. For investors with access to these sources, the central challenge is not data availability, rather, it is how to process a high-dimensional vector of public signals. Heterogeneity in information-processing ability is thus likely to be equally important as heterogeneity in *private* information.

Our work connects to different strands of the finance literature on (i) data-snooping and multiple testing in financial prediction; (ii) economically motivated bounds on attainable performance; and (iii) the virtue of complexity. Foundational warnings about overfitting and specification search include Lo and MacKinlay (1990), Sullivan et al. (1999), White (2000), Hansen (2005), and the factor-zoo view emphasizes that many "discoveries" may not be robust to distortions to inference resulting from multiple comparisons (Harvey et al. (2016)). In parallel, asset pricing theory provides limits on risk-adjusted returns and predictability, including the Hansen–Jagannathan bound (Hansen and Jagannathan (1991)), good-deal bounds on Sharpe ratios (Cochrane and Saa-Requejo (2000)), and explicit upper bounds on predictive R^2 (Zhou (2010), Huang and Zhou (2017), Potì (2018)). Our scaling-law approach complements these by making the role of compute explicit: conditional on a particular information set, we estimate how forecasting performance improves with compute and where it ultimately plateaus.

Our work is also related to the growing empirical literature using neural networks and other high-capacity learners in finance and macroeconomics. Early work argues deep architectures are well-suited for capturing nonlinear structure in portfolios and prediction (Heaton et al. (2017)). In asset pricing, Gu et al. (2020) document out-of-sample gains from flexible ML methods (including neural networks), while Chen et al. (2024) incorporate no-arbitrage restrictions directly into deep learning. Deep learning has also been applied to market mi-

crostructure and limit order books (Sirignano and Cont (2021)), and to macro forecasting when nonlinearities are important (Goulet Coulombe et al. (2022)). Finally, our paper speaks directly to the "virtue of complexity" debate (Kelly et al. (2024b), Nagel (2025)). In particular, our scaling laws are broadly consistent with recent work in Kelly et al. (2024b)[6] which has challenged the older view of parsimony, arguing that larger models tend to outperform smaller ones using Random Fourier Features ('RFF') in a variety of empirical and theoretical settings. However, our analysis goes well beyond the existing debate: rather than asking only whether bigger models help, we quantify *how* gains scale with compute (the exponent) and *what* remains learnable from the data (the asymptote), across a broad set of financial and economic forecasting problems.

3. Estimation of Scaling Laws

Consider a general forecasting problem which takes the form of a mapping g from a possibly high-dimensional vector of predictors x_t observed at time t to some dependent (target) variable r_{t+h} (such as a financial market return) observed at a future point in time $t+h$, where $h \geq 1$ denotes the forecast horizon. Assuming a squared error loss function, the optimal forecast takes the form of the conditional mean of r_{t+h} given the information x_t, $g(x_t) := E[r_{t+h}|x_t]$, and we can write

$$r_{t+h} = g(x_t) + \varepsilon_{t+h}, \qquad \varepsilon_{t+h} \sim (0, \sigma^2).$$

In practice, the functional form of g is unknown and could be highly complex, especially in situations in which x is high-dimensional. Theoretical results by Hornik et al. (1989) suggest that neural nets have universal approximating properties which means that the distance between the true conditional mean function and the best approximating neural net can be made arbitrarily small. However, the best approximating model may be highly complex and

[6] As well as their follow-on work in Kelly et al. (2022, 2024a), Didisheim et al. (2024).

require considerable compute to get close to the true conditional mean.

To explore the relationship between compute and return predictability, we organize the remainder of this section as follows. The first three subsections demonstrate how the scaling laws are modeled and estimated. Specifically, we first outline the scaling law for conventional measures of forecast loss such as root mean squared error (RMSE) and define equivalent economic measures such as the Sharpe ratio scaling law. With these objects defined, we cover how we estimate the power law curves. First, we explain how to estimate market efficiency through a scaling law. Second, we discuss how to compare scaling laws across different forecast problems. Finally, we introduce a statistical test for whether the information provided by new input variables alters the parameters of the scaling law in a meaningful way compared to a null scaling law.

3.1 Loss and Sharpe Ratio scaling law

Following the neural scaling law framework in Kaplan et al. (2020), Henighan et al. (2020), we hypothesize that the out-of-sample prediction error follows a power law relationship with compute. Specifically, we estimate the parameters (L_∞, c_0, α) by fitting Equation (1) to the empirical relationship between compute and test (MSE) loss using nonlinear least squares.[7]

In addition to statistical metrics such as MSE, investors ultimately care about the economic value of forecasts when used in trading strategies. A widely used summary of this value is the Sharpe ratio of the resulting portfolio. Theoretical results on high-dimensional return prediction emphasize that timing strategies can deliver strictly positive Sharpe ratios even when the associated predictive R^2 is negative, implying that R^2, and therefore also the underlying mean squared error (MSE), is an incomplete proxy for economic value. Motivated by this insight, we study how the Sharpe ratio scales with compute c across different portfolio constructions, and we refer to the resulting relationships as *Sharpe ratio scaling laws*.

[7]The same power law plus constant functional form is hypothesized for training loss $L_{\text{train}}(c)$ and validation loss $L_{\text{val}}(c)$, though the specific parameters may differ across these three loss measures. The relationship between these curves provides insight into the bias-variance tradeoff and the degree of overfitting as compute increases.

We examine simple panel-based long–short portfolios that map cross-sectional differences in forecasts into portfolios and compute the resulting Sharpe ratios as a function of compute c.

We hypothesize that the Sharpe ratio follows a scaling law but now is an increasing function of compute c:

$$\mathrm{SR}(c) = \mathrm{SR}_{\max} - \left(\frac{c_0'}{c}\right)^{\beta}, \tag{2}$$

where $\mathrm{SR}_{\max}$ represents the maximum attainable Sharpe ratio (an upper bound), $c_0' > 0$ is a scale parameter, and $\beta > 0$ is the scaling exponent. As compute increases, $\mathrm{SR}(c)$ approaches $\mathrm{SR}_{\max}$ from below, with the gap decreasing according to the power law. The parameter $\mathrm{SR}_{\max}$ reflects fundamental limits on risk-adjusted returns in the market.

We estimate Sharpe ratios for zero-investment portfolios based on the model return predictions. At each time t, we form equal-weighted portfolios based on the decile breakpoints of the forecasted next month returns across all stocks. The return for decile group D at time $t+1$ is computed as

$$r_{D,t+1} = \tfrac{1}{|D_t|} \sum_{i \in D_t} r_{i,t+1},$$

where the sum is over the D^{th} decile of stocks sorted by the forecasts $y_{i,t+1}$ made at time t. We then form portfolios by going long on a certain decile and short on another. Over the test period $\mathcal{S}$ which spans T time periods, the Sharpe ratio is calculated as

$$\mathrm{SR}_{\mathcal{S}} = \frac{\bar{r}_p}{\sigma(r_p)} = \frac{\frac{1}{T}\sum_{t \in \mathcal{S}} r_{p,t+1}}{\sqrt{\frac{1}{T-1}\sum_{t \in \mathcal{S}} (r_{p,t+1} - \bar{r}_p)^2}},$$

where $\bar{r}_p$ is the mean portfolio return and $\sigma(r_p)$ is its standard deviation. Most Sharpe ratios mentioned in the paper will refer to a strategy which we denote as LS_{50} that goes long the top half (deciles 6–10) and shorts the bottom half (1–5) unless otherwise specified. All Sharpe ratios reported in the paper are annualized.

3.2 Estimating scaling laws

We estimate the parameters of the scaling laws for both the MSE loss function and the Sharpe ratio using a separable non-linear least squares approach. Let $\{(c_i, y_i)\}_{i=1}^{M}$ denote the set of observed pairs of compute and the corresponding performance metric (either MSE loss or Sharpe ratio) for M trained models. We fit a generalized power law of the form

$$y_i = \theta_0 + \theta_1 c_i^{\gamma} + \epsilon_i, \quad i = 1, ..., M,$$

where ϵ_i represents the measurement error. This functional form encompasses both the loss scaling law in Equation (1), where $\gamma = -\alpha$, and the Sharpe ratio scaling law in Equation (2), where $\gamma = -\beta$. Direct estimation of the parameters $(\theta_0, \theta_1, \gamma)$ via standard gradient-based optimization tends to be numerically unstable due to the high correlation between the asymptotic term θ_0 and the power term. To address this, we employ a variable projection method which exploits the partial linearity of the model. For a fixed exponent γ, the optimal linear parameters $\hat{\theta}_0(\gamma)$ and $\hat{\theta}_1(\gamma)$ have a closed-form solution obtained via weighted linear least squares. The estimation problem thus reduces to a one-dimensional scalar optimization over the exponent γ to minimize the weighted residual sum of squares

$$\min_{\gamma} \sum_{i=1}^{M} w_i \left(y_i - \hat{\theta}_0(\gamma) - \hat{\theta}_1(\gamma) c_i^{\gamma} \right)^2.$$

We employ a log-space weighting scheme with weights defined as

$$w_i = \frac{\log(c_i) - \log(c_{\min}) + 1}{\sum_{j=1}^{M} \left(\log(c_j) - \log(c_{\min}) + 1 \right)},$$

where $c_{\min} = \min_j c_j$. This weighting scheme is chosen to prioritize the fit of the scaling curve for higher compute levels which are more indicative of the asymptotic behavior and the irreducible error L_∞ (or the maximum Sharpe ratio $\mathrm{SR}_{\max}$), while avoiding the extreme weight imbalance that arises from linear weighting when compute spans several orders of magnitude.

Constraints are imposed during the optimization to ensure the exponent corresponds to a decaying power law, consistent with the theoretical framework where performance converges to an asymptote as compute increases. Once the optimal exponent $\hat{\gamma}$ is found, the scale parameters c_0 or c_0' are recovered from the linear coefficient $\hat{\theta}_1$.

3.3 Market efficiency

The Sharpe ratio scaling law in Equation (2) provides a natural measure of market efficiency. It can be interpreted as the elasticity of risk-adjusted returns with respect to compute. Because the scaling relationship is concave, marginal returns to additional sophistication decline as aggregate computational capacity grows. Market efficiency, in this framework, corresponds to the region of the scaling curve where this elasticity is near zero.

To formalize this idea, consider an investor whose computational capacity exceeds that of the marginal investor at time t by a factor of $(1+\delta)$, where $\delta > 0$ measures relative sophistication. Evaluating the scaling law at $(1+\delta)c_t$ and c_t yields the absolute outperformance

$$\Delta\mathrm{SR}(\delta; c_t) = \mathrm{SR}((1+\delta)c_t) - \mathrm{SR}(c_t) = \left(\frac{c_0'}{c_t}\right)^{\beta} \left[1 - (1+\delta)^{-\beta}\right],$$

and the corresponding relative outperformance

$$\frac{\Delta\mathrm{SR}(\delta; c_t)}{\mathrm{SR}(c_t)} = \frac{\left(\frac{c_0'}{c_t}\right)^{\beta} \left[1 - (1+\delta)^{-\beta}\right]}{\mathrm{SR}_{\max} - \left(\frac{c_0'}{c_t}\right)^{\beta}}.$$

A first-order Taylor expansion for small δ makes the elasticity transparent:

$$\Delta\mathrm{SR}(\delta; c_t) \approx \beta\delta \left(\frac{c_0'}{c_t}\right)^{\beta}.$$

The gain from a δ-fold increase in compute scales linearly in δ but decays as a power law

in aggregate compute c_t. Expressing this as a percentage improvement,

$$\%\Delta\mathrm{SR}(\delta; c_t) \approx \frac{\beta\delta \left(\frac{c_0'}{c_t}\right)^{\beta}}{\mathrm{SR}_{\max} - \left(\frac{c_0'}{c_t}\right)^{\beta}} \times 100. \tag{3}$$

As $c_t \to \infty$, both ΔSR and $\%\Delta$SR converge to zero for any fixed δ. The market thus becomes increasingly efficient—not because predictability vanishes, but because the marginal investor operates closer to the asymptotic regime of the scaling law, leaving little room for proportionally more sophisticated investors to gain an edge.

To locate where the market currently sits on this curve, we introduce market-implied compute. Given the estimated structural parameters $(\mathrm{SR}_{\max}, c_0', \beta)$ and an empirical estimate of the aggregate market portfolio's Sharpe ratio $\mathrm{SR}_{\mathrm{mkt}}$, we invert Equation (2) to obtain

$$c_t^{\mathrm{implied}} = c_0' \left(\mathrm{SR}_{\max} - \mathrm{SR}_{\mathrm{mkt}}\right)^{-1/\beta}.$$

This quantity serves as a revealed-preference proxy for the sophistication of the average investor, conditional on the information set used to estimate $\mathrm{SR}_{\max}$. When c_t^{implied} is large, the market occupies the flat portion of the scaling curve: even substantial increases in δ yield negligible improvements in risk-adjusted returns. Substituting c_t^{implied} into Equation (3) then quantifies the degree of efficiency as the slope of the scaling curve at the market's operating point.

3.4 Hypothesis Testing With Scaling Laws

Because the scaling law parameters $(\theta_0, \theta_1, \gamma)$ are estimated via separable nonlinear least squares, their sampling distributions are unavailable in closed form. We base inference on a pairs bootstrap: in each of B replications, resample the K observed pairs $\{(c_k, y_k)\}_{k=1}^{K}$ with replacement, refit the variable projection estimator, and collect the parameter draws $\hat{\boldsymbol{\theta}}^{*b}$. Let $B_s \leq B$ denote the number of replications in which the optimizer converges. Standard

errors, percentile confidence intervals, and p-values are computed from these B_s draws in the usual way.

Test I: Comparing Two Scaling Laws

Let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^+$ denote the parameters of scaling laws estimated from a baseline and augmented information set, with K and K^+ models respectively. We test $H_0^{(1)} : L_\infty = L_\infty^+$ and $H_0^{(2)} : \alpha = \alpha^+$ marginally. In each replication, the two datasets are resampled independently and both scaling laws are refit, yielding bootstrap differences $\Delta_\infty^{*b} = \hat{L}_\infty^{+*b} - \hat{L}_\infty^{*b}$ and $\Delta_\alpha^{*b} = \hat{\alpha}^{+*b} - \hat{\alpha}^{*b}$. The two-sided p-value for $H_0^{(1)}$ is

$$p = 2\min\left(\frac{1}{B_s}\sum_{b=1}^{B_s}\mathbf{1}\{\Delta_\infty^{*b} \le 0\},\ \frac{1}{B_s}\sum_{b=1}^{B_s}\mathbf{1}\{\Delta_\infty^{*b} \ge 0\}\right), \tag{4}$$

and analogously for $H_0^{(2)}$. The same procedure applies to the Sharpe ratio scaling law, replacing L_∞ with $\mathrm{SR}_{\max}$ and α with β.

Test II: Level Test for the Asymptote

To test whether the estimated asymptote equals a benchmark h_0—for instance, the Sharpe ratio of a linear model—we test $H_0 : \mathrm{SR}_{\max} = h_0$ using the single-sample bootstrap distribution. Define $\delta^{*b} = \widehat{\mathrm{SR}}_{\max}^{*b} - h_0$; the two-sided p-value follows Equation (4) with δ^{*b} replacing Δ_∞^{*b}.

Test III: One-Sided Exceedance Test

When the alternative is directional, $H_0 : \mathrm{SR}_{\max} \le h_0$ versus $H_1 : \mathrm{SR}_{\max} > h_0$, the p-value is

$$p = \frac{1}{B_s}\sum_{b=1}^{B_s}\mathbf{1}\left\{\widehat{\mathrm{SR}}_{\max}^{*b} \le h_0\right\}.$$

This test evaluates whether the nonlinear scaling law delivers risk-adjusted returns that exceed what simpler specifications achieve on the same data.

4. Empirical results

This section provides details of the data that we use in our empirical analysis, followed by our main findings from applying scaling laws to this data, a detailed look into trading strategy scaling laws and, finally, an examination of market efficiency.

4.1 Data

We estimate scaling laws using the firm characteristics dataset of Jensen et al. (2021) (GFD), retaining all US exchange-traded securities after applying a less-than-20% missing-observation threshold for feature selection, yielding 64 features across 4,252,826 observations from just under 30,000 firms. The 30-year out-of-sample test period runs from December 1994 to December 2024, the validation sample from December 1992 to November 1994, and all earlier observations form the training sample; we refer to this specification as GFD-64. We also construct a nested subset with identical temporal cutoffs, GFD-32, by randomly eliminating half the features of GFD-64. Although this construction is straightforward and not particularly interesting, it serves as a useful benchmark for the marginal value of additional data: upgrading from GFD-32 to GFD-64 potentially improves forecasts at the cost of data acquisition and additional computational resources. We use these two datasets as examples of distinct but nested specifications. We further construct a GFD-Synthetic-64 dataset which takes the GFD-32 data and adds 32 irrelevant features drawn from the standard normal.

In all applications and for all models, we train models to generate one-month-ahead return forecasts using only the training and validation sample up to the point at which the forecast is generated, i.e., data up to month t is used to compute return forecasts for month $t + 1$. When we evaluate models we do so with the test sample using firm characteristics known at time t to predict returns earned at time $t + 1$.

4.2 Scaling Laws

Figure 2 summarizes performance across our GFD-32 (left) and GFD-64 (right) information sets, with RMSE test loss scaling laws in the upper panels and Sharpe ratio scaling laws in the lower panels. Table 1 reports pairwise differences in asymptotic performance and scaling exponents across specifications, with bootstrapped p-values. Panel (1) shows that doubling the input features from 32 to 64 raises the asymptotic Sharpe ratio from 0.937 to 1.438, a difference of 0.500 ($p < 0.01$). The scaling exponent also shifts: $\hat{\alpha}$ moves from -0.965 to -0.471, a difference of 0.495 ($p < 0.01$). Hence, the richer information set not only raises the attainable performance frontier but also flattens the scaling curve, implying a crossover: at low levels of compute the smaller feature set dominates, and only investors with sufficient compute can expect to benefit from the expanded predictor space, a point we return to below.

Table 1: Hypothesis Tests for Scaling Law Parameters

		Validation		Test		Sharpe Ratio	
	Comparison	L_∞	α	L_∞	α	S_∞	α
	GFD-32	15.748	−2.299	19.178	−2.565	0.937	−0.965
(1)	GFD-64	15.061	−1.917	18.904	−2.055	1.438	−0.471
	Diff	-0.687***	0.382	-0.274***	0.510	0.500***	0.495***
		(0.000)	(0.952)	(0.000)	(0.992)	(0.000)	(0.000)
	GFD-32	15.748	−2.299	19.178	−2.565	0.937	−0.965
(2)	GFD-Synthetic	15.749	−0.809	19.183	−0.806	0.951	−0.634
	Diff	0.001	1.490	0.005	1.758	0.014	0.331**
		(0.182)	(0.242)	(0.32)	(0.122)	(0.430)	(0.018)

Note: Pairwise differences in scaling law parameters. Scaling laws follow $L_\infty + (C/C_0)^{-\alpha}$ for loss and $S_\infty - (C/C_0)^{-\alpha}$ for Sharpe ratios, where L_∞ and S_∞ denote asymptotic performance and α the convergence rate. p-values in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The power law fits the data well at higher compute levels, although the fit deteriorates

where compute is low. For GFD-32 the R^2 is 72.6% for test RMSE and 80.3% for the Sharpe ratio; for GFD-64 these figures are 59.99% and 87.48%, respectively. The power law asymptote for the RMSE loss in the upper panels of Figure 2 is essentially reached near 10^{-7} and 10^{-5} PetaFLOP-days of compute for GFD-32 and GFD-64, respectively. Panel (1) of Table 1 confirms that the asymptotic loss L_∞ is significantly lower under GFD-64 than GFD-32 on both validation and test RMSE (-0.687 and -0.274, respectively; both $p < 0.01$), so the richer feature set reduces the irreducible prediction error. The RMSE scaling exponents, however, are not statistically distinguishable across the two specifications, indicating that additional compute reduces loss at a similar rate regardless of the information set.

To assess whether the gains under the larger GFD-64 information set stem from genuine predictive content rather than the mechanical expansion of the input dimension, we construct a synthetic 64-characteristic sample by augmenting GFD-32 with 32 features drawn from a standard normal distribution. These features contain, by construction, no information useful for forecasting returns, yet a neural network must learn their irrelevance, increasing the compute required to train the model. If the improvements under GFD-64 merely reflected a larger feature set—rather than the economic content of the added information—the synthetic sample should exhibit comparable gains.

Panel (2) of Table 1 shows that it does not. The asymptotic Sharpe ratio under GFD-Synthetic is $\widehat{\mathrm{SR}}_\infty = 0.951$, virtually identical to the GFD-32 estimate of 0.937; the difference of 0.014 ($p = 0.430$) is statistically indistinguishable from zero. By contrast, the GFD-64 ceiling of 1.438 lies far above both. This pattern confirms that the performance gains under GFD-64 arise from genuine cross-sectional information, not from overfitting or due to mechanical benefits of a higher-dimensional input. Although irrelevant features leave the asymptotic limit intact, they degrade the efficiency of compute investment: the Sharpe ratio scaling exponent shifts from -0.965 under GFD-32 to -0.634 under GFD-Synthetic, a difference of 0.331 ($p = 0.018$). Because a smaller exponent in magnitude means the scaling curve is less steep throughout its entire domain, each doubling of compute yields a smaller

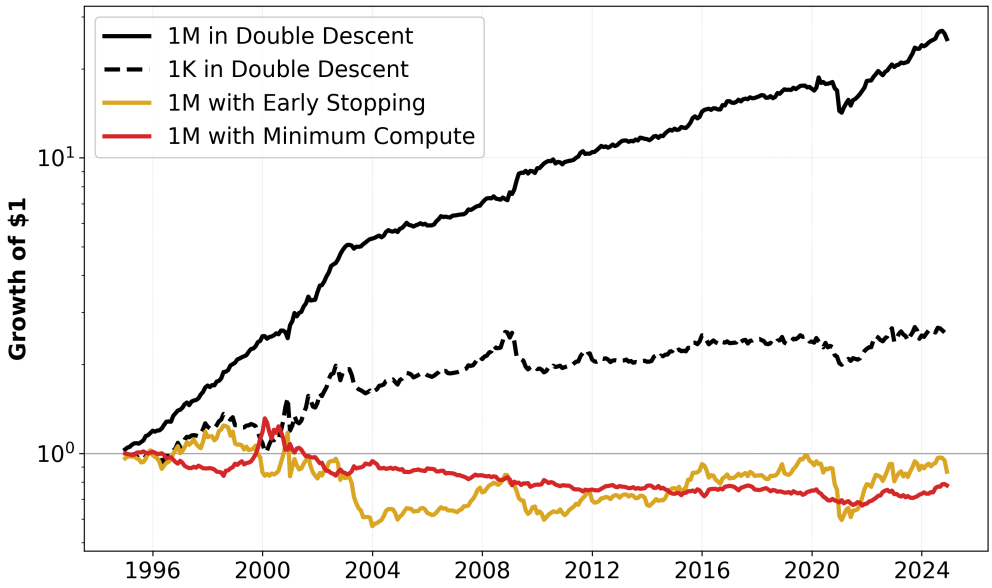
marginal improvement. Irrelevant features therefore do not foreclose the possibility of good predictions but make those predictions considerably more expensive to achieve as the model must learn to distinguish noise from genuine information.

Taken together, the evidence indicates that the composition of the dataset—specifically, its economic content—determines the attainable performance frontier, while irrelevant features serve only to inflate the compute cost of reaching that ceiling without any compensating gain in predictability. These scaling laws also carry implications for market efficiency. Figure A3 illustrates this construction for GFD-64 under the LS_{50} strategy. The estimated scaling law, $\text{SR} = 1.44 - (C/3.8 \times 10^{-9})^{-0.47}$ ($R^2 = 87.5\%$), places the market portfolio at an implied compute of approximately 5×10^{-9} PetaFLOP-days, corresponding to a Sharpe ratio of roughly 0.6. The tangent slope at this point is 6.92×10^7, implying that a 25% increase in compute yields a 14% increase in the Sharpe ratio—the market remains on the concave portion of the curve where computational superiority still commands rents at the margin.

4.3 Compute versus Complexity

Compute and model complexity capture different aspects of the learning process. Complexity is a feature of model design, often measured by parameter count: a neural network with one million parameters is typically more complex than one with 1,000. Compute, by contrast, measures the processing power used for training—a capital resource that determines investors' ability to extract information from a given dataset. This distinction matters because, as we argue in this paper, compute is the appropriate asset pricing primitive.

Figure 3: Double Descent in Practice



Note. This figure plots the cumulative return of a $1 investment in long-short portfolios formed from neural network return forecasts. The solid black line trains a one-million-parameter model past the interpolation threshold into double descent. The gold line applies classical early stopping to the same architecture. The red line trains the same architecture for a single epoch, representing the minimal compute required to use the model. These three specifications share identical architecture and parameter count; they differ only in the amount of compute used for training. The dashed black line represents a one-thousand-parameter model trained into double descent. The sample spans 1994–2024.

To see why, consider that not all parameters are made the same. We train a one-million-parameter model at different levels of compute and evaluate it at each level. Plotting the cumulative return of a $1 investment in long-short portfolios formed from neural network return forecasts, Figure 3 illustrates the performance of such a model trained at three levels: double descent, early stopping, and a single epoch representing the minimal compute one could expend to use a one-million-parameter model. If parameters alone determined asset pricing outcomes, these three versions should perform identically because they share the same parameter count. They do not. Performance strongly depends on how much the model is trained, that is, on compute. Moreover, a one-thousand-parameter model in Figure 3 outperforms the more complex one-million-parameter models when they are not trained into double descent.

Under the classical bias-variance tradeoff, early stopping leads one to conclude that the

one-million-parameter model suffers from severe overfitting, which would suggest adopting a simpler architecture. This interpretation is misleading: the model performs poorly under early stopping not because it is too complex, but because it is undertrained. When trained well past the interpolation threshold into double descent, the same model achieves a positive cumulative return with a Sharpe ratio well above 1. Parameters serve as a conduit through which the capital resource of compute is expended, making compute the natural primitive for analyzing asset pricing consequences. Parameters and compute remain closely linked—they map monotonically to one another, and larger models require substantially more compute—but it is ultimately compute that governs performance.

The relevant choice given a compute budget is therefore a parameter count large enough to be expressive and enjoy the virtues of complexity studied by Kelly et al. (2024b) yet small enough to reach double descent within the budget. We enforce this condition in training; larger models require proportionally more compute, spanning 10^{-10} to 10^{-3} PetaFLOP-days. This is illustrated by Figure 2 which plots test loss against compute for each trained model, with circles colored from deep purple (smallest) to yellow (largest), revealing a monotonic and heteroskedastic relation between forecast performance and compute.

This heteroskedasticity reveals a benefit of larger models beyond predictive performance: reduced model training risk. At low compute budgets, model performance displays large dispersion, whereas larger, more compute-intensive models converge more tightly. A natural explanation is geometric—as model size increases, the effective loss landscape near the learned solution becomes implicitly regularized and easier to navigate with first-order methods.

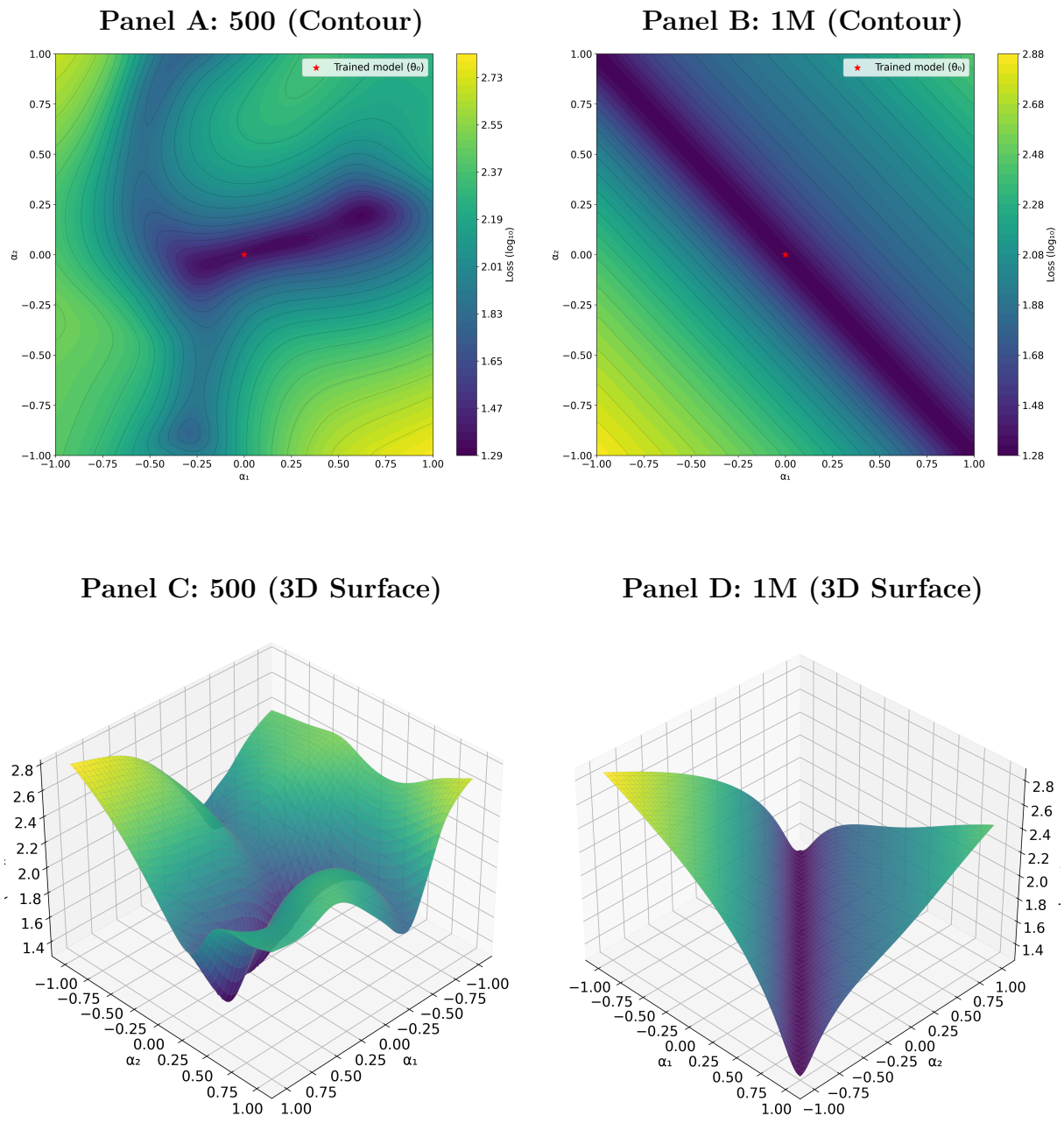
We make this mechanism concrete by visualizing the loss landscape at the extremes of our compute range. Specifically, we characterize the local optimization geometry by projecting the loss surface onto a two-dimensional affine subspace through the trained parameter vector

θ_0.[8] The resulting surface, visualized in Figure 4 as both a contour map and a three-dimensional plot for 500- and 1-million-parameter firm-level models with the origin at the trained model, reveals local curvature, the presence of nearby local minima, and the sharpness of the optimum.[9]

[8] We sample two random Gaussian direction vectors d_1 and d_2 in parameter space and evaluate the loss along perturbations $\theta(\alpha_1, \alpha_2) = \theta_0 + \alpha_1 d_1 + \alpha_2 d_2$ over a grid of (α_1, α_2) values, recomputing mean squared error on the test sample at each grid point.

[9] We report results only for GFD-64, though the patterns are largely the same for GFD-32.

Figure 4: Firm-Level Return Loss Landscapes Across Model Size



Note. Panels A and C show, respectively, a contour map and 3D surface of the loss landscape ($\log_{10}$ MSE) for the 500-parameter firm-level return model using GFD-64 in a two-dimensional random subspace around the trained parameters θ_0, panels B and D show the analogous plots for the 1M-parameter model. The red star marks the trained model at $(\alpha_1, \alpha_2) = (0, 0)$.

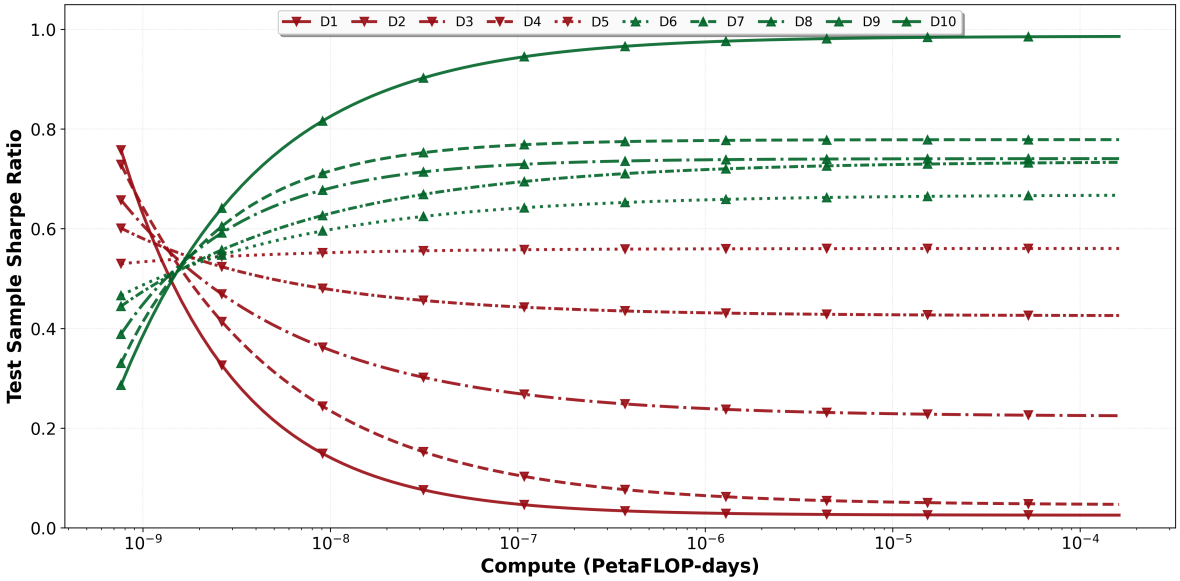
The small model's landscape contains more local minima in which optimization can become trapped, whereas the large model's loss surface forms a continuous ridge that makes navigation toward a global minimum via gradient descent considerably easier, consistent

with evidence that larger, overparameterized networks produce increasingly well-behaved loss surfaces (Li et al., 2018). For investors spending millions of dollars on compute, this expenditure purchases not only the predictive gains documented above but also a reduction in training risk: at the lower end of the compute distribution, a practitioner may arrive at a well-fitted model through favorable initialization alone—or become trapped in a saddle point, wasting the entire compute budget on a model that underperforms. At the upper end, the loss landscape is sufficiently smooth that performance becomes largely insensitive to initial conditions and hyperparameter choices.

4.4 Decomposing the Sharpe Ratio Scaling Law

Our results so far establish that Sharpe ratios scale as a power law in compute, but leave open the question of which components of the trading strategy drive this relationship. We address this by decomposing the long-short portfolio into decile portfolios sorted by predicted returns and by varying the breadth of the long-short rule.

Figure 5: Sharpe Ratio Scaling Laws by Decile

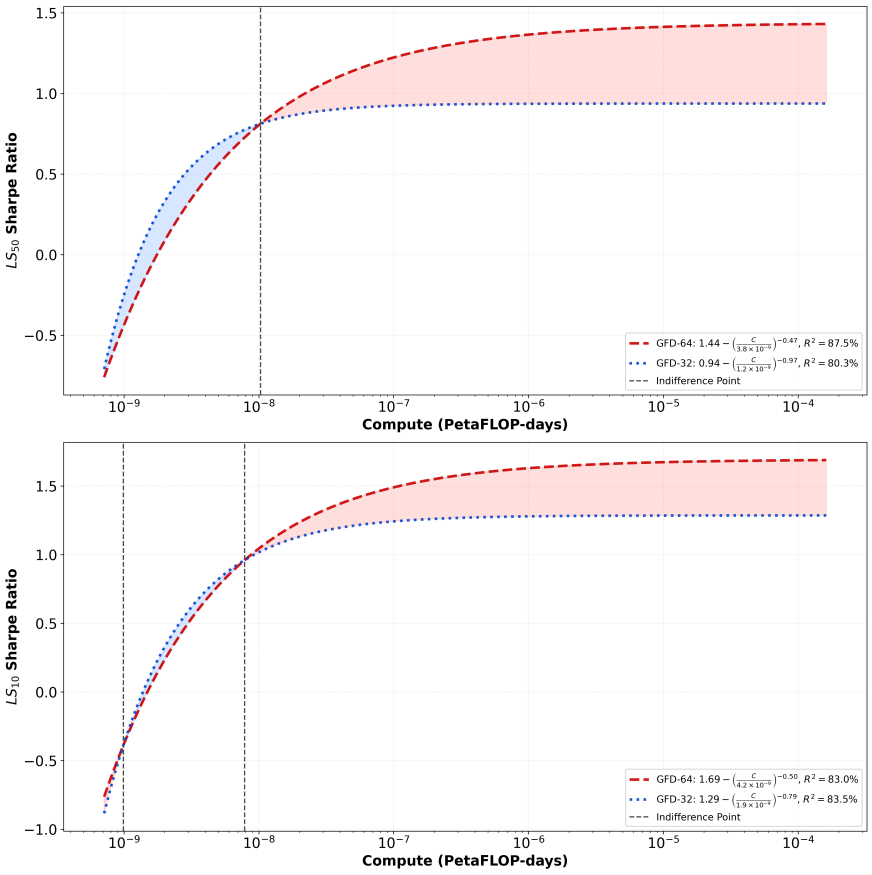


Note. This figure plots the annualized Sharpe ratios of the underlying decile portfolios from Figure 2 bottom right panel (GFD-64 Sharpe ratio scaling law).

Figure 5 plots compute against test-sample Sharpe ratios for decile portfolios, where

decile 1 contains stocks with the lowest predicted returns and decile 10 the highest. The compute–Sharpe ratio relation is clearly monotonically increasing for the top deciles and decreasing for the bottom deciles, with essentially flat curves for the middle deciles.[10] Additional compute therefore improves portfolio performance only for stocks with a large enough predictable return component to place them in the tails of the expected-return distribution. Stocks for which the models do not identify a strong predictive signal have return expectations close to their unconditional mean and are assigned to the middle portfolios, where compute has little marginal value.

Figure 6: Sharpe Ratio Scaling Laws by Breakpoint and Data



Note. This figure reports scaling laws for annualized Sharpe ratios of long-short portfolios formed from GFD-32 and GFD-64 firm-level return predictions. Each month, stocks are sorted into deciles by the model's forecast and portfolio returns are constructed by going long a set of upper deciles and short a set of lower deciles. For example, LS_{10} is long decile 10 and short decile 1, while LS_{50} is long the top half of deciles (6–10) and short the bottom half (1–5). Curves show fitted Sharpe scaling with compute/model size for each breakpoint specification.

[10]Table A2 reports parameter estimates for these decile portfolios.

Figure 6 extends this analysis by varying the breadth of the long-short rule. Each month, stocks are sorted into deciles by the model's forecast, and long-short portfolios are formed by going long one or more high-ranked deciles and shorting the corresponding low-ranked deciles: LS_{10} goes long decile 10 and shorts decile 1, while LS_{50} goes long the top half (deciles 6–10) and shorts the bottom half (1–5). We construct these portfolios for both GFD-32 and GFD-64 and compare their scaling laws.

Consistent with the crossover documented in the main results, investors operating under low compute budgets achieve higher Sharpe ratios by holding broad cross-sections rather than concentrating in the tails. At compute levels between 10^{-9} and 10^{-8}, the GFD-32 LS_{50} portfolio strictly dominates LS_{10}. Low-compute investors also benefit from smaller datasets: GFD-32 strictly dominates GFD-64 at lower compute levels across both LS_{50} and LS_{10}. Beyond approximately 10^{-8} PetaFLOP-days of compute this ordering reverses, with GFD-64 dominating and the Sharpe ratio differential widening monotonically in compute. These results underscore a broader principle: the marginal unit of compute is most productively deployed in lower-dimensional inference settings, so less-resourced investors optimally restrict attention to broader portfolios and more parsimonious datasets that are simpler to learn from.

4.5 Benchmarking

The preceding results document scaling laws for neural networks trained on firm-level return data. We next address how these asymptotic bounds compare with conventional linear methods.

4.5.1 Linear Benchmark

Table 2: Benchmarking to Ridge Regression

		λ							Scaling
		0	10^{-2}	10^{-1}	1	10	10^2	10^3	Law
GFD-32	Val.	15.75	15.75	15.75	15.75***	15.78***	15.79***	15.79***	15.75
	Test	110.20***	31.92***	31.02***	25.73***	19.64***	19.19***	19.19***	19.18
	LS_{50}	1.14	1.13	1.10	1.05	0.84***	0.77***	0.75***	0.94
	LS_{10}	1.31	1.31	1.28	1.23**	1.09***	1.01***	1.00***	1.29
	L_{10}	1.05	1.04	1.04	1.02***	0.95***	0.92***	0.92***	1.05
GFD-64	Val.	15.73***	15.73***	15.73***	15.74***	15.77***	15.79***	15.79***	15.06
	Test	397.4***	377.1***	284.5***	65.24***	21.39***	19.25***	19.19***	18.90
	LS_{50}	1.38**	1.39**	1.37***	1.28***	1.00***	0.76***	0.73***	1.44
	LS_{10}	1.53***	1.56***	1.50***	1.52***	1.24***	0.98***	0.95***	1.69
	L_{10}	1.17	1.17	1.15	1.11	0.98*	0.91***	0.90***	0.99

Note: This table compares pooled ridge regression benchmarks against the estimated asymptotic performance of neural scaling laws. Ridge regressions are trained on the pooled panel of firm-month observations. Columns index the regularization parameter λ, where $\lambda = 0$ corresponds to the ridgeless estimator. The final column reports L_∞ (for loss metrics) or SR_∞ (for Sharpe ratios). Validation (Val.) and test RMSE are reported in percentage points, LS_{50} and LS_{10} denote annualized Sharpe ratios. Significance stars on ridge coefficients are based on one-sided bootstrap hypothesis tests; for loss metrics, the test is $H_0 \colon L_\infty =$ Ridge RMSE against $H_1 \colon L_\infty <$ Ridge RMSE, so that rejection indicates the scaling law's irreducible loss lies significantly below the ridge estimate; for Sharpe ratios, the test is $H_0 \colon SR_\infty =$ Ridge SR against $H_1 \colon SR_\infty >$ Ridge SR, so that rejection indicates the scaling law's asymptotic Sharpe significantly exceeds the ridge benchmark. In both cases, the p-value is the fraction of bootstrap scaling-law asymptote draws on the "ridge side" of the hypothesized value. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 2 reports the RMSE and Sharpe ratio of ridge regression across a grid of penalty parameters λ. We vary λ across a wide range similar to Kelly et al. (2024b) and include the ridgeless case as well. A central feature of the linear model is the tension between prediction accuracy and portfolio performance. For GFD-32, the best test RMSE of 19.19 obtains at high regularization ($\lambda \geq 10^2$), where the Sharpe ratio falls to 0.77; conversely, the

best Sharpe ratio of 1.14 obtains at $\lambda = 0$, where the test RMSE explodes to 110.20 due to overfitting, illustrating that the ridge regression forecasts do not simultaneously perform well along the statistical and economic criteria.

The scaling laws face no such trade-off. As compute increases, validation RMSE, test RMSE, and Sharpe ratio all improve along their respective scaling curves. For the GFD-32 scaling law, these values converge to 15.748, 19.178, and 0.937, respectively. Neural networks pushed to the limits of predictability for GFD-32 therefore match the best-in-class ridge loss while delivering a 22% higher LS_{50} Sharpe ratio at the same ridge regularization level. Focusing solely on the Sharpe ratio, ridge regression achieves a significantly higher Sharpe only at the lowest level of penalization and, even then, only for the LS_{50} portfolio.

The pattern is starker for GFD-64, where the ridge trade-off is more severe: the best Sharpe of 1.39 ($\lambda = 10^{-2}$) coincides with a test RMSE of 377.11, while bringing test RMSE down to 19.19 ($\lambda = 10^{3}$) compresses the Sharpe ratio to 0.73. The scaling-law asymptotes of $L_\infty = 15.061$ and 18.904 surpass the best ridge loss, and the Sharpe ratio asymptotes significantly exceed the ridgeless benchmark under both portfolio sorts. Against the LS_{50} ridgeless Sharpe of 1.38, the estimated S_∞ of 1.44 yields $z = 2.17$ ($p = 0.022$; 95% CI $[0.009, 0.112]$). The gap widens for the LS_{10} portfolio: $S_\infty = 1.69$ exceeds the ridgeless Sharpe of 1.53 with $z = 4.41$ ($p < 0.001$; 95% CI $[0.099, 0.245]$). These results confirm that the richer feature set is where nonlinear models most clearly justify their computational cost.

4.6 Robustness to Architecture

We next address whether the scaling patterns uncovered in our main analysis are robust to our baseline architecture. All neural networks estimated thus far use five hidden layers, as detailed in Appendix A. To assess whether this choice materially affects the scaling laws, we re-estimate the GFD-64 specification with one hidden layer and ten hidden layers, holding all other hyperparameters fixed. As with the synthetic-feature test, we focus on Sharpe ratio scaling laws, though the conclusions are similar for RMSE.

The ten-layer specification yields an asymptotic Sharpe ratio of $\widehat{\mathrm{SR}}_\infty = 1.350$, compared with 1.438 for the five-layer baseline. The difference of -0.088 is not statistically significant (95% CI: $[-0.181, 0.046]$; $p = 0.168$), indicating that doubling the depth does not raise the performance ceiling. The scaling exponents do differ: $\hat{\alpha}$ shifts from -0.471 to -0.323, a difference of 0.147 that is significant at the 5% level (95% CI: [0.016, 0.282]; $p = 0.026$). Deeper architectures thus converge more slowly toward the same asymptote, consistent with the additional parameters requiring more compute without improving the limiting forecast.

The one-layer specification tells a different story. The asymptotic Sharpe ratio falls to $\widehat{\mathrm{SR}}_\infty = 1.133$, a difference of -0.304 from the baseline that is significant at the 1% level (95% CI: $[-0.443, -0.184]$; $p < 0.001$). The scaling exponent also differs sharply, moving from -0.471 to -1.364 (95% CI: $[-1.569, -0.552]$; $p < 0.001$). A single hidden layer thus faces a materially lower performance ceiling, suggesting that the nonlinear feature interactions captured by deeper architectures—which a shallow network cannot represent—account for a meaningful share of the predictability in the data.

Taken together, these results indicate that architecture depth matters up to a point. Sufficient depth is necessary to capture the nonlinear structure in firm characteristics that drives return predictability, but beyond that threshold additional layers do not raise the attainable Sharpe ratio.

5. Compute, Information Acquisition, and Market Equilibrium

To understand and explain the importance of compute for asset prices, in this section we develop an equilibrium asset pricing model that explicitly accounts for heterogeneity in investors' access to compute. Our setup assumes a single-period CARA-normal rational expectations model in which investors differ in computational capacity, optimally allocate compute across private and public information, and learn from the equilibrium price. To facilitate intuition and interpretability of our results, we intentionally choose a model that is close to the canonical Grossman-Stiglitz environment. High-compute investors can incur

a fixed access cost to use a private signal and then divide their total compute resources across private and public signals. Low-compute investors rely on the public signal and the information revealed by the market clearing price. Noisy supply prevents the price from fully revealing the private signal, so the equilibrium is only partially revealing of other investors' information or superior compute.

Our model delivers four core results that map directly to the empirical questions of the paper. First, the optimal partitioning of compute across private and public information is pinned down by a precision-weighted equimarginal condition. Second, the value of private information is reduced by price revelation but remains a nondecreasing function of compute when the anticipated price-learning terms are taken as given by individual investors, and it is strictly increasing under curvature conditions stated below. Third, conditional on a chosen stratified compute allocation, the linear rational expectations pricing block remains affine but is governed by a bilateral price-learning fixed point: the price transmits part of the low-compute investors' public realization to the high-compute investors and part of high-compute investors' private and public information to low-compute investors. Fourth, bilateral price learning changes the nature of residual information differences, because the direct high-compute precision advantage is now offset by a stronger low-compute recovery of that advantage through prices.

We develop an extension that sharpens the interpretation of the baseline REE. A benchmark with no private information acquisition shows that heterogeneity in compute alone still generates heterogeneous public posteriors when all investors observe the same underlying public information source but process distinct reduced-form realizations. Under power-law processing, the low-compute investors asymptotically recover the high-compute investors' pooled sufficient statistic from the price, so both types inherit the same RMSE and deterministic Sharpe-ratio exponents.

5.1 Environment

We assume that there is one risky asset with a liquidation value $\tilde{\theta} \sim \mathcal{N}(\bar{\theta}, \tau_0^{-1})$ and one risk-free asset whose return is normalized to be zero. The risky asset supply is noisy:

$$\tilde{z} = \bar{z} + \tilde{u},$$

where $\tilde{u} \sim \mathcal{N}(0, \tau_u^{-1})$ is independent of all other shocks. All investors have CARA utility $U(W) = -\exp(-AW)$ with common risk aversion $A > 0$.

A fraction $\lambda \in (0, 1)$ are high-compute investors with total compute budget c_{H}. The remaining fraction $1 - \lambda$ are low-compute investors with smaller compute budgets $c_{\mathrm{L}} < c_{\mathrm{H}}$. We fix these allocations of compute ex-ante, rather than deriving them as part of investors' optimizing decisions. The equilibrium we derive is therefore better interpreted as a short- to medium-term equilibrium with fixed compute allocations.

The economy features both private and public information sources about $\tilde{\theta}$. Access to the private source is contingent upon paying a fixed cost $\kappa > 0$, measured in certainty-equivalent wealth units at the acquisition stage. Public information is freely available. The baseline quality (precision) parameters of the two information sources are denoted by $\tau_{s_1} > 0$ and $\tau_{s_2} > 0$.

Investors' compute determines how effectively they process each information source. If an investor devotes c_1 units of compute to the private source and c_2 units of compute to the public source, the resulting processed information is summarized by Gaussian sufficient statistics whose posterior precision contributions are $f_1(c_1)\tau_{s_1}$ and $f_2(c_2)\tau_{s_2}$. We use this reduced-form effective-signal representation throughout the REE derivations, so the signal precision entering posterior beliefs and prices is always the compute-adjusted precision. The object that is common across investors is the underlying public information source or raw public dataset, not the realized reduced-form statistic itself. Different investors can therefore extract different effective public statistics from the same underlying source. In the stratified

equilibrium below we write these public statistics as $s_{2,\mathrm{H}}$ and $s_{2,\mathrm{L}}$: they share the same mean $\tilde{\theta}$ but generally have different realizations and different compute-adjusted precisions $q_{2,\mathrm{H}}$ and $q_{2,\mathrm{L}}$. We use analogous notation for the private processed statistic s_1 whenever the relevant compute allocation is clear from the context.

Assumption 1 (Compute transformations)**.** *For $k \in \{1, 2\}$, the mapping from compute c_k to the posterior precision takes the form $f_k(c_k)\tau_{s_k}$ where the function $f_k : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable on $(0, \infty)$ and satisfies $f_k(0) = 0$, $f_k'(c) > 0$ for all $c > 0$, and $f_k''(c) \leq 0$ for all $c > 0$.*

Assumption 1 imposes shape restrictions needed for the compute-allocation and comparative-static results. It does not impose any ranking between the private and public compute transformation functions and does not require processing capacity to be unbounded. The assumption therefore accommodates both unbounded and bounded saturating technologies.[11]

Timing in our model follows a standard setup: Investors are infinitesimal and therefore take the anticipated linear price law as given when choosing whether to acquire private data and how to allocate compute. Signals are then realized, the equilibrium price is posted, investors update their beliefs, and trading takes place.

5.2 Compute Allocation and Information Acquisition

Consider first the problem of a price-taking investor with a total compute budget c. The investor's posterior precision after acquiring the private signal and allocating compute (c_1, c_2) with $c_1 + c_2 = c$, but before using the price signal, is

$$\tau_0 + f_1(c_1)\tau_{s_1} + f_2(c_2)\tau_{s_2}.$$

[11] A leading unbounded example is the power-law family $f_k(c) = \kappa_k c^{\alpha_k}$ with $\kappa_k > 0$ and $\alpha_k \in (0, 1)$. An empirically aligned bounded example is $f_k(c) = \bar{f}_k\big[1 - (1 + c/c_{0k})^{-\alpha_k}\big]$ with $\bar{f}_k > 0$, $c_{0k} > 0$, and $\alpha_k > 0$ which approaches the finite ceiling $\bar{f}_k$ at a power-law rate. That bounded family is consistent with observed scaling laws in which performance gaps converge to finite asymptotes as compute rises.

In the stratified REE below, an acquiring investor uses the price to learn about the other type's public-data realization. Because a price-taking investor treats the price law as fixed when choosing (c_1, c_2), we represent the additional price-learning term by an anticipated constant $\tau_{P,H} \geq 0$. Thus the investor's post-price posterior precision from acquiring private information is

$$\tau^{\mathrm{H}}(c_1, c_2; \tau_{P,H}) = \tau_0 + f_1(c_1)\tau_{s_1} + f_2(c_2)\tau_{s_2} + \tau_{P,H}.$$

We assume throughout that posterior precision is additive across conditionally independent Gaussian signals. This is the same simplifying structure used in the baseline REE block and can be generalized if desired.[12]

Similarly, if the investor does not acquire the private signal and devotes all compute to the public source, the investor's post-price posterior precision is

$$\tau^{L}(c; \tau_{P,L}) = \tau_0 + f_2(c)\tau_{s_2} + \tau_{P,L},$$

where $\tau_{P,L}$ is the anticipated price-learning precision available to a non-acquiring investor. Appendix C.1 shows that, for a price-taking investor, the expected certainty equivalent is affine in the posterior precision. We model the acquisition stage in those same certainty-equivalent units, so investors compare expected certainty equivalent values net of the information acquisition cost, κ. Individual investors' compute and acquisition problem therefore can still be cast as a posterior-precision problem:

Proposition 1 (Optimal compute split)**.** *For an investor with a total compute budget c, the optimal allocation of compute to the private signal solves*

$$c_1^*(c) \in \arg\max_{0 \leq c_1 \leq c} \left\{ f_1(c_1)\tau_{s_1} + f_2(c - c_1)\tau_{s_2} \right\}.$$

[12]One could extend the analysis to more general posterior-precision aggregators without changing the central economic point that compute affects both direct signal extraction and the informativeness of the equilibrium price.

Any interior optimum satisfies

$$f_1'(c_1^*(c))\tau_{s_1} = f_2'(c - c_1^*(c))\tau_{s_2}. \tag{5}$$

If f_1 and f_2 are strictly concave, the optimum is unique and $c_1^*(c)$ is nondecreasing in c, with $0 < c_1^{*\prime}(c) < 1$ on any interval with an interior solution.

The first-order condition in (5) is a precision-weighted equimarginal condition: at the optimum, the marginal precision gained from a unit of compute devoted to the private source must equal that from devoting the same unit to the public source. The relative source-quality parameters τ_{s_1} and τ_{s_2} shift this interior solution, so a higher-quality private information source commands a larger compute share. Because c_1^* is nondecreasing in the total compute budget, investors with more compute devote weakly more absolute compute to private processing. This monotonicity property underpins the endogenous information stratification described below.

Next, define the maximized effective signal precision of an informed investor by

$$\Phi^*(c) \equiv f_1(c_1^*(c))\tau_{s_1} + f_2(c - c_1^*(c))\tau_{s_2} + \tau_{P,H}.$$

The net precision gain from private data acquisition is therefore

$$G(c) \equiv \Phi^*(c) - f_2(c)\tau_{s_2} - \tau_{P,H},$$

where $G(c)$ captures the extra posterior precision an investor can generate from optimally accessing private information versus devoting all compute to the public signal.

Next, consider the certainty-equivalent value of a unit of posterior precision:[13]

$$\Xi \equiv \frac{\mathbb{E}[(\tilde{\theta} - P)^2]}{2A} > 0.$$

[13]Note that the expectation is taken under the anticipated price law.

The scalar Ξ captures both the magnitude of residual uncertainty in the price and the investor's risk aversion, putting precision gains in the same units as access cost κ.

Proposition 2 (Acquisition threshold)**.** *An investor with compute budget c acquires the private signal if and only if $\kappa < \kappa^*(c; \tau_{P,H}, \tau_{P,L})$, where*

$$\kappa^*(c; \tau_{P,H}, \tau_{P,L}) \equiv \Xi \left[G(c) + \tau_{P,H} - \tau_{P,L} \right]_+,$$

where $[x]_+ = \max\{x, 0\}$. Moreover, $G(c)$ is nondecreasing in c, so $\kappa^*(c; \tau_{P,H}, \tau_{P,L})$ is nondecreasing in c for fixed anticipated price-learning terms. If f_2 is strictly concave, then $\kappa^*(c; \tau_{P,H}, \tau_{P,L})$ is strictly increasing on any interval where, for some c, $c_1^*(c) > 0$ and $G(c) + \tau_{P,H} > \tau_{P,L}$.

The acquisition threshold in Proposition 2 compares two price systems rather than a direct-information block against a single common price-implied signal. An acquiring investor anticipates that the equilibrium price will reveal some of the other type's public realization, while a non-acquiring investor anticipates that the price will reveal part of the informed type's aggregate statistic. The fixed access cost therefore must be compared with the direct precision gain from private acquisition, $G(c)$, plus the difference between these two anticipated price-learning terms. Crucially, however, the monotonicity in compute is unchanged: for fixed anticipated price terms, more total compute raises the value of having access to the private source.

Corollary 1 (Information stratification)**.** *In the acquisition-stage problem with common anticipated price-learning terms $(\tau_{P,H}, \tau_{P,L})$, high-compute investors acquire the private signal while low-compute investors do not if and only if*

$$\kappa^*(c_{\mathrm{L}}; \tau_{P,H}, \tau_{P,L}) \leq \kappa < \kappa^*(c_{\mathrm{H}}; \tau_{P,H}, \tau_{P,L}).$$

Corollary 1 formalizes the two-tier market structure that motivates the remainder of the

analysis. The stratified equilibrium arises endogenously from the conjunction of heterogeneity in compute and fixed costs of data acquisition: the same fixed cost that deters investors with low compute is worthwhile for those with access to high compute. This endogenous stratification is still driven entirely by differences in computational capacity rather than by exogenous signal access, but the subsequent REE now lets both types learn from prices because the public data arrive as distinct reduced-form realizations rather than a single common statistic.

Corollary 1 also helps explain cases in which low-compute investors rationally choose not to acquire the largest information set and instead use their limited compute to process a lower-dimensional data set consistent with the conclusion from Figure 5.

5.3 The Stratified Rational Expectations Equilibrium

In the stratified regime in Corollary 1, high-compute investors acquire the private signal and choose the optimal split

$$c_{1,\mathrm{H}} \equiv c_1^*(c_{\mathrm{H}}), \qquad c_{2,\mathrm{H}} \equiv c_{\mathrm{H}} - c_{1,\mathrm{H}}.$$

Low-compute investors allocate all compute to the public signal. Conditional on such a stratified allocation, we now solve for the affine price equations consistent with the induced effective precisions.

Define the components of the effective signal precision:

$$q_{1,\mathrm{H}} \equiv f_1(c_{1,\mathrm{H}})\tau_{s_1}, \qquad q_{2,\mathrm{H}} \equiv f_2(c_{2,\mathrm{H}})\tau_{s_2}, \qquad q_{2,\mathrm{L}} \equiv f_2(c_{\mathrm{L}})\tau_{s_2}.$$

The corresponding reduced-form signals can be represented as

$$s_1 = \tilde{\theta} + \tilde{\eta}_{1,\mathrm{H}}, \qquad s_{2,\mathrm{H}} = \tilde{\theta} + \tilde{\eta}_{2,\mathrm{H}}, \qquad s_{2,\mathrm{L}} = \tilde{\theta} + \tilde{\eta}_{2,\mathrm{L}},$$

where

$$\tilde{\eta}_{1,\mathrm{H}} \sim \mathcal{N}(0, q_{1,\mathrm{H}}^{-1}), \qquad \tilde{\eta}_{2,\mathrm{H}} \sim \mathcal{N}(0, q_{2,\mathrm{H}}^{-1}), \qquad \tilde{\eta}_{2,\mathrm{L}} \sim \mathcal{N}(0, q_{2,\mathrm{L}}^{-1})$$

are mutually independent and independent of $\tilde{u}$. Define the high-compute investors' pooled sufficient statistic by

$$Q_{\mathrm{H}} \equiv q_{1,\mathrm{H}} + q_{2,\mathrm{H}}, \qquad s_{\mathrm{H}} \equiv \frac{q_{1,\mathrm{H}} s_1 + q_{2,\mathrm{H}} s_{2,\mathrm{H}}}{Q_{\mathrm{H}}}.$$

Because $s_{\mathrm{H}} = \tilde{\theta} + \tilde{\eta}_{\mathrm{H}}^*$ with $\tilde{\eta}_{\mathrm{H}}^* \sim \mathcal{N}(0, Q_{\mathrm{H}}^{-1})$, it is a sufficient statistic for the high-compute investors' pre-price information. Their pre-price posterior precision and mean are therefore

$$\tau_{\mathrm{H}} = \tau_0 + Q_{\mathrm{H}},$$

and

$$\mu_{\mathrm{H}} = \frac{\tau_0 \bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}}}{\tau_{\mathrm{H}}} = \frac{\tau_0 \bar{\theta} + q_{1,\mathrm{H}} s_1 + q_{2,\mathrm{H}} s_{2,\mathrm{H}}}{\tau_{\mathrm{H}}}.$$

Low-compute investors observe only their own public realization before the price, so their pre-price posterior precision and mean are

$$\tau_{\mathrm{L}} = \tau_0 + q_{2,\mathrm{L}},$$

and

$$\mu_{\mathrm{L}} = \frac{\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}}}{\tau_{\mathrm{L}}}.$$

We conjecture an affine price of the form

$$P = a_0 + a_{\mathrm{H}} s_{\mathrm{H}} + a_{\mathrm{L}} s_{2,\mathrm{L}} - a_u \tilde{u},$$

with $a_{\mathrm{H}} > 0$, $a_{\mathrm{L}} > 0$, and $a_u > 0$. Appendix C.5 shows that both investor types now learn

from the price. High-compute investors can extract

$$s_{P,\mathrm{H}} \equiv \frac{P - a_0 - a_{\mathrm{H}} s_{\mathrm{H}}}{a_{\mathrm{L}}} = s_{2,\mathrm{L}} - \frac{a_u}{a_{\mathrm{L}}} \tilde{u}, \tag{6}$$

while low-compute investors can extract

$$s_{P,\mathrm{L}} \equiv \frac{P - a_0 - a_{\mathrm{L}} s_{2,\mathrm{L}}}{a_{\mathrm{H}}} = s_{\mathrm{H}} - \frac{a_u}{a_{\mathrm{H}}} \tilde{u}.$$

The corresponding price-implied precisions are

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \left(\frac{a_u}{a_{\mathrm{L}}} \right)^2 \tau_u^{-1} \right)^{-1},$$

and

$$\tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \left(\frac{a_u}{a_{\mathrm{H}}} \right)^2 \tau_u^{-1} \right)^{-1}.$$

Thus the post-price posterior precisions are

$$\tau_{\mathrm{H}}^R = \tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}},$$

and

$$\tau_{\mathrm{L}}^R = \tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}.$$

The corresponding post-price posterior means are

$$\mu_{\mathrm{H}}^R = \frac{\tau_0 \bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}} + \tau_{P,\mathrm{H}} s_{P,\mathrm{H}}}{\tau_{\mathrm{H}}^R},$$

and

$$\mu_{\mathrm{L}}^R = \frac{\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}} s_{P,\mathrm{L}}}{\tau_{\mathrm{L}}^R}.$$

The price transmits a noisy version of the low-compute investors' public realization to high-

compute investors and a noisy version of the high-compute investors' pooled statistic to low-compute investors.

Proposition 3 (Affine price equilibrium conditional on stratification)**.** *Suppose the economy is in the stratified regime of Corollary 1, and fix the induced effective precisions* $(q_{1,\mathrm{H}}, q_{2,\mathrm{H}}, q_{2,\mathrm{L}})$ *with* $q_{1,\mathrm{H}} > 0$*. Define* $\tau_{P,\mathrm{H}}(\beta_{\mathrm{H}}) \equiv \left(q_{2,\mathrm{L}}^{-1} + \beta_{\mathrm{H}}^2 \tau_u^{-1}\right)^{-1}$ *and* $\tau_{P,\mathrm{L}}(\beta_{\mathrm{L}}) \equiv \left(Q_{\mathrm{H}}^{-1} + \beta_{\mathrm{L}}^2 \tau_u^{-1}\right)^{-1}$*.*

Then there exists a positive pair $(\beta_{\mathrm{H}}, \beta_{\mathrm{L}})$ *satisfying*

$$\beta_{\mathrm{H}} = \frac{\dfrac{A}{1-\lambda} + \tau_{P,\mathrm{L}}(\beta_{\mathrm{L}})\beta_{\mathrm{L}}}{q_{2,\mathrm{L}}}, \tag{7}$$

and

$$\beta_{\mathrm{L}} = \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}(\beta_{\mathrm{H}})\beta_{\mathrm{H}}}{Q_{\mathrm{H}}}. \tag{8}$$

For any such pair, the affine rational expectations equilibrium price is

$$\begin{aligned} P = &\frac{\tau_0\bar{\theta} + \left[\lambda Q_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\right] s_{\mathrm{H}} + \left[\lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}\right] s_{2,\mathrm{L}} - A\bar{z}}{\bar{\tau}^R} \\ &- \frac{A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{\bar{\tau}^R}\tilde{u}, \end{aligned} \tag{9}$$

where $\tau_{P,\mathrm{H}} = \tau_{P,\mathrm{H}}(\beta_{\mathrm{H}})$*,* $\tau_{P,\mathrm{L}} = \tau_{P,\mathrm{L}}(\beta_{\mathrm{L}})$*, and*

$$\bar{\tau}^R = \tau_0 + \lambda Q_{\mathrm{H}} + \lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}} + (1-\lambda)\tau_{P,\mathrm{L}}. \tag{10}$$

Conversely, any affine REE consistent with the chosen stratified allocation must admit a positive pair $(\beta_{\mathrm{H}}, \beta_{\mathrm{L}})$ *satisfying* (7)–(8) *and a price equation of the form* (9)*.*

Proposition 3 shows that bilateral price learning replaces the one-dimensional coefficient formula of the baseline with a coupled fixed-point system. The ratio $\beta_{\mathrm{H}} = a_u/a_{\mathrm{L}}$ governs how precisely high-compute investors can recover the low-compute public draw from the price, while $\beta_{\mathrm{L}} = a_u/a_{\mathrm{H}}$ governs how precisely low-compute investors can recover the high-compute pooled statistic. The price remains affine and the feedback runs in both directions because

the public source generates distinct realizations across investor types.

5.4 Main Implications

Define the bilateral revelation ratios by

$$\Pi_{\mathrm{H}} \equiv \frac{\tau_{P,\mathrm{H}}}{q_{2,\mathrm{L}}}, \qquad \Pi_{\mathrm{L}} \equiv \frac{\tau_{P,\mathrm{L}}}{Q_{\mathrm{H}}}.$$

The ratio Π_{H} measures the fraction of the low-compute investors' direct public precision that is transmitted to the high-compute investors through the price. The ratio Π_{L} measures the fraction of the high-compute investors' pooled sufficient-statistic precision that is transmitted to the low-compute investors through the price.

Proposition 4 (Bilateral price learning)**.** *In the stratified affine rational expectations equilibrium,*

$$\Pi_{\mathrm{H}} = \frac{1}{1 + \beta_{\mathrm{H}}^2 q_{2,\mathrm{L}} \tau_u^{-1}}, \qquad \Pi_{\mathrm{L}} = \frac{1}{1 + \beta_{\mathrm{L}}^2 Q_{\mathrm{H}} \tau_u^{-1}}, \tag{11}$$

so

$$0 < \Pi_{\mathrm{H}} < 1, \qquad 0 < \Pi_{\mathrm{L}} < 1.$$

Proposition 4 shows that the price is only partially revealing on both sides. High-compute investors never recover the low-compute public draw perfectly because supply shocks contaminate the inversion in (6). Likewise, low-compute investors never recover the high-compute sufficient statistic perfectly for finite τ_u. Supply noise therefore prevents full revelation of either hidden object and keeps both price-learning terms strictly below the corresponding direct precisions.

We define the residual post-price information difference by

$$\Delta\tau^R \equiv \tau_{\mathrm{H}}^R - \tau_{\mathrm{L}}^R.$$

Proposition 5 (Residual information difference)**.** *In the stratified affine REE,*

$$\Delta\tau^R = Q_{\mathrm{H}} - q_{2,\mathrm{L}} + \tau_{P,\mathrm{H}} - \tau_{P,\mathrm{L}} = \Phi^*(c_{\mathrm{H}}) - f_2(c_{\mathrm{L}})\tau_{s_2} + \tau_{P,\mathrm{H}} - \tau_{P,\mathrm{L}}. \tag{12}$$

Proposition 5 decomposes the post-price information difference into three pieces. The term $\Phi^*(c_{\mathrm{H}})$ is the direct precision generated by the high-compute investors from their private and public processing. The term $f_2(c_{\mathrm{L}})\tau_{s_2}$ is the low-compute investors' direct public precision. The pair $(\tau_{P,\mathrm{H}}, \tau_{P,\mathrm{L}})$ captures the bilateral learning channel through the equilibrium price.

5.5 Benchmark with no Private Information Acquisition

To isolate the role of heterogeneous public-data processing more sharply, consider the benchmark in which no investor acquires the private signal, either because the private source is unavailable or because parameters support a no-acquisition equilibrium. High-compute investors then devote all compute to the public signal. Accordingly, define

$$q^0_{2,\mathrm{H}} \equiv f_2(c_{\mathrm{H}})\tau_{s_2}, \qquad q_{2,\mathrm{L}} \equiv f_2(c_{\mathrm{L}})\tau_{s_2}.$$

Proposition 6 (No-acquisition benchmark)**.** *In the no-acquisition benchmark, let*

$$s^0_{2,\mathrm{H}} = \tilde{\theta} + \tilde{\eta}^0_{2,\mathrm{H}}, \qquad s_{2,\mathrm{L}} = \tilde{\theta} + \tilde{\eta}_{2,\mathrm{L}},$$

with $\tilde{\eta}^0_{2,\mathrm{H}} \sim \mathcal{N}(0, (q^0_{2,\mathrm{H}})^{-1})$ *and* $\tilde{\eta}_{2,\mathrm{L}} \sim \mathcal{N}(0, q^{-1}_{2,\mathrm{L}})$ *independent. There exists an affine rational expectations equilibrium with price*

$$\begin{aligned} P^0 =& \frac{\tau_0\bar{\theta} + \left[\lambda q^0_{2,\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}}\right] s^0_{2,\mathrm{H}} + \left[\lambda\tau^0_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}\right] s_{2,\mathrm{L}} - A\bar{z}}{\bar{\tau}^0} \\ &- \frac{A + \lambda\tau^0_{P,\mathrm{H}}\beta^0_{\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}}\beta^0_{\mathrm{L}}}{\bar{\tau}^0}\tilde{u}, \end{aligned} \tag{13}$$

where $(\beta_{\mathrm{H}}^0, \beta_{\mathrm{L}}^0)$ *is a positive solution to*

$$\beta_{\mathrm{H}}^0 = \frac{\dfrac{A}{1-\lambda} + \tau_{P,\mathrm{L}}^0 \beta_{\mathrm{L}}^0}{q_{2,\mathrm{L}}}, \qquad \beta_{\mathrm{L}}^0 = \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0}{q_{2,\mathrm{H}}^0},$$

with

$$\tau_{P,\mathrm{H}}^0 = \left(q_{2,\mathrm{L}}^{-1} + (\beta_{\mathrm{H}}^0)^2 \tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}}^0 = \left((q_{2,\mathrm{H}}^0)^{-1} + (\beta_{\mathrm{L}}^0)^2 \tau_u^{-1}\right)^{-1},$$

and

$$\bar{\tau}^0 = \tau_0 + \lambda q_{2,\mathrm{H}}^0 + \lambda \tau_{P,\mathrm{H}}^0 + (1-\lambda) q_{2,\mathrm{L}} + (1-\lambda)\tau_{P,\mathrm{L}}^0. \tag{14}$$

The post-price information difference is

$$\Delta\tau^0 = q_{2,\mathrm{H}}^0 - q_{2,\mathrm{L}} + \tau_{P,\mathrm{H}}^0 - \tau_{P,\mathrm{L}}^0. \tag{15}$$

Proposition 6 shows that the no-acquisition benchmark is not a degenerate case in which prices merely re-encode one common public statistic. Even when only public information is processed, the two investor types observe different reduced-form realizations and the equilibrium price transmits part of each realization to the other side of the market. Heterogeneity in compute still matters through the direct public precisions $(q_{2,\mathrm{H}}^0, q_{2,\mathrm{L}})$, but the price now partially pools these realizations rather than leaving both types' post-price information unchanged.

5.6 Power-Law Processing and the Compute Frontier

We next show that the power-law formulation used in the empirical analysis fits naturally inside our REE framework and can be seen as a tractable special case of our more general processing structure that yields a closed-form compute frontier.[14] The split formula below is purely technological. The frontier and forecast-performance implications that follow are

[14]Bounded saturating specifications can also be studied within the same general framework, but the standard power-law case delivers the sharpest formulas.

equilibrium statements only on parameter regions, or along sequences of economies, for which the stratified REE continues to obtain as c_{H} varies.

Proposition 7 (Power-law processing)**.** *Consider the power-law family*

$$f_k(c) = \kappa_k c^{\alpha_k}, \qquad \kappa_k > 0, \qquad \alpha_k \in (0,1).$$

If $\alpha_1 = \alpha_2 = \alpha$, then the optimal high-compute split is

$$c_{1,\mathrm{H}} = \omega c_{\mathrm{H}}, \qquad c_{2,\mathrm{H}} = (1-\omega)c_{\mathrm{H}},$$

where

$$\omega = \frac{m}{1+m}, \qquad m \equiv \left(\frac{\kappa_1 \tau_{s_1}}{\kappa_2 \tau_{s_2}}\right)^{\frac{1}{1-\alpha}}.$$

Note that the parameter assumption in Proposition 7 satisfies Assumption 1. When both transformation functions share a common scaling exponent, the optimal compute allocation takes a remarkably simple form: high-compute investors devote a constant fraction ω of their total budget to private processing regardless of their total compute budget and this share depends only on the relative signal quality ratio $\kappa_1\tau_{s_1}/\kappa_2\tau_{s_2}$ and the common exponent α. This scale-invariance of the allocation is the theoretical counterpart of the stable scaling exponents estimated in the empirical analysis, where the rate at which forecast performance improves with compute is approximately constant across model sizes.

Under Proposition 7, the high-compute investors' pooled sufficient-statistic precision scales as

$$Q_{\mathrm{H}}(c_{\mathrm{H}}) = q_{1,\mathrm{H}}(c_{\mathrm{H}}) + q_{2,\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha},$$

where

$$C_{\mathrm{H}} \equiv \kappa_1 \tau_{s_1} \omega^{\alpha} + \kappa_2 \tau_{s_2} (1-\omega)^{\alpha}.$$

The relevant revelation object is now the low-compute investors' recovery of that pooled

sufficient statistic through the price. Defining the low-side opacity gap by $\Omega_{\mathrm{L}}(c_{\mathrm{H}}) \equiv 1 - \Pi_{\mathrm{L}}(c_{\mathrm{H}})$, we can show

Corollary 2 (Power-law revelation of the high-compute sufficient statistic)**.** *Suppose the common-exponent power-law specification in Proposition 7 holds. Keep c_{L} fixed and consider any sequence $c_{\mathrm{H}} \to \infty$ along which the economy remains in the stratified affine REE. Then*

$$\Pi_{\mathrm{L}}(c_{\mathrm{H}}) \to 1 \qquad \textit{and} \qquad \Omega_{\mathrm{L}}(c_{\mathrm{H}}) = O(c_{\mathrm{H}}^{-\alpha}) \qquad \textit{as } c_{\mathrm{H}} \to \infty. \tag{16}$$

Corollary 2 shows that, along stratified power-law sequences, low-compute investors asymptotically recover almost all of the high-compute investors' pooled sufficient statistic from the price. The remaining opacity decays at the same power-law rate as the reciprocal of high-compute precision. Asymptotically, the low-compute investors do not merely learn an increasing amount about the private component, they learn an increasing amount about the high-compute side's entire pooled statistic.

To connect the REE directly to forecast performance, define the post-price RMSE of each type's posterior mean forecast by

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) \equiv \left(\mathbb{E}\left[(\tilde{\theta} - \mu_{\mathrm{H}}^{R})^2\right]\right)^{1/2} = \left(\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})\right)^{-1/2},$$

and

$$\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \equiv \left(\mathbb{E}\left[(\tilde{\theta} - \mu_{\mathrm{L}}^{R})^2\right]\right)^{1/2} = \left(\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})\right)^{-1/2}.$$

Because the posterior precisions are constant for a given investor type, these RMSEs coincide with the corresponding conditional posterior standard deviations within any fixed stratified equilibrium.

Corollary 3 (Power-law RMSE frontier along stratified equilibria)**.** *Suppose the common-exponent power-law specification in Proposition 7 holds and the economy is in the stratified*

REE. Define

$$Q_{\mathrm{H}}(c_{\mathrm{H}}) \equiv C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}.$$

Then

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) = \left(\tau_0 + Q_{\mathrm{H}}(c_{\mathrm{H}}) + \tau_{P,\mathrm{H}}(c_{\mathrm{H}})\right)^{-1/2},$$

and

$$\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) = \left(\tau_0 + f_2(c_{\mathrm{L}})\tau_{s_2} + \tau_{P,\mathrm{L}}(c_{\mathrm{H}})\right)^{-1/2}.$$

If, in addition, c_{L} is fixed and $c_{\mathrm{H}} \to \infty$ along any sequence on which the stratified REE continues to obtain, then

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}, \qquad \mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2},$$

and

$$\frac{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})} \to 1 \qquad \textit{as } c_{\mathrm{H}} \to \infty.$$

Corollary 3 compresses the finite- and large-compute implications into a single frontier statement. Within any fixed stratified equilibrium, each type's RMSE is pinned down exactly by the inverse square root of its posterior precision. Along stratified power-law sequences, the high-compute side's direct precision $Q_{\mathrm{H}}(c_{\mathrm{H}})$ dominates the bounded extra information that high-compute investors learn from the low-compute public draw, while Corollary 2 implies that low-compute investors asymptotically recover that same pooled precision through the price. Consequently both types inherit the same RMSE exponent $\alpha/2$ and the asymptotic level gap disappears.

To connect the same power-law structure to trading performance, define the conditional Sharpe ratio of each type's optimal risky position by

$$\mathrm{SR}_{\mathrm{H}} \equiv \frac{\mathbb{E}\left[x_{\mathrm{H}}(\tilde{\theta} - P) \mid s_1, s_{2,\mathrm{H}}, P\right]}{\sqrt{\mathbb{V}\left[x_{\mathrm{H}}(\tilde{\theta} - P) \mid s_1, s_{2,\mathrm{H}}, P\right]}},$$

and

$$\mathrm{SR}_{\mathrm{L}}^{R} \equiv \frac{\mathbb{E}\big[x_{\mathrm{L}}(\tilde{\theta}-P) \mid s_{2,\mathrm{L}}, P\big]}{\sqrt{\mathbb{V}\big[x_{\mathrm{L}}(\tilde{\theta}-P) \mid s_{2,\mathrm{L}}, P\big]}}.$$

When the optimal risky position is zero, we set the corresponding conditional Sharpe ratio equal to zero. Because the investor chooses the sign of the position optimally, these conditional Sharpe ratios are nonnegative.

The raw Sharpe ratios remain state-dependent because they depend both on posterior precision and on the realized expected excess return term $|\mu - P|$. The deterministic compute component is the precision scale multiplying that realized price wedge, so define

$$\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}}) \equiv \sqrt{\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})} = \frac{1}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})},$$

and

$$\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \equiv \sqrt{\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})} = \frac{1}{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}.$$

Proposition 8 (Conditional Sharpe ratios and power-law scales)**.** *In the stratified rational expectations equilibrium,*

$$\mathrm{SR}_{\mathrm{H}} = \mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})\,|\mu_{\mathrm{H}}^{R} - P| = \frac{A|x_{\mathrm{H}}|}{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})},$$

and

$$\mathrm{SR}_{\mathrm{L}}^{R} = \mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})\,|\mu_{\mathrm{L}}^{R} - P| = \frac{A|x_{\mathrm{L}}|}{\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}.$$

If, in addition, the common-exponent power-law specification in Proposition 7 holds so that $Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}$*,* c_{L} *is fixed, and* $c_{\mathrm{H}} \to \infty$ *along any sequence on which the stratified REE continues to obtain, then*

$$\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}}) \sim C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2}, \qquad \mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2},$$

and

$$\frac{\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})} \to 1 \qquad \text{as } c_{\mathrm{H}} \to \infty.$$

The exact identities in Proposition 8 show that a higher Sharpe ratio for the high-compute investors requires more than a larger posterior precision: it also depends on the realized distance between each type's posterior mean and the equilibrium price. Because μ_{H}^{R}, μ_{L}^{R}, and P depend on realized signals, the next result is an ex-post, state-by-state comparison after information has been observed and the price has been revealed.

Proposition 9 (High compute investor Sharpe-ratio dominance)**.** *Let*

$$\Delta\mu^{R} \equiv \mu_{\mathrm{H}}^{R} - \mu_{\mathrm{L}}^{R}, \qquad \Delta\tau^{R} \equiv \tau_{\mathrm{H}}^{R} - \tau_{\mathrm{L}}^{R} = Q_{\mathrm{H}} - q_{2,\mathrm{L}} + \tau_{P,\mathrm{H}} - \tau_{P,\mathrm{L}}.$$

If

$$\Delta\tau^{R} \geq 0, \qquad (\mu_{\mathrm{L}}^{R} - P)\Delta\mu^{R} \geq 0, \qquad \Delta\mu^{R} \neq 0,$$

then

$$\mathrm{SR}_{\mathrm{H}} > \mathrm{SR}_{\mathrm{L}}^{R}.$$

Proposition 9 has a simple interpretation. The first condition says that, in the realized state under consideration, the high-compute investors have at least as much total post-price precision as the low-compute investors. The second condition says that realized posterior disagreement pushes the high-compute posterior mean farther away from the price in the same direction as the low-compute investors' own expected excess return. When both conditions hold and the disagreement is nondegenerate, the high-compute side obtains the higher conditional Sharpe ratio in that realized state.

6. Conclusion

We examine the role of compute for return predictability. Our analysis shows that scaling (power) laws, which govern the relation between compute and forecasting performance across a broad range of sciences, carry over to predictability of the cross-section of stock returns. We measure complexity using compute rather than raw parameter counts, which are not commensurable across model classes.

We propose a simple asset pricing model in which compute is a primitive resource that determines the precision gained from processing private or public information. Under concavity of the mapping from compute to posterior precision, the empirical power law relating compute to forecast accuracy is consistent with a partially revealing rational expectations equilibrium. In this equilibrium, heterogeneous access to compute drives price informativeness, with low-compute investors inferring a noisy signal correlated with high-compute investors' private information from the equilibrium price.

Low compute exposes modelers to local optima—what we call "training" or "processing" risk. Larger compute largely resolves this issue, at least for the information sets considered here. At the opposite end of the spectrum, the estimated scaling law bounds the limits to return predictability and quantifies the rate at which additional compute yields further gains in economic and statistical forecasting performance for a given dataset.

The slope of the power law measures the marginal value of additional compute. We propose and test how this slope depends on the underlying information set, finding empirically that low-compute investors are better served by smaller information sets, while high-compute investors benefit from larger ones.

Our results highlight a tension between modern machine learning and the tightly parameterized canonical asset pricing models. The capital asset pricing model is often described as having "one parameter"—the loading β on the market portfolio—but this description obscures the substantial structure embedded in the model. The CAPM imposes a fully specified

functional form, a particular choice of state variables, and a collection of strong identifying assumptions to empirically use the model—linearity, exogeneity, homoskedasticity, and serially uncorrelated disturbances—that together underpin the Gauss–Markov theorem. Its apparent parsimony thus reflects restrictions imposed ex ante and is suppressed when complexity is proxied by parameter counts.[15]

Modern deep learning models, by contrast, impose relatively little ex ante structure. A deep neural network defines a highly flexible function class subject only to weak regularity conditions, as formalized by universal approximation results such as Hornik et al. (1989), without presuming a linear factor structure or the orthogonality and homoskedasticity conditions underlying Gauss–Markov. Instead, it expends computational resources to search over a high-dimensional function space and approximate the conditional expectation. Consequently, even if the true model were exactly the CAPM, a generic neural network would not recover it with a one-dimensional parameter vector: the parsimony of the CAPM derives from restrictions imposed a priori, whereas in deep learning analogous structure must be learned from the data.

We view this proliferation of parameters not as a defect, but as the cost of relaxing strong model-specification assumptions—complexity substitutes for the parametric structure hard-coded in canonical asset pricing models. Modern machine learning replaces the assumption of a correctly specified model with an acknowledgment that the true specification is unknown but can be approximated arbitrarily well with sufficient compute. Because this trade-off is mediated through computational effort rather than nominal parameter counts (Belkin et al., 2019, Kaplan et al., 2020, Henighan et al., 2020), compute is the natural primitive for characterizing scaling laws in finance and economics.

[15]This perspective connects to a long-standing distinction between the data-modeling and algorithmic-modeling cultures in statistics Breiman (2001).

References

Admati, Anat R, 1985, A noisy rational expectations equilibrium for multi-asset securities markets, *Econometrica: Journal of the Econometric Society* 629–657.

Banerjee, Snehal, Ron Kaniel, and Ilan Kremer, 2009, Price drift as an outcome of differences in higher-order beliefs, *The Review of Financial Studies* 22, 3707–3734.

Belkin, Mikhail, Daniel Hsu, Siyuan Ma, and Soumik Mandal, 2019, Reconciling modern machine-learning practice and the classical bias–variance trade-off, *Proceedings of the National Academy of Sciences* 116, 15849–15854.

Breiman, Leo, 2001, Statistical modeling: The two cultures (with comments and a rejoinder by the author), *Statistical science* 16, 199–231.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2024, Deep learning in asset pricing, *Management Science* 70, 714–750.

Cochrane, John H, and Jesus Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of political economy* 108, 79–119.

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

Dugast, Jérôme, and Thierry Foucault, 2018, Data abundance and asset price informativeness, *Journal of Financial economics* 130, 367–391.

Farboodi, Maryam, and Laura Veldkamp, 2020, Long-run growth of financial data technology, *American Economic Review* 110, 2485–2523.

Farboodi, Maryam, and Laura Veldkamp, 2021, A model of the data economy, Technical report, National Bureau of Economic Research Cambridge, MA, USA.

Glosten, Lawrence R, and Paul R Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of financial economics* 14, 71–100.

Goulet Coulombe, Philippe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant, 2022, How is machine learning useful for macroeconomic forecasting?, *Journal of Applied Econometrics* 37, 920–964.

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

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

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of political economy* 99, 225–262.

Hansen, Peter Reinhard, 2005, A test for superior predictive ability, *Journal of Business & Economic Statistics* 23, 365–380.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Heaton, James B, Nick G Polson, and Jan Hendrik Witte, 2017, Deep learning for finance: deep portfolios, *Applied Stochastic Models in Business and Industry* 33, 3–12.

Hellwig, Martin F, 1980, On the aggregation of information in competitive markets, *Journal of economic theory* 22, 477–498.

Henighan, Tom, Jared Kaplan, Mor Katz, Mark Chen, Christopher Hesse, Jacob Jackson, Heewoo Jun, Tom B Brown, Prafulla Dhariwal, Scott Gray, et al., 2020, Scaling laws for autoregressive generative modeling, *arXiv preprint arXiv:2010.14701* .

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White, 1989, Multilayer feedforward networks are universal approximators, *Neural networks* 2, 359–366.

Huang, Dashan, and Guofu Zhou, 2017, Upper bounds on return predictability, *Journal of Financial and Quantitative Analysis* 52, 401–425.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, 2021, Is there a replication crisis in finance?, Technical report, National Bureau of Economic Research.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A rational theory of mutual funds' attention allocation, *Econometrica* 84, 571–626.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, 2020, Scaling laws for neural language models, *arXiv preprint arXiv:2001.08361* .

Kelly, Bryan, Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu, 2024a, Large (and deep) factor models, *arXiv preprint arXiv:2402.06635* .

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, 2024b, The virtue of complexity in return prediction, *The Journal of Finance* 79, 459–503.

Kelly, Bryan T, Semyon Malamud, and Kangying Zhou, 2022, The virtue of complexity everywhere, *Swiss Finance Institute Research Paper* .

Kyle, Albert S, 1985, Continuous auctions and insider trading, *Econometrica: Journal of the Econometric Society* 1315–1335.

Li, Hao, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein, 2018, Visualizing the loss landscape of neural nets, *Advances in neural information processing systems* 31.

Lo, Andrew W, and A Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431–467.

Mondria, Jordi, 2010, Portfolio choice, attention allocation, and price comovement, *Journal of Economic Theory* 145, 1837–1864.

Nagel, Stefan, 2025, Seemingly virtuous complexity in return prediction, *Chicago Booth Research Paper* .

Nakkiran, Preetum, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, and Ilya Sutskever, 2021, Deep double descent: Where bigger models and more data hurt, *Journal of Statistical Mechanics: Theory and Experiment* 2021, 124003.

Potì, Valerio, 2018, A new tight and general bound on return predictability, *Economics Letters* 162, 140–145.

Sims, Christopher A, 2003, Implications of rational inattention, *Journal of monetary Economics* 50, 665–690.

Sirignano, Justin, and Rama Cont, 2021, Universal features of price formation in financial markets: perspectives from deep learning, in *Machine learning and AI in finance*, 5–15 (Routledge).

Sullivan, Ryan, Allan Timmermann, and Halbert White, 1999, Data-snooping, technical trading rule performance, and the bootstrap, *The journal of Finance* 54, 1647–1691.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle, *The Journal of Finance* 64, 1187–1215.

Veldkamp, Laura L, 2006, Information markets and the comovement of asset prices, *The Review of Economic Studies* 73, 823–845.

Vives, Xavier, 2010, *Information and learning in markets: the impact of market microstructure* (Princeton University Press).

White, Halbert, 2000, A reality check for data snooping, *Econometrica* 68, 1097–1126.

Zhou, Guofu, 2010, How much stock return predictability can we expect from an asset pricing model?, *Economics Letters* 108, 184–186.

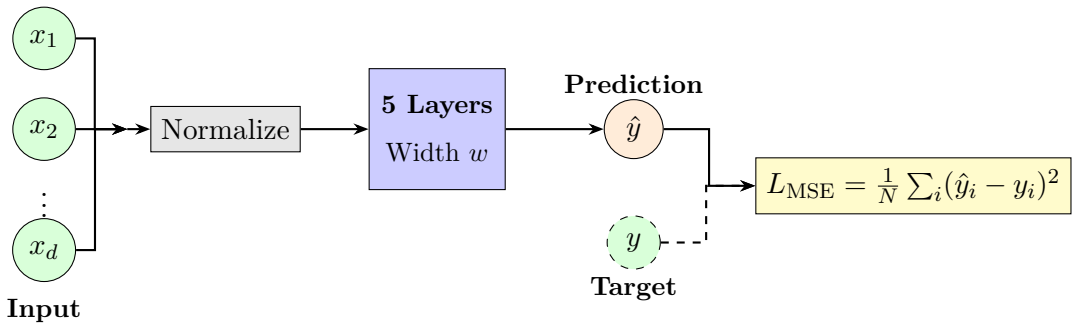
Internet Appendix

A. Neural Network Architecture & Training

In this section, we explain how we approximate the unknown functional mapping from the high-dimensional information set x_t to expected returns, $g(x_t)$.

Building on the universal approximators (Hornik et al. (1989)) theorem, we use neural nets as the class of models for this mapping. Specifically, we approximate the conditional mean function using feedforward neural networks of varying complexity. The architecture and training procedure are designed to isolate the effect of model capacity on forecasting performance while maintaining consistency across the wide range of model sizes required to estimate scaling laws. We provide details of the construction, training, and evaluation of these networks.

Figure A1: Mean Squared Error Neural Network Architecture



Note. Neural network architecture for return prediction. Input features $\mathbf{x}_t \in \mathbb{R}^d$ are first normalized to zero mean and unit variance, then processed through five hidden layers of equal width w. Each hidden layer applies a linear transformation (Dense), ReLU activation, and Layer Normalization. Dropout (rate 0.1) is applied to the middle three hidden layers only. The output layer produces a scalar prediction $\hat{y}_{t+h}$. The network is trained by minimizing mean squared error (MSE) between predictions and realized outcomes y_{t+h}. Model complexity is controlled by varying the width w, which determines the total parameter count.

A.1 Network architecture

Figure A1 provides a schematic overview of the network architecture for neural networks trained on mean squared error. Our neural network architecture follows a fixed-depth design with five hidden layers of equal width. This architectural choice serves two purposes. First,

holding depth constant while varying width isolates the effect of parameter count on model performance, avoiding confounding variation in network depth. Second, equal-width layers provide smoother transitions across parameter counts compared to tapered architectures which exhibit discrete jumps in effective capacity as layers are added or removed.

Given a target parameter count, we determine the layer width through binary search. For an input dimension of d features, five hidden layers of width w, and a scalar output, the total parameter count equals $dw + 4w^2 + w$ when counting only weight matrices. We search for the width w that matches the target parameter count within one percent, ensuring that models across the parameter range are comparable in their depth-to-width ratio.

Input features are first passed through a normalization layer that standardizes each feature to zero mean and unit variance using statistics computed from the training sample. This input normalization is critical for stable optimization across the wide range of model sizes we consider, as it ensures that gradient magnitudes remain comparable regardless of the original feature scales.

Each hidden layer consists of a linear transformation followed by a rectified linear unit activation function and layer normalization. We apply layer normalization rather than batch normalization because its statistics are computed per-sample rather than per-batch, making it more stable when batch sizes are large relative to the effective degrees of freedom in smaller models. The output layer applies a linear transformation without activation to produce a scalar prediction.

To reduce extreme double-descent fluctuations, we apply dropout at rates of 10% or 20%, depending on the application. For firm-level prediction tasks, we use a 10% dropout rate because the larger number of observations provides greater implicit regularization. Dropout is applied only to the middle three hidden layers, not the first or last, preserving the early-layer input representation and the late-layer output mapping while regularizing intermediate representations where overfitting is most likely. We initialize weights with He normal initialization to account for the variance scaling induced by ReLU activations, and we initialize all

biases to zero.

A.2 Training procedure

We train all networks by minimizing mean squared error using the Adam optimizer with an initial learning rate of 0.001. To prevent gradient explosion which can occur when training very large models on noisy financial data, we clip gradient norms at unity before each parameter update. This gradient clipping ensures stable training dynamics across the full range of model sizes without requiring size-specific hyperparameter tuning.

The learning rate is adjusted during training using a plateau-based scheduler. When validation loss fails to improve for a patience window equal to one-fifth of the total training epochs, the learning rate is reduced by half. The plateau scheduler allows the optimizer to escape shallow local minima early in training while enabling fine-grained convergence in later epochs.

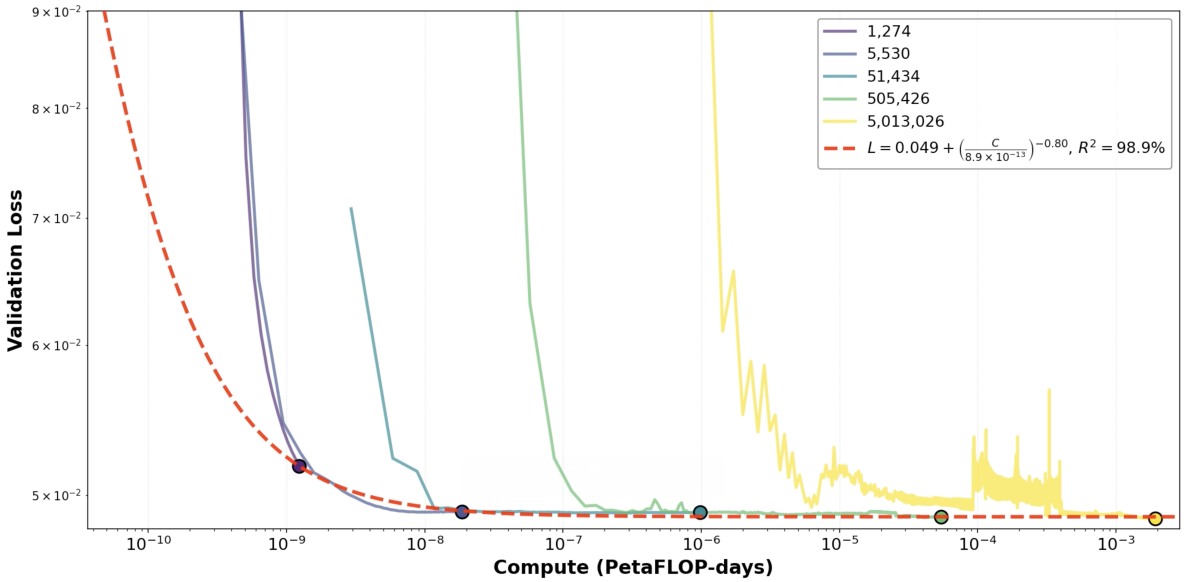
Training duration scales with model size to ensure that larger models have sufficient time to traverse the interpolation threshold and enter the double-descent regime. Specifically, we train firm-level return models for a number of epochs proportional to the parameter count raised to the 0.75 power. This sublinear scaling reflects the empirical observation that larger models require more epochs to converge, with diminishing marginal returns to additional training. For the firm-level return prediction application, we use $10 + \lfloor 0.1 \times p^{0.75} \rfloor$ epochs.

Our scaling choices reflect a practical trade-off: provided models are trained sufficiently long, we can still recover scaling laws, but excessively long training is computationally costly for the full set of experiments in this paper. We therefore set training schedules at levels where performance begins to asymptote across model sizes, without incurring unnecessary additional computation.

We deliberately do not employ early stopping based on validation loss. While early stopping can prevent overfitting in the classical bias-variance regime, it would prevent our models from passing through the interpolation threshold into the double-descent regime

where test loss decreases again. This can be empirically seen in Figure A2. If one were to implement early stopping, the 5M parameter model would not be able to achieve the asymptotic loss it could achieve. This is analogous to the stylized effect drawn in Figure 1. By training to completion, we allow sufficiently large models to benefit from the implicit regularization that emerges in the overparameterized regime.

Figure A2: Double Descent



Note. This figure plots validation MSE against cumulative compute (PetaFLOP-days) for a subset of firm-level return prediction networks. Colored lines trace validation loss over training for five model sizes (parameter counts in the legend); circled markers indicate each run's best validation loss, and the dashed red curve is the fitted scaling law. The non-monotone trajectories illustrate double descent around the interpolation regimes. Over-parameterized models must be trained sufficiently long enough (no early stopping) to pass the interpolation region and exploit implicit regularization that yields lower and more stable validation loss.

The batch size is fixed at 65,536 observations for both applications. This large batch size serves multiple purposes. It reduces gradient variance, which is particularly important given the high noise-to-signal ratio in financial return prediction. It also improves GPU throughput by maximizing parallelism in matrix operations. Most importantly, it ensures that the scaling laws we estimate reflect model capacity rather than optimization noise, as smaller batch sizes would introduce additional stochasticity that could obscure the relationship between compute and performance.

All models use 32-bit floating point precision with deterministic operations enabled. We fix random seeds for weight initialization, dropout masks, and data shuffling to ensure reproducibility. These choices prioritize consistency and reproducibility over computational efficiency, as our goal is to characterize the fundamental relationship between model capacity and forecasting performance rather than to minimize training time.

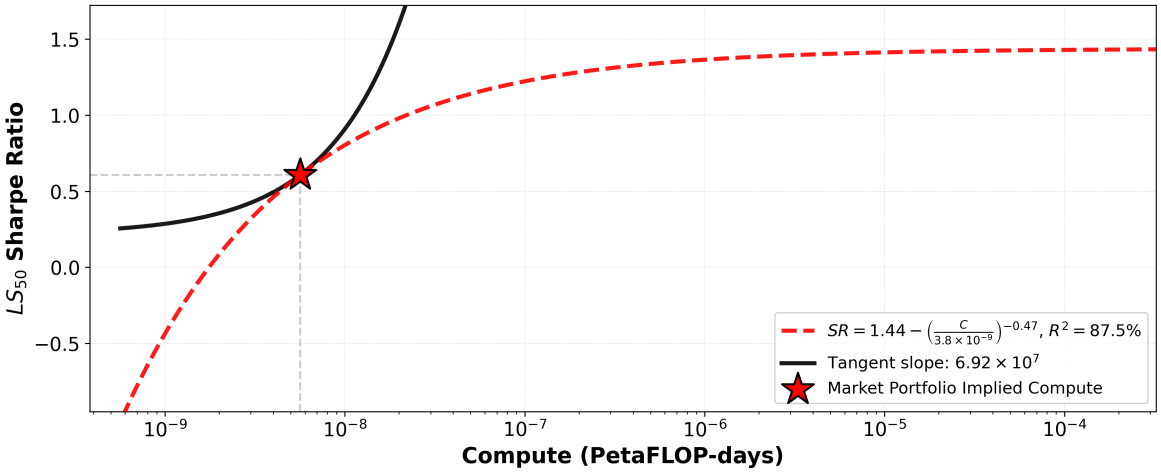
A.3 Compute measurement

To construct scaling laws, we measure computational cost in PetaFLOP-days, following the convention established in the machine learning literature. For each model, we estimate the floating point operations per training epoch as six times the parameter count times the number of training observations. This approximation accounts for the forward pass, backward pass, and parameter update, each of which requires approximately two floating point operations per parameter per observation. Cumulative compute is then computed as the sum of per-epoch compute across all training epochs, converted to PetaFLOP-days by dividing by 8.64×10^{19} operations per PetaFLOP-day.

This compute measure provides a hardware-agnostic way to compare models of different sizes and enables comparison with scaling laws estimated in other domains such as natural language processing and computer vision. By expressing performance as a function of compute rather than raw parameter count, we account for the fact that larger models require not only more parameters but also more training iterations to reach their asymptotic performance.

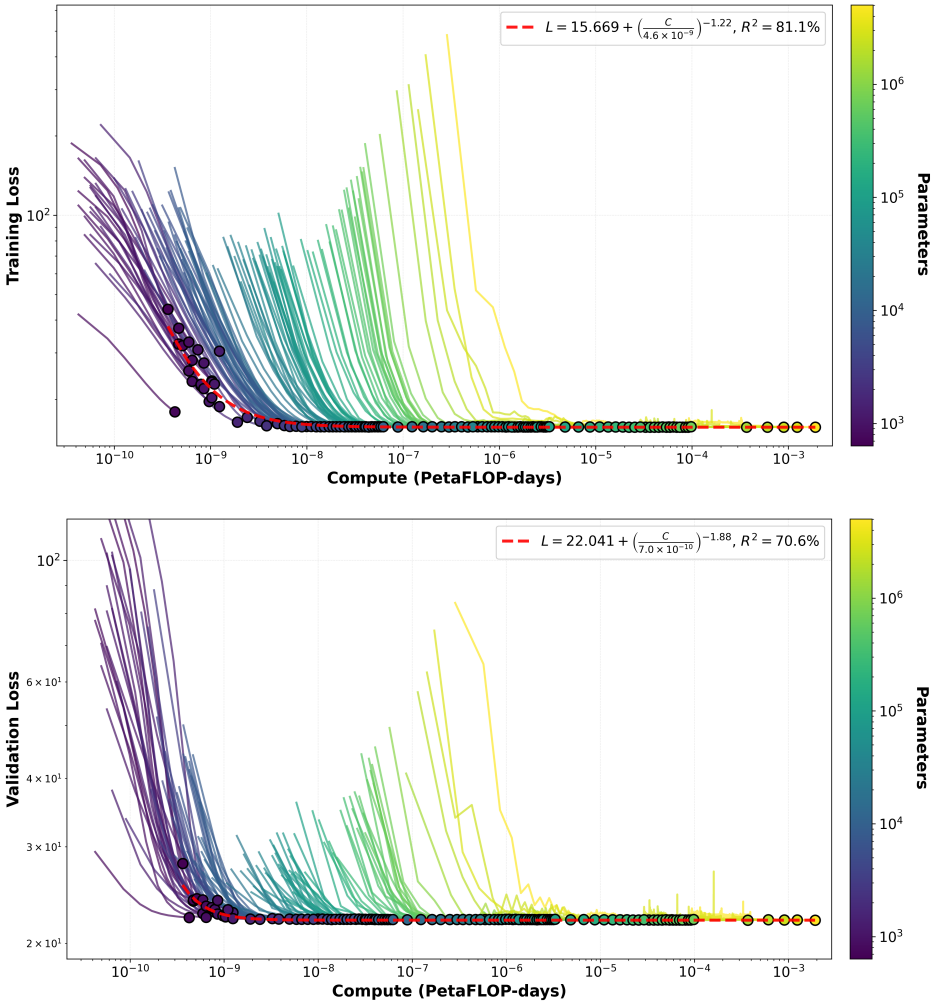
B. Additional Figures & Tables

Figure A3: Market Efficiency Example (1994-2024)



Note. This figure reports an example of how to estimate market efficiency from the scaling laws made with the GFD-64 data. The market buy and hold Sharpe ratio is denoted as a red star (Sharpe ratio is calculated from 1994 to 2024), which implies a level of market average compute. We then use this implied market compute to find the tangency line of the scaling law (black solid curve in log-linear space).

Figure A4: Scaling Laws in Training and Validation



Note. This figure plots training and validation RMSE against cumulative compute (PetaFLOP-days). The non-monotone trajectories illustrate double descent around the interpolation regimes. Over-parameterized models must be trained sufficiently long enough (no early stopping) to pass the interpolation region and exploit implicit regularization that yields lower and more stable validation loss.

Table A1: Public statements emphasizing compute and/or data scale.

Firm (linked)	**Public statement (verbatim excerpt)**
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**."
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**."
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"
D. E. Shaw	"... designs and maintains the firm's **largest compute infrastructure** ... **high-performance computing**, networking, and storage for research and trading."

Table A2: Estimated Scaling Laws by Firm-Level Forecast Decile

Decile	Equation	R^2 (%)
1	$0.03 + \left(\frac{C}{5.0 \times 10^{-10}}\right)^{-0.72}$	61.8
2	$0.05 + \left(\frac{C}{3.6 \times 10^{-10}}\right)^{-0.50}$	78.5
3	$0.22 + \left(\frac{C}{1.2 \times 10^{-10}}\right)^{-0.46}$	67.8
4	$0.43 + \left(\frac{C}{1.9 \times 10^{-11}}\right)^{-0.47}$	34.3
5	$0.56 - \left(\frac{C}{7.6 \times 10^{-13}}\right)^{-0.50}$	1.6
6	$0.67 - \left(\frac{C}{1.6 \times 10^{-11}}\right)^{-0.41}$	38.1
7	$0.74 - \left(\frac{C}{3.4 \times 10^{-11}}\right)^{-0.40}$	63.3
8	$0.74 - \left(\frac{C}{1.7 \times 10^{-10}}\right)^{-0.69}$	66.4
9	$0.78 - \left(\frac{C}{2.7 \times 10^{-10}}\right)^{-0.77}$	63.8
10	$0.99 - \left(\frac{C}{4.1 \times 10^{-10}}\right)^{-0.57}$	70.9
LS_{50}	$1.44 - \left(\frac{C}{3.8 \times 10^{-9}}\right)^{-0.47}$	87.5
LS_{30}	$1.52 - \left(\frac{C}{3.8 \times 10^{-9}}\right)^{-0.51}$	87.0
LS_{10}	$1.69 - \left(\frac{C}{4.2 \times 10^{-9}}\right)^{-0.50}$	83.0

Note. This table reports the estimated scaling-law fits $a - (C/b)^{\gamma}$ for each decile constructed on the firm-level forecasted return. The forecasts are generated with neural networks trained on the GFD-64 characteristics. The last three rows represent long-short portfolios. They are constructed by going long and short on specific deciles, for instance, LS_{10} goes long on decile 10 and short on decile 1. The scaling laws' R^2 values are reported in percent.

C. Proofs

Throughout the appendix, for any total compute budget c, write

$$H(c_1; c) \equiv f_1(c_1)\tau_{s_1} + f_2(c - c_1)\tau_{s_2}, \qquad 0 \leq c_1 \leq c.$$

This is the effective signal precision generated by allocating c_1 units of compute to the private signal and the remaining $c - c_1$ units to the public signal.

C.1 Expected certainty equivalent and posterior precision

Lemma 9.1 (Expected certainty equivalent and posterior precision)**.** *Fix an anticipated price law and let a price-taking investor have posterior distribution*

$$\tilde{\theta} \mid \mathcal{I} \sim \mathcal{N}(\mu, \tau^{-1}).$$

Then the investor's expected certainty equivalent under optimal trading is

$$\mathbb{E}[\mathrm{CE}] = W_0 - \frac{1}{2A} + \Xi\tau, \tag{17}$$

where

$$\Xi \equiv \frac{\mathbb{E}[(\tilde{\theta} - P)^2]}{2A} > 0.$$

In particular, for a fixed anticipated price law, expected certainty equivalent is strictly increasing in posterior precision. This is the acquisition-stage objective used in the main text, because the fixed private-data cost κ is measured in the same certainty-equivalent wealth units.

Proof. Suppose the investor buys x units of the risky asset at price P. Terminal wealth is

$$W = W_0 + x(\tilde{\theta} - P).$$

Conditional on the investor's information set $\mathcal{I}$, the random payoff $x(\tilde{\theta} - P)$ is Gaussian with mean $x(\mu - P)$ and variance $x^2\tau^{-1}$. With CARA utility, expected utility is therefore maximized by maximizing the certainty equivalent

$$\mathrm{CE}(x) = W_0 + x(\mu - P) - \frac{A}{2}x^2\tau^{-1}.$$

Differentiating with respect to x gives the first-order condition

$$\mu - P - Ax\tau^{-1} = 0,$$

so the optimal demand is

$$x^* = \frac{\tau(\mu - P)}{A}.$$

Substituting x^* back into the certainty equivalent yields

$$\mathrm{CE}^* = W_0 + \frac{\tau(\mu - P)^2}{2A}.$$

Take expectations. Because $\mu = \mathbb{E}[\tilde{\theta} \mid \mathcal{I}]$, the posterior error $\tilde{\theta} - \mu$ is orthogonal to every $\mathcal{I}$-measurable variable, and in particular to $\mu - P$. Hence

$$\mathbb{E}\Big[(\tilde{\theta} - P)^2\Big] = \mathbb{E}\Big[(\mu - P)^2\Big] + \mathbb{E}\Big[(\tilde{\theta} - \mu)^2\Big].$$

The posterior variance is τ^{-1}, so

$$\mathbb{E}\Big[(\mu - P)^2\Big] = \mathbb{E}\Big[(\tilde{\theta} - P)^2\Big] - \tau^{-1}.$$

Substituting into the expected certainty equivalent gives

$$\mathbb{E}[\mathrm{CE}^*] = W_0 + \frac{\tau}{2A}\left(\mathbb{E}\Big[(\tilde{\theta} - P)^2\Big] - \tau^{-1}\right) = W_0 - \frac{1}{2A} + \frac{\mathbb{E}\Big[(\tilde{\theta} - P)^2\Big]}{2A}\tau.$$

This is exactly (17). Since $\mathbb{E}[(\tilde{\theta} - P)^2] > 0$ in a nondegenerate equilibrium, $\Xi > 0$, and expected certainty equivalent is strictly increasing in posterior precision. ■

C.2 Proof of Proposition 1

Proof. Consider an investor with total compute budget c who has decided to acquire the private signal. If the investor allocates c_1 units of compute to the private signal and $c - c_1$ units of compute to the public signal, then in the acquisition-stage problem the investor treats the anticipated price-learning term $\tau_{P,H}$ as fixed. The post-price posterior precision is therefore

$$\tau^H(c_1, c - c_1; \tau_{P,H}) = \tau_0 + H(c_1; c) + \tau_{P,H}.$$

By Lemma 9.1, expected certainty equivalent is strictly increasing in posterior precision. Since both τ_0 and $\tau_{P,H}$ are constant with respect to the choice of c_1, the investor solves

$$\max_{0 \le c_1 \le c} H(c_1; c) = \max_{0 \le c_1 \le c} \left\{ f_1(c_1)\tau_{s_1} + f_2(c - c_1)\tau_{s_2} \right\}.$$

This proves the first claim.

The function $H(c_1; c)$ is continuous on the compact interval $[0, c]$, so a maximizer exists. Any interior maximizer must satisfy the first-order condition

$$\frac{\partial H(c_1; c)}{\partial c_1} = f_1'(c_1)\tau_{s_1} - f_2'(c - c_1)\tau_{s_2} = 0,$$

which is exactly

$$f_1'(c_1^*(c))\tau_{s_1} = f_2'(c - c_1^*(c))\tau_{s_2}.$$

This proves the interior first-order condition.

Now assume f_1 and f_2 are strictly concave. Then $H(c_1; c)$ is strictly concave in c_1, so the maximizer is unique. To study how the maximizer varies with c, note that

$$\frac{\partial^2 H(c_1; c)}{\partial c_1 \partial c} = -f_2''(c - c_1)\tau_{s_2} \ge 0.$$

Thus $H(c_1; c)$ has increasing differences in (c_1, c). Standard monotone comparative statics

therefore implies that the unique maximizer $c_1^*(c)$ is nondecreasing in c.

On any interval where the solution is interior, differentiate the first-order condition with respect to c:

$$f_1''(c_1^*(c))\tau_{s_1}\, c_1^{*\prime}(c) = f_2''(c - c_1^*(c))\tau_{s_2}\Big(1 - c_1^{*\prime}(c)\Big).$$

Solving for $c_1^{*\prime}(c)$ yields

$$c_1^{*\prime}(c) = \frac{f_2''(c - c_1^*(c))\tau_{s_2}}{f_1''(c_1^*(c))\tau_{s_1} + f_2''(c - c_1^*(c))\tau_{s_2}}.$$

Because both second derivatives are strictly negative, the numerator and denominator are both negative. Moreover, the denominator is more negative than the numerator. Hence

$$0 < c_1^{*\prime}(c) < 1.$$

This completes the proof. ■

C.3 Proof of Proposition 2

Proof. If the investor acquires the private signal and allocates compute optimally, then the direct effective precision is $\Phi^*(c)$. In the acquisition-stage problem, the investor also anticipates the post-price learning term $\tau_{P,H}$. Therefore the post-price posterior precision from acquiring is

$$\tau_0 + \Phi^*(c) + \tau_{P,H}.$$

If the investor does not acquire the private signal, then all compute is devoted to the public source and the direct precision is $f_2(c)\tau_{s_2}$. In that case the investor anticipates the post-price learning term $\tau_{P,L}$, so the post-price posterior precision from not acquiring is

$$\tau_0 + f_2(c)\tau_{s_2} + \tau_{P,L}.$$

Because the fixed access cost κ is measured in certainty-equivalent wealth units, Lemma 9.1 implies that the expected certainty equivalent from acquiring is

$$W_0 - \frac{1}{2A} + \Xi\Big[\tau_0 + \Phi^*(c) + \tau_{P,H}\Big] - \kappa,$$

while the expected certainty equivalent from not acquiring is

$$W_0 - \frac{1}{2A} + \Xi\Big[\tau_0 + f_2(c)\tau_{s_2} + \tau_{P,L}\Big].$$

Subtract the second expression from the first. The net gain from acquiring is

$$\Xi\Big[\Phi^*(c) - f_2(c)\tau_{s_2} + \tau_{P,H} - \tau_{P,L}\Big] - \kappa.$$

Using the definition of $G(c)$, this becomes

$$\Xi\Big[G(c) + \tau_{P,H} - \tau_{P,L}\Big] - \kappa.$$

Therefore the investor acquires the private signal if and only if

$$\kappa < \Xi\Big[G(c) + \tau_{P,H} - \tau_{P,L}\Big].$$

If the bracketed term is nonpositive, acquisition is never optimal for any nonnegative fixed cost. This is exactly captured by the positive-part threshold

$$\kappa^*(c; \tau_{P,H}, \tau_{P,L}) = \Xi\Big[G(c) + \tau_{P,H} - \tau_{P,L}\Big]_+.$$

This proves the threshold characterization.

It remains to show that $G(c)$ is nondecreasing in c. Write

$$G(c) = \max_{0 \le c_1 \le c} \left\{ f_1(c_1)\tau_{s_1} + \Big[f_2(c - c_1) - f_2(c) \Big] \tau_{s_2} \right\}.$$

Fix any feasible c_1. The derivative of the expression inside braces with respect to c is

$$\Big[f_2'(c - c_1) - f_2'(c) \Big] \tau_{s_2}.$$

Because f_2 is concave, its derivative is nonincreasing. Since $c - c_1 \le c$, we have $f_2'(c - c_1) \ge f_2'(c)$. Therefore the derivative above is nonnegative. This shows that, for every fixed feasible c_1, the expression inside braces is nondecreasing in c. The pointwise maximum over c_1 is therefore nondecreasing as well. Hence $G(c)$ is nondecreasing.

If f_2 is strictly concave and some optimal allocation satisfies $c_1^*(c) > 0$, then $c - c_1^*(c) < c$, so strict concavity implies

$$f_2'(c - c_1^*(c)) > f_2'(c).$$

Then the derivative of the maximizing expression is strictly positive at that allocation, so $G(c)$ is strictly increasing on that interval. Since $\Xi > 0$ and the positive-part operator preserves monotonicity, $\kappa^*(c; \tau_{P,H}, \tau_{P,L})$ is nondecreasing in c. Moreover, on any interval where some optimal allocation satisfies $c_1^*(c) > 0$ and

$$G(c) + \tau_{P,H} - \tau_{P,L} > 0,$$

the positive-part operator is inactive, so

$$\kappa^*(c; \tau_{P,H}, \tau_{P,L}) = \Xi \Big[G(c) + \tau_{P,H} - \tau_{P,L} \Big].$$

Because $G(c)$ is strictly increasing on that interval, the threshold is strictly increasing there as well. ■

C.4 Proof of Corollary 1

Proof. Proposition 2 states that an investor with compute budget c acquires the private signal if and only if

$$\kappa < \kappa^*(c; \tau_{P,H}, \tau_{P,L}).$$

Therefore the pair of inequalities

$$\kappa^*(c_{\mathrm{L}}; \tau_{P,H}, \tau_{P,L}) \leq \kappa < \kappa^*(c_{\mathrm{H}}; \tau_{P,H}, \tau_{P,L})$$

implies that low-compute investors do not acquire while high-compute investors do acquire. That is exactly the stratified regime.

Conversely, suppose the economy is in the stratified regime under the common anticipated price-learning terms $(\tau_{P,H}, \tau_{P,L})$. Then high-compute investors acquire, so Proposition 2 implies

$$\kappa < \kappa^*(c_{\mathrm{H}}; \tau_{P,H}, \tau_{P,L}).$$

Low-compute investors do not acquire, so the same proposition implies

$$\kappa \geq \kappa^*(c_{\mathrm{L}}; \tau_{P,H}, \tau_{P,L}).$$

Combining these two inequalities gives the stated condition. ■

C.5 Beliefs and demands in the linear REE

Lemma 9.2 (Post-price beliefs and demands)**.** *Suppose the price is conjectured to be affine:*

$$P = a_0 + a_{\mathrm{H}} s_{\mathrm{H}} + a_{\mathrm{L}} s_{2,\mathrm{L}} - a_u \tilde{u}.$$

Let

$$q_{1,\mathrm{H}} = f_1(c_{1,\mathrm{H}})\tau_{s_1}, \qquad q_{2,\mathrm{H}} = f_2(c_{2,\mathrm{H}})\tau_{s_2}, \qquad q_{2,\mathrm{L}} = f_2(c_{\mathrm{L}})\tau_{s_2},$$

and define

$$Q_{\mathrm{H}} \equiv q_{1,\mathrm{H}} + q_{2,\mathrm{H}}, \qquad s_{\mathrm{H}} \equiv \frac{q_{1,\mathrm{H}} s_1 + q_{2,\mathrm{H}} s_{2,\mathrm{H}}}{Q_{\mathrm{H}}}.$$

Then high-compute investors can extract

$$s_{P,\mathrm{H}} \equiv \frac{P - a_0 - a_{\mathrm{H}} s_{\mathrm{H}}}{a_{\mathrm{L}}} = s_{2,\mathrm{L}} - \frac{a_u}{a_{\mathrm{L}}} \tilde{u},$$

whose precision is

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \left(\frac{a_u}{a_{\mathrm{L}}} \right)^2 \tau_u^{-1} \right)^{-1}.$$

Low-compute investors can extract

$$s_{P,\mathrm{L}} \equiv \frac{P - a_0 - a_{\mathrm{L}} s_{2,\mathrm{L}}}{a_{\mathrm{H}}} = s_{\mathrm{H}} - \frac{a_u}{a_{\mathrm{H}}} \tilde{u},$$

whose precision is

$$\tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \left(\frac{a_u}{a_{\mathrm{H}}} \right)^2 \tau_u^{-1} \right)^{-1}.$$

Their post-price posterior precisions and means are therefore

$$\tau_{\mathrm{H}}^R = \tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}}, \qquad \mu_{\mathrm{H}}^R = \frac{\tau_0 \bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}} + \tau_{P,\mathrm{H}} s_{P,\mathrm{H}}}{\tau_{\mathrm{H}}^R},$$

and

$$\tau_{\mathrm{L}}^R = \tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}, \qquad \mu_{\mathrm{L}}^R = \frac{\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}} s_{P,\mathrm{L}}}{\tau_{\mathrm{L}}^R}.$$

Optimal demands are

$$x_{\mathrm{H}} = \frac{\tau_{\mathrm{H}}^R}{A} (\mu_{\mathrm{H}}^R - P), \qquad x_{\mathrm{L}} = \frac{\tau_{\mathrm{L}}^R}{A} (\mu_{\mathrm{L}}^R - P).$$

Proof. We begin with the high-compute investors. Before observing the price, they observe $(s_1, s_{2,\mathrm{H}})$. Because these two signals are conditionally independent Gaussian signals about

$\tilde{\theta}$, they can be summarized by the pooled sufficient statistic

$$s_{\mathrm{H}} = \frac{q_{1,\mathrm{H}} s_1 + q_{2,\mathrm{H}} s_{2,\mathrm{H}}}{Q_{\mathrm{H}}},$$

where $Q_{\mathrm{H}} = q_{1,\mathrm{H}} + q_{2,\mathrm{H}}$. The signal s_{H} has the representation

$$s_{\mathrm{H}} = \tilde{\theta} + \tilde{\eta}^*_{\mathrm{H}},$$

with $\tilde{\eta}^*_{\mathrm{H}} \sim \mathcal{N}(0, Q_{\mathrm{H}}^{-1})$. Therefore the high-compute investors' pre-price posterior precision is $\tau_0 + Q_{\mathrm{H}}$.

Now condition on the variables (s_{H}, P). Rearranging the conjectured price equation gives

$$s_{P,\mathrm{H}} \equiv \frac{P - a_0 - a_{\mathrm{H}} s_{\mathrm{H}}}{a_{\mathrm{L}}} = s_{2,\mathrm{L}} - \frac{a_u}{a_{\mathrm{L}}} \tilde{u}.$$

Since

$$s_{2,\mathrm{L}} = \tilde{\theta} + \tilde{\eta}_{2,\mathrm{L}},$$

the extracted object can be written as

$$s_{P,\mathrm{H}} = \tilde{\theta} + \left(\tilde{\eta}_{2,\mathrm{L}} - \frac{a_u}{a_{\mathrm{L}}} \tilde{u} \right).$$

The term in parentheses is Gaussian with variance

$$q_{2,\mathrm{L}}^{-1} + \left(\frac{a_u}{a_{\mathrm{L}}} \right)^2 \tau_u^{-1}.$$

Therefore the precision of the price-implied signal is

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \left(\frac{a_u}{a_{\mathrm{L}}} \right)^2 \tau_u^{-1} \right)^{-1}.$$

Because $\tilde{\eta}^*_{\mathrm{H}}$ is independent of both $\tilde{\eta}_{2,\mathrm{L}}$ and $\tilde{u}$, the signals s_{H} and $s_{P,\mathrm{H}}$ are conditionally

independent Gaussian signals about $\tilde{\theta}$. Standard Gaussian updating therefore gives

$$\tau_{\mathrm{H}}^{R} = \tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}}$$

and

$$\mu_{\mathrm{H}}^{R} = \frac{\tau_0 \bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}} + \tau_{P,\mathrm{H}} s_{P,\mathrm{H}}}{\tau_{\mathrm{H}}^{R}}.$$

The low-compute investors are analogous. Before observing the price, they observe only $s_{2,\mathrm{L}}$, so their pre-price posterior precision is $\tau_0 + q_{2,\mathrm{L}}$. Rearranging the conjectured price equation gives

$$s_{P,\mathrm{L}} \equiv \frac{P - a_0 - a_{\mathrm{L}} s_{2,\mathrm{L}}}{a_{\mathrm{H}}} = s_{\mathrm{H}} - \frac{a_u}{a_{\mathrm{H}}} \tilde{u}.$$

Because $s_{\mathrm{H}} = \tilde{\theta} + \tilde{\eta}_{\mathrm{H}}^{*}$, this becomes

$$s_{P,\mathrm{L}} = \tilde{\theta} + \left(\tilde{\eta}_{\mathrm{H}}^{*} - \frac{a_u}{a_{\mathrm{H}}} \tilde{u} \right).$$

The noise term is Gaussian with variance

$$Q_{\mathrm{H}}^{-1} + \left(\frac{a_u}{a_{\mathrm{H}}} \right)^2 \tau_u^{-1},$$

so its precision is

$$\tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \left(\frac{a_u}{a_{\mathrm{H}}} \right)^2 \tau_u^{-1} \right)^{-1}.$$

Because $\tilde{\eta}_{2,\mathrm{L}}$ is independent of both $\tilde{\eta}_{\mathrm{H}}^{*}$ and $\tilde{u}$, the signals $s_{2,\mathrm{L}}$ and $s_{P,\mathrm{L}}$ are conditionally independent Gaussian signals about $\tilde{\theta}$. Standard Gaussian updating therefore gives

$$\tau_{\mathrm{L}}^{R} = \tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}$$

and

$$\mu_{\mathrm{L}}^{R} = \frac{\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}} s_{P,\mathrm{L}}}{\tau_{\mathrm{L}}^{R}}.$$

Finally, consider an investor with posterior distribution $\mathcal{N}(\mu, \tau^{-1})$ after observing the relevant information set. The certainty equivalent from holding x units of the risky asset is

$$\mathrm{CE}(x) = W_0 + x(\mu - P) - \frac{A}{2}x^2\tau^{-1}.$$

The first-order condition is

$$\mu - P - Ax\tau^{-1} = 0,$$

so the optimal demand is $x = \tau(\mu - P)/A$. Applying this formula to the high-compute and low-compute investors yields the stated demand expressions. ■

C.6 Proof of Proposition 3

Proof. Suppose the economy is in the stratified regime and fix the induced effective precisions $(q_{1,\mathrm{H}}, q_{2,\mathrm{H}}, q_{2,\mathrm{L}})$ with $q_{1,\mathrm{H}} > 0$. Also define

$$Q_{\mathrm{H}} \equiv q_{1,\mathrm{H}} + q_{2,\mathrm{H}}.$$

We begin with a generic affine price conjecture

$$P = a_0 + a_{\mathrm{H}} s_{\mathrm{H}} + a_{\mathrm{L}} s_{2,\mathrm{L}} - a_u \tilde{u},$$

where $a_{\mathrm{H}} > 0$, $a_{\mathrm{L}} > 0$, and $a_u > 0$.

Define the coefficient ratios

$$\beta_{\mathrm{H}} \equiv \frac{a_u}{a_{\mathrm{L}}}, \qquad \beta_{\mathrm{L}} \equiv \frac{a_u}{a_{\mathrm{H}}}.$$

By Lemma 9.2, the price-implied precisions are

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \beta_{\mathrm{H}}^2 \tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \beta_{\mathrm{L}}^2 \tau_u^{-1}\right)^{-1},$$

and the post-price posterior means satisfy

$$\tau_{\mathrm{H}}^R \mu_{\mathrm{H}}^R = \tau_0 \bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}} + \tau_{P,\mathrm{H}} s_{P,\mathrm{H}},$$

$$\tau_{\mathrm{L}}^R \mu_{\mathrm{L}}^R = \tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}} s_{P,\mathrm{L}},$$

where

$$s_{P,\mathrm{H}} = s_{2,\mathrm{L}} - \beta_{\mathrm{H}} \tilde{u}, \qquad s_{P,\mathrm{L}} = s_{\mathrm{H}} - \beta_{\mathrm{L}} \tilde{u}.$$

The demands are

$$x_{\mathrm{H}} = \frac{\tau_{\mathrm{H}}^R}{A}(\mu_{\mathrm{H}}^R - P), \qquad x_{\mathrm{L}} = \frac{\tau_{\mathrm{L}}^R}{A}(\mu_{\mathrm{L}}^R - P).$$

Market clearing is

$$\lambda x_{\mathrm{H}} + (1-\lambda) x_{\mathrm{L}} = \bar{z} + \tilde{u}.$$

Multiply by A and collect terms in P:

$$\bar{\tau}^R P = \lambda \tau_{\mathrm{H}}^R \mu_{\mathrm{H}}^R + (1-\lambda) \tau_{\mathrm{L}}^R \mu_{\mathrm{L}}^R - A\bar{z} - A\tilde{u}, \tag{18}$$

where

$$\bar{\tau}^R \equiv \lambda \tau_{\mathrm{H}}^R + (1-\lambda) \tau_{\mathrm{L}}^R.$$

Using

$$\tau_{\mathrm{H}}^R = \tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}}, \qquad \tau_{\mathrm{L}}^R = \tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}},$$

we obtain

$$\bar{\tau}^R = \tau_0 + \lambda Q_{\mathrm{H}} + \lambda \tau_{P,\mathrm{H}} + (1-\lambda) q_{2,\mathrm{L}} + (1-\lambda) \tau_{P,\mathrm{L}}.$$

Next substitute the posterior means into (18). This gives

$$\bar{\tau}^R P = \lambda\Big[\tau_0\bar{\theta} + Q_{\mathrm{H}} s_{\mathrm{H}} + \tau_{P,\mathrm{H}}(s_{2,\mathrm{L}} - \beta_{\mathrm{H}}\tilde{u})\Big] + (1-\lambda)\Big[\tau_0\bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}(s_{\mathrm{H}} - \beta_{\mathrm{L}}\tilde{u})\Big] - A\bar{z} - A\tilde{u}. \tag{19}$$

Collect the constant, signal, and supply-noise terms. The coefficient on $\tau_0\bar{\theta}$ is $\lambda+(1-\lambda)=1$, so the constant part is simply $\tau_0\bar{\theta}$. The coefficient on s_{H} is

$$\lambda Q_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}.$$

The coefficient on $s_{2,\mathrm{L}}$ is

$$\lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}.$$

The coefficient on $\tilde{u}$ is

$$-A - \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} - (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}.$$

Therefore (19) can be rewritten as

$$P = \frac{\tau_0\bar{\theta} + \left[\lambda Q_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\right] s_{\mathrm{H}} + \left[\lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}\right] s_{2,\mathrm{L}} - A\bar{z}}{\bar{\tau}^R} - \frac{A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{\bar{\tau}^R}\tilde{u}. \tag{20}$$

For the conjectured price to be an equilibrium price, this expression must coincide with

$$P = a_0 + a_{\mathrm{H}} s_{\mathrm{H}} + a_{\mathrm{L}} s_{2,\mathrm{L}} - a_u\tilde{u}.$$

Matching coefficients therefore gives

$$a_{\mathrm{H}} = \frac{\lambda Q_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}}{\bar{\tau}^R}, \tag{21}$$

$$a_{\mathrm{L}} = \frac{\lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}}{\bar{\tau}^R}, \tag{22}$$

and

$$a_u = \frac{A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{\bar{\tau}^R}. \tag{23}$$

Now divide (23) by (22). Because $\beta_{\mathrm{H}} = a_u/a_{\mathrm{L}}$, we obtain

$$\beta_{\mathrm{H}} = \frac{A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{\lambda\tau_{P,\mathrm{H}} + (1-\lambda)q_{2,\mathrm{L}}}.$$

Multiply both sides by the denominator:

$$\beta_{\mathrm{H}}\lambda\tau_{P,\mathrm{H}} + \beta_{\mathrm{H}}(1-\lambda)q_{2,\mathrm{L}} = A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}.$$

The terms $\beta_{\mathrm{H}}\lambda\tau_{P,\mathrm{H}}$ on the left and $\lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}}$ on the right are identical, so they cancel. This leaves

$$\beta_{\mathrm{H}}(1-\lambda)q_{2,\mathrm{L}} = A + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}.$$

Dividing by $(1-\lambda)q_{2,\mathrm{L}}$ gives

$$\beta_{\mathrm{H}} = \frac{\dfrac{A}{1-\lambda} + \tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{q_{2,\mathrm{L}}}.$$

This is exactly (7).

Next divide (23) by (21). Because $\beta_{\mathrm{L}} = a_u/a_{\mathrm{H}}$, we obtain

$$\beta_{\mathrm{L}} = \frac{A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}}{\lambda Q_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}}.$$

Multiply both sides by the denominator:

$$\beta_{\mathrm{L}}\lambda Q_{\mathrm{H}} + \beta_{\mathrm{L}}(1-\lambda)\tau_{P,\mathrm{L}} = A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}} + (1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}.$$

Again the common $(1-\lambda)\tau_{P,\mathrm{L}}\beta_{\mathrm{L}}$ terms cancel, leaving

$$\beta_{\mathrm{L}}\lambda Q_{\mathrm{H}} = A + \lambda\tau_{P,\mathrm{H}}\beta_{\mathrm{H}}.$$

Dividing by λQ_{H} gives

$$\beta_{\mathrm{L}} = \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}\beta_{\mathrm{H}}}{Q_{\mathrm{H}}}.$$

This is exactly (8).

We now prove existence of a positive pair satisfying the fixed-point system. Define the maps

$$T_{\mathrm{H}}(y) \equiv \frac{\dfrac{A}{1-\lambda} + \tau_{P,\mathrm{L}}(y)y}{q_{2,\mathrm{L}}}, \qquad T_{\mathrm{L}}(x) \equiv \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}(x)x}{Q_{\mathrm{H}}},$$

where

$$\tau_{P,\mathrm{H}}(x) \equiv \left(q_{2,\mathrm{L}}^{-1} + x^2\tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}}(y) \equiv \left(Q_{\mathrm{H}}^{-1} + y^2\tau_u^{-1}\right)^{-1}.$$

The fixed-point system is equivalent to

$$\beta_{\mathrm{H}} = T_{\mathrm{H}}(\beta_{\mathrm{L}}), \qquad \beta_{\mathrm{L}} = T_{\mathrm{L}}(\beta_{\mathrm{H}}).$$

Define the one-dimensional map

$$F(x) \equiv T_{\mathrm{H}}(T_{\mathrm{L}}(x)).$$

Because $\tau_{P,\mathrm{H}}(x)$ and $\tau_{P,\mathrm{L}}(y)$ are continuous in their arguments, both T_{H} and T_{L} are continuous on $[0,\infty)$, and therefore F is continuous on $[0,\infty)$.

Next we show that F is bounded above. For any $x \geq 0$,

$$\tau_{P,\mathrm{H}}(x)x = \frac{x}{q_{2,\mathrm{L}}^{-1} + x^2\tau_u^{-1}} = \frac{q_{2,\mathrm{L}}\tau_u x}{\tau_u + q_{2,\mathrm{L}}x^2}.$$

Define

$$h_{\mathrm{H}}(x) \equiv \frac{q_{2,\mathrm{L}}\tau_u x}{\tau_u + q_{2,\mathrm{L}}x^2}.$$

Differentiating gives

$$h'_{\mathrm{H}}(x) = \frac{q_{2,\mathrm{L}}\tau_u(\tau_u - q_{2,\mathrm{L}}x^2)}{(\tau_u + q_{2,\mathrm{L}}x^2)^2}.$$

Thus $h'_{\mathrm{H}}(x) = 0$ when $x = \sqrt{\tau_u/q_{2,\mathrm{L}}}$, and that point is the global maximizer on $[0, \infty)$. Evaluating h_{H} there yields

$$\tau_{P,\mathrm{H}}(x)x \leq \frac{1}{2}\sqrt{q_{2,\mathrm{L}}\tau_u}.$$

Therefore

$$T_{\mathrm{L}}(x) \leq \frac{A/\lambda + \frac{1}{2}\sqrt{q_{2,\mathrm{L}}\tau_u}}{Q_{\mathrm{H}}}$$

for all $x \geq 0$. This shows that T_{L} is bounded above by a finite constant.

Apply the same argument to

$$\tau_{P,\mathrm{L}}(y)y = \frac{y}{Q_{\mathrm{H}}^{-1} + y^2\tau_u^{-1}} = \frac{Q_{\mathrm{H}}\tau_u y}{\tau_u + Q_{\mathrm{H}}y^2}.$$

The same derivative calculation implies

$$\tau_{P,\mathrm{L}}(y)y \leq \frac{1}{2}\sqrt{Q_{\mathrm{H}}\tau_u}$$

for all $y \geq 0$. Hence

$$T_{\mathrm{H}}(y) \leq \frac{A/(1-\lambda) + \frac{1}{2}\sqrt{Q_{\mathrm{H}}\tau_u}}{q_{2,\mathrm{L}}}$$

for all $y \geq 0$. Therefore $F(x) = T_{\mathrm{H}}(T_{\mathrm{L}}(x))$ is bounded above by the same finite constant for every $x \geq 0$.

Now evaluate F at zero. Because

$$T_{\mathrm{L}}(0) = \frac{A/\lambda}{Q_{\mathrm{H}}} > 0,$$

we have

$$F(0) = T_{\mathrm{H}}(T_{\mathrm{L}}(0)) > 0.$$

On the other hand, since F is bounded above while $x \to \infty$, we have

$$F(x) - x \to -\infty \qquad \text{as } x \to \infty.$$

Because F is continuous, the intermediate value theorem implies that there exists $x^* > 0$ such that

$$F(x^*) = x^*.$$

Set

$$\beta_{\mathrm{H}} \equiv x^*, \qquad \beta_{\mathrm{L}} \equiv T_{\mathrm{L}}(x^*).$$

Then $\beta_{\mathrm{L}} > 0$ and

$$\beta_{\mathrm{H}} = T_{\mathrm{H}}(\beta_{\mathrm{L}}), \qquad \beta_{\mathrm{L}} = T_{\mathrm{L}}(\beta_{\mathrm{H}}),$$

so $(\beta_{\mathrm{H}}, \beta_{\mathrm{L}})$ is a positive solution to the fixed-point system.

Finally, for any positive pair $(\beta_{\mathrm{H}}, \beta_{\mathrm{L}})$ satisfying the fixed-point system, define

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \beta_{\mathrm{H}}^2 \tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \beta_{\mathrm{L}}^2 \tau_u^{-1}\right)^{-1},$$

and let $\bar{\tau}^R$ be given by (10). Then (20) becomes exactly the price equation stated in (9). Conversely, every affine REE of the conjectured form must satisfy the coefficient-matching conditions above, so it must generate a positive pair $(\beta_{\mathrm{H}}, \beta_{\mathrm{L}})$ satisfying (7) and (8), and its price equation must have the form (9). This proves the proposition. ■

C.7 Proof of Proposition 4

Proof. By definition,

$$\Pi_{\mathrm{H}} = \frac{\tau_{P,\mathrm{H}}}{q_{2,\mathrm{L}}}, \qquad \Pi_{\mathrm{L}} = \frac{\tau_{P,\mathrm{L}}}{Q_{\mathrm{H}}}.$$

Proposition 3 gives

$$\tau_{P,\mathrm{H}} = \left(q_{2,\mathrm{L}}^{-1} + \beta_{\mathrm{H}}^2 \tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}} = \left(Q_{\mathrm{H}}^{-1} + \beta_{\mathrm{L}}^2 \tau_u^{-1}\right)^{-1}.$$

For the first ratio,

$$\Pi_{\mathrm{H}} = \frac{1}{q_{2,\mathrm{L}}} \left(q_{2,\mathrm{L}}^{-1} + \beta_{\mathrm{H}}^2 \tau_u^{-1}\right)^{-1} = \left(1 + \beta_{\mathrm{H}}^2 q_{2,\mathrm{L}} \tau_u^{-1}\right)^{-1}.$$

For the second ratio,

$$\Pi_{\mathrm{L}} = \frac{1}{Q_{\mathrm{H}}} \left(Q_{\mathrm{H}}^{-1} + \beta_{\mathrm{L}}^2 \tau_u^{-1}\right)^{-1} = \left(1 + \beta_{\mathrm{L}}^2 Q_{\mathrm{H}} \tau_u^{-1}\right)^{-1}.$$

This is exactly (11).

Because all precisions are strictly positive and $\beta_{\mathrm{H}}, \beta_{\mathrm{L}} > 0$, both denominators are strictly larger than one. Therefore

$$0 < \Pi_{\mathrm{H}} < 1, \qquad 0 < \Pi_{\mathrm{L}} < 1.$$

This proves the proposition. ■

C.8 Proof of Proposition 5

Proof. By definition,

$$\Delta\tau^R = \tau_{\mathrm{H}}^R - \tau_{\mathrm{L}}^R.$$

From Lemma 9.2,

$$\tau_{\mathrm{H}}^R = \tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}}, \qquad \tau_{\mathrm{L}}^R = \tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}.$$

Subtract the second expression from the first:

$$\Delta\tau^R = \left(\tau_0 + Q_{\mathrm{H}} + \tau_{P,\mathrm{H}}\right) - \left(\tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}\right) = Q_{\mathrm{H}} - q_{2,\mathrm{L}} + \tau_{P,\mathrm{H}} - \tau_{P,\mathrm{L}}.$$

Now use

$$Q_{\mathrm{H}} = q_{1,\mathrm{H}} + q_{2,\mathrm{H}} = \Phi^*(c_{\mathrm{H}})$$

and

$$q_{2,\mathrm{L}} = f_2(c_{\mathrm{L}})\tau_{s_2}.$$

Substituting these identities yields

$$\Delta\tau^R = \Phi^*(c_{\mathrm{H}}) - f_2(c_{\mathrm{L}})\tau_{s_2} + \tau_{P,\mathrm{H}} - \tau_{P,\mathrm{L}}.$$

This is exactly (12). ■

C.9 Proof of Proposition 6

Proof. In the no-acquisition benchmark, no investor observes the private signal. High-compute investors devote all compute to the public source, so define

$$q^0_{2,\mathrm{H}} \equiv f_2(c_{\mathrm{H}})\tau_{s_2}, \qquad q_{2,\mathrm{L}} \equiv f_2(c_{\mathrm{L}})\tau_{s_2}.$$

The two public reduced-form signals are

$$s^0_{2,\mathrm{H}} = \tilde{\theta} + \tilde{\eta}^0_{2,\mathrm{H}}, \qquad s_{2,\mathrm{L}} = \tilde{\theta} + \tilde{\eta}_{2,\mathrm{L}},$$

with

$$\tilde{\eta}^0_{2,\mathrm{H}} \sim \mathcal{N}(0, (q^0_{2,\mathrm{H}})^{-1}), \qquad \tilde{\eta}_{2,\mathrm{L}} \sim \mathcal{N}(0, q^{-1}_{2,\mathrm{L}}),$$

mutually independent and independent of $\tilde{u}$.

Conjecture an affine price equation of the form

$$P^0 = b_0 + b_{\mathrm{H}} s^0_{2,\mathrm{H}} + b_{\mathrm{L}} s_{2,\mathrm{L}} - b_u \tilde{u},$$

with $b_{\mathrm{H}} > 0$, $b_{\mathrm{L}} > 0$, and $b_u > 0$. Define

$$\beta^0_{\mathrm{H}} \equiv \frac{b_u}{b_{\mathrm{L}}}, \qquad \beta^0_{\mathrm{L}} \equiv \frac{b_u}{b_{\mathrm{H}}}.$$

Then high-compute investors can extract

$$s^0_{P,\mathrm{H}} \equiv \frac{P^0 - b_0 - b_{\mathrm{H}} s^0_{2,\mathrm{H}}}{b_{\mathrm{L}}} = s_{2,\mathrm{L}} - \beta^0_{\mathrm{H}} \tilde{u},$$

and low-compute investors can extract

$$s^0_{P,\mathrm{L}} \equiv \frac{P^0 - b_0 - b_{\mathrm{L}} s_{2,\mathrm{L}}}{b_{\mathrm{H}}} = s^0_{2,\mathrm{H}} - \beta^0_{\mathrm{L}} \tilde{u}.$$

The corresponding price-implied precisions are

$$\tau^0_{P,\mathrm{H}} = \left(q^{-1}_{2,\mathrm{L}} + (\beta^0_{\mathrm{H}})^2 \tau_u^{-1}\right)^{-1}, \qquad \tau^0_{P,\mathrm{L}} = \left((q^0_{2,\mathrm{H}})^{-1} + (\beta^0_{\mathrm{L}})^2 \tau_u^{-1}\right)^{-1}.$$

Therefore the post-price posterior precisions and means are

$$\tau^{0,R}_{\mathrm{H}} = \tau_0 + q^0_{2,\mathrm{H}} + \tau^0_{P,\mathrm{H}}, \qquad \mu^{0,R}_{\mathrm{H}} = \frac{\tau_0 \bar{\theta} + q^0_{2,\mathrm{H}} s^0_{2,\mathrm{H}} + \tau^0_{P,\mathrm{H}} s^0_{P,\mathrm{H}}}{\tau^{0,R}_{\mathrm{H}}},$$

and

$$\tau^{0,R}_{\mathrm{L}} = \tau_0 + q_{2,\mathrm{L}} + \tau^0_{P,\mathrm{L}}, \qquad \mu^{0,R}_{\mathrm{L}} = \frac{\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau^0_{P,\mathrm{L}} s^0_{P,\mathrm{L}}}{\tau^{0,R}_{\mathrm{L}}}.$$

Optimal demands are

$$x^0_{\mathrm{H}} = \frac{\tau^{0,R}_{\mathrm{H}}}{A}\left(\mu^{0,R}_{\mathrm{H}} - P^0\right), \qquad x^0_{\mathrm{L}} = \frac{\tau^{0,R}_{\mathrm{L}}}{A}\left(\mu^{0,R}_{\mathrm{L}} - P^0\right).$$

Market clearing is

$$\lambda x^0_{\mathrm{H}} + (1-\lambda) x^0_{\mathrm{L}} = \bar{z} + \tilde{u}.$$

Multiply by A and collect terms in P^0:

$$\bar{\tau}^0 P^0 = \lambda \tau_{\mathrm{H}}^{0,R} \mu_{\mathrm{H}}^{0,R} + (1-\lambda)\tau_{\mathrm{L}}^{0,R} \mu_{\mathrm{L}}^{0,R} - A\bar{z} - A\tilde{u}, \tag{24}$$

where

$$\bar{\tau}^0 \equiv \lambda \tau_{\mathrm{H}}^{0,R} + (1-\lambda)\tau_{\mathrm{L}}^{0,R}.$$

Using the definitions of the post-price precisions gives

$$\bar{\tau}^0 = \tau_0 + \lambda q_{2,\mathrm{H}}^0 + \lambda \tau_{P,\mathrm{H}}^0 + (1-\lambda) q_{2,\mathrm{L}} + (1-\lambda)\tau_{P,\mathrm{L}}^0.$$

Substitute the posterior means into (24). This yields

$$\begin{aligned}\bar{\tau}^0 P^0 = {} & \lambda\Big[\tau_0 \bar{\theta} + q_{2,\mathrm{H}}^0 s_{2,\mathrm{H}}^0 + \tau_{P,\mathrm{H}}^0 (s_{2,\mathrm{L}} - \beta_{\mathrm{H}}^0 \tilde{u})\Big] \\ & + (1-\lambda)\Big[\tau_0 \bar{\theta} + q_{2,\mathrm{L}} s_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}^0 (s_{2,\mathrm{H}}^0 - \beta_{\mathrm{L}}^0 \tilde{u})\Big] - A\bar{z} - A\tilde{u}.\end{aligned}$$

Collect the terms multiplying each object. The constant term is again $\tau_0 \bar{\theta}$. The coefficient on $s_{2,\mathrm{H}}^0$ is

$$\lambda q_{2,\mathrm{H}}^0 + (1-\lambda)\tau_{P,\mathrm{L}}^0.$$

The coefficient on $s_{2,\mathrm{L}}$ is

$$\lambda \tau_{P,\mathrm{H}}^0 + (1-\lambda) q_{2,\mathrm{L}}.$$

The coefficient on $\tilde{u}$ is

$$-A - \lambda \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0 - (1-\lambda)\tau_{P,\mathrm{L}}^0 \beta_{\mathrm{L}}^0.$$

Therefore

$$P^0 = \frac{\tau_0 \bar{\theta} + \left[\lambda q_{2,\mathrm{H}}^0 + (1-\lambda)\tau_{P,\mathrm{L}}^0\right] s_{2,\mathrm{H}}^0 + \left[\lambda \tau_{P,\mathrm{H}}^0 + (1-\lambda) q_{2,\mathrm{L}}\right] s_{2,\mathrm{L}} - A\bar{z}}{\bar{\tau}^0} - \frac{A + \lambda \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0 + (1-\lambda)\tau_{P,\mathrm{L}}^0 \beta_{\mathrm{L}}^0}{\bar{\tau}^0} \tilde{u}. \tag{25}$$

Matching this expression to the conjectured affine price yields

$$b_{\mathrm{H}} = \frac{\lambda q^0_{2,\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}}}{\bar{\tau}^0}, \tag{26}$$

$$b_{\mathrm{L}} = \frac{\lambda \tau^0_{P,\mathrm{H}} + (1-\lambda) q_{2,\mathrm{L}}}{\bar{\tau}^0}, \tag{27}$$

and

$$b_u = \frac{A + \lambda \tau^0_{P,\mathrm{H}} \beta^0_{\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}}{\bar{\tau}^0}. \tag{28}$$

Divide (28) by (27). Since $\beta^0_{\mathrm{H}} = b_u / b_{\mathrm{L}}$, we obtain

$$\beta^0_{\mathrm{H}} = \frac{A + \lambda \tau^0_{P,\mathrm{H}} \beta^0_{\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}}{\lambda \tau^0_{P,\mathrm{H}} + (1-\lambda) q_{2,\mathrm{L}}}.$$

Multiply both sides by the denominator:

$$\beta^0_{\mathrm{H}} \lambda \tau^0_{P,\mathrm{H}} + \beta^0_{\mathrm{H}} (1-\lambda) q_{2,\mathrm{L}} = A + \lambda \tau^0_{P,\mathrm{H}} \beta^0_{\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}.$$

Cancel the common $\lambda \tau^0_{P,\mathrm{H}} \beta^0_{\mathrm{H}}$ terms. This gives

$$\beta^0_{\mathrm{H}} (1-\lambda) q_{2,\mathrm{L}} = A + (1-\lambda)\tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}.$$

Dividing by $(1-\lambda) q_{2,\mathrm{L}}$ yields

$$\beta^0_{\mathrm{H}} = \frac{\dfrac{A}{1-\lambda} + \tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}}{q_{2,\mathrm{L}}}.$$

Next divide (28) by (26). Since $\beta^0_{\mathrm{L}} = b_u / b_{\mathrm{H}}$, we obtain

$$\beta^0_{\mathrm{L}} = \frac{A + \lambda \tau^0_{P,\mathrm{H}} \beta^0_{\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}} \beta^0_{\mathrm{L}}}{\lambda q^0_{2,\mathrm{H}} + (1-\lambda)\tau^0_{P,\mathrm{L}}}.$$

Multiply both sides by the denominator:

$$\beta_{\mathrm{L}}^0 \lambda q_{2,\mathrm{H}}^0 + \beta_{\mathrm{L}}^0 (1-\lambda)\tau_{P,\mathrm{L}}^0 = A + \lambda \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0 + (1-\lambda)\tau_{P,\mathrm{L}}^0 \beta_{\mathrm{L}}^0.$$

Cancel the common $(1-\lambda)\tau_{P,\mathrm{L}}^0 \beta_{\mathrm{L}}^0$ terms. This gives

$$\beta_{\mathrm{L}}^0 \lambda q_{2,\mathrm{H}}^0 = A + \lambda \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0.$$

Dividing by $\lambda q_{2,\mathrm{H}}^0$ yields

$$\beta_{\mathrm{L}}^0 = \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}^0 \beta_{\mathrm{H}}^0}{q_{2,\mathrm{H}}^0}.$$

Thus any affine REE must satisfy the fixed-point system stated in the proposition.

We next prove existence of a positive fixed point. Define

$$T_{\mathrm{H}}^0(y) \equiv \frac{\dfrac{A}{1-\lambda} + \tau_{P,\mathrm{L}}^0(y) y}{q_{2,\mathrm{L}}}, \qquad T_{\mathrm{L}}^0(x) \equiv \frac{\dfrac{A}{\lambda} + \tau_{P,\mathrm{H}}^0(x) x}{q_{2,\mathrm{H}}^0},$$

where

$$\tau_{P,\mathrm{H}}^0(x) \equiv \left(q_{2,\mathrm{L}}^{-1} + x^2 \tau_u^{-1}\right)^{-1}, \qquad \tau_{P,\mathrm{L}}^0(y) \equiv \left((q_{2,\mathrm{H}}^0)^{-1} + y^2 \tau_u^{-1}\right)^{-1}.$$

Define

$$F^0(x) \equiv T_{\mathrm{H}}^0(T_{\mathrm{L}}^0(x)).$$

Exactly the same continuity argument as in the proof of Proposition 3 shows that F^0 is continuous on $[0, \infty)$.

Exactly the same maximization argument also shows that

$$\tau_{P,\mathrm{H}}^0(x) x \leq \frac{1}{2}\sqrt{q_{2,\mathrm{L}} \tau_u}$$

for all $x \geq 0$, and

$$\tau^0_{P,\mathrm{L}}(y)y \leq \frac{1}{2}\sqrt{q^0_{2,\mathrm{H}}\tau_u}$$

for all $y \geq 0$. Therefore both T^0_{L} and T^0_{H} are bounded above by finite constants, so F^0 is bounded above by a finite constant as well.

Moreover,

$$T^0_{\mathrm{L}}(0) = \frac{A/\lambda}{q^0_{2,\mathrm{H}}} > 0,$$

so

$$F^0(0) = T^0_{\mathrm{H}}(T^0_{\mathrm{L}}(0)) > 0.$$

Because F^0 is bounded while $x \to \infty$, we also have

$$F^0(x) - x \to -\infty \qquad \text{as } x \to \infty.$$

The intermediate value theorem therefore implies the existence of $x^{0,*} > 0$ such that

$$F^0(x^{0,*}) = x^{0,*}.$$

Set

$$\beta^0_{\mathrm{H}} \equiv x^{0,*}, \qquad \beta^0_{\mathrm{L}} \equiv T^0_{\mathrm{L}}(x^{0,*}).$$

Then $(\beta^0_{\mathrm{H}}, \beta^0_{\mathrm{L}})$ is a positive solution to the fixed-point system.

For any such solution, define $\tau^0_{P,\mathrm{H}}$ and $\tau^0_{P,\mathrm{L}}$ from the fixed-point pair and define $\bar{\tau}^0$ by (14). Then (25) is exactly the price equation stated in (13). Hence an affine REE exists.

Finally, the post-price information difference is

$$\Delta\tau^0 = \tau^{0,R}_{\mathrm{H}} - \tau^{0,R}_{\mathrm{L}} = \left(\tau_0 + q^0_{2,\mathrm{H}} + \tau^0_{P,\mathrm{H}}\right) - \left(\tau_0 + q_{2,\mathrm{L}} + \tau^0_{P,\mathrm{L}}\right).$$

Therefore

$$\Delta\tau^0 = q^0_{2,\mathrm{H}} - q_{2,\mathrm{L}} + \tau^0_{P,\mathrm{H}} - \tau^0_{P,\mathrm{L}},$$

which is exactly (15). This completes the proof. ■

C.10 Proof of Proposition 7

Proof. Suppose

$$f_k(c) = \kappa_k c^{\alpha_k}, \qquad \kappa_k > 0, \qquad \alpha_k \in (0,1).$$

Then $f_k(0) = 0$. For $c > 0$,

$$f_k'(c) = \kappa_k \alpha_k c^{\alpha_k - 1} > 0$$

and

$$f_k''(c) = \kappa_k \alpha_k (\alpha_k - 1) c^{\alpha_k - 2} < 0.$$

Since $\alpha_k > 0$, the power-law family is also unbounded, but that extra property is not required by Assumption 1. The required conditions in the assumption are therefore satisfied.

Now assume $\alpha_1 = \alpha_2 = \alpha$. The high-compute investor's allocation problem becomes

$$\max_{0 \le c_1 \le c_{\mathrm{H}}} \left\{ \kappa_1 \tau_{s_1} c_1^{\alpha} + \kappa_2 \tau_{s_2} (c_{\mathrm{H}} - c_1)^{\alpha} \right\}.$$

Because $\alpha \in (0,1)$, the objective is strictly concave, so the optimum is unique. The derivative of each term tends to infinity as its argument approaches zero from above. Therefore the unique optimum is interior and satisfies the first-order condition

$$\kappa_1 \alpha \tau_{s_1} c_{1,\mathrm{H}}^{\alpha - 1} = \kappa_2 \alpha \tau_{s_2} (c_{\mathrm{H}} - c_{1,\mathrm{H}})^{\alpha - 1}.$$

Cancel α and rearrange:

$$\left(\frac{c_{1,\mathrm{H}}}{c_{\mathrm{H}} - c_{1,\mathrm{H}}} \right)^{\alpha - 1} = \frac{\kappa_2 \tau_{s_2}}{\kappa_1 \tau_{s_1}}.$$

Since $\alpha - 1 < 0$, this is equivalent to

$$\frac{c_{1,\mathrm{H}}}{c_{\mathrm{H}} - c_{1,\mathrm{H}}} = \left(\frac{\kappa_1 \tau_{s_1}}{\kappa_2 \tau_{s_2}}\right)^{\frac{1}{1-\alpha}} \equiv m.$$

Hence

$$c_{1,\mathrm{H}} = \frac{m}{1+m} c_{\mathrm{H}}, \qquad c_{2,\mathrm{H}} = \frac{1}{1+m} c_{\mathrm{H}}.$$

Writing $\omega = m/(1+m)$ yields the formulas in the proposition. ■

C.11 Proof of Corollary 2

Proof. Consider any sequence $c_{\mathrm{H}} \to \infty$ along which the common-exponent power-law specification holds and the economy remains in the stratified affine REE. Under Proposition 7,

$$Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha},$$

where

$$C_{\mathrm{H}} \equiv \kappa_1 \tau_{s_1} \omega^{\alpha} + \kappa_2 \tau_{s_2} (1-\omega)^{\alpha}.$$

Along the sequence under consideration, $Q_{\mathrm{H}}(c_{\mathrm{H}}) \to \infty$.

Define

$$X(c_{\mathrm{H}}) \equiv \beta_{\mathrm{H}}(c_{\mathrm{H}}), \qquad Y(c_{\mathrm{H}}) \equiv \beta_{\mathrm{L}}(c_{\mathrm{H}}) Q_{\mathrm{H}}(c_{\mathrm{H}}).$$

From the fixed-point equation (8),

$$Y(c_{\mathrm{H}}) = \frac{A}{\lambda} + \tau_{P,\mathrm{H}}(c_{\mathrm{H}}) X(c_{\mathrm{H}}). \tag{29}$$

Now compute the second term on the right-hand side:

$$\tau_{P,\mathrm{H}}(c_{\mathrm{H}}) X(c_{\mathrm{H}}) = \frac{X(c_{\mathrm{H}})}{q_{2,\mathrm{L}}^{-1} + X(c_{\mathrm{H}})^2 \tau_u^{-1}} = \frac{q_{2,\mathrm{L}} \tau_u X(c_{\mathrm{H}})}{\tau_u + q_{2,\mathrm{L}} X(c_{\mathrm{H}})^2}.$$

Define

$$h(X) \equiv \frac{q_{2,\mathrm{L}}\tau_u X}{\tau_u + q_{2,\mathrm{L}}X^2}.$$

Differentiating gives

$$h'(X) = \frac{q_{2,\mathrm{L}}\tau_u(\tau_u - q_{2,\mathrm{L}}X^2)}{(\tau_u + q_{2,\mathrm{L}}X^2)^2}.$$

Thus $h'(X) = 0$ when $X = \sqrt{\tau_u/q_{2,\mathrm{L}}}$, and that point is the global maximizer of h on $[0, \infty)$. Evaluating h there yields

$$h(X) \leq \frac{1}{2}\sqrt{q_{2,\mathrm{L}}\tau_u} \qquad \text{for all } X \geq 0.$$

Hence

$$\tau_{P,\mathrm{H}}(c_\mathrm{H})X(c_\mathrm{H}) \leq \frac{1}{2}\sqrt{q_{2,\mathrm{L}}\tau_u} \qquad \text{for all } c_\mathrm{H}.$$

Substituting this bound into (29) gives

$$0 < Y(c_\mathrm{H}) \leq \frac{A}{\lambda} + \frac{1}{2}\sqrt{q_{2,\mathrm{L}}\tau_u}.$$

Thus $Y(c_\mathrm{H})$ is uniformly bounded along the sequence.

Because $Y(c_\mathrm{H}) = \beta_\mathrm{L}(c_\mathrm{H})Q_\mathrm{H}(c_\mathrm{H})$, the boundedness of Y implies

$$\beta_\mathrm{L}(c_\mathrm{H}) = O\Big(Q_\mathrm{H}(c_\mathrm{H})^{-1}\Big) = O(c_\mathrm{H}^{-\alpha}).$$

Therefore

$$\beta_\mathrm{L}(c_\mathrm{H})^2 Q_\mathrm{H}(c_\mathrm{H}) = O\Big(Q_\mathrm{H}(c_\mathrm{H})^{-1}\Big) = O(c_\mathrm{H}^{-\alpha}) \to 0.$$

Now use Proposition 4:

$$\Pi_\mathrm{L}(c_\mathrm{H}) = \frac{1}{1 + \beta_\mathrm{L}(c_\mathrm{H})^2 Q_\mathrm{H}(c_\mathrm{H})\tau_u^{-1}}.$$

Since the term in the denominator after the 1 converges to 0, it follows that

$$\Pi_{\mathrm{L}}(c_{\mathrm{H}}) \to 1 \qquad \text{as } c_{\mathrm{H}} \to \infty.$$

Finally, the opacity gap is

$$\Omega_{\mathrm{L}}(c_{\mathrm{H}}) = 1 - \Pi_{\mathrm{L}}(c_{\mathrm{H}}) = \frac{\beta_{\mathrm{L}}(c_{\mathrm{H}})^2 Q_{\mathrm{H}}(c_{\mathrm{H}}) \tau_u^{-1}}{1 + \beta_{\mathrm{L}}(c_{\mathrm{H}})^2 Q_{\mathrm{H}}(c_{\mathrm{H}}) \tau_u^{-1}}.$$

The denominator is bounded below by 1, while the numerator is

$$O\Big(\beta_{\mathrm{L}}(c_{\mathrm{H}})^2 Q_{\mathrm{H}}(c_{\mathrm{H}})\Big) = O\Big(Q_{\mathrm{H}}(c_{\mathrm{H}})^{-1}\Big) = O(c_{\mathrm{H}}^{-\alpha}).$$

Therefore

$$\Omega_{\mathrm{L}}(c_{\mathrm{H}}) = O(c_{\mathrm{H}}^{-\alpha}).$$

This proves (16). ■

C.12 Proof of Corollary 3

Proof. Fix a stratified REE under the common-exponent power-law specification. Proposition 7 implies

$$Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}.$$

We first derive the exact RMSE formulas. Let

$$\mathcal{I}_{\mathrm{H}} \equiv \sigma(s_1, s_{2,\mathrm{H}}, P)$$

denote the high-compute investors' information set. Because

$$\mu_{\mathrm{H}}^{R} = \mathbb{E}\Big[\tilde{\theta} \mid \mathcal{I}_{\mathrm{H}}\Big],$$

the conditional forecast error has conditional mean zero:

$$\mathbb{E}\Big[\tilde{\theta} - \mu_{\mathrm{H}}^{R} \mid \mathcal{I}_{\mathrm{H}}\Big] = 0.$$

Therefore

$$\begin{aligned}
\mathbb{E}\Big[(\tilde{\theta} - \mu_{\mathrm{H}}^{R})^2 \mid \mathcal{I}_{\mathrm{H}}\Big] &= \mathbb{V}\Big(\tilde{\theta} - \mu_{\mathrm{H}}^{R} \mid \mathcal{I}_{\mathrm{H}}\Big) + \Big(\mathbb{E}\Big[\tilde{\theta} - \mu_{\mathrm{H}}^{R} \mid \mathcal{I}_{\mathrm{H}}\Big]\Big)^2 \\
&= \mathbb{V}\Big(\tilde{\theta} \mid \mathcal{I}_{\mathrm{H}}\Big) \\
&= (\tau_{\mathrm{H}}^{R})^{-1}. \qquad (30)
\end{aligned}$$

Lemma 9.2 gives

$$\tau_{\mathrm{H}}^{R} = \tau_0 + Q_{\mathrm{H}}(c_{\mathrm{H}}) + \tau_{P,\mathrm{H}}(c_{\mathrm{H}}).$$

For a fixed equilibrium, τ_{H}^{R} is deterministic, so taking unconditional expectations in (30) leaves the same value:

$$\mathbb{E}\Big[(\tilde{\theta} - \mu_{\mathrm{H}}^{R})^2\Big] = (\tau_{\mathrm{H}}^{R})^{-1}.$$

Taking square roots yields

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) = (\tau_{\mathrm{H}}^{R})^{-1/2} = \left(\tau_0 + Q_{\mathrm{H}}(c_{\mathrm{H}}) + \tau_{P,\mathrm{H}}(c_{\mathrm{H}})\right)^{-1/2}.$$

Now let

$$\mathcal{I}_{\mathrm{L}} \equiv \sigma(s_{2,\mathrm{L}}, P)$$

denote the low-compute investors' information set. Since

$$\mu_{\mathrm{L}}^{R} = \mathbb{E}\Big[\tilde{\theta} \mid \mathcal{I}_{\mathrm{L}}\Big],$$

the same argument gives

$$\begin{aligned}\mathbb{E}\Big[(\tilde{\theta}-\mu_{\mathrm{L}}^{R})^{2}\mid\mathcal{I}_{\mathrm{L}}\Big] &= \mathbb{V}\Big(\tilde{\theta}-\mu_{\mathrm{L}}^{R}\mid\mathcal{I}_{\mathrm{L}}\Big)+\Big(\mathbb{E}\Big[\tilde{\theta}-\mu_{\mathrm{L}}^{R}\mid\mathcal{I}_{\mathrm{L}}\Big]\Big)^{2}\\ &= \mathbb{V}\Big(\tilde{\theta}\mid\mathcal{I}_{\mathrm{L}}\Big)\\ &= (\tau_{\mathrm{L}}^{R})^{-1}.\end{aligned}$$

Lemma 9.2 gives

$$\tau_{\mathrm{L}}^{R}=\tau_{0}+f_{2}(c_{\mathrm{L}})\tau_{s_{2}}+\tau_{P,\mathrm{L}}(c_{\mathrm{H}}).$$

Again τ_{L}^{R} is deterministic within a fixed equilibrium, so unconditional expectation preserves the same value:

$$\mathbb{E}\Big[(\tilde{\theta}-\mu_{\mathrm{L}}^{R})^{2}\Big]=(\tau_{\mathrm{L}}^{R})^{-1}.$$

Therefore

$$\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}},c_{\mathrm{L}})=(\tau_{\mathrm{L}}^{R})^{-1/2}=\left(\tau_{0}+f_{2}(c_{\mathrm{L}})\tau_{s_{2}}+\tau_{P,\mathrm{L}}(c_{\mathrm{H}})\right)^{-1/2}.$$

We now prove the asymptotic claims. Fix c_{L} and consider any sequence $c_{\mathrm{H}}\to\infty$ along which the economy remains in the stratified REE. Because $C_{\mathrm{H}}>0$ and $\alpha\in(0,1)$,

$$Q_{\mathrm{H}}(c_{\mathrm{H}})=C_{\mathrm{H}}c_{\mathrm{H}}^{\alpha}\to\infty.$$

Consider first the high-compute investors. Proposition 4 implies

$$\Pi_{\mathrm{H}}(c_{\mathrm{H}})=\frac{\tau_{P,\mathrm{H}}(c_{\mathrm{H}})}{q_{2,\mathrm{L}}}\in(0,1).$$

Equivalently,

$$0<\tau_{P,\mathrm{H}}(c_{\mathrm{H}})<q_{2,\mathrm{L}}=f_{2}(c_{\mathrm{L}})\tau_{s_{2}}\qquad\text{for every } c_{\mathrm{H}}.$$

Hence the term $\tau_{0}+\tau_{P,\mathrm{H}}(c_{\mathrm{H}})$ is uniformly bounded along the sequence. Divide the posterior

precision by $Q_{\mathrm{H}}(c_{\mathrm{H}})$:

$$\frac{\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} = \frac{\tau_0 + Q_{\mathrm{H}}(c_{\mathrm{H}}) + \tau_{P,\mathrm{H}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} = 1 + \frac{\tau_0 + \tau_{P,\mathrm{H}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})}.$$

The numerator in the second term is bounded while the denominator diverges to infinity, so

$$\frac{\tau_0 + \tau_{P,\mathrm{H}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} \to 0.$$

Therefore

$$\frac{\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} \to 1,$$

that is,

$$\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}}) \sim Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}.$$

Because both $\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})$ and $Q_{\mathrm{H}}(c_{\mathrm{H}})$ are strictly positive,

$$\frac{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})}{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}} = \frac{(\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}}))^{-1/2}}{(Q_{\mathrm{H}}(c_{\mathrm{H}}))^{-1/2}} = \left(\frac{Q_{\mathrm{H}}(c_{\mathrm{H}})}{\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})}\right)^{1/2} \to 1.$$

Hence

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}.$$

Now consider the low-compute investors. Corollary 2 states that

$$\Pi_{\mathrm{L}}(c_{\mathrm{H}}) = \frac{\tau_{P,\mathrm{L}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} \to 1.$$

Equivalently,

$$\tau_{P,\mathrm{L}}(c_{\mathrm{H}}) \sim Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}.$$

Since c_{L} is fixed, the direct low-compute precision

$$q_{2,\mathrm{L}} = f_2(c_{\mathrm{L}})\tau_{s_2}$$

is constant along the sequence. Divide the low-compute posterior precision by $Q_{\mathrm{H}}(c_{\mathrm{H}})$:

$$\frac{\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} = \frac{\tau_0 + q_{2,\mathrm{L}} + \tau_{P,\mathrm{L}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} = \frac{\tau_0 + q_{2,\mathrm{L}}}{Q_{\mathrm{H}}(c_{\mathrm{H}})} + \frac{\tau_{P,\mathrm{L}}(c_{\mathrm{H}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})}.$$

The first term converges to 0 because its numerator is constant and its denominator diverges to infinity. The second term converges to 1 by Corollary 2. Hence

$$\frac{\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{Q_{\mathrm{H}}(c_{\mathrm{H}})} \to 1,$$

so

$$\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha}.$$

Again all quantities are strictly positive, so

$$\frac{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}} = \frac{(\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}))^{-1/2}}{(Q_{\mathrm{H}}(c_{\mathrm{H}}))^{-1/2}} = \left(\frac{Q_{\mathrm{H}}(c_{\mathrm{H}})}{\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}\right)^{1/2} \to 1.$$

Therefore

$$\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}.$$

Finally, divide both RMSEs by the same benchmark sequence $C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}$:

$$\frac{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})} = \frac{\dfrac{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}}}{\dfrac{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})}{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}}} \to \frac{1}{1} = 1.$$

This proves every claim in the corollary. ■

C.13 Proof of Proposition 8

Proof. We first prove the exact conditional Sharpe-ratio identities. Consider the high-compute investors and let

$$\mathcal{I}_{\mathrm{H}} \equiv \sigma(s_1, s_{2,\mathrm{H}}, P).$$

Conditional on $\mathcal{I}_{\mathrm{H}}$, the demand x_{H} is measurable because it is chosen after observing $(s_1, s_{2,\mathrm{H}}, P)$.

If $x_{\mathrm{H}} = 0$, then Lemma 9.2 implies

$$0 = x_{\mathrm{H}} = \frac{\tau_{\mathrm{H}}^R}{A}(\mu_{\mathrm{H}}^R - P).$$

Since $A > 0$ and $\tau_{\mathrm{H}}^R > 0$, this equality forces

$$\mu_{\mathrm{H}}^R - P = 0.$$

Therefore

$$\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})\,|\mu_{\mathrm{H}}^R - P| = 0 \qquad \text{and} \qquad \frac{A|x_{\mathrm{H}}|}{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})} = 0,$$

so the claimed identity is valid in the zero-position case.

Now suppose $x_{\mathrm{H}} \neq 0$. Because

$$\mu_{\mathrm{H}}^R = \mathbb{E}\Big[\tilde{\theta} \mid \mathcal{I}_{\mathrm{H}}\Big],$$

we have

$$\mathbb{E}\Big[\tilde{\theta} - P \mid \mathcal{I}_{\mathrm{H}}\Big] = \mu_{\mathrm{H}}^R - P.$$

Since x_{H} is $\mathcal{I}_{\mathrm{H}}$-measurable,

$$\mathbb{E}\Big[x_{\mathrm{H}}(\tilde{\theta} - P) \mid \mathcal{I}_{\mathrm{H}}\Big] = x_{\mathrm{H}}\mathbb{E}\Big[\tilde{\theta} - P \mid \mathcal{I}_{\mathrm{H}}\Big] = x_{\mathrm{H}}(\mu_{\mathrm{H}}^R - P).$$

Similarly,

$$\begin{aligned}\mathbb{V}\Big[x_{\mathrm{H}}(\tilde{\theta}-P)\mid\mathcal{I}_{\mathrm{H}}\Big] &= x_{\mathrm{H}}^2\mathbb{V}\Big(\tilde{\theta}-P\mid\mathcal{I}_{\mathrm{H}}\Big)\\ &= x_{\mathrm{H}}^2\mathbb{V}\Big(\tilde{\theta}\mid\mathcal{I}_{\mathrm{H}}\Big)\\ &= x_{\mathrm{H}}^2(\tau_{\mathrm{H}}^R)^{-1}. \qquad (31)\end{aligned}$$

Taking the square root of (31) yields the conditional standard deviation

$$\sqrt{\mathbb{V}\Big[x_{\mathrm{H}}(\tilde{\theta}-P)\mid\mathcal{I}_{\mathrm{H}}\Big]}=\frac{|x_{\mathrm{H}}|}{\sqrt{\tau_{\mathrm{H}}^R}}.$$

Hence the conditional Sharpe ratio is

$$\mathrm{SR}_{\mathrm{H}}=\frac{x_{\mathrm{H}}(\mu_{\mathrm{H}}^R-P)}{|x_{\mathrm{H}}|/\sqrt{\tau_{\mathrm{H}}^R}}=\sqrt{\tau_{\mathrm{H}}^R}\,\frac{x_{\mathrm{H}}}{|x_{\mathrm{H}}|}(\mu_{\mathrm{H}}^R-P).$$

Lemma 9.2 also gives the optimal-demand formula

$$x_{\mathrm{H}}=\frac{\tau_{\mathrm{H}}^R}{A}(\mu_{\mathrm{H}}^R-P).$$

Because $\tau_{\mathrm{H}}^R/A>0$, the sign of x_{H} is the same as the sign of μ_{H}^R-P. Therefore

$$\frac{x_{\mathrm{H}}}{|x_{\mathrm{H}}|}(\mu_{\mathrm{H}}^R-P)=|\mu_{\mathrm{H}}^R-P|,$$

which yields

$$\mathrm{SR}_{\mathrm{H}}=\sqrt{\tau_{\mathrm{H}}^R}\,|\mu_{\mathrm{H}}^R-P|=\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})\,|\mu_{\mathrm{H}}^R-P|.$$

Finally, multiply the optimal-demand formula by A and take absolute values:

$$A|x_{\mathrm{H}}|=\tau_{\mathrm{H}}^R|\mu_{\mathrm{H}}^R-P|.$$

Divide both sides by $\sqrt{\tau_{\mathrm{H}}^R} = \mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})$ to obtain

$$\frac{A|x_{\mathrm{H}}|}{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})} = \mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})\,|\mu_{\mathrm{H}}^R - P| = \mathrm{SR}_{\mathrm{H}}.$$

The low-compute investors are analogous. Let

$$\mathcal{I}_{\mathrm{L}} \equiv \sigma(s_{2,\mathrm{L}}, P).$$

Conditional on $\mathcal{I}_{\mathrm{L}}$, the demand x_{L} is measurable because it is chosen after observing $(s_{2,\mathrm{L}}, P)$. If $x_{\mathrm{L}} = 0$, Lemma 9.2 implies

$$0 = x_{\mathrm{L}} = \frac{\tau_{\mathrm{L}}^R}{A}(\mu_{\mathrm{L}}^R - P),$$

so $\mu_{\mathrm{L}}^R - P = 0$, and both claimed expressions equal 0. If $x_{\mathrm{L}} \neq 0$, then

$$\mu_{\mathrm{L}}^R = \mathbb{E}\big[\tilde{\theta} \mid \mathcal{I}_{\mathrm{L}}\big]$$

implies

$$\mathbb{E}\big[x_{\mathrm{L}}(\tilde{\theta} - P) \mid \mathcal{I}_{\mathrm{L}}\big] = x_{\mathrm{L}}(\mu_{\mathrm{L}}^R - P),$$

and

$$\mathbb{V}\big[x_{\mathrm{L}}(\tilde{\theta} - P) \mid \mathcal{I}_{\mathrm{L}}\big] = x_{\mathrm{L}}^2(\tau_{\mathrm{L}}^R)^{-1}.$$

Therefore

$$\mathrm{SR}_{\mathrm{L}}^R = \sqrt{\tau_{\mathrm{L}}^R}\,\frac{x_{\mathrm{L}}}{|x_{\mathrm{L}}|}(\mu_{\mathrm{L}}^R - P).$$

Using

$$x_{\mathrm{L}} = \frac{\tau_{\mathrm{L}}^R}{A}(\mu_{\mathrm{L}}^R - P),$$

the sign of x_{L} equals the sign of $\mu_{\mathrm{L}}^R - P$, so

$$\mathrm{SR}_{\mathrm{L}}^R = \sqrt{\tau_{\mathrm{L}}^R}\,|\mu_{\mathrm{L}}^R - P| = \mathcal{S}_{\mathrm{L}}^R(c_{\mathrm{H}}, c_{\mathrm{L}})\,|\mu_{\mathrm{L}}^R - P|.$$

Taking absolute values in the demand formula also gives

$$A|x_{\mathrm{L}}| = \tau_{\mathrm{L}}^{R}|\mu_{\mathrm{L}}^{R} - P|,$$

so

$$\frac{A|x_{\mathrm{L}}|}{\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})} = \mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})\,|\mu_{\mathrm{L}}^{R} - P| = \mathrm{SR}_{\mathrm{L}}^{R}.$$

We now prove the power-law scaling claims. Suppose additionally that the common-exponent power-law specification holds, that

$$Q_{\mathrm{H}}(c_{\mathrm{H}}) = C_{\mathrm{H}} c_{\mathrm{H}}^{\alpha},$$

that c_{L} is fixed, and that $c_{\mathrm{H}} \to \infty$ along a sequence on which the stratified REE continues to obtain. By definition,

$$\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}}) = \sqrt{\tau_{\mathrm{H}}^{R}(c_{\mathrm{H}})} = \frac{1}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})},$$

and

$$\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) = \sqrt{\tau_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})} = \frac{1}{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}.$$

Corollary 3 proves that

$$\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}$$

and

$$\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}.$$

Since all of these sequences are strictly positive,

$$\frac{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})}{C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2}} = \frac{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})} \to 1,$$

so

$$\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}}) \sim C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2}.$$

Likewise,

$$\frac{\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2}} = \frac{C_{\mathrm{H}}^{-1/2} c_{\mathrm{H}}^{-\alpha/2}}{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})} \to 1,$$

which gives

$$\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}}) \sim C_{\mathrm{H}}^{1/2} c_{\mathrm{H}}^{\alpha/2}.$$

Finally,

$$\frac{\mathcal{S}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{\mathcal{S}_{\mathrm{H}}(c_{\mathrm{H}})} = \frac{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})}{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})} \to 1$$

because Corollary 3 also shows

$$\frac{\mathrm{RMSE}_{\mathrm{L}}^{R}(c_{\mathrm{H}}, c_{\mathrm{L}})}{\mathrm{RMSE}_{\mathrm{H}}(c_{\mathrm{H}})} \to 1.$$

This proves the proposition. ■

C.14 Proof of Proposition 9

Proof. By Proposition 8,

$$\mathrm{SR}_{\mathrm{H}}^{2} = \tau_{\mathrm{H}}^{R}(\mu_{\mathrm{H}}^{R} - P)^{2}, \qquad \mathrm{SR}_{\mathrm{L}}^{R\,2} = \tau_{\mathrm{L}}^{R}(\mu_{\mathrm{L}}^{R} - P)^{2}.$$

Using

$$\mu_{\mathrm{H}}^{R} - P = (\mu_{\mathrm{L}}^{R} - P) + \Delta\mu^{R}, \qquad \tau_{\mathrm{H}}^{R} = \tau_{\mathrm{L}}^{R} + \Delta\tau^{R},$$

we obtain

$$\begin{aligned}
\mathrm{SR}_{\mathrm{H}}^{2} - \mathrm{SR}_{\mathrm{L}}^{R\,2} &= \tau_{\mathrm{H}}^{R}\Big[(\mu_{\mathrm{L}}^{R} - P) + \Delta\mu^{R}\Big]^{2} - \tau_{\mathrm{L}}^{R}(\mu_{\mathrm{L}}^{R} - P)^{2} \\
&= \Delta\tau^{R}(\mu_{\mathrm{L}}^{R} - P)^{2} + 2\tau_{\mathrm{H}}^{R}(\mu_{\mathrm{L}}^{R} - P)\Delta\mu^{R} + \tau_{\mathrm{H}}^{R}(\Delta\mu^{R})^{2}.
\end{aligned}$$

Under the stated assumptions, the first term is nonnegative, the second term is nonnegative, and the third term is strictly positive because $\tau_{\mathrm{H}}^{R} > 0$ and $\Delta\mu^{R} \neq 0$. Hence

$$\mathrm{SR}_{\mathrm{H}}^{2} - \mathrm{SR}_{\mathrm{L}}^{R\,2} > 0.$$

Therefore

$$\mathrm{SR}_{\mathrm{H}} > \mathrm{SR}_{\mathrm{L}}^{R}.$$

■