Behavioral Machine Learning? Regularization and Forecast Bias

Murray Z. Frank, Jing Gao, and Keer Yang*

Abstract

Machine learning forecasts of corporate earnings fail the Coibion–Gorodnichenko test of forecast efficiency, mirroring results for human analysts. Both show underreaction at the one-year horizon and overreaction at the two-year horizon. Because the algorithms have no psychology, behavioral bias cannot explain the rejections. We show theoretically that regularization and measurement noise generate exactly this horizon pattern. Varying regularization hyperparameters moves the test coefficient monotonically. Tuning to minimize out-of-sample error still produces significant rejections. After machine learning tools became widely accessible in 2013, technically trained analysts shifted sharply toward overreaction. The Coibion–Gorodnichenko test cannot distinguish behavioral bias from standard statistical practice.

JEL Classification: G17, G32, G40

Keywords: Overreaction, underreaction, regularization, machine learning, behavioral finance

*Frank is from University of Minnesota, Minneapolis, MN, E-mail: murra280@umn.edu. Gao is from Peking University, E-mail: jinggao@phbs.pku.edu.cn. Yang is from University of California at Davis, Davis, CA, E-mail: kkeyang@ucdavis.edu. None of the authors have any relevant or material financial interests related to the research described in this paper. We are grateful for helpful comments from Jiangze Bian, Nirmol Das, Wenting Liao (discussant), Loïc Maréchal (discussant), Melina Ludolph (discussant), Ke Tang (discussant), Rui Da (discussant), Colin Ward, Andy Winton, Moto Yogo and seminar participants at the University of Minnesota, 2023 JFDS Conference, 2023 ABFER, HKUST, and SUSTech Research Conference on Capital Market Research in the Era of AI, 2023 Cardiff Fintech Conference, 2023 FMA, 2023 KAIST AI Social Science BootCamp, 2024 UC Davis Finance Day, 2024 Econometric Society conference on Economics and AI+ML, 2025 CICF, 2025 Financial Intermediation in the 3rd Millennium conference at Essex, Renmin University Alumni Conference, and 2026 Early Career Women in Finance Conference. An earlier draft of this paper circulated with the title, "Behavioral Machine Learning? Computer Predictions of Corporate Earnings also Overreact". We used Claude Opus and ChatGPT to help find and fix analytical mistakes and improve wording. We alone are fully responsible for the content of the paper and for any remaining mistakes or misinterpretations.

1 Introduction

Expectations have a central position in modern finance and macroeconomics. Since Muth (1961) and Lucas Jr (1972), the testable content of rational expectations has been that forecast errors should be unpredictable from information available when the forecast was made. The Coibion and Gorodnichenko (2015) regression of forecast errors on forecast revisions has become the workhorse test of this condition (hereafter the CG test), and it routinely rejects: forecasts of corporate earnings, inflation, and output display underreaction at some horizons and overreaction at others (Bordalo et al., 2020; Afrouzi et al., 2023; De Silva and Thesmar, 2024). The dominant interpretation attributes these rejections to psychological factors such as anchoring, extrapolation, and diagnostic expectations. The behavioral parameters estimated from these regressions are increasingly embedded in models of asset prices, credit cycles, and corporate investment (Bordalo et al., 2024, 2026). A great deal therefore rides on whether the test identifies what it is claimed to identify.

We apply the CG test to earnings forecasts produced by standard machine learning methods: ridge regression, gradient boosting, and random forests. They are all trained on the same public data available to equity analysts. The predictions generated by the algorithms fail the CG test. Their forecast errors are strongly predictable from their own revisions. The pattern of failure mirrors the pattern documented for human analysts. We find positive coefficients at the one year horizon (the signature of underreaction), and negative coefficients at the two year horizon (the signature of overreaction). The algorithms have no brain structure, no emotions, and no capacity for regret. Whatever generates these rejections, it is not human brain structure and psychology. We argue that the source is regularization interacting with measurement noise in the signals that forecasters observe. Regularization is the deliberate shrinkage toward zero that these methods apply to improve out-of-sample accuracy (Hastie et al., 2009; Hastie, 2020).

We formalize this argument in a deliberately simple model. A forecaster predicts an AR(1) fundamental that she observes through a signal contaminated by transitory noise, and she forms linear forecasts subject to a ridge penalty. The CG coefficient has a closed

form that decomposes into two terms with opposite signs. Shrinkage damps the response to news. When a genuine fundamental shock arrives, the forecast moves too little, the realization overshoots it, and errors covary positively with revisions. Transitory noise works in the other direction. The forecaster cannot distinguish noise from fundamentals. So she revises in response to shocks that will not persist, and the subsequent error takes the opposite sign from the revision. At short horizons the shrinkage term dominates and the coefficient is positive. At longer horizons noise looms larger and the coefficient turns negative. The sign reversal across horizons is challenging. Standard behavioral models can only match this by allowing bias parameters to vary with horizon.

The model helps clarify our claims. In the population problem with known moments, shrinkage is not optimal. We do not derive the penalty from primitives. Ridge becomes valuable in the finite-sample, high-dimensional environments where machine learning actually operates (Stein, 1956; Hastie et al., 2009). The model's role is to show that stronger regularization shifts the CG coefficient upward in a predictable way. All widely used machine learning methods are constructed to use forms of regularization.

Our empirical work uses IBES analyst forecasts and machine learning predictions constructed following Zhang et al. (2025), covering 99,963 firm-year observations from 1986 to 2019. At the one year horizon, the consensus analyst CG coefficient is 0.114, evidence consistent with underreaction. The machine learning coefficients are 0.006 for ridge, 0.058 for gradient boosting, and 0.044 for random forests. These are closer to zero than the human coefficient, but significantly positive. At the two year horizon every coefficient flips sign. We estimate -0.606, -0.233, and -0.237 for the three machine learning methods, compared to -0.003 for analysts. At the one year horizon, all three ML methods are at least as accurate as analysts out-of-sample. Their out-of-sample mean squared error is up to 7% below the analysts' at the one year horizon. Accuracy and orthogonality are distinct dimensions of forecast quality. The methods that beat analysts out-of-sample fail the efficiency test strongly, and at long horizons more strongly.

We systematically vary the intensity of machine learning regularization by re-estimating the same forecasting models under different hyperparameter settings. Across models, stronger regularization leads to more positive CG coefficients. Specifically, the CG coef-

ficient increases monotonically as we raise the ridge penalty in ridge regression, lower the learning rate in gradient boosting, or reduce the maximum fraction of features considered at each split in random forests. When regularization is sufficiently weak, the CG coefficient can even become negative.

Regressions of estimated CG coefficients on ranked regularization intensity confirm a significant monotonic relationship for all three methods. Importantly, the hyperparameter settings that move the CG coefficient toward zero do not generally coincide with those that minimize out-of-sample prediction error. Across models, specifications with lower out-of-sample error can continue to exhibit significant CG coefficients, while further tuning toward CG orthogonality may come at some cost in predictive performance. These results highlight that minimizing prediction error and satisfying the CG orthogonality condition are distinct objectives, and tuning a model toward one need not optimize the other.

Does this framework describe human forecasters, or only machines? While there is no direct way to establish such a connection, the calibration provides suggestive evidence. We calibrate the learning rate of a gradient boosting model so that its one year CG coefficient matches that of analysts. The calibrated learning rate is 0.0545, compared with the software default of 0.1, implying substantially stronger regularization. This suggests that analyst forecasts behave as if they are generated under stronger effective regularization than the default gradient boosting specification. Although the calibration targets only this one moment, the resulting model also delivers out-of-sample forecast accuracy close to that of analysts at both the one and two year horizons, and its two year CG coefficient has the same sign as that of analysts. We view this as a magnitude discipline exercise rather than direct evidence that analysts implicitly use gradient boosting. A regularized statistical model calibrated to match a single feature of human forecasts generates several other broadly similar forecasting patterns.

The direction of the calibrated parameter also makes a further prediction. Suppose analysts behave as if more heavily regularized than off the shelf defaults. Then adopting machine learning tools should reduce their effective regularization and push their CG coefficients downward. That is what we find.

The scikit-learn library had its first stable release in 2012, making regularized forecasting tools substantially more accessible without requiring extensive programming investment. We hand-collect analysts' educational backgrounds from LinkedIn and FINRA BrokerCheck and classify them as technically trained or non-technical based on their exposure to advanced statistics and computation. After 2013, technical analysts shift sharply toward overreaction, exhibiting a substantially larger decline in the CG coefficient, while the change among non-technical analysts is much more modest and statistically insignificant. This pattern is consistent with greater adoption of machine learning tools among technically trained analysts after 2013. In our framework, such adoption may reduce effective regularization and thereby shift the CG coefficient downward. We cannot directly observe the forecasting techniques used by individual analysts so the link is not definitive. But the evidence is suggestive. Those best positioned to adopt the new statistical technology exhibit the largest change in measured "bias," and in the direction predicted by our framework.

The noise channel generates cross-sectional predictions that uniform psychological bias does not. A behavioral bias is an attribute of a human analyst. So it should apply broadly across the firms she covers. Measurement noise relates to the firm. Noisier firms are likely to be high R&D firms and young firms. Firms whose observables are noisier signals of fundamentals should have more negative CG coefficients. In fact, they do. That is true both for human and for machine-generated forecasts. The interaction of forecast revisions with R&D intensity is negative and significant in every specification. The interaction with firm age is positive and significant, with the same signs across analyst, ridge, boosting, and forest forecasts. The same statistical forces shape human and machine forecasts in this respect.

These patterns are not merely test statistic curiosities. Following the two step approach of Bordalo et al. (2026), we project forecast errors on revisions and ask whether the predicted component forecasts real activity. It does. In the post-2013 period, a one standard deviation increase in predicted forecast error is associated with subsequent investment rate changes of roughly 10% of the mean change for firms covered by technical analysts and 8% for firms covered by non-technical analysts. To the extent that managers

respond to analyst forecasts when allocating capital, regularization-induced forecast patterns propagate into real investment decisions.

Our results build most directly on two non-behavioral accounts of forecast error predictability. De Silva and Thesmar (2024) show that transitory noise in expectations generates negative CG coefficients that strengthen with horizon. We confirm this channel and rely on it for the long horizon sign. Our contribution is the regularization channel, which is new to this literature. It can be manipulated directly through varying related hyperparameters. It also generates the positive short horizon coefficient that noise alone cannot produce.

Where prior work documents that noise exists, our framework characterizes the forecaster's shrinkage response to it and tests whether forecasters implement that response. Farmer et al. (2024) generate under and overreaction from Bayesian agents learning about hard to learn long run parameters. Our mechanism is complementary and frequentist. We have no requirement of endowed beliefs. Our mechanism operates through a penalty that modern software imposes by construction, and therefore speaks directly to the behavior of machine forecasts as they are increasingly used by financial market participants.

Martin and Nagel (2022) study how regularization by investors alters inference in market efficiency tests in high-dimensional settings. We study forecast efficiency tests and supply direct experimental variation in the penalty together with adoption timing evidence. Information friction models (Mankiw and Reis, 2002; Sims, 2003; Woodford, 2020) predict underreaction. They do not explain why algorithms violate the CG test nor why the violations reverse sign with horizon.

Bianchi et al. (2022) are also interested in regularization in their study of macroeconomic surveys. They measure belief distortions against a regularized machine learning forecast constructed from a large real-time data set. They argue that shrinkage is necessary for such a benchmark, since without it estimation error and overfitting would contaminate the comparison. Our results identify what that shrinkage costs. Regularization improves the benchmark's out-of-sample accuracy while making the benchmark's own errors predictable from its own revisions. So a forecast measured against it will appear to

respond more to news than the benchmark does. This does not alter the accuracy comparisons, which are unaffected. It does mean that a measured departure from a regularized benchmark is not by itself evidence of psychological distortion.

Our findings suggest a caution for the growing literature that treats machine learning forecasts as a benchmark for rational expectations. Zhang et al. (2025) and Van Binsbergen et al. (2023) construct machine forecasts to measure the accuracy and conditional bias of human analysts. Our results imply that such benchmarks are not neutral. The same shrinkage that delivers the accuracy advantage produces systematically predictable errors. So defining "unbiased" expectations by reference to machine output imports predictable forecast errors of known sign. Machine forecasts are a useful accuracy benchmark. But they are not a good benchmark for forecast error orthogonality.

Two implications follow. First, a nonzero CG coefficient alone identifies neither irrationality nor incomplete information. A positive coefficient is consistent with cognitive constraints, sticky information, or strong shrinkage by a sophisticated forecaster. A negative coefficient is consistent with diagnostic expectations, or transitory measurement noise. The kinds of variation we exploit here help to distinguish these mechanisms: variation in regularization intensity, in signal quality, and in forecaster type.

Second, as machine learning adoption spreads through forecasting, measured CG coefficients should drift in a predictable direction. They are likely to shift toward overreaction because off the shelf tools relax the heavier implicit regularization that human forecasters appear to apply. That drift is already visible in the post-2013 behavior of technically trained analysts. Predictive accuracy and forecast error orthogonality measure different things. Low out-of-sample loss does not imply a zero CG coefficient. Inferences about rationality drawn from orthogonality tests rest on assumptions about the forecasting technology that no longer describe how forecasts are made.

The rest of the paper is organized as follows. Section 2 presents our theory showing how regularization affects forecasts. Section 3 provides the data sources and describes the basic empirical methodology. Section 4 gives our main evidence. Section 5 provides regularization evidence. Section 6 provides noise evidence. Section 7 discusses real economic impact. The conclusion is in Section 8.

2 Theoretical Framework

The Coibion and Gorodnichenko (2015) test is regarded as a workhorse test for full information rational expectations, and it is essentially an orthogonality test: it asks whether forecast errors are predictable from forecast revisions. When the test rejects, a standard interpretation is that the forecaster fails to process available information efficiently. This interpretation is appealing, but it rests on an implicit premise that a full-information rational-expectations forecaster's errors are always unpredictable from available information. Machine learning prediction methods violate this premise by design. They impose regularization (Stein, 1956; Hastie et al., 2009), deliberately shrinking coefficient estimates toward zero and accepting shrinkage bias in those estimates in exchange for lower variance. The objective is out-of-sample accuracy, not the orthogonality condition, and the two do not coincide.

Four properties of a forecast are used extensively. Our argument requires keeping the distinctions clear. Rational expectations and its full information version restrict beliefs. That is, the forecaster's subjective conditional distribution coincides with the objective distribution given her information set. We reserve 'rational' for this restriction on beliefs, not for a forecasting procedure that minimizes a statistical criterion. Forecast efficiency is the weaker requirement that forecast errors be orthogonal to forecast revisions. This is what the Coibion and Gorodnichenko (2015) regression tests. It is one implication of full information rational expectations rather than a restatement of it. Unbiasedness requires only that forecast errors average to zero. A forecast can satisfy this and still have errors that are predictable from its own revisions. Accuracy is predictive loss, measured here by out-of-sample mean squared error. A forecasting rule can rank first on accuracy and still fail the efficiency condition. That is the case we study.

Using a simple theoretical framework, this section shows that a forecaster using machine learning models with regularization will fail the Coibion and Gorodnichenko (2015) test in systematic ways. We proceed in four steps. We first describe the test. We then derive a baseline model with a forecaster using regularization to make forecasts in a noiseless environment. This baseline isolates the regularization channel and shows that

shrinkage alone is sufficient for the test to reject. We next add measurement noise to the signal and derive a closed-form expression for the test coefficient, which decomposes into a regularization component and a noise component that enter with opposite signs. We close with comparative statics and a list of empirical predictions that organize the analysis in Sections 4-6.

The theory is deliberately parsimonious. We do not derive a particular regularization parameter from primitives. The model is a simple representation of regularized forecasting, and we use it to clarify how shrinkage and noise enter the forecast efficiency regression. The point is not that any particular degree of regularization is rational. The point is that whenever a forecaster applies shrinkage in a noisy environment, the Coibion and Gorodnichenko (2015) coefficient is nonzero in a theoretically predictable way.[1]

2.1 The Coibion and Gorodnichenko (2015) Test

Consider a forecaster who wants to predict the future outcome y_{t+1} at date t. Let $F_t y_{t+1}$ denote the forecast made at time t and $F_{t-1} y_{t+1}$ the forecast made one period earlier. The Coibion and Gorodnichenko (2015) test regresses forecast errors on revisions, where the revision proxies for the news the forecaster receives at time t:

$$e_{t+1} = \gamma_0 + \gamma_{CG}\, r_t + \omega_t, \tag{1}$$

where $e_{t+1} = y_{t+1} - F_t y_{t+1}$ denotes the forecast error and $r_t = F_t y_{t+1} - F_{t-1} y_{t+1}$ the forecast revision.

Under rational expectations with perfect information and no regularization, $\gamma_{CG} = 0$, implying that forecast revisions are orthogonal to subsequent errors. A positive coefficient is commonly interpreted as 'underreaction', meaning that forecasts adjust too little to new information. Conversely, a negative coefficient indicates 'overreaction', where forecasts respond too aggressively to news. Our contribution is to show that such pat-

[1] Other mechanisms could also generate a nonzero CG coefficient. Our aim is not to enumerate them. The purpose of the model is to show that the regularization commonly used in machine learning produces forecasts that violate the Coibion and Gorodnichenko (2015) test. The sign and magnitude of the violation depend on signal structure in predictable ways.

terns need not reflect behavioral biases. The regularization built into machine learning algorithms shrinks forecasts away from the unregularized projection, trading shrinkage bias in the estimated coefficients against lower estimation variance. This attenuation produces a positive CG coefficient even when the forecaster is minimizing the penalized criterion that her forecasting method imposes.

2.2 Baseline Model: Regularization without Measurement Noise

We begin with the cleanest possible environment. Let y_{t+1} be the realized outcome variable to be forecast. The realized outcome variable depends on a fundamental s_{t+1} plus idiosyncratic noise:

$$y_{t+1} = \alpha s_{t+1} + \epsilon_{t+1}, \tag{2}$$

where $\alpha > 0$ and $\epsilon_{t+1} \sim$ i.i.d.$(0, \sigma_\epsilon^2)$ is independent of fundamentals.

The fundamental s_t follows a stationary AR(1) process,

$$s_{t+1} \;=\; \rho_s s_t \;+\; v_{t+1}, \qquad v_{t+1} \sim \text{i.i.d.}(0, \sigma_v^2), \qquad |\rho_s| < 1, \tag{3}$$

with unconditional variance $\sigma_s^2 = \sigma_v^2/(1-\rho_s^2)$.

In the baseline setting, the forecaster observes the fundamental without measurement noise and applies a ridge (L2) penalty to the predictive coefficient. Let z_t denote the signal that forecasters observe:

$$z_t = s_t. \tag{4}$$

At time t, forecasters observe z_t and use it to forecast y_{t+1}. In the absence of measurement noise, the current signal z_t equals the fundamental and captures all relevant information for predicting y_{t+1}. We restrict the forecasting rule to the linear form:

$$F_t y_{t+1} = \beta z_t. \tag{5}$$

Without regularization, the MSE-minimizing coefficient within this class is the population projection $\beta_{OLS} = \alpha\rho_s$ and $F_t y_{t+1} = \beta_{OLS} z_t$. With an L2 penalty, the coefficient

instead solves

$$\min_{\beta} \ \mathbb{E}\big[(y_{t+1} - \beta z_t)^2\big] \ + \ \lambda\beta^2, \tag{6}$$

producing

$$\beta_\lambda \ = \ \alpha\rho_s \frac{\sigma_s^2}{\sigma_s^2 + \lambda}. \tag{7}$$

Regularization shrinks the coefficient toward zero relative to the unregularized case, so that $|\beta_\lambda| < |\beta_{\text{OLS}}|$ for any $\lambda > 0$. The forecast under regularization is therefore $F_t y_{t+1} = \beta_\lambda z_t$.

In the population problem with known moments, there is no estimation variance to trade against bias: the unregularized projection minimizes mean squared forecast error, and ridge is not optimal. Ridge earns its keep in finite-sample, high-dimensional problems (Stein, 1956; Hastie et al., 2009), which is where machine learning algorithms operate. In this model, we treat λ as an exogenous shrinkage parameter and ask what this imposed rule implies for the Coibion and Gorodnichenko (2015) regression.

Proposition 2.1 (CG coefficient under regularization, noiseless signals)**.** *Under the environment of Section 2.2, the Coibion and Gorodnichenko (2015) coefficient defined by Equation (1) satisfies*

$$\gamma_{CG} \ = \ \frac{\lambda}{\sigma_s^2} \ \geq \ 0, \tag{8}$$

with equality if and only if $\lambda = 0$.

Proof. See Appendix A.2. □

Proposition 2.1 presents our main message. In the simplest forecasting environment, imposing the regularized forecasting rule is by itself enough to generate a nonzero Coibion and Gorodnichenko (2015) coefficient. The coefficient scales linearly in the penalty λ and inversely in the variance of the fundamental σ_s^2: stronger shrinkage produces a larger coefficient, and a more variable fundamental produces a smaller one.

Two implications follow. First, the rejection has nothing to do with belief formation. The forecaster in Proposition 2.1 is a linear projection with a ridge penalty. She does not anchor, extrapolate, or overweight salient signals. She applies a standard shrinkage rule and, in doing so, generates forecast errors that are predictable from her own revisions.

Second, the coefficient is bounded below by zero. Regularization alone can produce apparent underreaction, but not overreaction. Yet the data deliver overreaction at longer horizons, with γ_{CG} turning negative for both human and machine forecasts (Table 2). Thus, that sign must come from a force outside the penalty term. The next subsection shows that measurement noise supplies it.

2.3 Regularization with Measurement Noise

The outcome process is the same as in the baseline case. However, we now assume that the forecaster cannot observe the fundamental directly; instead, she observes only a noisy signal equal to the fundamental plus measurement noise. The signal is given below

$$z_t \;=\; s_t \;+\; \eta_t, \tag{9}$$

where the noise follows an AR(1) process,

$$\eta_t \;=\; \rho_\eta \eta_{t-1} \;+\; u_t, \qquad u_t \sim \text{i.i.d.}(0, \sigma_u^2), \qquad |\rho_\eta| < 1, \tag{10}$$

with unconditional variance $\sigma_\eta^2 = \sigma_u^2/(1-\rho_\eta^2)$. The signal-to-noise ratio σ_s^2/σ_η^2 captures the quality of information. All innovations (v_t, u_t, ϵ_t) are mutually independent, so $\mathrm{Cov}(s_t, \eta_s) = 0$ for all t, s. We retain the AR(1) structure for simplicity. The AR(1) coefficients ρ_s and ρ_η govern the persistence of fundamentals and noise respectively. Empirically, measurement noise in earnings signals is typically transient and resembles white noise, so $\rho_s \gg \rho_\eta$ is the relevant case.

At time t, forecasters observe z_t and use it to forecast y_{t+1}. As in the noiseless case, we again restrict the forecasting rule to the linear form $F_t y_{t+1} = \beta z_t$. This specification restricts the forecaster to depend on z_t alone, and linearly so, rather than on the full history of past signals $(z_t, z_{t-1}, \dots)$. This is a restriction we impose for analytical tractability.[2]

[2]Imposing this restriction for analytical tractability does not drive our main conclusions. First, our baseline framework, which assumes full information and no measurement noise, generates our central prediction that regularization induces underreacting forecasts and results in nonzero Coibion and Gorodnichenko (2015) coefficients whose magnitudes vary systematically with the degree of regularization. Second, we conducted a numerical simulation of the model with measurement noise in which the machine learning

With this caveat, we show that a nonzero Coibion and Gorodnichenko (2015) coefficient does not necessarily arise from irrationality or incomplete information: regularization and noise alone can generate both under- and overreaction, depending on the relative magnitude of these two forces.

The forecaster applies the same shrinkage logic as in Section 2.2, with the only change being that the variance of the signal now includes the noise term. The ridge coefficient becomes

$$\beta_\lambda \;=\; \alpha\rho_s \frac{\sigma_s^2}{\sigma_s^2+\sigma_\eta^2+\lambda}. \tag{11}$$

Proposition 2.2 (CG coefficient under regularization with measurement noise)**.** *Under the environment of Section 2.3, the Coibion and Gorodnichenko (2015) coefficient defined by Equation (1) at the one-period horizon is*

$$\gamma_{CG} \;=\; \frac{(1-\rho_s^2)\lambda \;-\; \rho_s(\rho_s-\rho_\eta)\sigma_\eta^2}{(1-\rho_s^2)\sigma_s^2 \;+\; (1+\rho_s^2-2\rho_s\rho_\eta)\sigma_\eta^2}. \tag{12}$$

The first term in the numerator is nonnegative and is strictly positive when $\lambda > 0$. It captures the contribution of regularization. The second term is subtracted from the coefficient and is positive whenever $\rho_s > \rho_\eta$. It captures the contribution of measurement noise. The sign of γ_{CG} depends on the relative magnitudes of the two terms. The expression reduces to Proposition 2.1 when $\sigma_\eta^2 = 0$.

Proof. See Appendix A.3. The derivation expands $\mathrm{Cov}(e_{t+1}, r_t)$ and $\mathrm{Var}(r_t)$ using the AR(1) structure and applies independence to eliminate cross-terms. □

Proposition 2.2 shows that two forces are at work and operate in opposite directions. The first is regularization, captured by the term $(1-\rho_s^2)\lambda$ in the numerator of Equation (12). By shrinking her response to the signal, the forecaster underreacts. When a true fundamental shock arrives, the forecaster revises the forecast upward, but by too little. The

forecasts are given full information (reported in Internet Appendix Table IA10). We again find that greater regularization increases the CG coefficient, whereas higher noise decreases it. Together, these findings indicate that our main conclusions are not driven by the simplifying assumptions imposed for analytical tractability. Third, on the empirical side, Table 3 reports the CG test results when the machine learning forecasts are constructed using both contemporaneous and lagged predictors. The resulting patterns of underreaction and overreaction are similar to those in our main specification, consistent with the alternative theoretical framework that allows historical information to enter the prediction model.

realized outcome exceeds the forecast, producing a positive forecast error. Revisions and errors therefore covary positively, and regularization contributes positively to γ_{CG}.

The second force is measurement noise, captured by $-\rho_s(\rho_s - \rho_\eta)\sigma_\eta^2$ in the numerator of Equation (12). Given persistent fundamentals ($\rho_s > 0$), its sign is set by the persistence gap $\rho_s - \rho_\eta$: the force is negative when $\rho_s > \rho_\eta$ and grows in magnitude as the gap widens.

In this paper, we focus on corporate earnings, for which, empirically, fundamentals are persistent while measurement noise is close to white noise ($\rho_\eta \approx 0$). When such transient noise enters the signal, the forecaster cannot distinguish it from a persistent fundamental shock and revises her forecast accordingly. Because the noise does not persist into the outcome, the revision overshoots, and the subsequent error takes the opposite sign from the revision: a noise-driven upward revision is followed by a negative error, a downward revision by a positive one. Noise therefore makes revisions and errors covary negatively, contributing negatively to γ_{CG}.[3]

Overall, this model incorporates both regularization and measurement noise. The regularization channel is new to this literature and speaks to the growing adoption of machine learning in earnings forecasting. Because the choice of penalty is an explicit modeling decision that forecasters make and can vary, the channel is directly observable: we can use hyperparameters governing regularization intensity, change it, and trace its effect on γ_{CG}. The noise channel is not new. It is central to De Silva and Thesmar (2024), and it belongs to a broader set of non-behavioral explanations for forecast error predictability, including the Bayesian learning mechanism of Farmer et al. (2024). We include it for two reasons. It is necessary: regularization alone yields $\gamma_{CG} \geq 0$ and cannot account for the negative coefficients we observe at longer horizons. And it is separable: because the two forces enter Equation (12) through distinct terms, we can isolate the regularization channel empirically by holding signal quality fixed and varying the penalty. Our contri-

[3]The negative effect of noise on the Coibion and Gorodnichenko (2015) coefficient is the mechanism emphasized by De Silva and Thesmar (2024). Our framework differs primarily in how this noise is modeled. Specifically, De Silva and Thesmar (2024) introduce reporting (or output) noise, whereas we model noise in the inputs to the forecasting algorithm. Incorporating both reporting noise and regularization into the analytical model substantially increases the algebraic complexity, so our main theoretical framework focuses on input noise. In Internet Appendix B Table IA11, we extend the model by incorporating reporting noise following De Silva and Thesmar (2024). The qualitative results remain unchanged: regularization increases the CG coefficient, while both reporting noise and input noise reduce it.

bution to the noise channel is not to document that noise exists, but to characterize how a regularized forecasting rule responds to it and to test whether forecasters' errors behave accordingly. Neither channel excludes behavioral forces. Our claim is that a nonzero γ_{CG} cannot by itself distinguish among them.

2.3.1 Comparative Statics and Implications for Empirical Tests

The closed-form expression for γ_{CG} in Proposition 2.2 provides a set of testable predictions on how γ_{CG} varies with regularization and noise. Below we discuss two key implications that guide our empirical analysis.

Proposition 2.3 (Regularization Intensity)**.** *As λ increases, the CG coefficient increases.*

Proposition 2.3 implies that stronger regularization leads to more pronounced underreaction and thus more positive γ_{CG}. Intuitively, greater shrinkage causes forecasters to place less weight on new information, systematically attenuating revisions.

Proposition 2.4 (Noise Volatility)**.** *The CG coefficient is more negative when noise volatility σ_η is higher.*

Arithmetically, a higher σ_η^2 decreases the numerator, $-\rho_s(\rho_s-\rho_\eta)\sigma_\eta^2$, and increases the denominator, $(1+\rho_s^2-2\rho_s\rho_\eta)\sigma_\eta^2$, resulting in a smaller CG coefficient. Under the empirically relevant case where noise is transient (i.e., $\rho_\eta \approx 0$), greater noise volatility worsens signal quality. Consequently, forecasters place excessive weight on contemporaneous signals that are noisier measures of the fundamental. This overreliance induces overreaction, leading to a more negative correlation between forecast revisions and subsequent forecast errors.

Our model is static. Proposition 2.2 characterizes the CG coefficient for a single forecasting environment, summarized by the signal-to-noise ratio and the penalty. We do not model the one- and two year horizons as iterates of one dynamic system. Following De Silva and Thesmar (2024), who impose no restrictions across horizons, we treat each horizon as a separate environment and let the horizon enter through the parameters: transitory measurement noise is a larger share of the observed signal at longer horizons

— their upward-sloping term structure of noise, and by Proposition 2.4 the CG coefficient is lower. The empirical reversal from positive at one year to negative at two years is thus a comparative static in the noise ratio, not a claim about the dynamics of a fixed-parameter model.

2.3.2 Empirical Implications

Our theoretical framework demonstrates that a nonzero γ_{CG} does not necessarily reflect forecasters' irrationality. In much of the existing literature, $\gamma_{CG} \neq 0$ estimated at the individual firm level is interpreted as evidence of behavioral bias (Bordalo et al., 2020; Gulen et al., 2024). In contrast, we demonstrate that such patterns can arise from regularization, a standard statistical procedure used to improve out-of-sample performance. Researchers must therefore distinguish regularization from behavioral bias. A bias-variance tradeoff resolved in the presence of noisy signals is fundamentally different from systematically misusing available information or reacting to news with excessive optimism or pessimism. The distinction matters more as financial institutions increasingly delegate forecasting to machine learning algorithms, whose shrinkage is built in by construction rather than chosen by a forecaster.

These theoretical insights guide our empirical analysis in Sections 4-6. We test the model's predictions using three sources of variation based on the predictions discussed above: (i) changes in regularization hyperparameters of machine learning models, (ii) time-series variation in the adoption of machine learning tools and thus the degree of regularization, and (iii) cross-sectional differences in firm characteristics related to the magnitude of measurement noise.

3 Data and Empirical Methods

Our empirical work uses analyst earnings forecasts from IBES, firm financial data from Compustat and CRSP, macroeconomic variables from the Federal Reserve Bank of Philadelphia, and manually collected analysts' educational backgrounds and employment histories from LinkedIn and FINRA BrokerCheck records. The sample starts in 1986, reflecting

the availability of IBES data, and ends in 2019 to avoid the effects of the COVID period.

3.1 Data Sources and Sample Construction

Annual earnings per share (EPS) forecasts and realizations are obtained from IBES. In particular, realized earnings are obtained from the IBES Actual files rather than Compustat due to differences in accounting methods for measuring actual earnings. The IBES database provides firm-level forecasts made by individual equity analysts. We use analyst forecasts of annual EPS at horizons of one, two, and three years. The forecasting horizon is defined as the number of months between the fiscal year end of the realized annual EPS to be predicted and the month in which the forecast is issued. For each firm and month, we compute the consensus forecast as the median of individual analyst forecasts. We exclude financial firms (SIC codes 6000-6999) and require non-missing data for prediction variables. Our final sample includes 99,963 firm-year observations covering 1986 to 2019.

In our empirical evidence, we examine whether analyst technical training affects the characteristics of their forecasts following ML adoption. To do this, we hand-collected analysts' educational backgrounds from LinkedIn profiles and FINRA BrokerCheck records. We restrict the sample to analysts who issued at least one earnings forecast in 2018 and then identify their LinkedIn profiles,[4] cross-referencing with FINRA BrokerCheck to improve matching accuracy. This procedure yields 858 successfully matched analysts.

We classify analysts as "technical" based on their educational background. Specifically, we manually review the curriculum requirements for each major and designate a major as technical if it involves training in advanced statistics, linear algebra, and computational methods. The majors classified as technical are primarily STEM majors, though with some variation. For example, general biology programs typically do not require advanced statistical coursework and are therefore excluded, whereas degrees in computational biology are classified as technical because they incorporate substantial training in statistical and computational methods. We further augment this classification with self-reported technical skills from LinkedIn profiles, such as machine learning, artificial

[4] Figure A1 provides an example of an analyst's LinkedIn profile.

intelligence, and advanced statistical analysis.

In total, we identify 173 analysts with technical education backgrounds and 685 non-technical analysts. For each firm month, we calculate separate median consensus forecasts for technical analysts ($F_t^{Tech} y_{i,t+1}$) and non-technical analysts ($F_t^{Non-Tech} y_{i,t+1}$). For this part, our sample covers 1994 to 2018, producing 10,377 firm-year observations for forecasts made by technical analysts and 11,294 firm-year observations for forecasts made by non-technical analysts.

3.2 Machine Learning Forecasts

ML-based earnings forecasts are constructed following the methodology in Zhang et al. (2025). Our predictor variables include: (i) financial ratios from the Wharton Research Data Services (WRDS) Financial Ratio suites, (ii) the last available annual earnings per share from IBES, (iii) the consensus analyst forecast of EPS, and (iv) macroeconomic variables from the Philadelphia Fed's real-time dataset (Real Gross Domestic Product, Real Personal Consumption Expenditures, Industrial Production Index, and Civilian Unemployment Rate). For variables with a large number of missing observations, we replace missing values with the industry average.[5]

We implement three machine learning forecasting methods, each of which applies regularization either explicitly or implicitly: ridge regression, gradient boosting, and random forests. Ridge regression shrinks coefficient estimates through an L2 penalty parameter λ. Its implementation parallels the theoretical framework in Section 2, where the coefficient minimizing the penalized population criterion is

$$\beta_\lambda = \alpha \rho_s \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\eta^2 + \lambda}, \tag{13}$$

where $\lambda \geq 0$ is the regularization penalty.

Gradient boosting builds forecasts sequentially, with each tree correcting errors from previous trees. The learning rate controls regularization intensity. Random forests average predictions across many decision trees, with regularization operating through the

[5]Internet Appendix Table IA2 reports the predictors used for the machine learning forecasts.

maximum number of features considered at each split, the minimum leaf size, and the maximum tree depth, etc. For each method, we estimate models using rolling windows to avoid look-ahead bias.

In our main specification, we use a combination of hyperparameter values commonly adopted in the literature and values selected based on our own tuning procedure. Specifically, the hyperparameters governing regularization are as follows. (1) For ridge regression, the regularization penalty is set to alpha = 30 (i.e. λ in the equation above),[6] as it provides the best overall out-of-sample performance across one- and two-year-ahead earnings forecasts in our hyperparameter tuning procedure. (2) For gradient boosting, the learning rate is set to 0.1, the default value in scikit-learn and a commonly used choice in empirical finance applications (Gu et al., 2020; Li et al., 2025). (3) For random forest, the minimum number of samples at a leaf node is set to 5, and the maximum features is 1.0, following Zhang et al. (2025).[7] We rely on these hyperparameter values because our forecasting tasks span multiple horizons, including both short- and long-term earnings forecasts, and the values reported in the literature are intended to perform well across a broad range of prediction settings. In subsequent analyses, we vary these hyperparameters to examine the extent to which predictive performance improves under alternative tuning choices. All remaining hyperparameters are set to their default values in scikit-learn (version 1.3.0).

The machine learning training and prediction procedures are designed to mimic the forecasting process of equity analysts, who issue monthly forecasts of firms' annual EPS. For each month, we train a machine learning model for the one-year-ahead forecast using the previous 12 months of predictor variables and EPS data, and then apply the trained model to the current month's predictor variables to generate a one-year-ahead earnings forecast. We use a similar procedure for the two-year-ahead and three-year-ahead fore-

[6] The ridge penalty, denoted λ in Section 2, is the parameter scikit-learn calls alpha. When reporting hyperparameter settings in the tables and figures, we retain the name alpha to facilitate replication with scikit-learn; alpha and λ refer to the same quantity and should not be confused with the loading α in the theoretical model.

[7] Ridge regression provides a simple benchmark model, and consequently its default regularization parameter is rarely adopted in the literature. In contrast, gradient boosting regression trees are well suited to financial prediction problems, and the default learning rate of 0.1 is commonly employed in empirical finance studies.

casts, relying on the previous 24 months and 72 months of data, respectively, to ensure sufficiently large training samples.[8] Internet Appendix A provides additional details on the machine learning forecasting procedure.

3.3 Variable Definitions

Forecast error. For firm i in fiscal year t, the one year ahead forecast error is

$$e_{i,t+1} = y_{i,t+1} - F_t y_{i,t+1} \tag{14}$$

where $y_{i,t+1}$ denotes realized earnings per share from the IBES Actual files and $F_t y_{i,t+1}$ represents the consensus forecast for year $t+1$ earnings made in fiscal year t. We measure $F_t y_{i,t+1}$ as the first forecast of $y_{i,t+1}$ released in fiscal year t after the announcement of year t's actual EPS, $y_{i,t}$. Two year ahead forecast errors are defined analogously.

Forecast revision. The forecast revision is the change in the consensus forecast of fiscal year $t+1$ earnings, from the forecast made in year $t-1$ to that made in year t:

$$r_{i,t} = F_t y_{i,t+1} - F_{t-1} y_{i,t+1}. \tag{15}$$

We measure $F_{t-1} y_{i,t+1}$ as the first forecast of $y_{i,t+1}$ released in fiscal year $t-1$ after the announcement of year $t-1$'s actual EPS, $y_{i,t-1}$.

Signal quality proxies. Following the literature, we proxy for signal quality using firm age and R&D intensity (R&D expenditure scaled by total assets). Higher R&D intensity and lower firm age are interpreted as proxies for noisier signals, corresponding to higher σ_η in Equation 12.

3.4 Empirical Strategy

Our empirical strategy uses complementary sources of variation to distinguish the regularization and measurement-noise channels developed in Section 2. We first document

[8] We use longer training windows for longer-horizon forecasts because IBES provides fewer observations at these horizons, so a wider window is needed to ensure sufficient training data.

the baseline CG-test patterns for human and machine learning forecasts. We then examine whether these patterns vary with regularization and signal quality in the directions predicted by Propositions 2.3 and 2.4.

Our primary test of the regularization channel exploits controlled variation in machine learning hyperparameters. Holding the algorithm, predictor set, and training sample fixed, we vary the ridge penalty, the gradient boosting learning rate, and the maximum features parameter in random forests. These hyperparameters govern regularization intensity in their respective models. Proposition 2.3 predicts that stronger regularization should produce a more positive γ_{CG}. Because this exercise changes regularization while holding the forecasting environment otherwise fixed, it provides our most direct evidence on the regularization mechanism.

We then examine whether the same mechanism can help explain time-series variation in human analysts' forecasts associated with the adoption of machine learning tools. We examine changes in analysts' CG coefficients around the widespread availability of machine learning tools after 2013 and use analysts' technical training to proxy for differential exposure to these tools. If technically trained analysts adopt machine learning methods earlier or more intensively, their forecasts should become more similar to machine learning forecasts after 2013. These analysts should therefore exhibit sharper changes in forecasting behavior, with their CG coefficients moving more strongly in the direction predicted by machine learning models.

Finally, we test the measurement-noise channel using cross-sectional variation in signal quality. Proposition 2.4 predicts more negative CG coefficients when signals are noisier. We proxy for signal quality using R&D intensity and firm age and test whether forecast revisions predict subsequent errors differently across these firm characteristics for both human and machine learning forecasts.

3.5 Summary Statistics

Table 1 reports summary statistics for the main sample. Panel A shows firm characteristics including EPS, book-to-market ratio, return on assets, and debt-to-assets ratio. Panel

B shows forecast levels for human analysts and the three ML methods. Mean forecasts range from 1.322 (random forests) to 1.436 (human analysts). Panel C shows forecast errors, defined as realized minus predicted EPS. The mean errors are -0.214 for human analysts, -0.169 for ridge regression, -0.110 for gradient boosting, and -0.100 for random forests over the full 1986 to 2019 period, suggesting that all are overly optimistic. Panel D shows forecast revisions, defined as the EPS forecast at time t minus that at time $t-1$. The mean revisions are -0.266 for human analysts, -0.053 for ridge regression, -0.020 for gradient boosting, and -0.003 for random forests over the full 1986 to 2019 period.

4 Machine Learning Forecasts Fail the CG Test

In this section, we show that ML forecasts systematically violate forecast efficiency tests in ways that closely mirror the patterns observed for human analysts.

4.1 Forecast Efficiency Violations

In this section, we test whether ML forecasts satisfy the Coibion and Gorodnichenko (2015) forecast efficiency condition, given in Equation 1. Under rational expectations with perfect information and no regularization, forecast revisions should be uncorrelated with subsequent forecast errors. All specifications include firm and year fixed effects with standard errors clustered by firm.[9]

Table 2 Panel A shows the results for one-year-ahead forecasts. Column (1) shows human analysts have a significant positive coefficient of 0.114 (t-statistic 14.04). This pattern is commonly interpreted in the literature as underreaction. It says that forecasters do not revise enough when new information arrives.

Columns (2) through (4) apply the same test to forecasts produced by ML algorithms. Ridge regression has a coefficient of 0.006 (t-statistic 1.72) which is nearly zero and barely significant. Gradient boosting gives 0.058 (t-statistic 6.37). Random forests gives 0.044

[9]We cluster standard errors at the firm level following Bordalo et al. (2026). The results remain robust when standard errors are not clustered (Internet Appendix Table IA7) and when fixed effects are excluded (Internet Appendix Table IA8).

(t-statistic 5.36). All ML methods exhibit significantly smaller coefficients than human analysts.

The economic magnitudes are large. A one-unit increase in human forecast revision predicts a $0.114-per-share increase in the EPS forecast error. For gradient boosting, the same revision predicts only a $0.058-per-share increase in forecast error, a 49% reduction. The out-of-sample MSE values at the bottom of Table 2 show that these efficiency patterns are not artifacts of overfitting. Ridge, gradient boosting, and random forests produce MSE of 0.918, 0.901, and 0.892 respectively. All of these out-of-sample evaluations are lower than human analysts' 0.959. The lower MSE supports the use of regularization, which both theory and empirical evidence suggest can lead to violations of the CG test.[10]

Table 2 Panel B extends the analysis to two-year-ahead forecasts, examining the characteristics of forecasts over a longer horizon. Human analysts produce a negative coefficient, but close to zero (-0.003, t-statistic -0.24). ML models also yield negative coefficients, though more significant: -0.606 for ridge (t-statistic -27.41), -0.233 for gradient boosting (t-statistic -19.05), and -0.237 for random forests (t-statistic -19.68). These negative coefficients imply that both human analysts and machine learning predictions produce stronger overreaction (weaker underreaction) at a longer horizon.

Both human and machine predictions exhibit a systematic sign change in their CG coefficients: positive at the one year horizon and more negative at the two year horizon. This is consistent with the comparative statics in Section 2.3.1, where two opposing forces affect the coefficient. First, the regularization shrinks coefficients toward zero, underweighting current signals. This contributes to a positive γ_{CG}, i.e. 'underreaction' to fundamentals. Second, noise in the inputs induces a mechanical negative correlation between forecast errors and forecast revisions, generating 'overreaction' and contributing to a negative γ_{CG}.

At the one year horizon, the regularization force is relatively stronger, producing pos-

[10] Although our primary focus is not on forecast accuracy, these results indicate that regularized forecasting methods deliver competitive performance, providing a clear incentive for forecasters to adopt them. Figure 1 plots the share of observations for which machine learning forecasts outperform human analysts. On average, machine learning algorithms outperform analysts, although this advantage diminishes in the later years of the sample, when analysts may have begun adopting machine learning tools more widely. This convergence highlights the importance of understanding the systematic forecast error patterns generated by machine learning methods as these tools become more widely adopted.

itive coefficients. At the two year horizon, by contrast, the link between current signals and future earnings is more complex, and measurement noise rises substantially, as supported by De Silva and Thesmar (2024). The noise effect then dominates, yielding a more negative γ_{CG}. Table 2 Panel B corroborates this interpretation: both human and ML forecast errors are roughly twice as large at the two year horizon, with MSE increasing nearly twofold to fourfold across the ML methods. In addition, the large difference in γ_{CG} magnitudes across machine learning methods also provides suggestive evidence for the measurement noise argument. The ridge regression coefficient of -0.606 is more than 2.5 times larger in absolute value than the coefficient of the gradient boosting model (-0.233). This gap likely reflects ridge's linear structure, which applies uniform shrinkage and lacks the nonlinear adaptivity of tree-based methods. Gradient boosting and random forests, by contrast, can learn different relationships at different horizons, partially offsetting the noise-driven overreaction.

These horizon effects are consistent with De Silva and Thesmar (2024), who show that measurement noise produces negative γ_{CG} that strongly decreases with forecast horizon. We incorporate measurement noise because regularization alone does not explain negative γ_{CG} at longer horizons. We contribute to the literature by showing that machine learning forecasts have the tendency to underreact to current news, which could distort capital allocation outcomes if firms' and investors' decisions are guided by these 'biased' expectations.[11]

4.2 Robustness to Information Sets

To examine whether the CG-test failures documented above are driven by a particular set of predictors, we re-estimate the tests using alternative information sets. Table 3 reports the results: Panel A for one-year-ahead forecasts, Panel B for two-year-ahead.

In Columns (1) to (3) of both Panels A and B, machine learning predictions are gener-

[11]The Coibion and Gorodnichenko (2015) test is the dominant testing approach in the recent literature, but other tests have also been used. For example, Bordalo et al. (2018) use a simple levels test that regresses forecast errors on the level of signals rather than on changes in signals. In Appendix Table IA12, we implement this levels test using investment rates and net debt issuance as signals. The results show that machine learning forecasts also exhibit overreaction at longer horizons.

ated using a more limited set of predictors that excludes WRDS financial ratios, in contrast with Table 2, which uses the full set. At the one year horizon, limiting the predictor set has modest effects on CG coefficients. The CG coefficients of gradient boosting and random forest predictions are particularly stable across specifications. The gradient boosting CG coefficient is 0.058 with the full predictor set and about 0.06 with limited predictors. Similarly, the random forest coefficient is 0.044 with full predictors and 0.042 with limited predictors. The ridge coefficient is also relatively stable and remains close to zero, but varies somewhat more across specifications. This greater stability of the tree-based models is consistent with their capacity to internally select relevant features. These patterns extend to the two year horizon, where the machine learning predictions continue to exhibit overreaction. The ridge CG coefficient becomes less negative, increasing from –0.606 to –0.190. The gradient boosting coefficient changes only marginally, from –0.233 to –0.226, while the random forest coefficient remains similarly stable, changing from –0.237 to –0.248.

In Columns (4) through (9) of both Panels A and B, the machine learning predictions are generated using a richer set of predictors. In Columns (4), (5), and (6), the predictor set includes the baseline predictors from Table 2 as well as realized EPS lagged one year relative to the latest available EPS. In Columns (7), (8), and (9), the predictor set includes the baseline predictors from Table 2 together with one year lags of all baseline predictors. At the one year horizon, the CG coefficients remain positive and statistically significant, with similar magnitudes across specifications, except for the ridge forecasts, where the coefficient is close to zero.[12] At the two year horizon, the CG coefficients remain negative and statistically significant, again with similar magnitudes.

Overall, the results remain similar when we vary the information set provided to the machine learning algorithms. This suggests that our results are robust to allowing the machine learning model to use prior observations to distinguish slow-moving latent components from noise, as well as to incorporate additional firm-level information. Longer time series also allow the model to learn persistent historical patterns that may help mit-

[12]This is likely due to the substantially larger set of predictors and the lack of adjustment to the regularization parameter. Ridge regression may therefore be more prone to overfitting in this setting than more complex machine learning models.

igate potential data biases. The relatively stable coefficients suggest that the observed CG test violations are structural features inherent to the machine learning models rather than spurious artifacts driven by specific specifications.

4.3 Removing Analyst Forecast

Throughout different specifications, we always include the consensus analyst EPS forecast because it is part of the information set available to market participants and more importantly, it contains substantial predictive information (Cao et al., 2024). One potential concern is that machine learning forecasts may mechanically inherit analysts' forecast biases. To examine how this affects our results, Table 4 re-estimates the Coibion and Gorodnichenko (2015) regressions using machine learning predictions constructed without analyst forecasts.

At the two year horizon, the results are similar to the baseline: both the CG coefficient and the out-of-sample MSE remain largely unchanged. At the one year horizon, however, excluding analyst forecasts changes the CG coefficients and MSE of machine learning forecasts in noticeable ways. In particular, it increases the MSE of ridge forecasts by more than 50%, from approximately 0.92 to 1.46. Similar patterns emerge for gradient boosting and random forest forecasts. This finding reflects that analyst forecasts contain powerful predictive information compared to other public information at the shorter horizon. The deterioration in forecast accuracy due to excluding analyst forecasts makes the forecasts noisier and leads to a more negative CG coefficient compared to the baseline results in Table 2. By contrast, at the longer horizon analyst forecasts provide considerably less incremental predictive information, consistent with the nearly unchanged MSE and CG coefficient.

The negative CG coefficient at the one year horizon does not contradict our main finding regarding the effect of regularization on forecast underreaction. Later in Table 6, we systematically vary the regularization hyperparameters for models estimated without analyst forecasts. The results show that stronger regularization continues to increase the CG coefficient and can shift it into positive territory.

More generally, there is no theoretical reason to expect the inclusion of analyst forecasts to systematically increase or decrease the CG coefficient. A model trained to minimize out-of-sample prediction error can adjust for stable and predictable components of analyst forecast errors rather than mechanically inherit them (Cao et al., 2024). In Internet Appendix C, we use simulated data to show that the CG coefficient does not systematically inherit the bias embedded in analyst forecasts.[13]

5 The Effects of Regularization

Table 2 shows that empirically, ML forecasts violate the Coibion and Gorodnichenko (2015) efficiency test. Section 2 identifies two forces that can drive this pattern: measurement noise, documented in prior work (De Silva and Thesmar, 2024), and the regularization built into machine learning algorithms. This section isolates the role of regularization in three steps, extending beyond the ridge model of Section 2 to the tree-based methods used in practice. First, we systematically vary the regularization parameter of each ML method and examine how the CG coefficient responds. Second, we calibrate the regularization parameter to match human analysts' forecasts. Third, we examine whether analysts' adoption of ML tools moves their forecast errors closer to the patterns machine algorithms produce.

5.1 Varying Regularization Hyperparameters

To examine the effect of regularization on forecast underreaction, we systematically vary hyperparameters governing regularization and examine how the CG coefficients respond. Our theory predicts that stronger regularization leads to more underreacting forecasts, raising γ_{CG} (Proposition 2.3).

Table 5 presents the results. Panel A varies ridge regression's penalty parameter α from 0.1 (low regularization) to 40 (high regularization). As α rises, the CG coefficient increases monotonically from 0.001 to 0.006, moving from statistically insignificant to sig-

[13] In Internet Appendix C, we conduct a simulation analysis in which we vary the fraction of analysts whose signals are mean-undistorted. The resulting CG coefficients remain largely unchanged.

nificant (t-statistic 1.89). Although the magnitudes are modest, the monotonic relationship strongly supports our proposed mechanism.

Panel B examines gradient boosting, whose regularization is governed by the learning rate. In gradient boosting, each successive tree corrects only a fraction of the residual error set by the learning rate. A smaller learning rate therefore requires more trees to fit the training data, effectively imposing stronger regularization by limiting how aggressively the model adapts to any individual observation. A smaller learning rate thus means stronger regularization. Panel B shows that as the learning rate increases from 0.01 to 0.4, the CG coefficient decreases monotonically from 0.675 (column 1) to -0.040 (column 6). At very low learning rates, gradient boosting produces large positive CG coefficients, showing strong underreaction. At higher learning rates, where regularization is weaker, the coefficient turns negative, consistent with the theoretical prediction in Proposition 2.3.

Panel C examines random forests by varying the maximum features hyperparameter, which limits the number of features considered when determining a split at each node of a tree. A lower maximum number of features restricts the information set considered at each split and increases randomization across trees, so a lower maximum features means stronger regularization. As maximum features increases from 0.25 (stronger regularization) to 1.00 (weak regularization), the CG coefficient decreases monotonically from 0.151 (column 1) to 0.044 (column 6). The coefficient γ_{CG} increases as regularization becomes stronger, again consistent with our hypothesis.

The monotonic relationships in Panels A through C confirm that stronger regularization leads to more positive CG coefficients, as predicted. Panel D formally tests this relationship by regressing the estimated CG coefficients on a rank measure of regularization intensity. For ridge regression, we generate predictions from 40 models with α ranging from 0.1 to 40, rank the models by α, and regress the corresponding CG coefficients on this rank. The estimated coefficient is 0.00009, positive and statistically significant. For gradient boosting, we similarly estimate 40 models with learning rates ranging from 0.01 to 0.4. For random forests, we estimate 40 models with the maximum-features parameter ranging from 0.25 to 1. In each case, we rank the models by regularization intensity and regress the corresponding CG coefficients on the rank measure. The estimated relation-

ships are again positive and statistically significant. Together, these results support our hypothesis that stronger regularization systematically produces larger positive CG coefficients.

The regularization channel also raises an obvious question: if adjusting the degree of regularization can move the CG coefficient toward zero, why not simply tune the model to eliminate it? The answer is that doing so can be costly, because the hyperparameter values that minimize the CG coefficient do not necessarily deliver the best out-of-sample predictive performance. For each model specification, we calculate the out-of-sample MSE and compare it with the corresponding CG coefficient. Figure 2 plots this relationship over a wider range of hyperparameter values than those reported in Panels A–C of Table 5. Panels A, B, and C plot the CG coefficients against the regularization hyperparameters for ridge regression, gradient boosting, and random forests, respectively. Panels D, E, and F plot the corresponding out-of-sample MSEs against the same hyperparameters.

For ridge regression, there is a clear tradeoff between obtaining a CG coefficient close to zero and minimizing out-of-sample MSE. Reducing the ridge penalty moves the CG coefficient toward zero; at $\alpha = 0.1$, the coefficient becomes statistically indistinguishable from zero. This, however, comes at the cost of substantially higher out-of-sample MSE. A similar tradeoff appears for gradient boosting. At a learning rate of 0.3, the CG coefficient becomes statistically indistinguishable from zero, but the corresponding out-of-sample MSE is above its minimum, which occurs at a learning rate of 0.1 and therefore under stronger regularization. For random forests, the lowest out-of-sample MSE is reached at a maximum-features value of approximately 0.9, whereas further reducing regularization by increasing the maximum-features parameter moves the CG coefficient closer to zero. Thus, eliminating the CG coefficient again comes at some cost in predictive performance, although this tradeoff is less pronounced than for ridge regression and gradient boosting.

Our results illustrate a fundamental tradeoff: forecasters have to choose between satisfying the CG orthogonality condition and minimizing out-of-sample prediction error. A forecaster judged on predictive accuracy will prioritize the latter, so the efficiency violation is not a mistake to be corrected but a by-product of forecasting accurately.

Moreover, we find that the effect of regularization on the CG coefficient across ma-

chine learning algorithms is robust to excluding analyst forecasts from the predictor set. In Table 6, we vary the same hyperparameters governing regularization as in Table 5. The results are consistent with our baseline findings: stronger regularization, implemented through a higher ridge penalty alpha, a lower gradient boosting learning rate, or a lower maximum features parameter in random forests, systematically leads to more positive CG coefficients even when analyst forecasts are excluded. With analyst forecasts excluded and the hyperparameters held at their baseline values, the CG coefficient at the one year forecast horizon is negative. In Panels B and C, however, stronger regularization shifts the one year CG coefficient from negative to positive.[14] These findings further support regularization, rather than the inclusion of analyst forecasts, as an important driver of the CG patterns in machine learning forecasts.

Overall, results above support our proposed mechanism. By exogenously varying regularization intensity while holding all else constant, we cleanly identify its predicted effect on forecast underreaction and thus CG coefficient γ_{CG}.

5.2 Calibration to Analyst Forecasts

Section 5.1 establishes that the CG coefficient of machine learning forecasts varies systematically with the strength of regularization. This raises a natural question: is the underreaction of human analysts also a signature of regularization? This subsection reports a calibration exercise that addresses this directly.

We calibrate the learning rate of the gradient boosting model because it most closely resembles human forecasts quantitatively and is widely used in financial prediction settings (Rossi, 2018; Li and Rossi, 2020). Specifically, we calibrate the model so that its one year CG coefficient matches the corresponding coefficient for human analysts in column (1) of Table 2 (0.114). We then evaluate the resulting model on three moments that were not targeted in the calibration, including one year out-of-sample MSE, the two year CG

[14] Multiple hyperparameters are related to regularization in random forests, we evaluate an alternative set of hyperparameters in Internet Appendix Table IA6 and Figure IA3. We also vary three random-forest hyperparameters jointly along a single regularization path: maximum tree depth decreases, the minimum number of observations per leaf increases, and the fraction of predictors considered at each split decreases. We obtain similar results, with specifications imposing stronger regularization producing more positive and statistically significant CG coefficients.

coefficient, and the two year out-of-sample MSE.

Table 7 compares moments from analyst forecasts with those from the calibrated gradient boosting model. The calibrated gradient boosting model targets the one year CG coefficient (0.114) and matches it exactly by construction. The one year MSE of the calibrated model is 0.946, close to the human MSE of 0.959. The two year MSE of the calibrated model is 2.117, also close to the human MSE of 1.976. The two year CG coefficients of the calibrated model and human forecasts do not line up in magnitude, but have the same sign. Overall, the calibration matches both targeted and untargeted moments reasonably well.[15]

We avoid the strong interpretation that human forecasters are implementing gradient boosting. At a minimum, however, the calibration results indicate that the magnitude patterns generated by varying regularization overlap with those observed in human analyst forecasts at the one year horizon. This overlap also extends to out-of-sample predictive accuracy across both horizons. Overall, the framework in Section 2 appears to be a plausible description of human forecasting behavior, not only of machine learning algorithms.

The learning rate that matches analyst behavior is 0.0545, below the baseline gradient boosting value of 0.1. As a smaller learning rate implies stronger regularization, the reasonable fit across both targeted and untargeted moments, together with this lower rate, suggests that human analysts may regularize more than sophisticated machine learning models. This calibrated model provides a useful benchmark for assessing the effect of analysts' ML adoption. If human analysts implicitly apply stronger regularization than the baseline gradient boosting model that generates low MSE, then adopting machine learning tools should reduce effective regularization. We take this prediction to our analysis of machine learning adoption in Section 5.3 and find consistent results.

[15]The two year CG coefficient of the calibrated model (-0.189) is larger than the human coefficient of -0.003 in magnitude. A more demanding test would calibrate jointly to multiple moments, or would compare the calibrated three-moment match against the corresponding match for a behavioral model with a single bias parameter. We do not formalize that test here.

5.3 Time Series Variation in Machine Learning Adoption

Section 5.1 provides empirical evidence that regularization contributes to underreacting forecasts and therefore violates the CG test. This section, together with the results in Section 5.2, offers corroborating evidence by exploiting the real-world adoption of ML models among finance practitioners.

Machine learning gained substantial traction in the finance industry around 2013, when machine learning tools became considerably more accessible and widely deployed. There are several reasons for us to use 2013 as a critical inflection point in ML adoption. The popular open source Python library scikit-learn reached its first stable version (0.10) in 2012. That made previously specialized statistical methods easily accessible to analysts without extensive programming expertise or effort. That library includes ridge regression, random forests, gradient boosting and other machine learning tools with standardized APIs. This dramatically reduced the cost of implementing machine learning forecasting. Many other papers adopt a similar timing in studying machine learning adoption. For example, Chen et al. (2019) show that this timing coincides with the rapid growth in FinTech patents. It is also consistent with the evidence in Gu et al. (2020) that the predictive advantage of machine learning models becomes more pronounced in recent decades.

As analysts began incorporating regularization techniques around 2013, forecast patterns should exhibit a structural shift in that period. Adoption, however, was unlikely to be uniform across analysts. Those with quantitative training were positioned to embrace ML methods earlier and more aggressively, given their familiarity with the underlying methodology. Accordingly, such analysts should exhibit more pronounced changes in forecast patterns around the adoption threshold.

We classify analysts as 'technical' if they hold undergraduate or graduate degrees in statistics, mathematics, computer science, engineering, or related quantitative fields. We identified this training by manually collecting and verifying individual LinkedIn profiles and FINRA BrokerCheck records. Our sample includes 173 technical analysts and 685 non-technical analysts covering 10,377 and 11,294 firm-year observations respectively. For each firm-month, we calculate median consensus forecasts for technical analysts and

non-technical analysts separately. We lack data to tell who actually used ML. What we examine is whether the associated forecasts make sense under the assumption that technical training proxies for ML use.

Table 8 compares the CG test results of tech and non-tech analysts before and after 2013.[16] Columns (1) and (2) report the CG coefficients estimated in pre-2013 samples. In this period, technical analysts have a small and statistically insignificant CG coefficient of 0.020, whereas non-technical analysts exhibit a much larger and statistically significant coefficient of 0.107, indicating stronger underreaction among non-technical analysts.[17] A clear shift emerges after 2013, as shown in Columns (3) and (4). The CG coefficient for technical analysts turns sharply negative to -0.147 and is statistically significant, indicating a pronounced drop in underreaction. In contrast, the coefficient for non-technical analysts also declines but more modestly, reaching -0.014.

Using seemingly unrelated regressions, we formally test these changes over time. The decline in the CG coefficient is large and statistically significant for technical analysts (from 0.020 to -0.147, p-value = 0.001), while the change for non-technical analysts (from 0.107 to -0.014) is not statistically significant (p-value = 0.16). Taken together, these results indicate that the post-2013 shift in forecast behavior is both economically larger and statistically more pronounced among technical analysts.

The overall drop in CG coefficient, i.e. weaker underreaction, and a larger change for tech analysts are consistent with our hypothesis that regularization contributes to forecast underreaction. Recall that in our calibration results in Table 7, a gradient boosting model with a learning rate of 0.0545 closely matches analysts' empirical CG coefficients, compared to a learning rate of 0.1 for the gradient boosting model. The lower learning rate required to fit analyst behavior provides suggestive evidence that the degree of regularization embedded in human forecasting exceeds that of standard ML defaults. The shift toward overreaction after 2013 in Table 8 is consistent with a reduction in regularization intensity as analysts began using advanced ML algorithms with more moderate reg-

[16] We do not implement a difference-in-differences design because only a small number of firms in our sample are covered by both technical and non-technical analysts, limiting the statistical power of such an analysis.

[17] Our focus is not on the relative magnitude of γ_{CG} between tech and non-tech analysts, but on the relative *change* in γ_{CG} before and after 2013 which we will discuss below.

ularization. Furthermore, under the reasonable assumption that technically trained analysts adopted ML-based forecasting methods more aggressively following the widespread availability of accessible tools, the differential response between tech and non-tech analysts is consistent with regularization used in forecasts contributing to the seemingly "behavioral" bias observed in ML predictions.

A reasonable concern is that technical and non-technical analysts may cover systematically different firms, confounding the interpretation. All specifications therefore include firm fixed effects, so that comparisons are drawn within firm, holding fixed time-invariant characteristics that might otherwise correlate with both analyst type and forecast accuracy. The time-series variation, a shift in forecasting technology with heterogeneous effects across analysts, further helps separate statistical regularization from psychological biases, which need not change around 2013, and from gradual shifts in firm-specific fundamentals. We acknowledge that identification rests on low-frequency variation, which makes it hard to rule out confounding events in the same period; we therefore do not claim a causal relationship. But the consistency of Table 8 with the cross-sectional evidence above makes it unlikely that any single alternative accounts for the full pattern. Taken together, the converging evidence supports the hypothesis that regularization plays a central role in the CG-test failures of machine learning forecasts documented in Section 4.1.

6 Cross-Sectional Evidence: Signal Quality

According to our theory, forecast efficiency violations (nonzero γ_{CG}) vary systematically with the quality of signals. Specifically, Proposition 2.4 predicts more negative CG coefficients when noise volatility σ_η is larger. Testing these cross-sectional predictions is helpful for distinguishing regularization, measurement noise, and behavioral biases. Because behavioral explanations are rooted in the analyst's psychology, they typically predict uniform effects across all firms the analyst covers. In contrast, the measurement noise mechanism predicts heterogeneity tied to firm characteristics that proxy for noise volatility.

To test this, we interact forecast revisions with two firm characteristics that proxy for signal quality: R&D expenditure scaled by total assets, and firm age. Higher R&D intensity reflects greater uncertainty about future cash flows, and younger firms have noisier signals because their operations are less stable. Both characteristics are associated with higher noise volatility. The empirical specification is

$$e_{i,t+1} = \gamma_1 r_{i,t} + \gamma_2 r_{i,t} \times \text{Characteristic}_i + \alpha_i + \delta_t + \varepsilon_{i,t+1}. \tag{16}$$

Table 9 reports the results. Panel A shows the results exploiting heterogeneity in R&D expenditure. We split firms into two groups based on whether they are above or below the sample average of R&D expenditure. For human analysts, Column (1) shows a CG coefficient of 0.125 for low R&D firms, while Column (2) reports 0.103 for high R&D firms. The smaller coefficient for high-R&D firms indicates stronger 'overreaction' (more negative coefficient) in that subsample.

ML forecasts exhibit similar patterns. For ridge regression, the CG coefficient is 0.033 for low R&D firms and -0.000 for high R&D firms. For gradient boosting, coefficients are 0.092 for low R&D firms and 0.034 for high R&D firms. For random forests, coefficients are 0.068 for low R&D firms and 0.027 for high R&D firms. Columns (3), (6), (9), and (12) include interaction terms between forecast revisions and R&D expenditure. The coefficients are -0.538, -0.301, -0.454, and -0.450 respectively. All of them are negative and statistically significant. These findings confirm the prediction that firms with higher R&D intensity have more negative CG coefficients and stronger 'overreaction'. This is consistent with our hypothesis.

Panel B uses firm age as another proxy for noise volatility. Younger firms are presumably more uncertain and therefore have higher noise volatility. For analyst predictions, the CG coefficient is 0.049 for younger firms and 0.139 for older firms. For ridge regression, the coefficient is -0.034 for younger firms and 0.011 for older firms. Noise can even flip the sign of the CG coefficient, amplifying the magnitude of the effect. For gradient boosting, coefficients are 0.026 for younger firms and 0.069 for older firms. For random forest, coefficients are 0.006 for younger firms and 0.057 for older firms.

Columns (3), (6), (9), and (12) include interaction terms between forecast revisions and firm age. The coefficients are 0.002 for all four forecasting methods, all positive and statistically significant. These findings show that as predicted, older firms have less negative (or more positive) CG coefficients.

7 Economic Consequences

Our analysis so far shows that regularization produces underreacting forecasts and systematic violations of the Coibion and Gorodnichenko (2015) test. We now ask whether these features have economic consequences for corporate investment. If managers respond to analyst forecasts when allocating capital, and those forecasts have errors that are systematically predictable from revisions, partly because of regularization, then predicted forecast errors should covary with subsequent investment changes.

To see if that happens, we follow the two-step approach used by Bordalo et al. (2026). The first step is to regress forecast errors on forecast revisions. This is intended to isolate the component of forecast errors that is predictable from revisions. The second step is to see if these predicted forecast errors correlate with subsequent investment changes. This approach allows us to isolate the component of forecast errors that is predictable from information available at the forecast date; at the same time, it reduces concerns about omitted variables that affect both forecast errors and investment.

We study the post-2013 sample since that is when machine learning tools became widely accessible. This period is suitable for distinguishing between technical analysts who adopted ML methods more aggressively, and non-technical analysts who did not. We restrict the second-stage sample to firms for which more than 40% of covering analysts have traceable educational backgrounds. This is intended to allow adequate representation of both analyst types without losing an excessive number of observations.

Table 10 presents the results. Columns (1) and (2) show the first-stage regressions. These just replicate the findings in Table 8. For technical analysts post-2013, the CG coefficient is -0.147 (t-statistic -3.90), showing strong overreaction. For non-technical analysts, the coefficient is -0.014 (t-statistic -0.19) and it is statistically insignificant.

Columns (3) and (4) provide the second stage results. The coefficient on predicted forecast errors is 0.007 (t-statistic 2.21) for technical analysts, and 0.006 (t-statistic 2.42) for non-technical analysts. Both coefficients are positive and statistically significant, which implies that the predicted forecast errors are associated with subsequent investment changes in the same direction.

These effects are large enough to be economically significant. The average investment rate change is 0.034. The standard deviations of predicted forecast errors are 0.50 and 0.47 for tech and non-tech analysts, respectively. For tech analysts, one standard deviation increase in predicted forecast error leads to a $10\%(= 0.007 \times 0.50/0.034)$ increase in investment changes. For non-tech analysts, one standard deviation increase in predicted forecast error leads to an $8\%(= 0.006 \times 0.47/0.034)$ increase in investment changes.

This shows that technical analysts post-2013 exhibit overreaction (negative CG coefficients). The positive second stage coefficient of 0.007 means that these negative forecast errors after positive revisions (good news) are associated with subsequent decreases in investment growth. Positive forecast errors following downward revisions are associated with increases in investment growth.

How should this evidence be interpreted? Managers may initially respond to analyst forecast revisions when making investment decisions. Then they adjust when realized earnings reveal errors in the direction predicted by the earlier revision. This would be a true reversal driven by managers correcting initial mistakes. However, it is also possible that both investment changes and forecast errors are responding to common underlying information. Again, regularization creates predictable patterns in how analysts process that information. Our two stage procedure is an attempt to isolate the first channel by instrumenting forecast errors with forecast revisions. However, given the complexity involved in actual firm investment decisions, we cannot entirely rule out the second interpretation. Overall, we view Table 10 as suggestive evidence that predictable forecast errors affect firms' real decisions.

8 Conclusion

A central finding of this paper is that machine learning forecasts systematically violate the Coibion and Gorodnichenko (2015) test. This is not incidental. Across all three ML methods we study, ridge regression, gradient boosting, and random forests, forecast errors are strongly predictable from forecast revisions. These violations are theoretically predicted, empirically stable, and monotonically related to the strength of regularization.

We demonstrate that these violations need not arise from cognitive limitations or emotional influences. They can arise from standard statistical practice. We develop a simple unified framework in which a forecaster uses regularization in the presence of measurement noise. The regularized forecaster shrinks the weight placed on new signals, as the textbook bias-variance tradeoff prescribes in finite samples (Hastie et al., 2009). This shrinkage systematically produces forecast revisions and errors that violate the Coibion and Gorodnichenko (2015) test. The sign and magnitude of the apparent bias depend on the intensity of regularization and the relative magnitude of measurement noise. A forecaster who shrinks in order to minimize out-of-sample MSE will look behaviorally biased under the standard tests, even though nothing in her procedure involves belief distortion.

These results have two sets of implications worth emphasizing. First, machine forecasts are a useful accuracy benchmark but not a benchmark for forecast error orthogonality: the same shrinkage that delivers their accuracy advantage makes their errors predictable from their own revisions, so measured departures from a machine benchmark are not by themselves evidence of psychological distortion. Second, our results bear on the interpretation of the CG test, which is widely used to test for rationality and information frictions such as sticky information and rational inattention. Our findings show that the test cannot distinguish among these mechanisms. Tests of rationality or information frictions that rely solely on the CG statistic are therefore underidentified in environments where forecasters use regularization. More refined tests, exploiting variation in regularization intensity, signal quality, or forecaster type, are needed to distinguish these mechanisms.

Future research extending this framework to other forecasting environments, such as

macroeconomic conditions, credit ratings, and asset prices, would be valuable. Developing diagnostic tools that distinguish regularization from psychological bias will require more refined tests than are currently standard. As algorithmic decision-making grows, understanding the statistical origins of apparent behavioral anomalies may prove at least as important as documenting them.

References

Afrouzi, H., S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar (2023). Overreaction in expectations: Evidence and theory. *Quarterly Journal of Economics 138*(3), 1713–1764.

Begenau, J., M. Farboodi, and L. Veldkamp (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics 97*, 71–87.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022). Belief distortions and macroeconomic fluctuations. *American Economic Review 112*(7), 2269–2315.

Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer (2024). Belief overreaction and stock market puzzles. *Journal of Political Economy 132*(5), 1450–1484.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020). Overreaction in macroeconomic expectations. *American Economic Review 110*(9), 2748–2782.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *Journal of Finance 73*(1), 199–227.

Bordalo, P., N. Gennaioli, A. Shleifer, and S. J. Terry (2026). Real credit cycles. *American Economic Review*.

Cao, S., W. Jiang, J. Wang, and B. Yang (2024). From man vs. machine to man + machine: The art and AI of stock analyses. *Journal of Financial Economics 160*, 103910.

Chen, M. A., Q. Wu, and B. Yang (2019). How valuable is fintech innovation? *The Review of Financial Studies 32*(5), 2062–2106.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–2678.

De Silva, T. and D. Thesmar (2024). Noise in expectations: Evidence from analyst forecasts. *Review of Financial Studies 37*(5), 1494–1537.

Dechow, P. M. and I. D. Dichev (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review 77*(s-1), 35–59.

Dichev, I. D. and V. W. Tang (2009). Earnings volatility and earnings predictability. *Journal of Accounting and Economics 47*(1-2), 160–181.

Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

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

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Gulen, H., M. Ion, C. E. Jens, and S. Rossi (2024). Credit cycles, expectations, and corporate investment. *The Review of Financial Studies 37*(11), 3335–3385.

Hastie, T. (2020). Ridge regularization: An essential concept in data science. *Technometrics 62*(4), 426–433.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2 ed.). Springer Science & Business Media.

Li, B. and A. G. Rossi (2020). Selecting mutual funds from the stocks they hold: A machine learning approach. *Working Paper*. Available at https://ssrn.com/abstract=3737667.

Li, B., A. G. Rossi, X. S. Yan, and L. Zheng (2025). Machine learning from a "universe" of signals: The role of feature engineering. *Journal of Financial Economics 172*, 104138.

Lucas Jr, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory 4*(2), 103–124.

Mankiw, N. G. and R. Reis (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics 117*(4), 1295–1328.

Martin, I. W. R. and S. Nagel (2022). Market efficiency in the age of big data. *Journal of Financial Economics 145*(1), 154–177.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica 29*(3), 315–335.

Rossi, A. G. (2018). Predicting stock market returns with machine learning. Technical report, Working Paper.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics,* Berkeley, Calif., pp. 197–206. University of California Press.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man versus machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies 36*(6), 2361–2396.

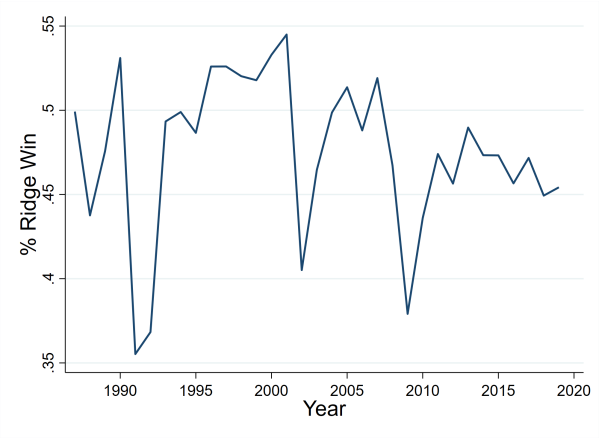
Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics 12*(1), 579–601.

Zhang, Y., Y. Zhu, and J. T. Linnainmaa (2025). Man versus machine learning revisited. *The Review of Financial Studies 38*(12), 3768–3790.

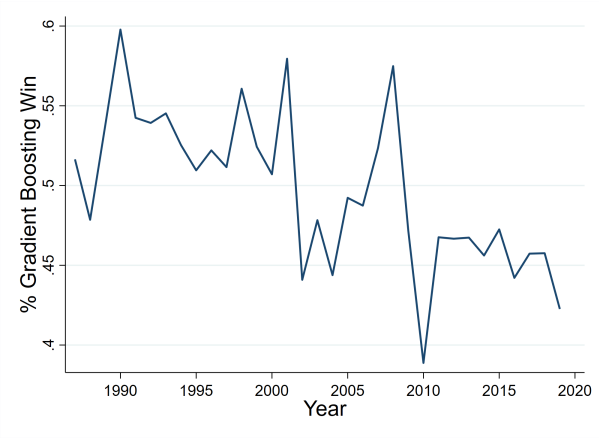
Figures and Tables

Figure 1: **ML Win Rates Over Time**

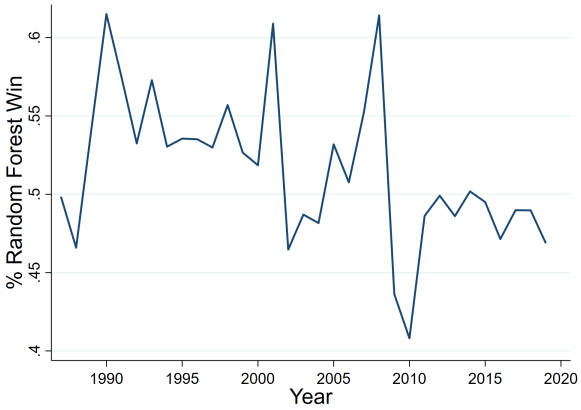
Panel A: Ridge Regression



Panel B: Gradient Boosting



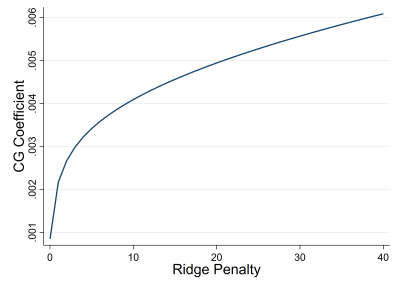
Panel C: Random Forests



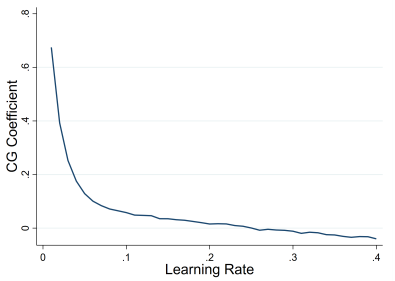
This figure reports the annual share of firm-year observations for which machine learning models outperform human analysts in terms of squared errors. Panel A presents results for Ridge regression, Panel B for Gradient Boosting, and Panel C for Random Forests.

Figure 2: **Regularization Intensity, CG Coefficients and out-of-sample Performance**

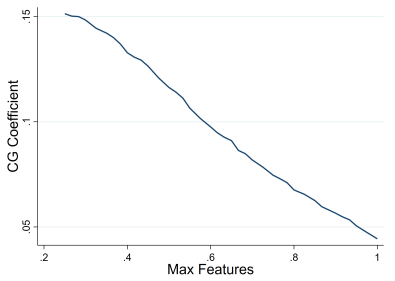
Panel A: Ridge Regression γ_{CG}



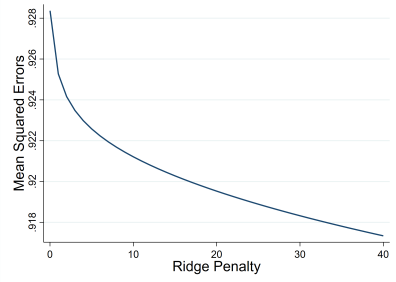
Panel B: Gradient Boosting γ_{CG}



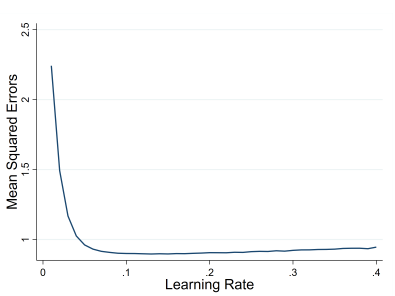
Panel C: Random Forests γ_{CG}



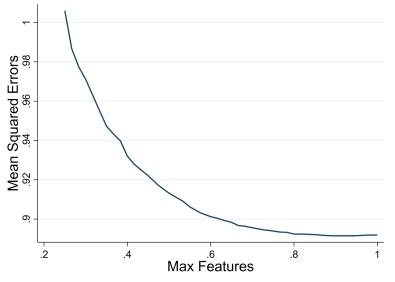
Panel D: Ridge Regression MSE



Panel E: Gradient Boosting MSE



Panel F: Random Forests MSE



This figure plots the Coibion and Gorodnichenko (2015) coefficient against regularization intensity at the one year forecast horizon, along with the mean squared errors (MSE) of these forecasts. Panel A varies the ridge penalty alpha in ridge regression, from 0.1 to 40. Panel B varies the learning rate in gradient boosting, from 0.01 to 0.4. Panel C varies the maximum features in random forests, from 0.25 to 1. Panels D, E and F report the corresponding out-of-sample MSE for ridge, gradient boosting and random forests, respectively.

Table 1: **Summary Statistics**

Variable	Mean	Std Dev	P25	Median	P75	Min	Max	N
Panel A: Firm Characteristics								
EPS	1.222	2.022	0.130	0.990	2.070	-9.547	18.276	99,963
Book to Market	0.618	0.473	0.302	0.518	0.806	0.014	6.813	99,963
Return on Assets	0.082	0.198	0.030	0.110	0.177	-1.491	0.581	99,963
Debt to Assets	0.214	0.200	0.042	0.172	0.334	0.000	1.139	99,963
Panel B: Forecast Levels								
Human Analyst Forecast	1.436	1.784	0.400	1.150	2.150	-6.417	18.331	99,963
Ridge Forecast	1.391	1.840	0.321	1.120	2.147	-7.389	18.932	99,963
Gradient Boosting Forecast	1.332	1.758	0.286	1.078	2.073	-7.360	15.776	99,963
Random Forests Forecast	1.322	1.805	0.251	1.070	2.088	-7.047	15.992	99,963
Panel C: Forecast Errors								
Human Forecast Error	-0.214	0.956	-0.369	-0.050	0.120	-21.951	11.160	99,963
Ridge Forecast Error	-0.169	0.943	-0.356	-0.045	0.168	-22.165	11.375	99,963
Gradient Boosting Error	-0.110	0.943	-0.287	0.001	0.206	-19.651	11.235	99,963
Random Forests Error	-0.100	0.939	-0.270	0.011	0.209	-19.856	11.212	99,963
Panel D: Forecast Revisions								
Human Forecast Revision	-0.266	0.784	-0.450	-0.100	0.050	-13.334	9.571	99,963
Ridge Forecast Revision	-0.053	1.455	-0.438	-0.008	0.356	-19.856	45.156	99,963
Gradient Boosting Revision	-0.020	0.901	-0.323	0.037	0.341	-12.974	10.374	99,963
Random Forests Revision	-0.003	0.917	-0.289	0.058	0.348	-13.360	10.023	99,963

This table reports summary statistics for the main sample from 1986 to 2019. Panel A shows firm characteristics. Panel B shows forecast levels. Panel C shows forecast errors (realized minus predicted). Panel D shows forecast revisions (prediction at time t minus that at time $t-1$). Detailed variable definitions are provided in the Appendix.

Table 2: **Machine Learning Forecast Errors Are Predictable**

Panel A: Underreaction Patterns at one year Horizon				
	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$			
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.114***	0.006*	0.058***	0.044***
	(14.04)	(1.72)	(6.37)	(5.36)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963
Adj R^2	0.16	0.11	0.11	0.10
out-of-sample MSE	0.959	0.918	0.901	0.892

Panel B: Overreaction Patterns at two year Horizon				
	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
Dependent variable:	$y_{i,t+2} - F_t y_{i,t+2}$			
$F_t y_{i,t+2} - F_{t-1} y_{i,t+2}$	-0.003	-0.606***	-0.233***	-0.237***
	(-0.24)	(-27.41)	(-19.05)	(-19.68)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	58,864	58,864	58,864	58,864
Adj R^2	0.12	0.36	0.18	0.18
out-of-sample MSE	1.976	4.070	2.096	2.086

This table reports OLS regression results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). In Panel A, the dependent variable is the one-year-ahead forecast error, defined as one-year-ahead realized EPS minus predicted EPS. In Panel B, the dependent variable is the two-year-ahead forecast error, defined as two-year-ahead realized minus predicted EPS. The main independent variable is the change in forecasts from $t-1$ to t. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 3: **Robustness of the Underreaction and Overreaction Results**

Panel A: Underreaction Patterns at one year Horizon

	(1) Ridge	(2) Gradient Boosting	(3) Random Forest	(4) Ridge	(5) Gradient Boosting	(6) Random Forest	(7) Ridge	(8) Gradient Boosting	(9) Random Forest
				$y_{i,t+1} - F_t y_{i,t+1}$					
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.049***	0.060***	0.042***	0.006*	0.059***	0.045***	-0.001	0.061***	0.053***
	(6.74)	(7.36)	(5.56)	(1.88)	(5.78)	(4.96)	(-0.52)	(5.76)	(5.85)
Predictors		Limited			Add lagged EPS			Add lagged baseline predictors	
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	82,511	82,511	82,511	68,424	68,424	68,424
Adj R^2	0.13	0.12	0.11	0.09	0.09	0.09	0.07	0.08	0.08
out-of-sample MSE	0.921	0.909	0.899	0.898	0.881	0.876	0.889	0.884	0.876

Panel B: Overreaction Patterns at two year Horizon

	(1) Ridge	(2) Gradient Boosting	(3) Random Forest	(4) Ridge	(5) Gradient Boosting	(6) Random Forest	(7) Ridge	(8) Gradient Boosting	(9) Random Forest
				$y_{i,t+2} - F_t y_{i,t+2}$					
$F_t y_{i,t+2} - F_{t-1} y_{i,t+2}$	-0.190***	-0.226***	-0.248***	-0.568***	-0.218***	-0.224***	-0.806***	-0.221***	-0.229***
	(-17.35)	(-18.46)	(-21.18)	(-29.44)	(-16.43)	(-16.92)	(-40.06)	(-13.31)	(-13.81)
Predictors		Limited			Add lagged EPS			Add lagged baseline predictors	
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	58,864	58,864	58,864	49,647	49,647	49,647	29,844	29,844	29,844
Adj R^2	0.15	0.18	0.18	0.35	0.18	0.18	0.68	0.19	0.20
out-of-sample MSE	2.076	2.100	2.121	4.221	2.203	2.190	14.615	2.449	2.444

This table presents OLS regression results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The dependent variable in Panel A is the one-year-ahead forecast error, defined as one-year-ahead realized minus predicted EPS. In Panel B, the dependent variable is the two-year-ahead forecast error, defined as two-year-ahead realized minus predicted EPS. The main independent variable is the change in forecasts from t-1 to t. In columns (1), (2), and (3) of both Panels A and B, the machine learning predictions are generated using a more limited set of predictors that excludes WRDS financial ratios. In Columns (4), (5), and (6), the predictor set includes the baseline predictors from Table 2 as well as realized EPS from the year prior to the latest available EPS. In Columns (7), (8), and (9), the predictor set includes the baseline predictors from Table 2 together with one year lags of all baseline predictors. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: **Machine Learning Forecast Errors Are Predictable without Analyst Forecasts**

Panel A: one year Horizon			
	(1) Ridge	(2) Gradient Boosting	(3) Random Forest
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$		
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	-0.060*** (-10.50)	-0.065*** (-6.03)	-0.065*** (-6.06)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	99,963	99,963	99,963
Adj R^2	0.11	0.11	0.09
out-of-sample MSE	1.461	1.348	1.310

Panel B: two year Horizon			
	(1) Ridge	(2) Gradient Boosting	(3) Random Forest
Dependent variable:	$y_{i,t+2} - F_t y_{i,t+2}$		
$F_t y_{i,t+2} - F_{t-1} y_{i,t+2}$	-0.440*** (-27.22)	-0.251*** (-18.40)	-0.249*** (-17.15)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
N	58,864	58,864	58,864
Adj R^2	0.26	0.20	0.19
out-of-sample MSE	4.230	2.400	2.387

This table reports OLS regression results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). Analysts' forecasts are excluded from the predictor set used to train the machine learning model. In Panel A, the dependent variable is the one-year-ahead forecast error, defined as one-year-ahead realized EPS minus predicted EPS. In Panel B, the dependent variable is the two-year-ahead forecast error, defined as two-year-ahead realized minus predicted EPS. The main independent variable is the change in forecasts from $t-1$ to t. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: **Regularization Intensity and Forecast Underreaction**

Panel A: Ridge Penalty Alpha						
Dependent variable:			$y_{i,t+1} - F_t y_{i,t+1}$			
	(1)	(2)	(3)	(4)	(5)	(6)
	0.1	1	10	20	30	40
	(Low Reg)					(High Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.001	0.002	0.004	0.005	0.006*	0.006*
	(0.24)	(0.61)	(1.21)	(1.50)	(1.72)	(1.89)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.11	0.11	0.11	0.11	0.11	0.11
out-of-sample MSE	0.928	0.925	0.921	0.920	0.918	0.917
Panel B: Gradient Boosting Learning Rate						
Dependent variable:			$y_{i,t+1} - F_t y_{i,t+1}$			
	(1)	(2)	(3)	(4)	(5)	(6)
	0.01	0.05	0.1	0.2	0.3	0.4
	(High Reg)					(Low Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.675***	0.129***	0.058***	0.015*	-0.012	-0.040***
	(30.42)	(12.99)	(6.37)	(1.65)	(-1.37)	(-4.65)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.48	0.17	0.11	0.09	0.09	0.09
out-of-sample MSE	2.242	0.962	0.901	0.907	0.923	0.946
Panel C: Random Forests Max Features						
Dependent variable:			$y_{i,t+1} - F_t y_{i,t+1}$			
	(1)	(2)	(3)	(4)	(5)	(6)
	0.25	0.40	0.60	0.70	0.80	1.00
	(High Reg)					(Low Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.151***	0.133***	0.098***	0.082***	0.068***	0.044***
	(14.12)	(13.52)	(10.90)	(9.31)	(7.96)	(5.36)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.12	0.11	0.10	0.10	0.10	0.10
out-of-sample MSE	1.006	0.932	0.901	0.896	0.892	0.892

Table 5: **Regularization Intensity and Forecast Underreaction (continued)**

Panel D: Formal Tests			
	(1) Ridge	(2) Gradient Boosting	(3) Random Forests
		γ_{CG}	
Regularization	0.00009***	0.005***	0.003***
	(11.31)	(6.78)	(70.69)
N	40	40	40
Adj R^2	0.87	0.55	0.99

This table reports OLS regression results examining whether the predictability of forecast errors is related to the degree of regularization in machine learning models. Panels A to C report results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The dependent variable is the one-year-ahead forecast error. In Panel A, the hyperparameter varied in Ridge Regression is the ridge penalty alpha, ranging from 0.1 to 40. In Panel B, the hyperparameter varied in Gradient Boosting is the learning rate, ranging from 0.01 to 0.4. In Panel C, the hyperparameter varied in Random Forests is the maximum features, ranging from 0.25 to 1.00. The t-statistics in parentheses in Panels A-C are computed using standard errors clustered by firm. Panel D regresses the CG coefficient (γ_{CG}) on the ranked order of regularization intensity, where a higher rank indicates stronger regularization, separately for ridge, gradient boosting and random forests. Because γ_{CG} is itself a first-stage estimate, for regressions in Panel D, observations are weighted by the inverse of their first-stage sampling variance $1/se^2$ to account for estimation error. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: **Regularization Intensity and Forecast Underreaction (Excluding Analyst Forecasts)**

Panel A: Ridge Penalty Alpha						
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
	0.1	1	10	20	30	40
	(Low Reg)					(High Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	-0.070^{***}	-0.067^{***}	-0.061^{***}	-0.060^{***}	-0.060^{***}	-0.060^{***}
	(-11.58)	(-11.24)	(-10.69)	(-10.56)	(-10.50)	(-10.48)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.11	0.11	0.11	0.11	0.11	0.11
out-of-sample MSE	1.475	1.470	1.464	1.463	1.461	1.460
Panel B: Gradient Boosting Learning Rate						
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
	0.01	0.05	0.1	0.2	0.3	0.4
	(High Reg)					(Low Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.158***	-0.027^{**}	-0.065^{***}	-0.103^{***}	-0.126^{***}	-0.151^{***}
	(6.83)	(-2.20)	(-6.03)	(-10.39)	(-13.49)	(-18.41)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.45	0.17	0.11	0.09	0.09	0.09
out-of-sample MSE	2.619	1.463	1.348	1.335	1.343	1.356
Panel C: Random Forests Max Features						
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
	0.25	0.40	0.60	0.70	0.80	1.00
	(High Reg)					(Low Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.011	0.001	-0.022^{**}	-0.032^{***}	-0.043^{***}	-0.065^{***}
	(0.95)	(0.09)	(-2.00)	(-2.95)	(-4.04)	(-6.06)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.15	0.12	0.10	0.10	0.09	0.09
out-of-sample MSE	1.334	1.289	1.285	1.289	1.295	1.310

This table reports OLS regression results examining whether the predictability of forecast errors is related to the degree of regularization in machine learning models when analyst forecast is excluded from the predictor set. Panels A to C report results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The dependent variable is the one-year-ahead forecast error. In Panel A, the hyperparameter varied in Ridge Regression is the ridge penalty alpha, ranging from 0.1 to 40. In Panel B, the hyperparameter varied in Gradient Boosting is the learning rate, ranging from 0.01 to 0.4. In Panel C, the hyperparameter varied in Random Forests is the maximum features, ranging from 0.25 to 1.00. The t-statistics in parentheses are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: **Calibrating Gradient Boosting to Human Analyst Forecasts**

	one year Forecast		two year Forecast	
Forecast method	γ_{CG}	MSE	γ_{CG}	MSE
Analysts	0.114***	0.959	-0.003	1.976
Default gradient boosting	0.058***	0.901	-0.233***	2.096
Calibrated gradient boosting	0.114***	0.946	-0.189***	2.117

This table reports mean squared errors (MSE) and OLS regression coefficients (γ_{CG}) testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The analyses are conducted at both the one year and two year horizons, corresponding to Panels A and B of Table 2. The row "Analysts" reports the results using human analyst forecasts. The row "Default gradient boosting" reports the results using predictions generated by a gradient boosting model with default hyperparameters. The row "Calibrated gradient boosting" reports the results using gradient boosting predictions where the learning rate is calibrated to 0.0545 in order to match the CG coefficient on human analysts' one-year-ahead forecast (Table 2 Column (1)). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: **Coibion and Gorodnichenko (2015) Test Results by Analyst Technical Background**

	Pre-2013 (1994–2012)		Post-2013 (2013–2018)	
	(1) Technical	(2) Non-Tech	(3) Technical	(4) Non-Tech
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$			
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.020	0.107**	−0.147***	−0.014
	(0.43)	(2.35)	(−3.90)	(−0.19)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	4,813	5,282	5,564	6,012
Adj R^2	0.15	0.15	0.16	0.09
SUR p-value (pre vs. post)			0.001	0.16

This table reports Coibion and Gorodnichenko (2015) regressions of forecast errors on forecast revisions for analyst subsamples defined by technical background and time period. Technical analysts have undergraduate or graduate degrees in statistics, computer science, mathematics, engineering, or related quantitative fields, identified from LinkedIn and FINRA BrokerCheck records. The sample spans 1994 to 2018 and is split at 2013, when ML tools became widely accessible through powerful open-source libraries such as scikit-learn. The dependent variable is the analyst forecast error. All regressions include firm and year fixed effects. The t-statistics (in parentheses) are computed using standard errors clustered by firm. The table also reports p-values from seemingly unrelated regressions testing whether the coefficients differ significantly across the two periods, separately for tech and non-tech analysts. *, **, *** indicate significance at 10%, 5%, and 1% levels.

Table 9: **Cross-sectional Heterogeneity**

Panel A: Heterogeneity in R&D Expenditure												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Analysts			Ridge			Gradient Boosting			Random Forest		
	Low	High		Low	High		Low	High		Low	High	
	$y_{i,t+1} - F_t y_{i,t+1}$											
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.125***	0.103***	0.123***	0.033***	-0.000	0.030***	0.092***	0.034***	0.091***	0.068***	0.027**	0.070***
	(10.06)	(9.73)	(13.14)	(3.18)	(-0.13)	(3.86)	(8.72)	(2.58)	(11.15)	(6.50)	(2.30)	(8.75)
$(F_t y_{i,t+1} - F_{t-1} y_{i,t+1})$			-0.538***			-0.301***			-0.454***			-0.450***
$\times R\&D Expenditure$			(-7.82)			(-5.38)			(-6.94)			(-6.65)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	40,913	59,050	99,963	40,913	59,050	99,963	40,913	59,050	99,963	40,913	59,050	99,963
Adj R^2	0.15	0.16	0.14	0.11	0.11	0.10	0.12	0.11	0.10	0.10	0.11	0.09
Panel B: Heterogeneity in Firm Age												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Analysts			Ridge			Gradient Boosting			Random Forest		
	Young	Old		Young	Old		Young	Old		Young	Old	
	$y_{i,t+1} - F_t y_{i,t+1}$											
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.049***	0.139***	0.063***	-0.034***	0.011***	-0.028***	0.026**	0.069***	0.033***	0.006	0.057***	0.011
	(4.16)	(13.51)	(5.07)	(-3.66)	(3.10)	(-2.84)	(2.23)	(5.83)	(2.82)	(0.52)	(5.42)	(0.97)
$(F_t y_{i,t+1} - F_{t-1} y_{i,t+1})$			0.002***			0.002***			0.002***			0.002***
$\times Age$			(3.79)			(5.46)			(4.24)			(4.81)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	40,464	59,499	99,963	40,464	59,499	99,963	40,464	59,499	99,963	40,464	59,499	99,963
Adj R^2	0.16	0.17	0.14	0.12	0.12	0.10	0.12	0.12	0.10	0.11	0.12	0.09

This table reports OLS regression results testing whether the predictability of forecast errors, based on forecast revisions following Coibion and Gorodnichenko (2015), differs by R&D expenditure and firm age. In Panel A, the sample is split into two groups based on firms' average R&D expenditure. Columns (1), (4), (7), and (10) include firms below the sample-average R&D expenditure, while columns (2), (5), (8), and (11) include firms above the average. Columns (3), (6), (9), and (12) additionally include the interaction term between R&D expenditure and forecast revisions. In Panel B, the sample is similarly split into two groups based on firms' average age. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, *** indicate significance at 10%, 5%, and 1% levels.

Table 10: **Economic Consequences: Forecast Errors and Investment Reversals**

	(1) Technical	(2) Non-Tech	(3) Technical	(4) Non-Tech
Dependent Variable:	$y_{i,t+1} - F_t y_{i,t+1}$		ΔInvestment$_{t,t+1}$	
Predicted $y_{i,t+1} - F_t y_{i,t+1}$			0.007**	0.006**
			(2.21)	(2.42)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	-0.147***	-0.014		
	(-3.90)	(-0.19)		
Profitability$_{i,t}$			0.021***	0.028***
			(4.57)	(8.42)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	5,564	6,012	1,914	2,076
Adj R^2	0.16	0.09	0.05	0.06

This table examines whether forecast errors predict subsequent changes in corporate investment from 2013 to 2018. Columns (1) and (2) report the first-stage estimates, regressing forecast errors on forecast revisions. The dependent variable is the forecast error, $y_{i,t+1} - F_t y_{i,t+1}$, and the independent variable is forecast revision at time t, $F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$. Columns (3) and (4) report the second-stage regression, showing how instrumented forecast errors affect investment. The dependent variable is the change in investment rate from year t to year $t+1$. Predicted forecast errors are estimated from first stage regressions of forecast errors on forecast revisions. We restrict the second stage sample to firms for which we can trace the educational background of more than 40% of covering analysts. The mean value of investment rate change is 0.034. The standard deviations of predicted forecast errors are 0.50 and 0.47 for tech and non-tech analysts, respectively. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, *** indicate significance at 10%, 5%, and 1% levels.

Appendix A Theoretical Derivations

This appendix provides complete proofs of all propositions in the main text. We consider two models: (1) the baseline model without measurement error and (2) the extended model with measurement error. We begin with preliminary results on AR(1) processes that are used throughout.

Appendix A.1 Preliminaries: AR(1) Process Properties

Consider a stationary AR(1) process.

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2). \tag{A1}$$

Lemma A1 (AR(1) Moments)**.** *For the stationary AR(1) process above*

1. *Unconditional variance:* $\mathrm{Var}(x_t) = \sigma_x^2 = \sigma_\varepsilon^2/(1-\rho^2)$
2. *Autocovariance:* $\mathrm{Cov}(x_t, x_{t-k}) = \rho^k \sigma_x^2$ *for* $k \geq 0$
3. *Variance of first difference:* $\mathrm{Var}(x_t - x_{t-1}) = 2(1-\rho)\sigma_x^2$
4. *Covariance with first difference:* $\mathrm{Cov}(x_t, x_t - x_{t-1}) = (1-\rho)\sigma_x^2$

Proof. (1) Taking variance of $x_t = \rho x_{t-1} + \varepsilon_t$ and using independence of ε_t from x_{t-1},

$$\mathrm{Var}(x_t) = \rho^2 \mathrm{Var}(x_{t-1}) + \mathrm{Var}(\varepsilon_t) \tag{A2}$$

$$\sigma_x^2 = \rho^2 \sigma_x^2 + \sigma_\varepsilon^2 \tag{A3}$$

$$\sigma_x^2(1-\rho^2) = \sigma_\varepsilon^2 \tag{A4}$$

$$\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}. \tag{A5}$$

(2) We prove by induction. Base case $k = 0$: $\text{Cov}(x_t, x_t) = \sigma_x^2 = \rho^0 \sigma_x^2$. Base case $k = 1$.

$$\text{Cov}(x_t, x_{t-1}) = \text{Cov}(\rho x_{t-1} + \varepsilon_t, x_{t-1}) \tag{A6}$$

$$= \rho \text{Var}(x_{t-1}) + \text{Cov}(\varepsilon_t, x_{t-1}) \tag{A7}$$

$$= \rho\sigma_x^2 + 0 = \rho\sigma_x^2. \tag{A8}$$

Inductive step: Assume $\text{Cov}(x_t, x_{t-k}) = \rho^k \sigma_x^2$. Then

$$\text{Cov}(x_t, x_{t-k-1}) = \text{Cov}(\rho x_{t-1} + \varepsilon_t, x_{t-k-1}) \tag{A9}$$

$$= \rho \text{Cov}(x_{t-1}, x_{t-k-1}) \tag{A10}$$

$$= \rho \cdot \rho^k \sigma_x^2 = \rho^{k+1} \sigma_x^2. \tag{A11}$$

(3) Expanding

$$\text{Var}(x_t - x_{t-1}) = \text{Var}(x_t) + \text{Var}(x_{t-1}) - 2\text{Cov}(x_t, x_{t-1}) \tag{A12}$$

$$= \sigma_x^2 + \sigma_x^2 - 2\rho\sigma_x^2 \tag{A13}$$

$$= 2(1 - \rho)\sigma_x^2. \tag{A14}$$

(4) Expanding

$$\text{Cov}(x_t, x_t - x_{t-1}) = \text{Cov}(x_t, x_t) - \text{Cov}(x_t, x_{t-1}) \tag{A15}$$

$$= \sigma_x^2 - \rho\sigma_x^2 \tag{A16}$$

$$= (1 - \rho)\sigma_x^2. \tag{A17}$$

□

Appendix A.2 Proof for the Baseline Model without Measurement Noise

At time t the forecaster observes all historical information $\{z_\tau\}_{\tau=0}^t$. Because given z_t all past realizations are irrelevant for y_{t+1}, the forecasting rule reduces to $F_t y_{t+1} = \beta\, z_t$.

Without regularization. Because signals are perfect ($\sigma_\eta^2 = 0$), the OLS coefficient collapses to:

$$\beta^{\mathrm{OLS}} = \alpha\rho_s. \tag{A18}$$

This is the exact Kalman gain under noiseless observation: the forecaster weights the signal by the persistence of the fundamental and the scaling parameter α.

With ridge regularization. Imposing an L_2 penalty $\lambda \geq 0$ on the coefficient yields:

$$\beta_\lambda = \alpha\rho_s \, \frac{\sigma_s^2}{\sigma_s^2 + \lambda}. \tag{A19}$$

Note $|\beta_\lambda| < |\beta^{\mathrm{OLS}}|$ whenever $\lambda > 0$, and $\beta_\lambda \to \beta^{\mathrm{OLS}}$ as $\lambda \to 0$.

To compute forecast revisions, we begin by computing the forecaster's two-period-ahead forecast.

$$F_t y_{t+2} = \rho_s \, \beta_\lambda \, z_t. \tag{A20}$$

***Proof of Proposition** 2.1.* The one-period-ahead forecast revision is

$$r_t = F_t y_{t+1} - F_{t-1} y_{t+1} = \beta_\lambda z_t - \rho_s \beta_\lambda z_{t-1} = \beta_\lambda (s_t - \rho_s s_{t-1}) = \beta_\lambda v_t,$$

The forecast error is

$$\begin{aligned} e_{t+1} &= y_{t+1} - F_t y_{t+1} = \alpha s_{t+1} + \epsilon_{t+1} - \beta_\lambda s_t \\ &= \alpha(\rho_s s_t + v_{t+1}) + \epsilon_{t+1} - \beta_\lambda s_t \\ &= (\alpha\rho_s - \beta_\lambda)\, s_t + \alpha\, v_{t+1} + \epsilon_{t+1}. \end{aligned}$$

Since v_t is independent of v_{t+1}, ϵ_{t+1}, and $s_{t-1} = \sum_{k=1}^{\infty} \rho_s^{k-1} v_{t-k}$, while $\mathrm{Cov}(s_t, v_t) = \sigma_v^2$, we have

$$\mathrm{Cov}(e_{t+1}, r_t) = (\alpha\rho_s - \beta_\lambda)\, \beta_\lambda \, \underbrace{\mathrm{Cov}(s_t, v_t)}_{\sigma_v^2} = (\alpha\rho_s - \beta_\lambda)\, \beta_\lambda \, \sigma_v^2.$$

The denominator is

$$\mathrm{Var}(r_t) = \beta_\lambda^2 \, \mathrm{Var}(v_t) = \beta_\lambda^2 \, \sigma_v^2.$$

Hence,

$$\gamma_{CG} = \frac{\mathrm{Cov}(e_{t+1}, r_t)}{\mathrm{Var}(r_t)} = \frac{\alpha\rho_s - \beta_\lambda}{\beta_\lambda}.$$

Substituting $\beta_\lambda = \alpha\rho_s\,\sigma_s^2/(\sigma_s^2+\lambda)$ from Equation (A19):

$$\gamma_{CG} = \frac{\alpha\rho_s\left(1 - \dfrac{\sigma_s^2}{\sigma_s^2+\lambda}\right)}{\alpha\rho_s\,\dfrac{\sigma_s^2}{\sigma_s^2+\lambda}} = \frac{\dfrac{\lambda}{\sigma_s^2+\lambda}}{\dfrac{\sigma_s^2}{\sigma_s^2+\lambda}} = \frac{\lambda}{\sigma_s^2}.$$

□

Appendix A.3 Proof for Model with Measurement Noise

We incorporate measurement noise into our baseline model to illustrate its effect on the CG coefficient. For simplicity, we assume that at time t, the forecaster observes z_t only, and the forecast is $F_t y_{t+1} = \beta\, z_t$.

The ridge coefficient with regularization parameter λ is

$$\beta_\lambda = \frac{\alpha\rho_s\sigma_s^2}{\sigma_s^2 + \sigma_\eta^2 + \lambda}. \tag{A21}$$

Proof. The ridge estimator solves

$$\min_\beta \mathbb{E}\left[(y_{t+1} - \beta z_t)^2\right] + \lambda\beta^2, \tag{A22}$$

where $y_{t+1} = \alpha s_{t+1} + \varepsilon_{t+1}$ and $z_t = s_t + \eta_t$. Differentiating the objective with respect to β and setting equal to zero yields

$$\mathbb{E}[y_{t+1} z_t] = \beta\left(\mathbb{E}[z_t^2] + \lambda\right). \tag{A23}$$

Because $\mathbb{E}[y_{t+1}z_t] = \mathbb{E}[(\alpha s_{t+1} + \varepsilon_{t+1})(s_t + \eta_t)] = \alpha\rho_s\sigma_s^2$ and $\mathbb{E}[z_t^2] = \sigma_s^2 + \sigma_\eta^2$, the solution is

$$\beta_\lambda = \frac{\alpha\rho_s\sigma_s^2}{\sigma_s^2 + \sigma_\eta^2 + \lambda}. \tag{A24}$$

□

To compute forecast revisions, we compute the forecaster's two-period-ahead forecast

in a similar way.

$$F_{t-1}y_{t+1} = \rho_s \frac{\alpha\rho_s\sigma_s^2}{\sigma_s^2 + \sigma_\eta^2 + \lambda} z_{t-1} = \rho_s \beta_\lambda z_{t-1} \tag{A25}$$

***Proof of Proposition** 2.2.* Under the regularized rule, $F_t y_{t+1} = \beta_\lambda z_t$ and $F_{t-1}y_{t+1} = \rho_s\beta_\lambda z_{t-1}$. The revision is

$$r_t = \beta_\lambda z_t - \rho_s\beta_\lambda z_{t-1} = \beta_\lambda(z_t - \rho_s z_{t-1}). \tag{A26}$$

Define $\tilde{\Delta}z_t = z_t - \rho_s z_{t-1}$. The CG coefficient is

$$\gamma_{CG} = \frac{\text{Cov}(e_{t+1}, r_t)}{\text{Var}(r_t)} = \frac{\text{Cov}(e_{t+1}, \tilde{\Delta}z_t)}{\beta_\lambda \text{Var}(\tilde{\Delta}z_t)}. \tag{A27}$$

To compute $\text{Cov}(e_{t+1}, \tilde{\Delta}z_t)$, the forecast error is $e_{t+1} = y_{t+1} - F_t y_{t+1} = \alpha s_{t+1} - \beta_\lambda s_t - \beta_\lambda \eta_t + \epsilon_{t+1}$.

The persistence-adjusted difference is

$$\begin{aligned}
\tilde{\Delta}z_t &= z_t - \rho_s z_{t-1} \\
&= (s_t + \eta_t) - \rho_s(s_{t-1} + \eta_{t-1}) \\
&= (s_t - \rho_s s_{t-1}) + (\eta_t - \rho_s\eta_{t-1}).
\end{aligned}$$

Notice from $s_t = \rho_s s_{t-1} + v_t$, the fundamental component is exactly the innovation:

$$s_t - \rho_s s_{t-1} = v_t. \tag{A28}$$

Therefore, by linearity and independence,

$$\begin{aligned}
\text{Cov}(e_{t+1}, \tilde{\Delta}z_t) = \alpha\text{Cov}(s_{t+1}, v_t) + \alpha\text{Cov}(s_{t+1}, \eta_t - \rho_s\eta_{t-1}) \\
- \beta_\lambda\text{Cov}(s_t, v_t) - \beta_\lambda\text{Cov}(s_t, \eta_t - \rho_s\eta_{t-1}) \\
- \beta_\lambda\text{Cov}(\eta_t, v_t) - \beta_\lambda\text{Cov}(\eta_t, \eta_t - \rho_s\eta_{t-1}) \\
+ \text{Cov}(\epsilon_{t+1}, \tilde{\Delta}z_t).
\end{aligned}$$

By independence

$$\text{Cov}(e_{t+1}, \tilde{\Delta} z_t) = \alpha \text{Cov}(s_{t+1}, v_t) - \beta_\lambda \text{Cov}(s_t, v_t) - \beta_\lambda \text{Cov}(\eta_t, \eta_t - \rho_s \eta_{t-1}). \tag{A29}$$

Since $s_{t+1} = \rho_s s_t + v_{t+1} = \rho_s(\rho_s s_{t-1} + v_t) + v_{t+1} = \rho_s^2 s_{t-1} + \rho_s v_t + v_{t+1}$

$$\text{Cov}(s_{t+1}, v_t) = \rho_s \text{Var}(v_t) = \rho_s \sigma_v^2 = \rho_s \sigma_s^2 (1 - \rho_s^2). \tag{A30}$$

Since $s_t = \rho_s s_{t-1} + v_t$ and v_t is independent of s_{t-1}:

$$\text{Cov}(s_t, v_t) = \text{Var}(v_t) = \sigma_v^2 = \sigma_s^2 (1 - \rho_s^2). \tag{A31}$$

$\text{Cov}(\eta_t, \eta_t - \rho_s \eta_{t-1})$ is given by:

$$\begin{aligned} \text{Cov}(\eta_t, \eta_t - \rho_s \eta_{t-1}) &= \text{Var}(\eta_t) - \rho_s \text{Cov}(\eta_t, \eta_{t-1}) \\ &= \sigma_\eta^2 - \rho_s \rho_\eta \sigma_\eta^2 \\ &= (1 - \rho_s \rho_\eta) \sigma_\eta^2. \end{aligned}$$

Combining terms above, we have

$$\text{Cov}(e_{t+1}, \tilde{\Delta} z_t) = \alpha \rho_s \sigma_s^2 (1 - \rho_s^2) - \beta_\lambda \sigma_s^2 (1 - \rho_s^2) - \beta_\lambda (1 - \rho_s \rho_\eta) \sigma_\eta^2 \tag{A32}$$

$$= (1 - \rho_s^2) \sigma_s^2 (\alpha \rho_s - \beta_\lambda) - \beta_\lambda (1 - \rho_s \rho_\eta) \sigma_\eta^2. \tag{A33}$$

To calculate γ_{CG}, we still need $\text{Var}(\tilde{\Delta} z_t)$. The fundamental component $(s_t - \rho_s s_{t-1})$ has variance

$$\text{Var}(s_t - \rho_s s_{t-1}) = \text{Var}(v_t) = \sigma_s^2 (1 - \rho_s^2). \tag{A34}$$

The noise component $(\eta_t - \rho_s \eta_{t-1})$ has variance

$$\text{Var}(\eta_t - \rho_s \eta_{t-1}) = \text{Var}(\eta_t) + \rho_s^2 \text{Var}(\eta_{t-1}) - 2\rho_s \text{Cov}(\eta_t, \eta_{t-1}) \tag{A35}$$

$$= \sigma_\eta^2 + \rho_s^2 \sigma_\eta^2 - 2\rho_s \rho_\eta \sigma_\eta^2 \tag{A36}$$

$$= (1 + \rho_s^2 - 2\rho_s \rho_\eta)\sigma_\eta^2. \tag{A37}$$

By independence of fundamentals and noise

$$\text{Var}(\tilde{\Delta} z_t) = (1 - \rho_s^2)\sigma_s^2 + (1 + \rho_s^2 - 2\rho_s \rho_\eta)\sigma_\eta^2. \tag{A38}$$

Therefore, γ_{CG} is as follows

$$\begin{aligned} \gamma_{CG} &= \frac{(1 - \rho_s^2)\sigma_s^2(\alpha\rho_s - \beta_\lambda) - \beta_\lambda(1 - \rho_s\rho_\eta)\sigma_\eta^2}{\beta_\lambda[(1 - \rho_s^2)\sigma_s^2 + (1 + \rho_s^2 - 2\rho_s\rho_\eta)\sigma_\eta^2]} \\ &= \frac{(1 - \rho_s^2)\sigma_s^2(\frac{\alpha\rho_s}{\beta_\lambda} - 1) - (1 - \rho_s\rho_\eta)\sigma_\eta^2}{(1 - \rho_s^2)\sigma_s^2 + (1 + \rho_s^2 - 2\rho_s\rho_\eta)\sigma_\eta^2}. \end{aligned}$$

Plug in

$$\beta_\lambda = \alpha\rho_s \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\eta^2 + \lambda}, \tag{A39}$$

we have

$$\gamma_{CG} = \frac{(1 - \rho_s^2)(\sigma_\eta^2 + \lambda) - (1 - \rho_s\rho_\eta)\sigma_\eta^2}{(1 - \rho_s^2)\sigma_s^2 + (1 + \rho_s^2 - 2\rho_s\rho_\eta)\sigma_\eta^2} \tag{A40}$$

$$= \frac{(1 - \rho_s^2)\lambda - \rho_s(\rho_s - \rho_\eta)\sigma_\eta^2}{(1 - \rho_s^2)\sigma_s^2 + (1 + \rho_s^2 - 2\rho_s\rho_\eta)\sigma_\eta^2}. \tag{A41}$$

□

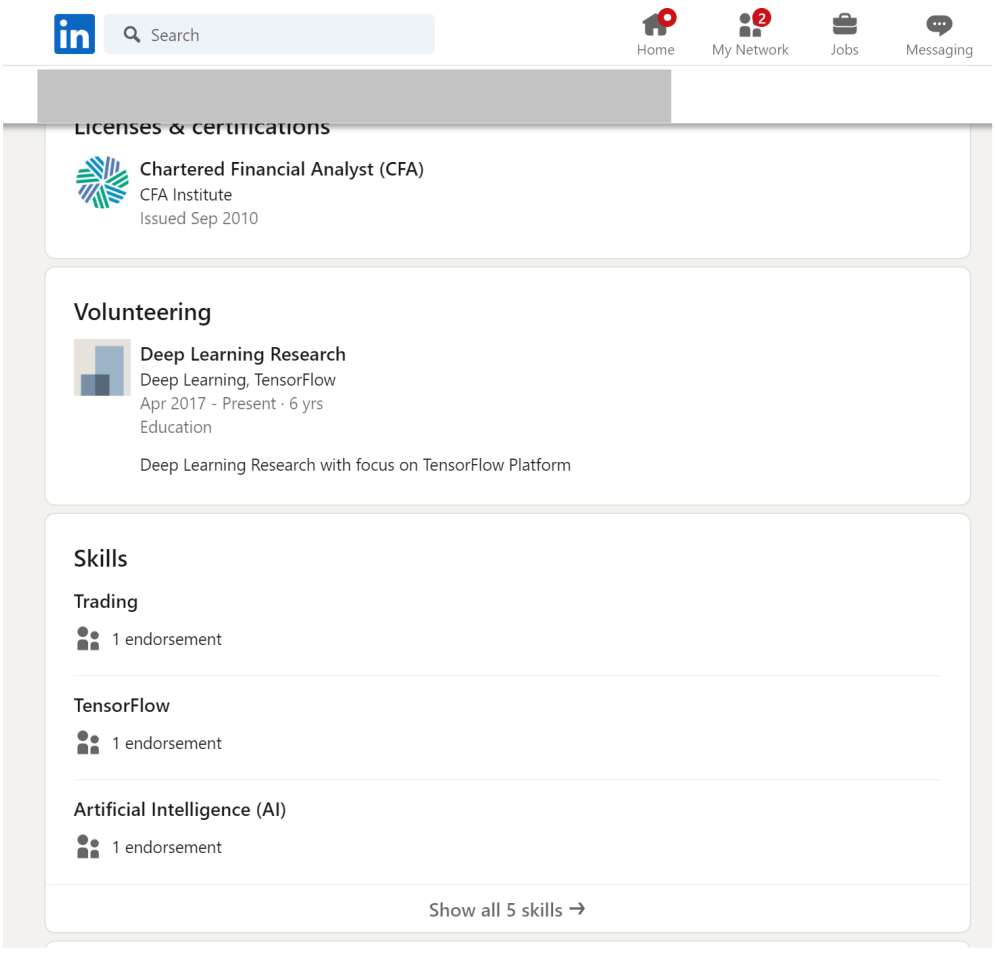
Appendix B Variable Definitions

Table A1: Variable Definitions

Variable	Definition	Source
Panel A: Dependent Variables		
EPS	Earnings Per Share, manually adjusted for the share split factor.	IBES Unadjusted Detail
Human Forecast	Consensus EPS Forecast, manually adjusted for share split factor	IBES Unadjusted Summary
ML Forecast	EPS forecast made by machine learning models	
Forecast Error	Realized EPS minus forecasted EPS	IBES Unadjusted Summary
one year Forecast Revision	(EPS for year t forecasted in year t-1) - (EPS for year t forecasted in year t-2)	
two year Forecast Revision	(EPS for year t forecasted in year t-2) - (EPS for year t forecasted in year t-3)	
Panel B: Firm Characteristics		
Investment Rate	capx/at	Compustat
Total Assets	at	Compustat
Age	Current fiscal year - first fiscal year observed	Compustat
R&D Expenditure	xrd/at	Compustat
Book-to-Market	Book/Market	WRDS Financial Ratios Suite
Return on Assets	Return on assets	WRDS Financial Ratios Suite
Debt to Assets	Total long-term debt plus debt in current liabilities/total assets	WRDS Financial Ratios Suite
Panel C: Analyst Characteristics		
Technical Background	Coded manually	

This table provides detailed definitions and data sources for all variables used in the empirical analysis. All accounting variables are from Compustat or WRDS financial ratios. Analyst forecasts are from IBES. ML forecasts are constructed following Zhang et al. (2025).

Figure A1: Example LinkedIn Profile Used to Classify Analyst Educational Background



This figure gives an example of a technical analyst's LinkedIn page.

Internet Appendices for *"Behavioral Machine Learning? Regularization and Forecast Bias"*

Internet Appendix A Machine Learning Prediction Detailed Steps

Our machine learning prediction model strictly follows the existing literature on earnings forecasts. This ensures replicability and allows our results to be jointly evaluated alongside other research questions. Our data and sample are the same as Van Binsbergen et al. (2023) and Zhang et al. (2025). Table IA1 summarizes the data-cleaning procedures.

Internet Appendix A.1 Data and Sample

Our data construction mirrors the approach in Van Binsbergen et al. (2023) and Zhang et al. (2025) and relies on four underlying sources:

1. CRSP monthly stock-level files,
2. IBES Unadjusted Summary and Detail files for analysts' consensus forecasts and reported earnings,
3. The WRDS *Financial Ratios Suite* for firm fundamentals, and
4. Real-time macroeconomic releases from the Federal Reserve Bank of Philadelphia.

The sample consists of U.S. common equities (CRSP share codes 10 and 11) listed on the NYSE, AMEX, or Nasdaq (exchange codes 1–3). The sample period runs from January 1986 to December 2019. We analyze analysts' one-, two- and three-year-ahead annual EPS forecasts (FPI $= 1, 2, 3$).

Following Van Binsbergen et al. (2023) and Zhang et al. (2025), the predictors used in our earnings-forecasting exercise fall into three broad categories: firm-level variables, macroeconomic indicators, and analysts' forecasts. Only the first firm-level variable listed below is potentially subject to look-ahead bias. As in Zhang et al. (2025), when we use firm-level variables "Most recently realized earnings", we ensure that the most recent realized earnings we used are not subject to look-ahead bias. The full list of predictors is provided below in Table IA2. This table is taken directly from the Internet Appendix (Table A.1) of Zhang et al. (2025); we reproduce it here for completeness.

Internet Appendix A.2 Model Specification and Implementation

Each month t, we estimate each machine learning model to forecast firms' earnings per share (EPS) at different horizons h.

$$F_t(y_{i,t+h}) = f_t^h(X_{i,t}|\Gamma) \tag{IA1}$$

Let $y_{i,t+h}$ denote firm i's realized earnings per share (EPS) at time $t+h$. Let $f(\cdot)$ represent the machine learning function with hyperparameters Γ applied to the feature set $X_{i,t}$. The forecasting horizons we consider include one year (A1), two years (A2), and three years (A3).

At each year-month t, we train the model in Equation IA1 using a 12-month rolling window (from $t-12$ to $t-1$) when forecasting one-year-ahead earnings. We then apply the trained model $\tilde{f}_t^h(\cdot \mid \Gamma)$ to the feature vector $X_{i,t}$ to predict EPS at time $t+h$, $F_t(y_{i,t+h})$. For two-year-ahead forecasts, we use a 24-month rolling window, and for three-year-ahead forecasts, we use a 72-month rolling window.

Our alternative specification is consistent with Zhang et al. (2025). We report the detailed forecast comparison in Table IA3, and the results are very similar.

Internet Appendix A.3 Software Package

We use scikit-learn version 1.3.0. For Ridge Regression, we import Ridge from sklearn.linear_model; for Gradient Boosting, we import GradientBoostingRegressor from sklearn.ensemble; and for Random Forest, we import RandomForestRegressor from sklearn.ensemble.

Internet Appendix B Numerical Exercises of Theoretical Model

Internet Appendix B.1 Are These Effects Large Enough to Matter?

To assess the quantitative relevance of our theoretical predictions, we conduct a numerical exercise by choosing the model's parameters to match key features of earnings forecast data. Table IA9 reports predicted Coibion and Gorodnichenko (2015) test coefficients for economically relevant parameter configurations.

The numerical exercise proceeds as follows. The noise-to-signal ratio σ_η^2/σ_s^2 spans values from 0.25 to 1.0, corresponding to the interquartile range of annual earnings volatility (EPS scaled by lagged total assets) in Compustat over 1986–2019.[18] We set fundamental persistence $\rho_s = 0.9$ to match the average first-order autocorrelation of annual ROA for mature S&P 500 firms with at least 20 years of data (Fama and French, 2006, Table 3). Noise persistence $\rho_\eta = 0.5$ matches the average autocorrelation of IBES one-year-ahead forecast errors at the firm-year level in our sample (1994–2018). The regularization parameter λ_{norm} represents the penalty normalized by total signal variance $(\sigma_s^2 + \sigma_\eta^2)$, ranging from 0 (no regularization) to 0.5 (strong shrinkage).

Internet Appendix B.2 Numerical Tests for Proposition 2.2

Table IA9 provides the main results. The numerical exercise generates CG coefficients ranging from -0.327 to $+0.069$. These ranges fully cover the empirical estimates in the literature. Coibion and Gorodnichenko (2015) report CG coefficients between -0.10 and $+0.20$ in macroeconomic forecasts. Bordalo et al. (2020) reports values between -0.15 and $+0.30$ across different contexts. Afrouzi et al. (2023) reports $\gamma_{CG} \approx -0.12$ for high-uncertainty firms. Our numerical exercise manages to get this empirical coverage despite using only parameters directly estimated from earnings data. We do not need further free

[18] See Dichev and Tang (2009) and Dechow and Dichev (2002) for similar volatility ranges. The Coibion and Gorodnichenko coefficient depends on the measurement noise only through the ratio σ_η^2/σ_s^2 (noise relative to the fundamental variance), and is invariant to the idiosyncratic outcome-noise variance σ_ϵ^2, which cancels from the error–revision covariance ratio. We therefore report scenarios in terms of $\sigma_\eta^2/\sigma_s^2 \in \{0.25, 0.50, 1.00\}$ for Low, Medium, and High noise. These correspond to the noise-to-signal range $\sigma_\eta^2/\sigma_\epsilon^2 \in \{0.5, 1.0, 2.0\}$ used in our calibration—the interquartile range of annual earnings volatility—under $\sigma_\epsilon^2 = 0.5\,\sigma_s^2$, since $\sigma_\eta^2/\sigma_\epsilon^2 = (\sigma_\eta^2/\sigma_s^2)(\sigma_s^2/\sigma_\epsilon^2)$.

parameters chosen to fit forecast coefficients.

As predicted by the model, the CG coefficient depends critically and monotonically on the level of regularization. Across all specifications, an increase in regularization leads to an increase in the CG coefficient. In the specification with low noise volatility, the sign shifts from negative (–0.216) to positive (0.069). When reducing noise volatility, proxied by the noise-to-signal ratio σ^2/σ_s^2, the CG coefficient also increases, consistent with the theoretical prediction.

The numerical exercise results show that, when the model is calibrated to match key features of earnings forecast data, it generates variation in the CG coefficient that is qualitatively consistent with our theoretical predictions, with economically meaningful magnitude.

Internet Appendix B.3 Numerical Results with Full Signal History

Section 2.3 restricts the forecaster to the contemporaneous signal z_t for analytical tractability. In this section, we relax that restriction and let the forecaster use the *full signal history* $(z_t, z_{t-1}, \dots)$. The forecast is the ridge distributed-lag projection of y_{t+1} onto the signal history,

$$F_t\, y_{t+1} = b' Z_t, \qquad Z_t = (z_t, z_{t-1}, \dots, z_{t-p})', \qquad b = (\Gamma + \lambda I)^{-1} c,$$

where $\Gamma_{ij} = \gamma_z(|i-j|)$ is the Toeplitz autocovariance matrix of the signal, $\gamma_z(k) = \sigma_s^2 \rho_s^{|k|} + \sigma_\eta^2 \rho_\eta^{|k|}$, and $c_j = \alpha \sigma_s^2 \rho_s^{j+1}$. The Coibion and Gorodnichenko (2015) coefficient γ_{CG} is computed numerically. All parameterizations follow Internet Appendix B.2: fundamental persistence $\rho_s = 0.9$, noise persistence $\rho_\eta = 0.5$, and the normalized penalty $\lambda_{\text{norm}} = \lambda/(\sigma_s^2 + \sigma_\eta^2)$ ranging from 0 to 0.5. The measurement noise-to-signal ratio σ_η^2/σ_s^2 takes the values 0.25 (Low), 0.50 (Medium), and 1.00 (High).

Table IA10 reports the CG coefficients across different noise scenarios and different values of penalty parameter λ_{norm}. The table shows that γ_{CG} *increases monotonically* in the penalty λ_{norm}: regularization shrinks the forecaster's response to news, producing underreaction and a positive CG coefficient, and more shrinkage produces more underreaction. Thus, under full information, the regularization channel still holds.

Internet Appendix B.4 Numerical Results with De Silva and Thesmar (2024) Noise

The negative effect of noise on the Coibion and Gorodnichenko (2015) (CG) coefficient is the mechanism emphasized by De Silva and Thesmar (2024). Our theoretical framework in Section 2.3 differs primarily in how this noise is modeled. Specifically, De Silva and Thesmar (2024) introduce reporting (or output) noise, whereas we model noise in the inputs to the forecasting algorithm. Incorporating both reporting noise and regularization into the analytical model substantially increases the algebraic complexity, so our main theoretical framework focuses on input noise. In this section, we extend the model by incorporating reporting noise following De Silva and Thesmar (2024). Since our goal is to provide a numerical illustration rather than a closed-form analytical characterization, we allow the predictors to satisfy the full-information assumption. The qualitative results remain unchanged: regularization increases the CG coefficient, while both reporting noise and input noise reduce it.

In this exercise the reported forecast carries an idiosyncratic reporting-noise term ξ_t independent of other innovations, following De Silva and Thesmar (2024). The forecast with the reporting noise is therefore:

$$F_t y_{t+1} = b' Z_t + \xi_t^{t+1}, \qquad F_{t-1} y_{t+1} = \rho_s \, b' Z_{t-1} + \xi_{t-1}^{t+1}. \tag{IA2}$$

where ξ_n^m denotes the i.i.d. reporting noise in the forecast made at time n for earnings at time m, with mean zero and variance σ_ξ^2.

The same noise term enters both sides of the CG regression: with a positive sign in the forecast revision on the right-hand side, and with a negative sign in the forecast error on the left-hand side. This opposite sign mechanically leads to a more negative CG coefficient. The expression of CG coefficient is

$$\gamma_{CG} = \frac{\gamma_{CG}^\star - \kappa}{1 + 2\kappa}, \qquad \kappa \equiv \frac{\sigma_\xi^2}{\mathrm{Var}(\bar{r}_t)},$$

where $\gamma_{CG}^\star$ and the reporting-noiseless revision variance $\mathrm{Var}(\bar{r}_t)$ are the full-information

quantities of Table IA10, and κ is the reporting-noise variance normalized by the (reporting-noiseless) revision variance. Table IA11 reports γ_{CG} across the noise scenarios, the penalty grid, and the reporting-noise grid $\kappa \in \{0, 0.5, 1, 2\}$; the $\kappa = 0$ column is Table IA10.

The exercise shows that reporting noise generates overreaction. The coefficient decreases in κ and crosses zero once κ exceeds the reporting-noiseless value $\gamma^{\star}_{CG}$, turning negative even though the underlying full-information forecast underreacts. For example, at Medium noise and $\lambda_{\text{norm}} = 0.3$, γ_{CG} falls from $+0.511$ at $\kappa = 0$ to $+0.005$ at $\kappa = 0.5$, -0.163 at $\kappa = 1$, and -0.298 at $\kappa = 2$.

Figure IA2 plots the predicted Coibion and Gorodnichenko (2015) coefficient γ_{CG} against the reporting-noise ratio $\kappa = \sigma^2_{\xi}/\text{Var}(\bar{r}_t)$ for a full-information forecaster, for four levels of the normalized regularization penalty $\lambda_{\text{norm}} \in \{0.0, 0.1, 0.3, 0.5\}$ (light to dark). The measurement noise-to-signal ratio is fixed at $\sigma^2_{\eta}/\sigma^2_{s} = 0.50$ (Medium). As the noise level increases, γ_{CG} decreases monotonically, whereas as the regularization penalty λ increases, γ_{CG} increases monotonically.

Internet Appendix C An Economic Model of Analyst Forecasts

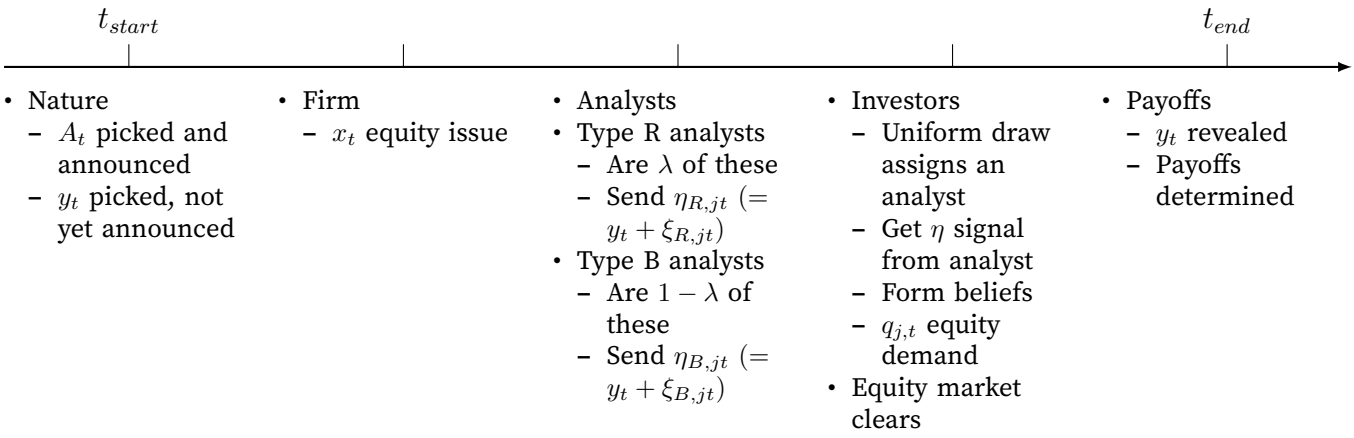
In this section, we build an economic model, based on which we generate simulated data to understand the behavioral biases documented in the real data.

Internet Appendix C.1 Setup

Based on Begenau et al. (2018), we build up a repeated static model with one firm. In each period, the firm has a one-period investment project whose payoffs depend on the realizations of both the aggregate and idiosyncratic profitability. At the beginning of each period, the aggregate profitability realizes and is public to all agents in the economy. Then the firm chooses the number of shares to issue in order to maximize its net revenue. The firm uses the raised capital to finance its one-period investment opportunity with a stochastic payoff affected by both idiosyncratic and aggregate shocks. Then investors make portfolio choice decisions. Importantly, investors do not know the idiosyncratic profitability of the firm's investment project. They receive private signals about the profitability from two types of analysts randomly, i.e. rational and biased analysts. Observing the private signal and equity price in the market, each investor updates her belief about idiosyncratic profitability and forms posterior belief. Investors' decisions are made based on their posterior beliefs. At the end of each period, payoffs and utilities are realized. In the next period, new investors arrive and the same sequence repeats. Figure IA1 illustrates the timing within each period t.

Firm There is one firm that chooses the number of shares x_t to issue in order to raise funds for investment. The capital raised from each share is invested in the firm's investment project whose payoff is $A_t y_t$. A_t is the aggregate profitability, and y_t is the idiosyncratic profitability. We introduce A_t to conceptually allow for aggregate economic state. Another advantage is that including aggregate profitability increases the dimension of predictors we could use in simulations, which mimics the high dimensionality of predic-

Figure IA1: Order of Events Within Period t



tors in the real data. Otherwise, A_t functions exactly the same as y_t in our current model.[19] Because we are interested in the analysts' forecasts on the *firm-level* profitability, we assume that A_t is public information and is realized at the beginning of each period. The logarithm of A_t follows an AR(1) process

$$logA_t = \delta_A logA_{t-1} + \epsilon_{At}, \tag{IA3}$$

with $\epsilon_{At} \sim N(0, \rho_A^{-1}),\ \delta_A > 0$.

The idiosyncratic profitability y_t is not in the firm's information set, that is, the firm does not know the realized value of y_t when it makes decisions. We assume that y_t follows an AR(1) process which is conditionally normal.

$$y_t = (1-\delta)\mu_y + \delta y_{t-1} + \epsilon_{yt}, \tag{IA4}$$

with $\epsilon_{yt} \sim N(0, \rho_y^{-1}),\ \delta > 0,\ \mu_y > 0$.

As will become clearer later, from the firm's perspective, y_t matters only through equity price p_t. Thus, it is investors' belief about y_t, not the actual value of y_t, that matters. In this economy, the firm understands the equilibrium behaviors of investors. Conditional on all information realized in the previous period and current period aggregate profitabil-

[19] This still holds even when we have more than one firm with different ex post idiosyncratic shocks. As we will see below, this is because CARA utility does not have a wealth effect and investors face no purchasing constraints.

ity A_t, the firm owner maximizes its current period net revenue from the sale of the firm. That is, she solves the following problem

$$\max_{x_t} E[x_t p_t - \frac{\phi}{2}(\Delta x_t)^2 | \mathcal{I}_{t-1}, A_t], \tag{IA5}$$

where $\frac{\phi}{2}(\Delta x_t)^2$ is equity issuance cost and $\Delta x_t = x_t - x_{t-1}$. $\{\mathcal{I}_{t-1}, A_t\}$ is the information set of the firm owner when she makes equity issuance decision, therein, $\mathcal{I}_{t-1}$ represents all the information realized prior to time t.

Given that the firm is the only supplier of stocks, its equity issuance affects the price of the stock p_t. Thus, the firm makes decisions knowing that its issuance affects the price of each share.

Analysts There are two types of analysts with a total measure of 1. $\lambda \in [0, 1]$ of them are rational investors who can make unbiased predictions (type R), $1-\lambda$ are analysts with biased expectations (type B). We could also assume that the former investors are biased but to a lesser extent, and all the conclusions would be the same.[20]

Investors There is a continuum of investors with a measure of 1. Each investor is endowed with wealth W_t and decides how to allocate her wealth between riskless and risky assets, taking prices as given.

Investor information The way that investors understand information is key to the model. When making decisions, investors do not know the firm's idiosyncratic profitability y_t, but they have a common prior that $y_t \sim N(\mu_{yt}, \rho_y^{-1})$ with $\mu_{yt} = E[y_t | \mathcal{I}_{t-1}, A_t] = (1-\delta)\mu_y + \delta y_{t-1}$.

Investors update their beliefs on y_t based on private signals they receive from analysts and the stock price. Before making investment decisions, each investor randomly draws an analyst forecast from the pool of analysts. By the law of large numbers, λ investors end up receiving the forecast of type R analysts, and $1-\lambda$ from type B analysts. We call the former investors type R (rational) investors, and the latter type B (biased) investors. Investors do not know the exact type of signal they receive.

Investor j only receives one private signal from one analyst: 1) if the signal is provided

[20] We can also think of the 'rational' investors as having bias ϵ which is infinitesimal.

by a type R analyst, the signal is $\eta_{R,jt} = y_t + \xi_{R,jt}$ with $\xi_{R,jt} \sim N(0, \rho_\eta^{-1})$; 2) if the signal is provided by type B analysts, it is $\eta_{B,jt} = y_t + \xi_{B,jt}$ with $\xi_{B,jt} \sim N(\theta\delta\epsilon_{t-1}, \rho_\eta^{-1})$. $\xi_{R,jt}$ ($\xi_{B,jt}$) is the difference between the true profitability and the signal value, that is, the total forecast error, and is uncorrelated across investors. The mean of the error term of type R analysts' forecasts is normalized to 0, and that of type B analysts is $\theta\delta\epsilon_{t-1}$, which captures the behavioral bias.

Investors do not know about the potential biasedness of signals and therefore believe the private signal they receive is $\eta_{jt} = y_t + \xi_{jt}$, where ξ_{jt} follows a normal distribution with mean 0

$$\xi_{jt} \sim N(0, \rho_\eta^{-1}).$$

Apart from the private signals, investors also learn about y_t from the market price p_t, we assume and later on verify that observing this price is equivalent to observing a normally distributed signal $\eta_{pt} = y_t + e_t$ with $e_t \sim N(0, \rho_{pt}^{-1})$.

Then by Bayes' law, the perceived posterior distribution of y_t is

$$y_t|\eta_{jt}, \eta_{pt} \sim N(\mu_{jt}, \rho_t^{-1}), \tag{IA6}$$

where $\rho_t^{-1} \equiv (\rho_y + \rho_\eta + \rho_{pt})^{-1}$ and $\mu_{jt} \equiv E[y_t|\mathcal{I}_{t-1}, A_t, \eta_{jt}, p_t] = \rho_t^{-1}(\rho_y\mu_{yt} + \rho_\eta\eta_{jt} + \rho_{pt}\eta_{pt})$. As investors do not know about the potential bias of analyst signals, after a positive shock ϵ_{yt-1}, $1-\lambda$ investors who have received type B signals overestimate the firm's profitability.

Investor portfolio choice After forming expectations about the firm's profitability, investors decide how to allocate their wealth between riskless and risky assets, taking prices as given.

$$\max_{q_{jt}} U_{jt} = \alpha E_{jt}[W_{jt}|\mathcal{I}_{t-1}, A_t, \eta_{jt}, p_t] - \frac{\alpha^2}{2} V_{jt}[W_{jt}|\mathcal{I}_{t-1}, A_t, \eta_{jt}, p_t], \tag{IA7}$$

$$s.t.\ W_{jt} = r_t(W_t - q_{jt}p_t) + q_{jt}A_t y_t, \tag{IA8}$$

where $\{\eta_{jt}, p_t, \mathcal{I}_{t-1}, A_t\}$ is investors' information set.

Plug the budget constraint into investor's utility, the maximization problem transforms

into

$$\max_{q_{jt}} U_{jt} = \alpha E_j \left[r_t W_t + q_{jt}(A_t y_t - p_t r_t) | \mathcal{I}_{t-1}, A_t, \eta_{jt}, p_t \right] - \frac{\alpha^2 A_t^2}{2} q_{jt}^2 V_j(y_t | \mathcal{I}_{t-1}, A_t, \eta_{jt}, p_t). \tag{IA9}$$

Asset market clearing condition The equity market clears:

$$\int_0^1 q_{jt} dj = x_t + \tilde{x}_t, \tag{IA10}$$

where $\tilde{x}_t$ is noisy supply and is distributed normally $N(0, \rho_x^{-1})$. The introduction of noisy supply prevents fully revealing y_t with a continuum of traders.

Internet Appendix C.2 Equilibrium

Definition IA1. *An equilibrium is a stock price $\{p_t\}$, firm equity issuance x_t, and investor demand for equity $\{q_{jt}\}_j$, such that:*

1. *Exogenous states evolve according to Equations IA3 and IA4.*
2. *Each period t, given all the information realized before t, $\mathcal{I}_{t-1}$, and current aggregate profitability A_t, the firm owner chooses equity issuance amount x_t to maximize the firm's net revenue as in Equation IA5.*
3. *Then, each investor $j \in [0, 1]$ receives a private signal about the firm's idiosyncratic profitability y_t, which is randomly drawn from a continuum of analysts (meaning that investor j does not know the type of signal she receives). In this economy, there are two types of analysts: λ type R analysts and $1 - \lambda$ type B analysts. The former produces unbiased signals $\eta_{R,jt}$, while the signal from the latter $\eta_{B,jt}$ suffers from overreaction and model misspecification error. Importantly, investor j does not know the biasedness of $\eta_{B,jt}$, and her perceived signal is denoted as η_{jt}.*
4. *Given all previous information and current aggregate profitability shock $\{\mathcal{I}_{t-1}, A_t\}$, a common prior about the distribution of y_t, a perceived private signal η_{jt} about the firm's profitability y_t and stock price p_t, investor j updates her perceived distribution of y_t according*

to Bayes' rule. Based on her updated posterior belief, she chooses the amount of equity to purchase, q_{jt}, to maximize her expected utility as in Equation IA7.

5. *Asset market clears following Equation IA10.*

To solve the model, we first solve for investors' equity demand as a function of the firm's equity issuance. We begin by solving for the investors' demand as a function of share price. Then we express their demand as a function of equity supply using the asset market clearing condition. Lastly, we plug investors' demand function into the firm's maximization problem and derive the equilibrium solution. The equilibrium value of equity issuance x_t is:

Proposition IA2. *In equilibrium, the number of shares issued is*

$$x_t = \frac{\frac{A_t}{\rho_t r_t}[(\rho_y + \rho_\eta + \rho_{pt})\mu_{yt} + \rho_\eta(1-\lambda)\theta\delta\epsilon_{yt-1}] + \phi x_{t-1}}{2\frac{\alpha A_t^2}{\rho_t r_t} + \phi}, \tag{IA11}$$

where $\rho_t = (\rho_y + \rho_\eta + \rho_{pt})$, $\rho_{pt} = \frac{\rho_\eta^2}{\alpha^2 A_t^2}\rho_x$, $\mu_{yt} = (1-\delta)\mu_y + \delta y_{t-1}$.

Equation IA11 gives the total shares of equity the firm issues in equilibrium. The first term in the numerator captures the effect of expected y_t which combines the information from the prior belief, private signals and stock price. The remaining terms in the numerator represent the marginal equity issuance costs. The denominator captures the sensitivity of price to demand plus the marginal equity issuance cost. High risk aversion (α) leads to a lower equilibrium issuance: The intuition is that from the demand side, when investors are more risk averse, the amount of risky assets they would like to hold is lower for a given price level. Besides, high equity issuance cost (ϕ) leads to less equity issuance when previous equity issuance is low (ϕx_{t-1} is small): The intuition is that from the supply side, when equity issuance is more costly, the firm is less willing to issue more equity for a given price level.

Internet Appendix C.3 Simulation Results

In this section, we conduct the Coibion and Gorodnichenko (2015) test in data simulated based on the economic model above. The simulation results are presented in Table IA5, and the baseline parameters used in the simulations are given in Table IA4.

Internet Appendix D Alternative Tests of Forecast Error Predictability

The Coibion and Gorodnichenko (2015) test is the dominant testing method in the modern literature. But there are other tests that have also been used in the literature. Bordalo et al. (2018) use a simple levels test that regresses forecast errors on the level of signals rather than their changes.

$$e_{i,t+2} = \delta_B z_{i,t} + \alpha_i + \delta_t + \nu_{i,t}, \tag{IA12}$$

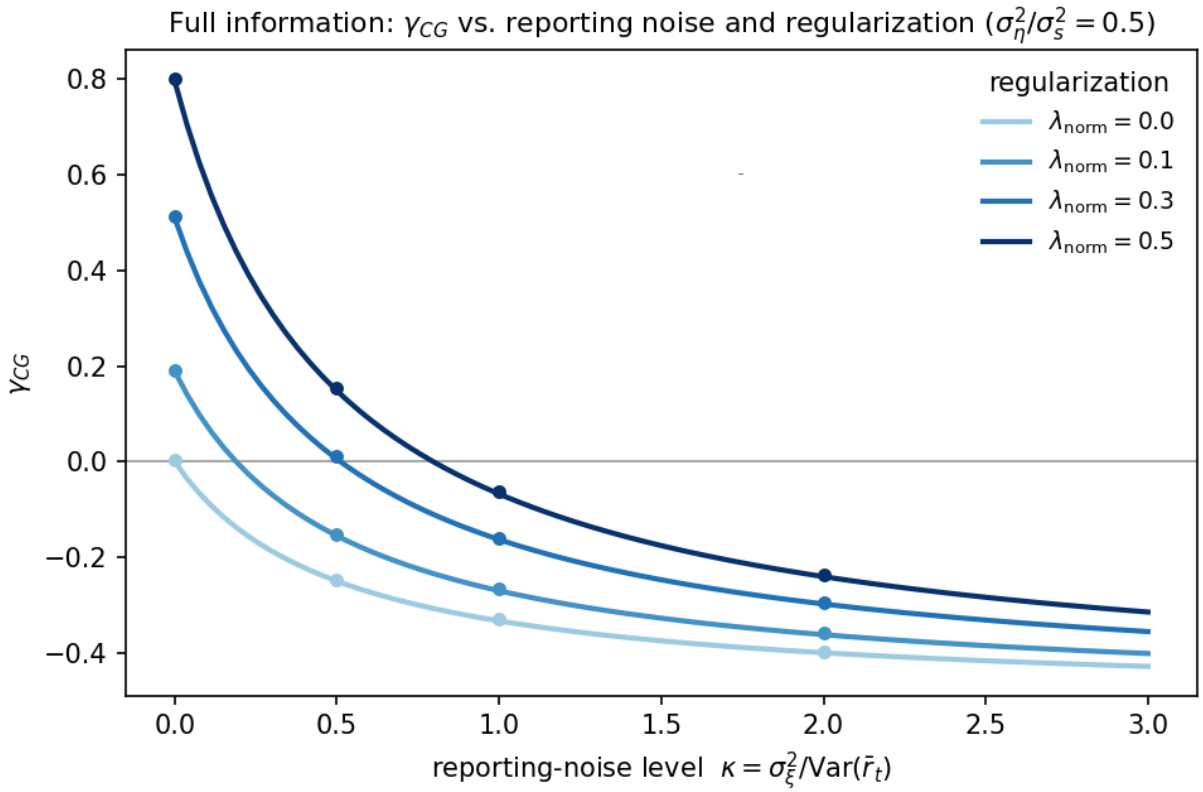
where $z_{i,t}$ represents salient signals. Under rational expectations, $\delta_B = 0$. A positive coefficient is interpreted as forecasters underweighting salient signals (underreaction). A negative coefficient is interpreted as overreaction. Table IA12 carries out these tests using investment and net debt issuance as signals.

Table IA12 Panel A uses the investment rate as the signal. Column (1) shows human forecasts have a coefficient of -1.860 (t-statistic -13.947). This confirms the widely documented finding that human forecasts overreact at longer horizons. Columns (2) to (4) show ML forecasts have even stronger overreaction, with coefficients of -2.285, -2.061, and -2.154 respectively. All of them are highly significant.

Panel B uses net debt issuance as another signal. The results are similar. Human forecasts have a coefficient of -0.729. ML forecasts range from -0.783 to -0.800. These findings confirm that ML forecasts exhibit overreaction at longer horizons. This provides further support for our main results.

Internet Appendix E Additional Figures and Tables

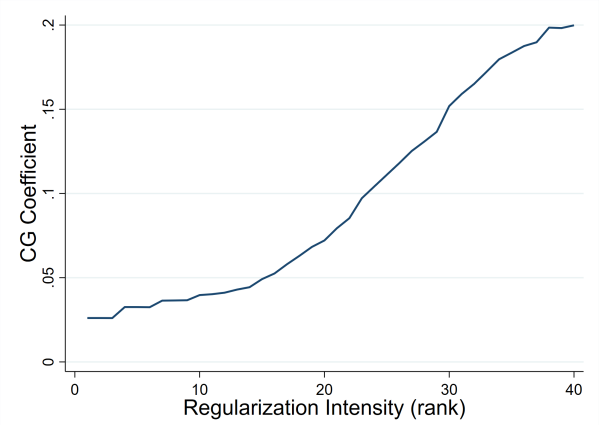
Figure IA2: Numerical Exercise With Full Information and De Silva and Thesmar (2024) Noise



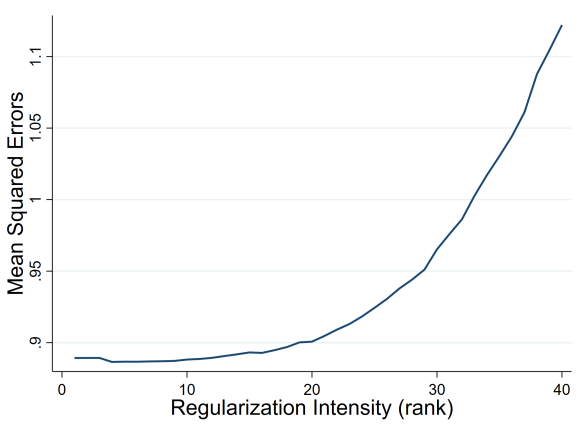
This figure plots the predicted Coibion and Gorodnichenko (2015) coefficient γ_{CG} against the reporting-noise ratio $\kappa = \sigma_\xi^2/\text{Var}(\bar{r}_t)$ for a full-information forecaster, for four levels of the normalized regularization penalty $\lambda_{\text{norm}} \in \{0.0, 0.1, 0.3, 0.5\}$ (light to dark). The measurement noise-to-signal ratio is fixed at $\sigma_\eta^2/\sigma_s^2 = 0.50$ (Medium).

Figure IA3: Joint Regularization Intensity in Random Forests, CG Coefficients and Out-of-Sample Performance

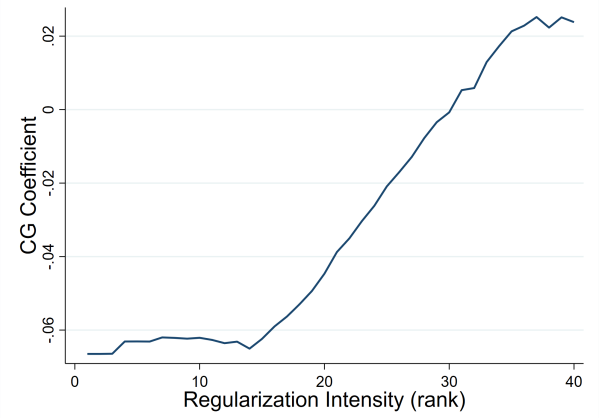
Panel A: γ_{CG}, analyst forecast included



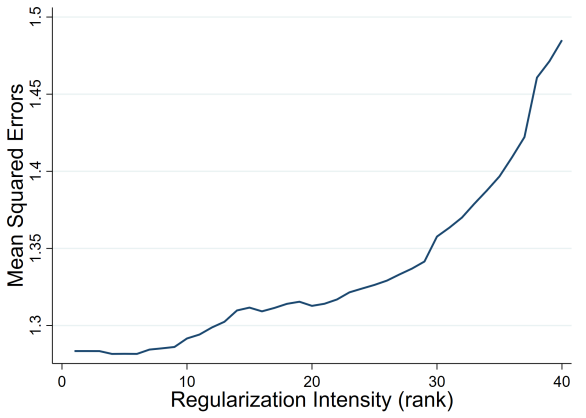
Panel B: MSE, analyst forecast included



Panel C: γ_{CG}, analyst forecast excluded



Panel D: MSE, analyst forecast excluded



This figure plots the Coibion and Gorodnichenko (2015) coefficient and the out-of-sample MSE against joint regularization intensity in random forests, at the one-year forecast horizon. Unlike Figure 2, which varies a single hyperparameter at a time, here the three hyperparameters that govern regularization in a random forest, the maximum tree depth, the minimum number of samples per leaf, and the maximum share of features considered at each split, are varied jointly over 40 specifications. The specifications are ordered from the weakest regularization (rank 1: maximum depth 50, minimum leaf size 1, maximum features 1.00) to the strongest (rank 40: maximum depth 4, minimum leaf size 32, maximum features 0.35), so regularization intensity increases from left to right along the horizontal axis. Panels A and B use the baseline predictor set, which includes the consensus analyst forecast. Panels C and D exclude the consensus analyst forecast from the predictor set, holding the estimation sample fixed. Panels A and C plot γ_{CG} from the regression of the one-year forecast error on the one-year forecast revision, with firm and year fixed effects and standard errors clustered by firm. Panels B and D plot the corresponding out-of-sample MSE. All specifications are estimated on the same sample of 99,963 firm-year observations.

Table IA1: Data Cleaning

Step	Detail	Date	Obs
Construct Training Sample			
1	CRSP Data Monthly	1980 Jan to 2023 Dec	2,704,599
2	Compustat Annual	1964 to 2025 (fyear)	496,758
3	Macro Data (RGDP RCON INDPROD UNEMP)	1948 Jan to 2024 Apr	915
4	Financial Ratios Suite by WRDS	1980 Jan to 2024 Dec (public_date)	2,619,622
5	IBES Actual EPS Detail Unadjusted Annual	1976-01-15 to 2025-04-30 (ANNDATS)	937,986
6	IBES EPS Summary Unadjusted Annual	1980-01-31 to 2025-02-28 (FPEDATS)	3,517,900
7	Merge all file	1984 Jan to 2019 Dec	1,364,372
Construct Final Sample			
8	Start with merging all files	1984 Jan to 2019 Dec (YearMonth)	1,364,372
9	Conduct ML Training	1984 Jan to 2019 Dec (YearMonth)	1,312,517
10	Select Firm-Year Level Prediction	1983 to 2020 (FPEDATS)	124,491
11	Merge in Forecast Revision	1984 to 2020 (FPEDATS)	107,745
12	Keep Observation with NonMissing ML Forecast Error	1985 to 2020 (FPEDATS)	102,296
13	Keep Year from 1986 to 2019	1986 to 2019 (FPEDATS)	101,737
14	Drop Firms with only 1 year observation	1986 to 2019 (FPEDATS)	99,963

This table reports the data cleaning step.

Table IA2: Machine Learning Predictors Definitions

Variable	Definition	Variable	Definition
A. Firm Fundamental			
IBES actual (1)			
Last EPS	Last Year Actual EPS		
WRDS Financial Ratios (67)			
accrual	Accruals/Average Assets	intcov_ratio	Interest Coverage Ratio
adv_sale	Advertising Expenses/Sales	inv_turn	Inventory Turnover
aftret_eq	After-tax Return on Average Common Equity	invt_act	Inventory/Current Assets
aftret_equity	After-tax Return on Total Stockholders Equity	lt_debt	Long-term Debt/Total liabilities
aftret_invcapx	After-tax Return on Invested Capital	lt_ppent	Total Liabilities/Total Tangible Assets
at_turn	Asset Turnover	npm	Net Profit Margin
bm	Book/Market	ocf_lct	Operating CF/Current Liabilities
capei	Shiller's Cyclically Adjusted P/E Ratio	opmad	Operating Profit Margin After Depreciation
capital_ratio	Capitalization Ratio	opmbd	Operating Profit Margin Before Depreciation
cash_conversion	Cash Conversion Cycle (Days)	pay_turn	Payables Turnover
cash_debt	Cash Flow/Total Debt	pcf	Price/Cash flow
cash_lt	Cash Balance/Total Liabilities	pe_exi	P/E (Diluted, Excl. EI)
cash_ratio	Cash Ratio	pe_inc	P/E (Diluted, Incl. EI)
cfm	Cash Flow Margin	PEG_trailing	Trailing P/E to Growth (PEG) ratio
curr_debt	Current Liabilities/Total Liabilities	pretret_earnat	Pre-tax Return on Total Earning Assets
curr_ratio	Current Ratio	pretret_noa	Pre-tax return on Net Operating Assets
de_ratio	Total Debt/Equity	profit_lct	Profit Before Depreciation/Current Liabilities
debt_assets	Total Debt/Total Assets (1)	ps	Price/Sales
debt_at	Total Debt/Total Assets (2)	ptb	Price/Book
debt_capital	Total Debt/Capital	ptpm	Pre-tax Profit Margin
debt_ebitda	Total Debt/EBITDA	quick_ratio	Quick Ratio (Acid Test)
debt_invcap	Long-term Debt/Invested Capital	RD_SALE	Research and Development/Sales
divyield	Dividend Yield	rect_act	Receivables/Current Assets
dltt_be	Long-term Debt/Book Equity	rect_turn	Receivables Turnover
dpr	Dividend Payout Ratio	roa	Return on Assets

Table IA2: Machine Learning Predictors Definitions (continued)

Variable	Definition	Variable	Definition
efftax	Effective Tax Rate	roce	Return on Capital Employed
equity_invcap	Common Equity/Invested Capital	roe	Return on Equity
evm	Enterprise Value Multiple	sale_equity	Sales/Stockholders Equity
fcf_ocf	Free Cash Flow/Operating Cash Flow	sale_invcap	Sales/Invested Capital
gpm	Gross Profit Margin	sale_nwc	Sales/Working Capital
GProf	Gross Profit/Total Assets	short_debt	Short-Term Debt/Total Debt
int_debt	Interest/ Average Long-term Debt	staff_sale	Labor Expenses/Sales
int_totdebt	Interest/Average Total Debt	totdebt_invcap	Total Debt/Invested Capital
intcov	After-tax Interest Coverage		
CRSP (2)			
ret	Monthly Return	prc	Closing Price
Fundamental Per Share (26, optional)			
dividend_p	Dividend Per Share	inventory_p	Inventory Per Share
BE_p	Book Equity Per Share	receivables_p	Receivables Per Share
Liability_p	Total Liability Per Share	cur_debt_p	Short-term Debt Per Share
cur_liability_p	Current Liabilities Per Share	interest_p	Interest Per Share
LT_debt_p	Long-term Debt Per Share	fcf_ocf_p	Free Cash Flow Per Share
cash_p	Cash Balance Per Share	evm_p	Enterprise Value Per Share
total_asset_p	Total Asset Per Share	sales_p	Sales Per Share
tot_debt_p	Total Debt Per Share	invcap_p	Invested Capital Per Share
accrual_p	Accruals Per Share	c_equity_p	Current Equity Per Share
EBIT_p	EBIT Per Share	rd_p	RD Per Share
cur_asset_p	Current Asset Per Share	opmad_p	Operating Income After Depreciation Per Share
pbda_p	Operating Income before DA Per Share	gpm_p	Gross Profit Per Share
ocf_p	Operating Cash Flow Per Share	ptpm_p	Pretax Income Per Share
B. Macroeconomic variables (4)			
RCON	Log Difference of Real Consumption	INDPROD	Log Difference of Industrial Production Index
RGDP	Log Difference of Real GDP	UNEMP	Unemployment rate
C. Analyst forecasts (1)			
AF	Analyst Forecasts		

Table IA3: Forecast Errors

Panel A: Summary Statistics from Our Manuscript

	RF	AF	AE	RF-AE	t(RF-AE)	AF-AE	t(AF-AE)	$(\text{RF-AE})^2$	$(\text{AF-AE})^2$	Obs
Y1	1.197	1.314	1.168	0.029	2.018	0.146	6.195	0.594	0.643	1,246,821
Y2	1.368	1.731	1.367	0.001	0.016	0.364	8.280	1.790	1.839	1,127,459

Panel B: Summary Statistics from Code Used in Zhang et al. (2025)

	RF	AF	AE	RF-AE	t(RF-AE)	AF-AE	t(AF-AE)	$(\text{RF-AE})^2$	$(\text{AF-AE})^2$	Obs
Y1	1.184	1.306	1.157	0.027	1.887	0.149	6.136	0.602	0.655	1,265,472
Y2	1.356	1.727	1.359	-0.004	-0.072	0.368	8.261	1.805	1.856	1,136,233

This table tabulates the forecasting errors of Random Forests using a procedure closely aligned with the code published by Zhang et al. (2025). Panel A reports our results, while Panel B reproduces the results without look-ahead bias from Panel C of Table 3, output cell 8, from https://github.com/yandi-zhu/Man-versus-Machine-Learning-Revisited/blob/main/code/03_Main.ipynb.

Table IA4: Parameter Values for Simulation

This table shows the baseline parameter values used for the simulation results. The value of θ is chosen based on the diagnostic parameter used in the literature (Bordalo et al., 2026). The riskless rate r is normalized to 1. In the baseline case, the fraction of rational analysts is set to $\lambda = 1$, to illustrate that we could obtain a non-zero Coibion and Gorodnichenko (2015) test coefficient, γ_{CG}, even when all agents are rational. The rest of the parameters are not calibrated to match real-world data. Instead, they are chosen to test the mechanism of overreaction of ML forecasts. Specifically, parameters associated with 'Per share return', μ_y, ϕ and α are chosen to generate large variations in this economy and increase the sensitivity of the firm's investment to investors' expectations.

Per share return	δ_A	$\sigma(\epsilon_A), \rho_A^{-\frac{1}{2}}$	δ	$\sigma(\epsilon_z), \rho_y^{-\frac{1}{2}}$	$\sigma(\eta), \rho_\eta^{-\frac{1}{2}}$	$\sigma(\tilde{x}), \rho_x^{-\frac{1}{2}}$
	0.50	1.00	0.50	20.00	0.80	1.00
Investor decision	θ	λ	α	r	μ_y	
	1.00	1.00	0.50	1.00	10.00	
Firm decision	ϕ					
	0.10					

Table IA5: CG Coefficients in Simulated Data

		(1) Average γ_{CG}	(2) Percentage of positive γ_{CG}
Baseline ($\lambda = 1$)		-0.0072	50%
λ	0.5	-0.0073	50%
	0	-0.0074	50%

This table presents the Coibion and Gorodnichenko (2015) test coefficients estimated using data simulated from the economic model. For each set of parameters, we simulate 100 panels, each of which consists of 1000 firms spanning 300 periods (with burn-in periods of 201). For each panel, we estimate the Coibion and Gorodnichenko (2015) test coefficient, γ_{CG}, from the equation below

$$e_{t+1} = \gamma_0 + \gamma_{CG}\, r_t + \omega_t, \tag{IA13}$$

where $e_t = y_t - F_{t-1,t}$ is the forecast error and $r_t = F_{t-1,t} - F_{t-2,t}$ is the revision. The forecasts are made using ridge. Predictors include the idiosyncratic profitability y in the previous period, aggregate profitability A and investment x. Column 1 shows γ_{CG} (times 100 for display) averaged across the 100 panels. Column 2 shows the percentage of γ_{CG} that are positive. The regularization penalty α is set to 1000. In the baseline scenario, all analysts are rational (λ=1). We test the results when we change the percentage of rational analysts in the economy. When $\lambda = 0.5$, half of the analysts are rational, and when $\lambda = 0$, all analysts are overreacting. The remaining parameter values are given in Table IA4.

Table IA6: Joint Regularization Intensity and Forecast Predictability: Random Forests

Panel A: Random Forests, Including Analyst Forecasts						
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Max depth	50	7	7	6	5	4
Min samples per leaf	1	5	10	16	25	32
Max features	1.00	1.00	0.80	0.60	0.40	0.35
	(Low Reg)					(High Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.026***	0.044***	0.079***	0.131***	0.188***	0.200***
	(3.10)	(5.36)	(9.76)	(15.03)	(19.04)	(19.00)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.10	0.10	0.11	0.12	0.15	0.17
Out-of-Sample MSE	0.889	0.892	0.905	0.944	1.044	1.122
Panel B: Random Forests, Excluding Analyst Forecasts						
Dependent variable:	$y_{i,t+1} - F_t y_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Max depth	50	7	7	6	5	4
Min samples per leaf	1	5	10	16	25	32
Max features	1.00	1.00	0.80	0.60	0.40	0.35
	(Low Reg)					(High Reg)
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	−0.067***	−0.065***	−0.039***	−0.008	0.023**	0.024**
	(−6.46)	(−6.06)	(−3.63)	(−0.74)	(2.14)	(2.19)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	99,963	99,963	99,963	99,963	99,963	99,963
Adj R^2	0.08	0.09	0.10	0.12	0.16	0.19
Out-of-Sample MSE	1.284	1.310	1.314	1.337	1.409	1.485

This table reports OLS regression results testing whether machine learning forecast errors are predictable from forecast revisions, following Coibion and Gorodnichenko (2015), as the intensity of regularization in Random Forests varies. The dependent variable is the one-year-ahead forecast error. Unlike Tables 5 and 6, which vary a single hyperparameter at a time, here all three Random Forest hyperparameters move jointly along a single regularization path: from column (1) to column (6) the maximum tree depth falls, the minimum number of observations per leaf rises, and the fraction of predictors considered at each split falls, so that regularization increases monotonically from left to right. Panel A includes the analyst consensus forecast in the predictor set, while Panel B excludes it holding the estimation sample fixed. The t-statistics in parentheses are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table IA7: Machine Learning Forecast Errors Are Robustly Predictable (Without Clustered Standard Errors)

Panel A	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+1} - F_t y_{i,t+1}$			
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.114***	0.006***	0.058***	0.044***
	(29.80)	(2.71)	(16.61)	(12.84)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster at Firm	No	No	No	No
N	99,963	99,963	99,963	99,963
Adj R^2	0.16	0.11	0.11	0.10
Panel B	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+1} - F_t y_{i,t+1}$			
$F_t y_{i,t+1} - F_{t-1} y_{i,t+1}$	0.122***	0.009***	0.064***	0.047***
	(15.13)	(2.68)	(7.46)	(6.17)
Firm FE	Yes	Yes	Yes	Yes
Year FE	No	No	No	No
Cluster at Firm	No	No	No	No
N	99,963	99,963	99,963	99,963
Adj R^2	0.14	0.09	0.09	0.09

This table reports OLS regression results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The dependent variable is the one-year-ahead forecast error, defined as realized minus one-year-ahead predicted EPS. The key independent variable is the change in forecasts from t-1 to t. Standard errors and fixed effects are indicated. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table IA8: Machine Learning Forecast Errors Are Robustly Predictable at Two-Year Horizon (Without Clustered Standard Errors)

Panel A	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+2} - F_t y_{i,t+2}$			
$F_t y_{i,t+2} - F_{t-1} y_{i,t+2}$	-0.003	-0.606***	-0.233***	-0.237***
	(-0.43)	(-159.30)	(-37.30)	(-38.90)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster at Firm	No	No	No	No
N	58,864	58,864	58,864	58,864
Adj R^2	0.12	0.36	0.18	0.18
Panel B	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+2} - F_t y_{i,t+2}$			
$F_t y_{i,t+2} - F_{t-1} y_{i,t+2}$	-0.004	-0.496***	-0.316***	-0.319***
	(-0.31)	(-20.58)	(-28.58)	(-29.40)
Firm FE	Yes	Yes	Yes	Yes
Year FE	No	No	No	No
Cluster at Firm	No	No	No	No
N	58,864	58,864	58,864	58,864
Adj R^2	0.07	0.27	0.11	0.10

This table reports OLS regression results testing whether forecast errors are predictable using forecast revisions, following Coibion and Gorodnichenko (2015). The dependent variable is the two-year-ahead forecast error, defined as realized minus two-year-ahead predicted EPS. The key independent variable is the change in forecasts from t-1 to t. Standard errors and fixed effects are indicated. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table IA9: Numerical Exercise: Predicted CG Coefficients

Scenario	Parameters				
	σ_η^2/σ_s^2	ρ_s	ρ_η	λ_{norm}	γ_{CG}
Low Noise	0.25	0.9	0.5	0.0	-0.216
				0.1	-0.159
				0.3	-0.045
				0.5	0.069
Medium Noise	0.50	0.9	0.5	0.0	-0.279
				0.1	-0.235
				0.3	-0.147
				0.5	-0.058
High Noise	1.00	0.9	0.5	0.0	-0.327
				0.1	-0.293
				0.3	-0.224
				0.5	-0.155

This table reports predicted Coibion and Gorodnichenko (2015) test coefficients γ_{CG}. Parameters calibrated to match Compustat earnings volatility and IBES forecast patterns. The scenarios set the measurement noise-to-signal ratio σ_η^2/σ_s^2 to 0.25 (Low), 0.50 (Medium), and 1.00 (High). λ_{norm} is the regularization penalty normalized by total signal variance $\sigma_s^2 + \sigma_\eta^2$, and fundamental and noise persistence are fixed at $\rho_s = 0.9$ and $\rho_\eta = 0.5$.

Table IA10: Numerical Exercise: Predicted CG Coefficients with Full Information

Scenario	Parameters				
	σ_η^2/σ_s^2	ρ_s	ρ_η	λ_{norm}	γ_{CG}
Low Noise	0.25	0.9	0.5	0.0	$+0.000$
				0.1	$+0.239$
				0.3	$+0.623$
				0.5	$+0.954$
Medium Noise	0.50	0.9	0.5	0.0	-0.000
				0.1	$+0.189$
				0.3	$+0.511$
				0.5	$+0.797$
High Noise	1.00	0.9	0.5	0.0	-0.000
				0.1	$+0.152$
				0.3	$+0.421$
				0.5	$+0.668$

This table reports predicted Coibion and Gorodnichenko (2015) test coefficients γ_{CG} when the forecaster uses the full signal history $(z_t, z_{t-1}, \dots)$ rather than the contemporaneous signal z_t alone. Forecasts are formed by ridge regression of y_{t+1} on the signal history, and γ_{CG} is computed numerically because the ridge coefficient has no closed form. The scenarios set the measurement noise-to-signal ratio σ_η^2/σ_s^2 to 0.25 (Low), 0.50 (Medium), and 1.00 (High). λ_{norm} is the regularization penalty normalized by total signal variance $\sigma_s^2 + \sigma_\eta^2$, and fundamental and noise persistence are fixed at $\rho_s = 0.9$ and $\rho_\eta = 0.5$.

Table IA11: Numerical Exercise: Predicted CG Coefficients with Full Information and Reporting Noise

		Reporting noise $\kappa = \sigma_\xi^2/\mathrm{Var}(\bar{r}_t)$			
		(1)	(2)	(3)	(4)
σ_η^2/σ_s^2	λ_{norm}	0.0	0.5	1.0	2.0
0.25	0.0	$+0.000$	-0.250	-0.333	-0.400
	0.1	$+0.239$	-0.131	-0.254	-0.352
	0.3	$+0.623$	$+0.062$	-0.126	-0.275
	0.5	$+0.954$	$+0.227$	-0.015	-0.209
0.50	0.0	-0.000	-0.250	-0.333	-0.400
	0.1	$+0.189$	-0.155	-0.270	-0.362
	0.3	$+0.511$	$+0.005$	-0.163	-0.298
	0.5	$+0.797$	$+0.148$	-0.068	-0.241
1.00	0.0	-0.000	-0.250	-0.333	-0.400
	0.1	$+0.152$	-0.174	-0.283	-0.370
	0.3	$+0.421$	-0.040	-0.193	-0.316
	0.5	$+0.668$	$+0.084$	-0.111	-0.266

This table reports predicted Coibion and Gorodnichenko (2015) test coefficients γ_{CG} when the forecaster uses the full signal history $(z_t, z_{t-1}, \dots)$ with reporting noise. λ_{norm} is the regularization intensity normalized by the total signal variance. γ_{CG} is the Coibion-Gorodnichenko test coefficient. Parameters are calibrated to match Compustat earnings volatility and IBES forecast patterns. The definitions of κ, σ_η, σ_s, λ, σ_ξ are given in Internet Appendix B.4.

Table IA12: Level Tests for Overreaction

Panel A: Investment as Signal				
	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+2} - F_t y_{i,t+2}$			
Investment	-1.860***	-2.285***	-2.061***	-2.154***
	(-13.947)	(-16.598)	(-15.715)	(-16.236)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	83,084	83,084	83,084	83,084
Adj R^2	0.14	0.12	0.15	0.14
Panel B: Debt Issuance as Signal				
	(1) Human Analysts	(2) Ridge	(3) Gradient Boosting	(4) Random Forest
	$y_{i,t+2} - F_t y_{i,t+2}$			
Debt Net Issuance	-0.729***	-0.783***	-0.785***	-0.800***
	(-17.089)	(-17.486)	(-17.922)	(-18.306)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	83,672	83,672	83,672	83,672
Adj R^2	0.14	0.11	0.15	0.13

This table reports OLS regression results testing whether forecast errors are predictable using the levels of signals, following Bordalo et al. (2018). The dependent variable is the two-year-ahead forecast error. In Panel A, the signal is investment rate. In Panel B, the signal is net debt issuance. The t-statistics (in parentheses) are computed using standard errors clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.