Nobody Trades the Ensemble: Feature Endowments and Disagreement in Machine-Learning Asset Pricing

Qian Yang*

August 12, 2026

Abstract

Machine-learning asset-pricing studies typically evaluate models using the same comprehensive set of return predictors. We examine a setting in which investors observe equal-sized but differently composed subsets of predictors. Using 94 firm characteristics and five model classes, we decompose forecast disagreement into feature-endowment, model-choice, and interaction components. Under random endowments, feature composition explains 46% to 49% of forecast disagreement, compared with 12% to 16% for model choice. The feature channel remains twice as large under publication-vintage endowments. The best-performing model varies across endowments. Averaging forecasts across investors produces an ensemble forecast that substantially outperforms the median investor, although no individual investor observes the full ensemble. Strategies based on an investor's deviation from the ensemble earn returns centered near zero, whereas deviations of the full-feature forecast are informative. These results show that heterogeneous predictor access is an important source of disagreement in machine-learning asset pricing.

JEL classification: G11, G12, G14, C45, C58.

Keywords: machine learning, return prediction, disagreement, information sets, feature endowments, ensembles

---

Preliminary. Please do not cite or circulate without permission. Acknowledgment: I have no conflict of interest to disclose.

*DeGroote School of Business, McMaster University. E-mail: yangq109@mcmaster.ca.

1. Introduction

Machine-learning studies in asset pricing commonly evaluate forecasting models with a comprehensive set of firm characteristics, macroeconomic interactions, and industry indicators (Gu et al., 2020). This design provides a useful benchmark, but it assigns the same predictor set to every model and every investor. In practice, investors may construct different subsets because they use different data vendors, research procedures, and computing systems. The reported performance of a full-feature model may therefore differ from the performance available to an investor with a more limited predictor set. In our sample, the full-feature benchmark earns an annualized long-short return of 15.8%, whereas the median feature-constrained investor earns between 1.4% and 4.9%.

The full library is itself incomplete. The 94 characteristics comprise the subset of candidate signals documented in the academic literature. The broader signal space is substantially larger. Yan and Zheng (2017) construct more than 18,000 fundamental signals from financial statements, while Chordia et al. (2020) examine more than two million trading strategies randomly generated from real data. Proprietary signals do not enter the academic record. Thus, neither investors nor econometricians observe the complete signal space. Every information set, including the full-feature benchmark, is an endowment defined by a particular subset of signals. The benchmark differs only because it contains the union of published characteristics. Studying these endowments does not impose an artificial constraint on otherwise fully informed agents. Rather, it describes the general case, with the standard full-library evaluation as one special case. Restricting the analysis to the 94 published characteristics keeps all endowments observable and replicable. Because a broader signal space is likely to produce at least as much cross-investor dispersion and no smaller attainability gap, the estimates we report are conservative.

We study heterogeneity in *feature endowments*, defined as the set of predictors available to an investor. We address three questions. First, how much do investors disagree when they observe the same number of predictors but different predictor sets? Second, does the best-performing model class depend on the investor's feature endowment? Third, how does the performance of an individual investor compare with the performance of a forecast that aggregates information across investors?

Our empirical design uses the panel constructed by Gu et al. (2020). We create 20 synthetic investors. Each investor receives k firm characteristics drawn from a universe of 94 characteristics, and the draw is held fixed over time. The investor combines these characteristics with eight aggregate predictors and industry indicators. Each investor then estimates five model classes: ordinary least squares, principal-components regression, partial

least squares, elastic net, and a three-layer neural network. The models use identical rolling estimation windows and produce forecasts from 1987 to 2021. The resulting crossed design contains 20 investors, five model classes, and 35 test years. It allows us to decompose forecast disagreement into a feature-endowment component, a model-choice component, and their interaction.

We obtain four main findings. First, investors with equal-sized endowments produce materially different forecasts. Two investors who each hold ten characteristics have an average cross-sectional rank correlation of about 0.2. Because endowment size is fixed, this dispersion reflects differences in predictor composition rather than differences in the number of predictors. The predictor sets can nevertheless differ in informational value, and we examine this distinction in the tests of deviation returns.

Second, the relative importance of feature composition and model choice depends on the extent to which investors share the same predictors. Under uniform random endowments, feature composition accounts for 46% to 49% of forecast disagreement at $k = 5$, 10, and 20. Model choice accounts for 12% to 16%. The feature component is therefore about three to four times as large as the model component. When all investors share size, book-to-market, and momentum, the feature and model components are similar at 32% and 30%, respectively. When illiquidity and return volatility are added to the common set, the feature component falls to 19%, and the model component rises to 48%. Thus, feature composition is more important when predictor access is fragmented, whereas model choice is more important when investors share the strongest predictors.

We also construct publication-vintage endowments. In each year, an investor can hold only characteristics published by that date. Investors adopt newly published characteristics according to fixed random priorities. This design uses historically available predictor sets and avoids assigning a recently published characteristic to an investor in an earlier year. Under this design, the feature component is twice as large as the model component. The publication-vintage results therefore place the historically feasible case in the feature-dominated region. Machine forecast disagreement, which varies the forecasting rule while holding the predictor set fixed, corresponds to the opposite case in which investors share a common feature set (Bali et al., 2026).

Third, the best-performing model class varies across feature endowments. Under uniform endowments, each of the five model classes produces the highest out-of-sample Sharpe ratio for at least one investor at every value of k. Under the common-core designs, fewer model classes attain the highest Sharpe ratio, but the model that performs best with all 94 characteristics is best for only 3 of 20 investors. These results imply that model comparisons based on a single common predictor set need not identify the best model for an investor

with a different predictor set. A dynamic model-selection rule also performs poorly. An investor who selects the model with the highest trailing Sharpe ratio underperforms both a fixed-model benchmark and an equal-weight average of model forecasts in every endowment design.

Fourth, averaging forecasts across investors produces a random-subspace ensemble that no individual investor observes. For four of the five model classes, the investor consensus earns an annualized long-short return between 8.6% and 9.2% at $k = 10$, compared with 15.8% for the full-feature benchmark. The median investor earns between 1.4% and 4.9% across the endowment sizes we study. The ambiguity decomposition of Krogh and Vedelsby (1994) provides a direct relation in forecast-error space: the difference between the mean individual squared error and the consensus squared error equals cross-investor forecast dispersion. Portfolio returns are nonlinear functions of forecasts, so the portfolio performance differences are related empirical results rather than consequences of the identity.

We build a stylized linear model to rationalize the empirical findings. Returns depend on firm characteristics, and each investor forms the population best linear forecast using the characteristics in their endowment. When characteristics are correlated, included predictors absorb part of the loadings on omitted predictors. Investors can therefore use their own data optimally and still produce different forecasts. The model also distinguishes feature-endowment disagreement from disagreement based on private information. In the model, an investor's deviation from the consensus depends on the random composition and quality of their feature set. Across investors, the expected return to trading these deviations is zero. In the data, the mean annualized return across investor-model deviation strategies is -1.2%. 88% of the strategies are statistically indistinguishable from zero, and the significant estimates include both positive and negative returns. In contrast, a strategy based on the full-feature forecast's deviation from the same consensus earns between 5% and 12% annually, with t-statistics as high as 3.7.

The publication-vintage design provides an additional test of the mechanism. We compare two settings. In the first setting, characteristic inclusion rates are held fixed by construction. Forecast dispersion shows no decline over 35 years. In the second setting, inclusion rates change with publication dates. Dispersion rises as the set of published predictors expands. The model is intended to explain this long-run contrast. It does not explain the annual timing of changes in dispersion.

The feature-endowment interpretation requires persistent differences in the predictor sets available to investors. Several features of the setting make this assumption plausible. First, the full set of 94 characteristics was not available throughout the sample. Only 17 had been published by 1990, and 37 by 2000. Second, characteristic construction can require

point-in-time accounting data, analyst forecasts, intraday quotations, and other source files that are obtained from different vendors. Third, investors may implement the same published characteristics differently because of choices concerning reporting lags, restatements, missing values, and industry adjustments. Fourth, the construction and joint evaluation of many predictors require research and computing resources. The analysis does not require every investor to hold a small feature set. It requires only persistent variation in feature composition across market participants. We therefore report results for endowment sizes of 5, 10, and 20, as well as designs that vary the size of a common predictor set.

This paper contributes to several strands of literature. The closest counterpart is machine forecast disagreement (Bali et al., 2026). That approach holds the input data fixed and compares forecasts from different model specifications. We study variation in input sets and cross it with variation in model classes. The two sources of disagreement are therefore measured within one design. Li et al. (2025) compare machine-learning strategies based on a broad, real-time universe of signals with strategies based on curated signal sets and show that feature engineering materially affects performance. We study a different implication of feature selection: investors who use different subsets of predictors can disagree even when they use the same forecasting method.

The paper also relates to evidence on the out-of-sample and post-publication performance of return predictors (McLean and Pontiff, 2016; Chen and Velikov, 2023; Jensen et al., 2023; Harvey et al., 2016). These studies evaluate the reliability and attenuation of predictor returns. We instead hold the predictor library fixed within each design and study variation in predictor composition across investors. The publication-vintage analysis adds a time-series result: the expansion of the published predictor set is associated with greater forecast dispersion among feature-constrained investors. Research on characteristic selection asks which predictors contain incremental information (Kozak et al., 2020; Freyberger et al., 2020). We take the predictor library as given and study how its allocation across investors affects forecasts. Finally, theories of disagreement commonly emphasize heterogeneous signals, beliefs, or short-sale constraints (Miller, 1977; Diether et al., 2002; Hong and Stein, 2007). Our mechanism instead uses heterogeneous subsets of public predictors. The random-subspace method and the ambiguity decomposition are established statistical results (Ho, 1998; Breiman, 2001; Krogh and Vedelsby, 1994). Our contribution is to apply them to heterogeneous investor information sets.

The remainder of the paper is organized as follows. Section 2 describes the data and the feature-endowment designs. Section 3 presents the variance decomposition, model-performance comparisons, and ensemble results. Section 4 develops the model. Section 5 reports the empirical tests. Section 6 concludes.

2. Data and Feature-Endowment Designs

2.1. Panel, characteristics, and models

We use the data set of Gu et al. (2020). The panel combines 94 firm characteristics with monthly CRSP returns from 1957 to 2021. Within each month, each characteristic is transformed to its cross-sectional rank and scaled to $[-1, 1]$. We remove stock-month observations for which fewer than 20% of the characteristics are available. For the remaining observations, missing values are replaced with the monthly cross-sectional median, which is zero after the rank transformation. The predictor set also includes eight aggregate variables from Welch and Goyal (2008), lagged by one month, and 74 indicators for two-digit SIC industries. Table 1 reports summary statistics.

[Table 1 about here.]

We estimate five model classes considered by Gu et al. (2020). The linear classes are ordinary least squares (OLS), principal-components regression (PCR), partial least squares (PLS), and elastic net (ENet). The nonlinear class is a neural network with three hidden layers (NN3). The NN3 architecture contains 32, 16, and 8 units and is estimated with five random seeds. Its forecast is the average across the five estimated networks. All investors use the same expanding training windows, twelve-year validation periods, and hyperparameter grids. The models produce annual test samples from 1987 through 2021. Appendix A.1 provides the implementation details.

2.2. Feature endowments

The unit of an endowment is a base firm characteristic. We treat the aggregate variables as common information and their interactions with firm characteristics as mechanical transformations. Thus, an investor who can construct book-to-market can also construct its interaction with the term spread. Each synthetic investor $i \in \{1, \ldots, 20\}$ receives k firm characteristics drawn uniformly without replacement from the set of 94 characteristics. Draws are independent across investors and remain fixed over the sample. This design represents persistent differences in predictor availability. Every investor observes the aggregate variables and industry indicators. Investor i therefore uses the design matrix $[\, x_{S_i} \mid x_{S_i} \otimes z \mid \text{SIC} \,]$, which contains $9k + 74$ columns, while the full-feature design contains 920 columns.

The baseline analysis sets $k = 10$. We use this value as a reference point rather than as an estimate of the number of signals held by a typical investor. Two findings from the

characteristic-selection literature provide useful benchmarks. Green et al. (2017) jointly examine 94 characteristics and identify 12 reliable independent determinants of returns among non-microcap stocks from 1980 to 2014, with only two remaining after 2003. DeMiguel et al. (2020) identify six jointly significant characteristics without transaction costs and 15 when transaction costs are included. These results place ten characteristics within the range of estimated effective dimensions in related work. Survey evidence also indicates that institutional factor portfolios usually combine a limited number of strategies. For example, the 2016 FTSE Russell survey reports that 21% of smart-beta adopters used at least five strategies, compared with 2% in 2014.[1] Finally, only 17 of the 94 characteristics had been published by 1990. A ten-characteristic endowment therefore represents more than half of the published set near the beginning of the test period.

We repeat the analysis with $k = 5$ and $k = 20$ using new independent draws. We also estimate each model with all 94 characteristics to construct the full-feature benchmark. At $k = 10$, two investors share 1.11 characteristics on average, so the baseline endowments have little overlap.

Uniform draws isolate the effect of predictor composition, but they may produce less overlap than is present among sophisticated investors. We therefore construct two common-core designs at $k = 10$. In the three-variable core design, every investor holds size, book-to-market, and momentum and draws seven additional characteristics from the remaining set. In the five-variable core design, every investor also holds an illiquidity characteristic and a return-volatility characteristic, and draws five additional characteristics. Endowment size remains fixed at ten. The three designs vary only in the share of predictors that is common across investors.

The uniform and common-core designs assign characteristics from the complete 94-variable library in every year. This convention permits an investor in an early test year to hold a characteristic that was published later. We address this issue with a publication-vintage design. We use the publication year of each characteristic's originating study, as reported by Green et al. (2017). In test year t, the feasible set contains only characteristics published by year $t - 1$. Each investor receives a fixed random ordering of the 94 characteristics. In each year, her endowment contains the first k characteristics in that ordering that have already been published. This procedure uses no future publication information in the investor's annual predictor set and preserves a persistent investor-specific priority ordering.

The published set increases from six characteristics in 1987 to 94 in 2020. From 1987

[1] FTSE Russell, *Smart Beta: 2016 Global Survey Findings from Asset Owners*, 2016. This survey statistic describes implemented strategies and is not a direct estimate of the number of firm characteristics used by an investor.

through 1989, fewer than ten characteristics are available, and all investors therefore hold the same published set. Cross-investor dispersion is mechanically zero in these common-endowment years. Beginning in 1990, every investor holds ten characteristics. Appendix A.2 lists the baseline uniform endowments and reports their overlap statistics.

For each stock-month, the crossed design produces a 20×5 matrix of forecasts. We standardize each investor-model forecast within the monthly cross-section before measuring disagreement. This transformation removes forecast-scale differences and focuses the analysis on differences in the cross-sectional ranking of stocks.

2.3. Interpretation of constrained endowments

The endowment designs represent several sources of persistent variation in predictor availability. First, many characteristics require source data that are obtained from separate databases, including point-in-time accounting files, analyst forecasts, and intraday quotations. Different database access can therefore produce different feasible predictor sets. Second, characteristic construction requires choices concerning reporting lags, restatements, missing values, and industry adjustments. Evidence from anomaly replication studies shows that implementation and data choices can affect the measured performance of return predictors (Chen and Velikov, 2023; Jensen et al., 2023). Third, the joint evaluation of a large predictor set requires research, computing, and model-maintenance resources. Fourth, publication dates provide a common benchmark for when a predictor became publicly documented. Post-publication changes in anomaly returns are consistent with investors learning about published predictors (McLean and Pontiff, 2016).

These considerations motivate heterogeneous feature sets but do not imply that all investors are highly constrained. Large quantitative institutions may construct most of the 94 characteristics, though this library is itself the published subset of a far larger candidate space (Yan and Zheng, 2017; Chordia et al., 2020), so even an investor holding all 94 characteristics holds an endowment, distinguished only by being the published union. The empirical design requires only that the predictor composition differs across some investors. The uniform, common-core, and publication-vintage designs cover a range of overlap patterns. The results can therefore be interpreted at different points in the distribution of predictor sharing rather than at a single value of k.

3. Forecast Disagreement across Feature Endowments

3.1. Variance decomposition

For each stock-month with a complete 20×5 forecast matrix, we decompose the variance of standardized forecasts into an investor main effect, a model main effect, and an investor-model interaction. The investor main effect measures the contribution of feature-endowment composition. The model main effect measures the contribution of model class. The interaction measures the extent to which the effect of model choice varies across feature endowments. The baseline $k = 10$ sample contains about 2.0 million complete stock-month observations. Table 2 reports the decomposition for each endowment size.

[Table 2 about here.]

The feature-endowment component accounts for 46% to 49% of forecast variance at each value of k. The model-choice component accounts for 12% to 16%, and the interaction accounts for about 38%. The ratio of the feature main effect to the model main effect ranges from 2.9 to 3.9. These estimates are conditional on the feature and model spaces used in the analysis. The feature space contains 94 base characteristics, whereas the model space contains five model classes. The NN3 forecast averages across seeds, so seed-level variation is not included in the model component. A broader model space that included additional learning methods, hyperparameter specifications, or individual neural-network seeds would increase the measured model component.

Because endowment size is fixed within each design, the cross-investor component is not generated by differences in the number of predictors. It can still reflect differences in the quality of the selected predictors. In more direct units, the median cross-investor standard deviation of standardized forecasts ranges from 0.68 to 0.77. The average pairwise rank correlation between investors increases from 0.23 at $k = 5$ and $k = 10$ to 0.35 at $k = 20$. Forecast disagreement therefore declines as endowments become larger, but it remains substantial when each investor holds 20 of the 94 characteristics. Figure 1 summarizes these results.

[Fig. 1 about here.]

Figure 2 presents the variance shares and the underlying distributions of cross-investor and cross-model dispersion for the baseline $k = 10$ design.

[Fig. 2 about here.]

3.2. Best-performing model classes

For each investor, we define the best-performing model as the model class with the highest out-of-sample Sharpe ratio of the value-weighted decile long-short portfolio. The best-performing model for the full-feature benchmark is defined in the same way. Under this criterion, OLS is the full-feature benchmark, with a Sharpe ratio of 0.82. Table 3 reports the model counts.

Each model class is best for at least one investor at every endowment size. The distribution changes with k. PCR, PLS, and ENet are selected more frequently when the design is small. OLS is selected for 12 of 20 investors at $k = 20$. The full-feature best-performing model is selected for only 5 of 20 investors at $k = 10$. Thus, the relative performance of the model classes depends on the feature endowment.[2]

[Table 3 about here.]

3.3. Feature sharing across investors

Table 4 compares the uniform, common-core, and publication-vintage designs at $k = 10$. The results change systematically with the degree of feature sharing. From the uniform design to the three-variable and five-variable core designs, the feature share falls from 48.7% to 32.3% and then to 18.7%. The model share rises from 12.6% to 29.7% and then to 48.0%. The two main effects are similar when investors share three characteristics, and the model effect is larger when they share five.

Greater feature sharing also increases agreement in stock rankings. The average pairwise rank correlation rises from 0.23 under uniform endowments to 0.54 under the three-variable core and 0.78 under the five-variable core. The cross-investor standard deviation falls by about one-half. The number of model classes that is best for at least one investor falls from five to two. Even in the common-core designs, however, the full-feature best-performing model is best for only 3 of 20 investors.

The publication-vintage design provides a historically based comparison. The feature share is 45.6%, and the model share is 22.8%. Thus, the feature main effect is twice the model main effect. Four model classes are best for at least one investor, and OLS is best for 12 of 20 investors. These results are consistent with an early published set that contains several widely used predictors and with changes in investor endowments as new characteristics are published. Figure 3 summarizes the comparison across designs.

[2]The main analysis defines the best-performing model by the Sharpe ratio. If annualized return is used instead, ENet is the full-feature best-performing model with a return of 15.8%. The investor counts shift modestly toward ENet, but the conclusion that model rankings vary across endowments is unchanged. Appendix A.3 reports the criterion audit.

[Fig. 3 about here.]

[Table 4 about here.]

The difference between the predicted and observed attenuation provides information about the underlying mechanism. Under the orthogonal-characteristic benchmark in Section 4, assigning a characteristic to every investor removes only its own $\beta_j^2 \pi_j (1-\pi_j)$ contribution to expected dispersion. Using squared univariate characteristic information coefficients as proxies for squared loadings, the predicted dispersion ratio relative to the uniform design is 0.97 for Core-3 and 0.67 for Core-5. The observed ratios are 0.66 and 0.49.

The larger observed decline is consistent with the proxy-loading mechanism in Lemma 4.1. Under uniform endowments, an included characteristic can proxy for omitted correlated characteristics, and the resulting projection loadings vary across investors. A common core removes both the direct contribution of the common characteristics and part of the variation in proxy loadings on correlated characteristics. This interpretation is subject to two limitations. Squared univariate information coefficients are imperfect proxies for multivariate loadings, and the population model does not include the estimation effects that may also become smaller when investors share strong predictors.

3.4. Individual and aggregate forecast performance

Table 5 compares the performance of individual investors, the equal-weight average of the 20 investor forecasts, and the full-feature benchmark. The first two investor rows use NN3 for every value of k, which separates the effect of endowment size from model selection. The aggregate forecast is formed at the stock-month level before portfolio construction.

[Table 5 about here.]

At $k = 10$, the median NN3 investor earns 1.4% annually. For four of the five model classes, the aggregate investor forecast earns between 8.6% and 9.2%. The NN3 aggregate earns 2.4%, consistent with the weaker performance of NN3 at these design widths. The full-feature annualized-return benchmark earns 15.8%. Aggregation therefore recovers a substantial part of full-feature performance even though the individual investors use only a fraction of the predictor set.

The ambiguity identity in Section 4 explains the corresponding forecast-error comparison. The reduction in mean squared error from the average individual forecast to the aggregate forecast equals the measured cross-investor dispersion. We verify this identity directly in Appendix A.4. Relative to the pooled variance of individual forecasts, the dispersion term

ranges from 59% to 91% across model classes. Thus, a large share of an individual forecast's variation is specific to the feature endowment that generated it. In raw mean-squared-error units, the dispersion term is small for regularized models because realized-return variance dominates forecast variance in this low-R^2 setting.

The portfolio results are not implied by the ambiguity identity because decile portfolio formation is a nonlinear function of forecast ranks. They provide a separate economic comparison. For NN3, the difference between the aggregate return and the median investor return changes from -0.2 percentage points at $k = 5$ to 2.1 percentage points at $k = 20$. The corresponding difference also increases when investors share stronger common predictors. Aggregation becomes more valuable when the individual forecasts contain more common signal and less unrelated variation.[3]

4. A Model of Feature-Endowment Disagreement

This section develops a linear framework in which investors observe public information but have access to different subsets of the predictor set. Each investor forms the population best linear forecast using their available characteristics. The model has two objectives. First, it shows how differences in predictor composition generate forecast dispersion when endowment size is held fixed. Second, it relates cross-investor dispersion to the forecast-error advantage of an aggregate forecast through the ambiguity decomposition of Krogh and Vedelsby (1994).

4.1. Setup

A stock's excess return is

$$r = f(x) + \varepsilon, \qquad f(x) \equiv x^\top \beta, \qquad \varepsilon \perp x, \quad \mathbb{E}[\varepsilon] = 0, \tag{1}$$

where $x \in \mathbb{R}^P$ is a vector of firm characteristics, $\mathbb{E}[x] = 0$, $\mathrm{Var}(x) = \Sigma$, and $\beta \in \mathbb{R}^P$ contains the predictive loadings. All characteristics are public in principle.

There are B investors indexed by i. Investor i observes a subset $S_i \subseteq \{1, \ldots, P\}$ with fixed cardinality $|S_i| = k$. The subsets are drawn uniformly without replacement and independently across investors. The subset is an exogenous feature endowment. Because every investor observes the same number of characteristics, investors differ only in endowment

[3] Full-feature OLS and NN3 have weak out-of-sample forecast R^2 but strong decile-portfolio performance. Their forecast magnitudes are unstable in wide designs, while their cross-sectional rankings remain informative. The same scale effect contributes to their larger raw dispersion shares in Appendix A.4.

composition. Equal cardinality does not imply equal predictive content. Endowments that contain high-loading characteristics can be more informative than other endowments of the same size.

Investor i forms the population best linear forecast of r using the characteristics in S_i:

$$m_i(x) \;=\; x_{S_i}^{\top}\beta_{S_i}, \qquad \beta_{S_i} \;=\; \Sigma_{S_iS_i}^{-1}\Sigma_{S_i\cdot}\,\beta, \tag{2}$$

where $\Sigma_{S_iS_i}$ is assumed to be nonsingular and $\Sigma_{S_i\cdot}$ is the covariance between the included characteristics and the full characteristic vector. The model contains no estimation error or behavioral bias. Each investor uses the available characteristics optimally within the class of linear forecasts. Under joint normality, (2) also equals $\mathbb{E}[r \mid x_{S_i}]$. The results below require only the linear-projection interpretation.

Lemma 4.1 (Proxy loading)**.** *Partition the characteristic vector into included and omitted components. If $\Sigma_{S_i,-S_i}\beta_{-S_i} \neq 0$, then $\beta_{S_i} \neq \beta|_{S_i}$. The projection loading on an included characteristic therefore contains part of the predictive loading on correlated omitted characteristics.*

Proof. For $x = (x_S, x_{-S})$, the projection coefficient is $\beta_S^{\text{proj}} = \beta_S + \Sigma_{SS}^{-1}\Sigma_{S,-S}\beta_{-S}$. The second term is nonzero under the condition in the lemma. □

Lemma 4.1 identifies a source of disagreement beyond the direct omission of predictors. Investors with different feature sets also use different projection loadings because each included predictor can proxy for a different collection of omitted predictors.

4.2. Disagreement with equal-sized endowments

Define forecast dispersion at x as the variance of the population forecast over the endowment distribution:

$$D(x) \;\equiv\; \mathrm{Var}_S[m_S(x)]\,. \tag{3}$$

The endowment size k is fixed. This object is the population analogue of the cross-investor forecast dispersion measured in Section 3.1.

Proposition 1 (Composition effect)**.** *Let $\Sigma = I_P$ and define the common inclusion rate as $\pi = k/P$. Then*

$$D(x) \;=\; \pi(1-\pi)\sum_{j=1}^{P}\beta_j^2x_j^2 \;+\; \left(\pi_2 - \pi^2\right)\sum_{j\neq\ell}\beta_j\beta_\ell x_j x_\ell, \qquad \pi_2 \equiv \frac{k(k-1)}{P(P-1)}, \tag{4}$$

and

$$\mathbb{E}_x[D(x)] \;=\; \pi(1-\pi)\sum_{j=1}^{P}\beta_j^2. \tag{5}$$

Expected dispersion is positive whenever at least one characteristic has a nonzero predictive loading and $0 < k < P$. Holding total predictive strength $\sum_j \beta_j^2$ fixed, expected dispersion does not depend on whether the loadings are concentrated or diffuse. Expected dispersion is hump-shaped in π, with a maximum at $\pi = 1/2$, and converges to zero as π approaches zero or one.

Proof. Under $\Sigma = I_P$, $m_S(x) = \sum_j \mathbf{1}\{j \in S\}\beta_j x_j$. For a uniformly drawn subset of size k, $\mathrm{Var}(\mathbf{1}_j) = \pi(1-\pi)$ and $\mathrm{Cov}(\mathbf{1}_j, \mathbf{1}_\ell) = \pi_2 - \pi^2$ for $j \neq \ell$. These expressions yield (4). Taking expectations over x eliminates the cross terms because $\mathbb{E}[x_j x_\ell] = \mathbf{1}\{j = \ell\}$, which yields (5). □

Proposition 1 establishes that feature composition can generate forecast dispersion even when all investors observe the same number of predictors and use them optimally. The proposition concerns the raw population forecasts. It does not imply that the standardized empirical dispersion measure must be monotonic in k over the range used in the empirical analysis.

The expected dispersion in (5) is independent of signal sparsity, but the dispersion for an individual stock depends on how the stock's signal is distributed across characteristics.

Corollary 1 (Signal concentration)**.** *For a fixed stock, let $s_j \equiv \beta_j x_j$, so that $f(x) = \sum_j s_j$. Under the assumptions of Proposition 1,*

$$D(x) \;=\; (\pi - \pi_2)\sum_{j=1}^{P} s_j^2 \;+\; (\pi_2 - \pi^2)f(x)^2, \qquad \pi - \pi_2 > 0, \quad \pi_2 - \pi^2 < 0. \tag{6}$$

Conditional on $f(x)$, dispersion increases with $\sum_j s_j^2$. It is highest when the signal is concentrated in a small number of characteristics and lowest when the same total signal is distributed evenly across characteristics. If characteristics are included independently with rates π_j, the first term becomes $\sum_j \pi_j(1-\pi_j)s_j^2$.

Proof. Substitute $\sum_{j\neq\ell} s_j s_\ell = f(x)^2 - \sum_j s_j^2$ into (4). Under sampling without replacement, $\pi - \pi_2 = k(P-k)/[P(P-1)] > 0$ and $\pi_2 - \pi^2 < 0$. With independent, characteristic-specific inclusion, the variance contribution of characteristic j is $\pi_j(1-\pi_j)s_j^2$. □

The corollary separates the magnitude of a stock's total expected-return signal from the concentration of its sources. If the signal is concentrated in one characteristic, forecasts differ

according to whether that characteristic is included. If the same total signal is distributed across many characteristics, a larger share of the signal is observed by each random subset. Correlation among characteristics adds the proxy-loading mechanism in Lemma 4.1. The following two-characteristic example illustrates this effect.

Remark 1 (Correlation and proxy loadings)**.** *Let* $P = 2$*,* $k = 1$*,* $\mathrm{Corr}(x_1, x_2) = \rho$*, and* $\mathrm{Var}(x_1) = \mathrm{Var}(x_2) = 1$*. Let* $\beta_1 = \beta_2 = b$*. An investor who observes* x_1 *forecasts* $m_1 = b(1+\rho)x_1$*, and an investor who observes* x_2 *forecasts* $m_2 = b(1+\rho)x_2$*. Expected squared disagreement is*

$$\mathbb{E}[(m_1 - m_2)^2] \;=\; 2b^2(1+\rho)^2(1-\rho). \tag{7}$$

This expression is maximized at $\rho = 1/3$*. It exceeds the orthogonal benchmark* $2b^2$ *for* $0 < \rho < (\sqrt{5}-1)/2$*. Moderate positive correlation can therefore increase disagreement because each investor places a larger loading on the observed characteristic. As* ρ *approaches one, the two characteristics become substitutes and disagreement converges to zero.*

4.3. Aggregate forecasts and the ambiguity decomposition

Define the equal-weight aggregate forecast as $\bar{m}(x) = B^{-1}\sum_{i=1}^{B} m_i(x)$. Random feature subsets make this forecast a random-subspace ensemble (Ho, 1998; Breiman, 2001). As B increases, $\bar{m}(x)$ converges to $\mathbb{E}_S[m_S(x)]$. Under orthogonal characteristics, this limit is $\pi f(x)$. The aggregate forecast uses information distributed across investors, although no individual investor observes all ensemble components.

The forecast-error comparison follows from the ambiguity decomposition of Krogh and Vedelsby (1994). This is an established statistical identity. We apply it to forecasts generated by heterogeneous feature endowments.

Proposition 2 (Ambiguity decomposition, Krogh and Vedelsby, 1994)**.** *For any* x *and target* $f(x)$*,*

$$(\bar{m}(x) - f(x))^2 \;=\; \underbrace{\frac{1}{B}\sum_{i=1}^{B}(m_i(x) - f(x))^2}_{\textit{mean individual squared error}} \;-\; \underbrace{\frac{1}{B}\sum_{i=1}^{B}(m_i(x) - \bar{m}(x))^2}_{\textit{cross-investor dispersion}}. \tag{8}$$

After taking expectations over x *and the endowment distribution,*

$$\mathrm{MSE}(\bar{m}) \;=\; \mathbb{E}_i[\mathrm{MSE}(m_i)] \;-\; \mathbb{E}_x[D(x)]. \tag{9}$$

Proof. Write $m_i - f = (m_i - \bar{m}) + (\bar{m} - f)$, square the expression, and average over i. The cross term is zero because $B^{-1}\sum_i (m_i - \bar{m}) = 0$. □

The identity applies to the mean individual squared error. It does not provide an identity for long-short returns, Sharpe ratios, rank correlations, or the median investor. We therefore verify (9) in forecast-error space and interpret the portfolio-performance comparison as a separate empirical result.

Corollary 2 (Deviation from the population consensus)**.** *Assume $\Sigma = I_P$ and let the population consensus be $\mathbb{E}_S[m_S(x)] = \pi f(x)$. For investor i,*

$$m_i(x) - \pi f(x) = \sum_j (\mathbf{1}\{j \in S_i\} - \pi)\beta_j x_j, \tag{10}$$

and

$$\mathrm{Cov}(m_i - \pi f, f) \;=\; \sum_{j \in S_i} \beta_j^2 \;-\; \pi \sum_{j=1}^{P} \beta_j^2. \tag{11}$$

The covariance is positive for endowments with above-average total squared loadings and negative for endowments with below-average total squared loadings. Its expectation over the endowment distribution is zero.

The corollary predicts that the average return to a deviation-based strategy is zero across random endowments, while individual strategies can have positive or negative expected returns. The prediction concerns the distribution across investors, not a zero return for every investor. If investors do not observe the true loading vector, they also do not know the sign of (11) before observing out-of-sample returns.

Remark 2 (Potential pricing implication)**.** *Suppose prices place greater weight on optimistic forecasts because pessimistic investors face short-sale constraints. In a Miller-type mechanism (Miller, 1977), greater feature-endowment disagreement can then be associated with overpricing and lower subsequent returns. This implication is conditional on the price-formation mechanism. It is not a consequence of the linear projection model alone.*

4.4. Empirical implications

The model provides four empirical implications. First, equal-sized feature endowments can generate positive forecast dispersion. The model predicts convergence as investors approach the full feature set, but it does not impose monotonicity on standardized dispersion over the empirical range of k. Second, the aggregate forecast's mean-squared-error advantage over the average investor equals cross-investor dispersion. Appendix A.4 verifies this identity in raw forecast-error units. Third, the average return to trading an investor's deviation from the population consensus is zero across random endowments, while individual

deviation returns can have either sign. Section 5.1 examines this prediction. Fourth, when a characteristic's inclusion rate moves from an interior value to one, its direct contribution $\pi_j(1-\pi_j)s_j^2$ converges to zero. The common-core designs implement this comparison across endowment structures. The publication-vintage design studies how dispersion changes as the set of available characteristics expands over time.

The model has two important limitations. It excludes estimation error and therefore does not explain why the best-performing model class varies across endowments. It also does not specify an equilibrium price-formation process. The potential return implications of disagreement are therefore stated conditionally rather than derived from a complete equilibrium model.

5. Empirical Tests and Additional Results

This section examines three implications of the feature-endowment mechanism using the forecasts generated by the crossed design. We study returns to investor deviations from the aggregate forecast, performance-based model switching, and changes in dispersion under fixed and publication-vintage inclusion rates. We then report the relation between aggregate forecast accuracy and cross-investor dispersion.

5.1. Returns to deviations from the aggregate forecast

Corollary 2 predicts a distribution of deviation returns across investors. An endowment with above-average total squared loadings can produce a positive covariance between the investor's deviation and the return signal, whereas a below-average endowment can produce a negative covariance. The average covariance across random endowments is zero. The corresponding empirical prediction is that returns to deviation-based strategies are centered near zero across investors, not that every investor has a zero return.

For each investor-model pair, we sort stocks each month on $m_i - \bar{m}$, where m_i is the investor's forecast and $\bar{m}$ is the equal-weight average of the 20 investor forecasts. We form value-weighted decile long-short portfolios. Table 6 reports the results. Across the 100 investor-model portfolios, the mean annualized return is -1.2%. Eighty-eight percent of the return estimates have absolute t-statistics below 1.96. The statistically significant estimates include both positive and negative values. Figure 4 reports the full distribution.

The model also predicts that individual deviation returns depend on endowment quality. We examine two feasible quality measures: the sum of squared univariate characteristic information coefficients and the joint cross-sectional R^2 of the characteristics in an endow-

ment. Neither measure predicts an investor's own out-of-sample performance across the 20 endowments. The sample therefore does not provide a reliable test of the cross-investor ordering prediction. Appendix A.5 reports these results. The main deviation test relies on the cross-investor mean and on the full-feature comparison described below.

[Fig. 4 about here.]

[Table 6 about here.]

A return estimate near zero is informative only if the portfolio construction can detect additional return information. The full-feature comparison provides this assessment. The full-feature forecast uses characteristics that are not contained in any individual $k = 10$ endowment. Its deviation from the 20-investor aggregate forecast earns annualized returns from 4.8% to 11.8%, with t-statistics as high as 3.7. The aggregate forecast itself earns between 8.6% and 9.2% for four model classes. The same portfolio procedure therefore produces positive returns when the forecast deviation contains incremental predictor information, while the investor deviations remain centered near zero.

5.2. Performance-based model switching

Section 3.2 shows that the best-performing model class varies across feature endowments. We next examine whether an investor can improve performance by selecting models on the basis of trailing returns. At the end of each year, the switching rule ranks the five model classes by the Sharpe ratio of the investor's realized long-short returns over an expanding window. The investor uses the highest-ranked model during the following year. During a five-year initial period, the strategy uses the equal-weight average of model forecasts.

We compare the switching rule with three benchmarks. The first commits to a single model class over the full sample. The second averages the five model forecasts before portfolio formation. The third selects the best fixed model with full-sample hindsight and is not attainable in real time. Table 7 reports the results.

The switching rule underperforms the median fixed model in every endowment design. The median paired difference in annualized Sharpe ratio ranges from -0.03 to -0.13. It also underperforms the equal-weight model average in every design, with differences from -0.07 to -0.14. The loss is large relative to the switching strategy's Sharpe ratio. Figure 5 reports the median Sharpe ratios across strategies and endowment designs.

[Fig. 5 about here.]

[Table 7 about here.]

A prior prediction was that the switching loss would become smaller as the common feature set increased. The results do not support this prediction. The largest loss relative to the pooled forecast occurs in the five-variable core design. In that design, switching underperforms the pooled forecast for 95% of investors and the median fixed benchmark for all investors. The median number of model changes also rises to seven.

These results distinguish cross-investor variation in the identity of the best-performing model from within-investor predictability of future model performance. When model classes have similar expected performance, small trailing-return differences can produce unstable rankings and frequent model changes. The results indicate that these rankings do not provide a reliable basis for annual model selection. Averaging model forecasts or maintaining a fixed model performs better in the designs considered here.

5.3. Inclusion rates and forecast dispersion

The characteristic-specific term in Corollary 1 is proportional to $\pi_j(1-\pi_j)$. A predictor contributes no cross-investor dispersion when it is held by every investor. More generally, changes in inclusion rates can change dispersion. We examine two designs. The first holds inclusion rates fixed over time. The second allows inclusion rates to change with publication dates.

Fixed inclusion rates. Under the uniform $k = 10$ design, endowments remain fixed over the sample. The annual median of cross-investor dispersion shows no significant decline from 1987 through 2021 for any model class. OLS and ENet have positive estimated trends, while the remaining trends are statistically insignificant at conventional levels. Table 8 and Figure 6 report the results.

[Fig. 6 about here.]

[Table 8 about here.]

Publication-vintage inclusion rates. Under publication-vintage endowments, the feasible set expands as characteristics are published. After excluding the common-endowment years, the mean annual median of D_{it} is 0.66 in the first half of the sample and 0.75 in the second half. Forecast dispersion therefore increases as the published predictor set expands. Figure 7 reports the time series.

[Fig. 7 about here.]

The orthogonal-characteristic index provides a limited quantitative fit. Its correlation with observed dispersion is 0.65 in levels and 0.71 in non-overlapping five-year means. The correlation is 0.10 in annual changes and -0.02 after removing linear trends. The relation is therefore concentrated at low frequencies. Observed dispersion rises mainly during the 1990s, when the published set increases from 17 to 37 characteristics and endowment overlap declines. The orthogonal index increases more strongly during the publication wave in the mid-2000s.

The timing difference is consistent with mechanisms omitted from the orthogonal benchmark, including correlated proxy loadings and estimation variation. We therefore interpret the evidence as support for the long-run contrast between fixed and publication-vintage inclusion rates. We do not interpret the index as an explanation of annual changes in dispersion.

[Table 9 about here.]

5.4. Aggregate forecast accuracy and dispersion

We next describe how the accuracy of the aggregate forecast varies with cross-investor dispersion. Within each month, we sort stocks into quintiles of D_{it}. For each quintile, we calculate the Spearman rank correlation between the aggregate forecast and the realized return. Table 10 reports the time-series average of these monthly correlations.

For all five model classes, the correlation is lower in the highest-dispersion quintile than in the lowest-dispersion quintile. The difference ranges from -0.0042 for OLS to -0.0152 for PLS. This pattern is descriptive. The ambiguity decomposition is an identity for squared forecast errors and does not impose a cross-sectional monotonic relation between rank accuracy and dispersion.

[Table 10 about here.]

5.5. Relation to other disagreement measures

Feature-endowment disagreement and existing disagreement measures capture different sources of forecast variation. Machine forecast disagreement holds the predictor set fixed and varies the forecasting specification (Bali et al., 2026). It is high when models map a common set of predictors to returns differently. Analyst forecast dispersion combines forecasts from analysts whose information sets and forecasting procedures are generally not observed (Diether et al., 2002). Our measure holds the model class fixed and varies the set of observed predictors.

Corollary 1 gives feature-endowment disagreement a stock-level interpretation under the orthogonal benchmark. Conditional on the magnitude of a stock's expected-return signal, dispersion is higher when the signal is concentrated in a small number of characteristics. With heterogeneous inclusion rates, dispersion is also higher when the signal depends on characteristics held by fewer investors. A stock can therefore have high model disagreement but low feature-endowment disagreement, or the reverse.

A potential pricing implication arises when pessimistic investors face short-sale constraints. If a negative expected-return signal depends on a narrowly held characteristic, relatively few investors observe that signal and cannot fully express negative positions. The corresponding positive signal can be reflected in prices through purchases by investors who observe it. This conditional mechanism is consistent with evidence that many anomaly returns are concentrated in short portfolios (Stambaugh et al., 2012). A direct test would require a measure of the concentration of each stock's expected-return sources and its interaction with short-sale constraints. Such a measure should also separate concentration from the magnitude of the expected-return signal, as indicated by (6). We leave this analysis to future work.

6. Conclusion

Machine-learning asset-pricing studies commonly compare forecasting models using a comprehensive and common set of return predictors. We examine a different setting in which investors observe equal-sized but differently composed predictor sets. The results show that feature-endowment heterogeneity is an important source of forecast disagreement.

Under uniform random endowments, feature composition explains 46% to 49% of forecast disagreement, compared with 12% to 16% for model class. When investors share a common set of strong predictors, the feature component falls and the model component rises. Under publication-vintage endowments, which restrict annual predictor sets to characteristics published by that date, the feature component remains twice the model component. The best-performing model class also varies across feature endowments.

Averaging forecasts across investors produces a random-subspace ensemble that uses information distributed across feature sets. The ambiguity decomposition shows that the ensemble's mean-squared-error advantage over the average investor equals cross-investor forecast dispersion. In the portfolio tests, the aggregate forecast substantially outperforms the median investor and recovers part of the performance of the full-feature benchmark.

The deviation tests distinguish random feature composition from an account based solely on superior private information. Returns to investor deviations from the aggregate forecast

are centered near zero across investors. In contrast, the full-feature forecast's deviation from the same aggregate forecast earns positive returns. Performance-based annual model switching also underperforms both a fixed-model benchmark and an equal-weight average of model forecasts in every endowment design.

The publication-vintage analysis provides evidence on changes in disagreement over time. Dispersion shows no decline when characteristic inclusion rates are held fixed. It increases when the feasible predictor set expands with publication dates. An orthogonal-characteristic index captures the long-run increase but not annual changes, which indicates an important role for correlated proxy loadings and estimation effects outside the model.

Several limitations remain. The estimated feature-to-model ratio is conditional on the feature and model spaces included in the experiment. The model excludes estimation error and does not explain endowment-specific model rankings. It also treats price formation in reduced form. The data do not directly identify the distribution of predictor access among actual institutions. Future work could measure institutional feature sets, study changes in predictor accessibility, and test whether the concentration of a stock's expected-return sources predicts returns in the presence of short-sale constraints.

The main conclusion is that the performance of a full-feature forecasting system need not describe the performance available to an individual investor. Forecasts and model rankings depend on the predictor set available to the investor, while aggregation across heterogeneous feature sets can produce a more accurate forecast than the one available to any single participant.

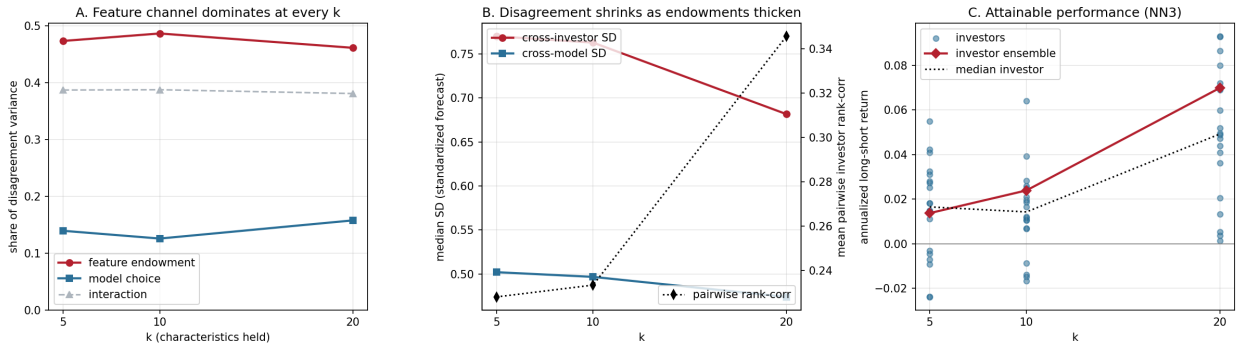


Fig. 1. Forecast disagreement by endowment size

Panel A reports the feature-endowment, model-choice, and interaction shares for uniform endowments with $k \in \{5, 10, 20\}$. Panel B reports the median cross-investor and cross-model standard deviations of standardized forecasts on the left axis and the average pairwise Spearman rank correlation between investors on the right axis. Panel C reports each investor's annualized NN3 long-short return and the return from the equal-weight average of investor forecasts. Forecast disagreement declines as endowment size increases, while the difference between aggregate and median investor performance becomes larger.

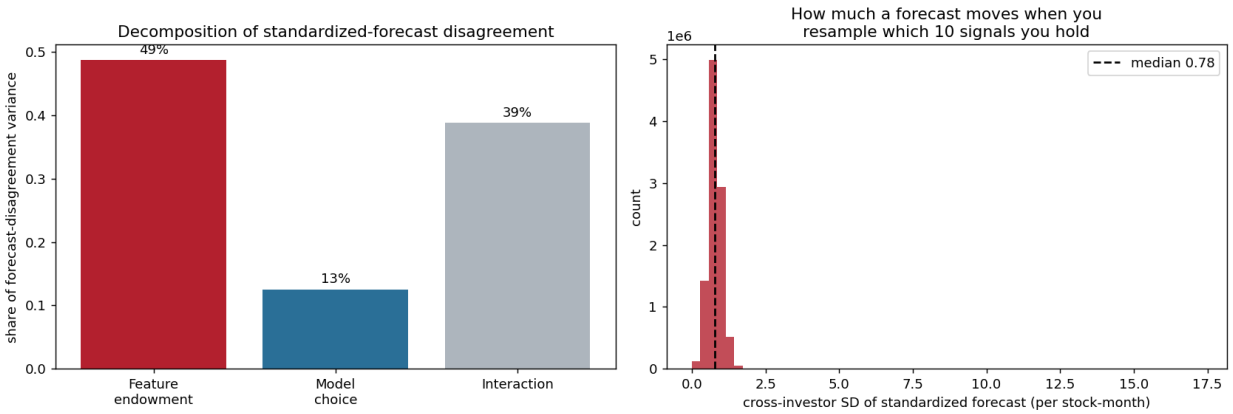


Fig. 2. Decomposition of forecast disagreement for $k = 10$

The left panel reports the shares of forecast variance attributable to feature endowments, model class, and their interaction. The decomposition uses within-month standardized forecasts for complete 20×5 stock-month matrices from 1987 through 2021. The right panel reports the distribution across stock-months of the cross-investor standard deviation, holding model class fixed, and the cross-model standard deviation, holding the investor fixed.

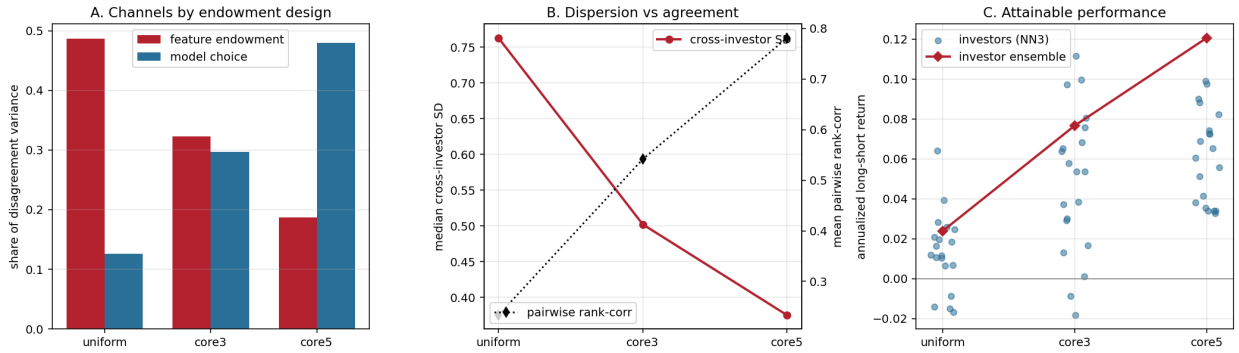


Fig. 3. Forecast disagreement across feature-sharing designs

Panel A reports the feature-endowment and model-choice variance shares for the uniform, three-variable core, and five-variable core designs at $k = 10$. Panel B reports the median cross-investor standard deviation on the left axis and the average pairwise investor rank correlation on the right axis. Panel C reports each investor's annualized NN3 long-short return and the return from the equal-weight average of investor forecasts. Greater feature sharing reduces cross-investor disagreement and increases both individual and aggregate forecast performance.

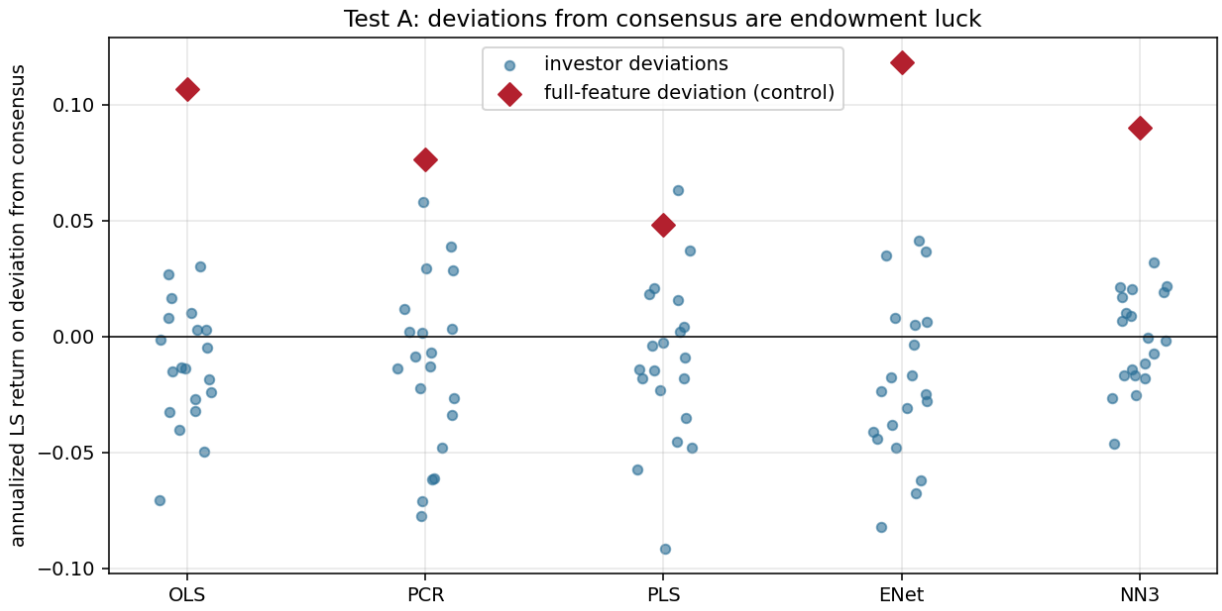


Fig. 4. Returns to deviations from the aggregate forecast

Each blue point is the annualized return of one investor's value-weighted decile long-short portfolio sorted on $m_i - \bar{m}$, reported by model class for the uniform $k = 10$ design. Red diamonds report the same portfolio construction using the full-feature forecast's deviation from the 20-investor aggregate forecast. Investor deviation returns are distributed around zero, whereas the full-feature comparison is positive for each model class.

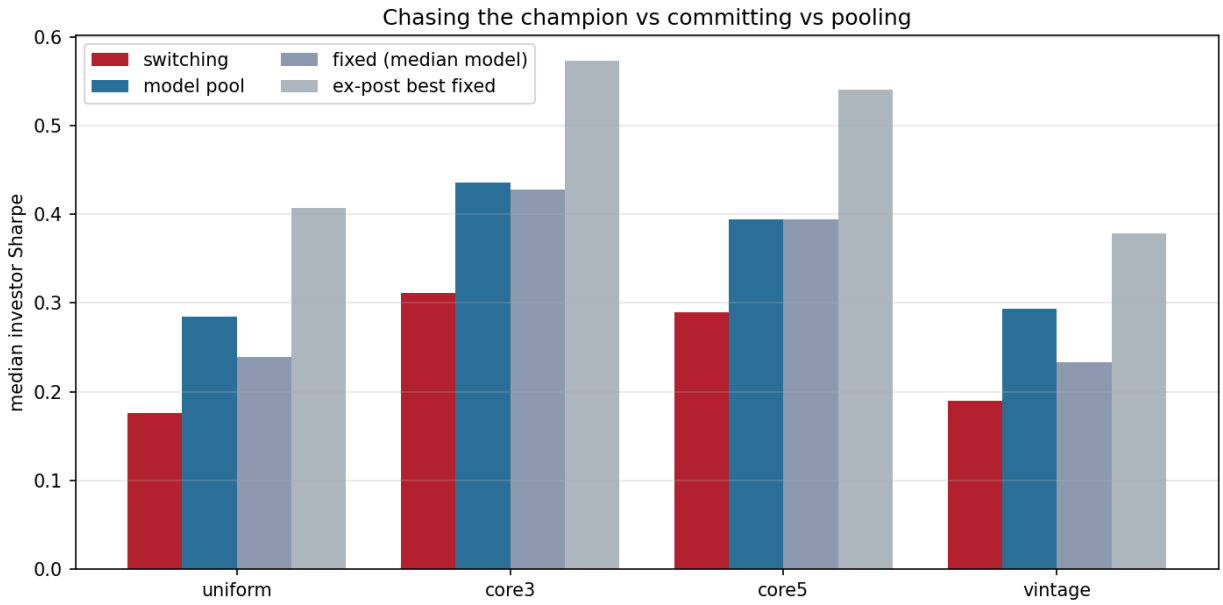


Fig. 5. Performance-based model switching

The figure reports the median investor's annualized Sharpe ratio for four model-selection procedures under each endowment design. Switching selects the model class with the highest trailing expanding-window Sharpe ratio. Pool averages the five model forecasts before portfolio formation. Fixed reports the median across five strategies that each use one model class throughout the sample. Ex-post best reports the best fixed model selected with full-sample hindsight. The switching rule underperforms both the pooled forecast and the median fixed model in every design.

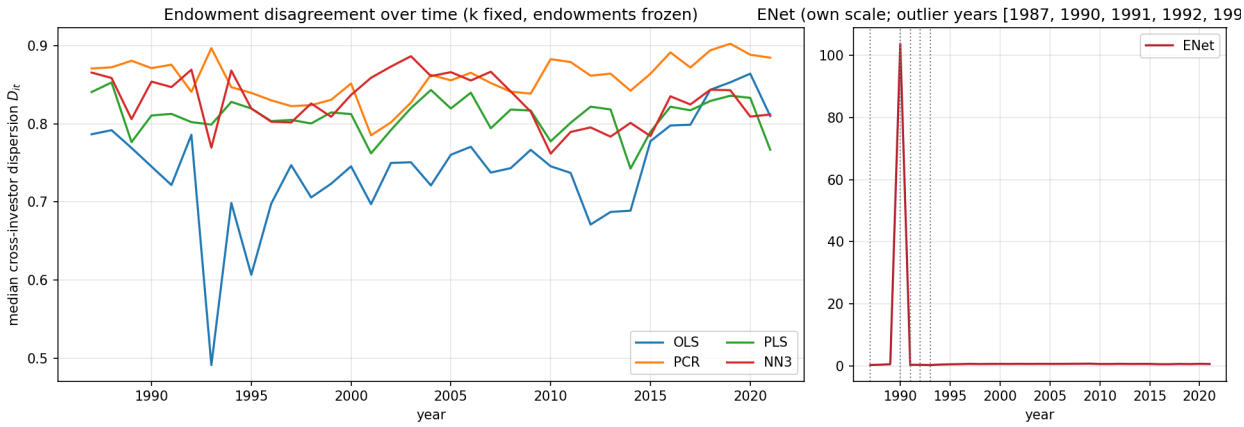


Fig. 6. Forecast dispersion with fixed inclusion rates

The figure reports the annual median of D_{it}, defined as the cross-investor standard deviation of within-month standardized forecasts, for each model class from 1987 through 2021 under the uniform $k = 10$ design. Endowments remain fixed, so characteristic inclusion rates are constant. ENet is reported on a separate scale. Dotted vertical lines identify years in which validation selects near-total shrinkage for many ENet investors. No model class has a significant negative time trend.

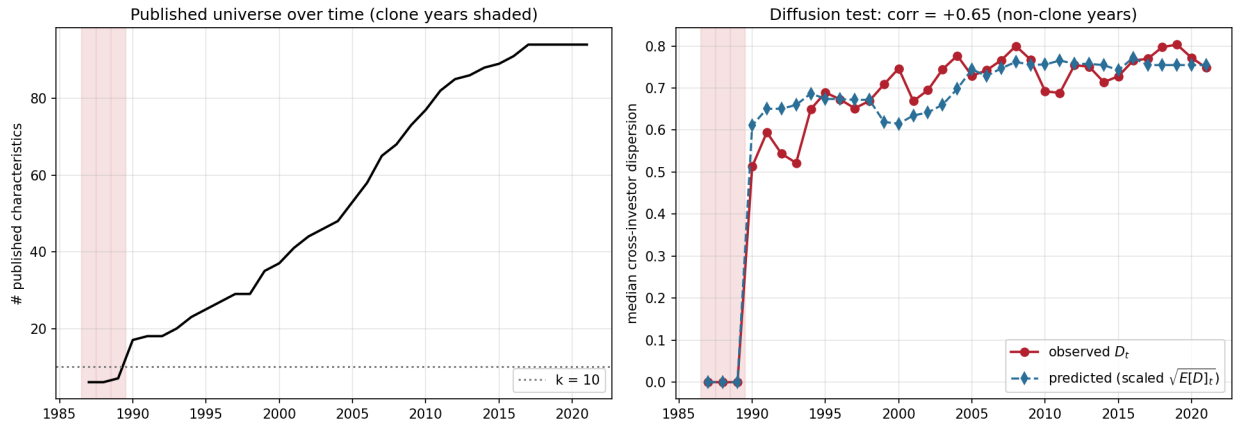


Fig. 7. Forecast dispersion under publication-vintage inclusion rates

The left panel reports the number of characteristics available in each year under the publication-vintage design, using publication year plus a one-year lag. The shaded period from 1987 through 1989 contains fewer than ten published characteristics, so all investors hold the same set. The right panel reports the annual median of observed dispersion D_t and the scaled orthogonal-characteristic index $\widehat{D}_t = \sum_j IC_j^2 \pi_{jt}(1 - \pi_{jt})$. The index uses the realized inclusion rate of each characteristic and its squared univariate information coefficient. Both series increase over the sample, although the index does not match annual changes in observed dispersion.

Table 1: Summary statistics

The table reports summary statistics for the data set used in the analysis. The panel combines 94 firm characteristics with monthly CRSP returns from February 1957 through December 2021. Statistics for continuous variables are based on values before the cross-sectional rank transformation. The estimation sample excludes stock-month observations for which fewer than 20% of the characteristics are available. For the remaining observations, missing characteristics are replaced with the monthly cross-sectional median after the rank transformation to $[-1, 1]$. Market capitalization is measured in millions of dollars. The characteristic with the lowest coverage is *realestate*, which is available for 29.3% of the observations.

Sample period	1957-02 to 2021-12
Stock-month observations (raw panel)	3,304,061
Obs. passing $\geq$ 20%-characteristics filter	2,987,794
Stock-month observations (test years 1987 to 2021)	2,010,176
Unique stocks (PERMNOs)	25,743
Stocks per month: mean / min / max	4,241 / 1,030 / 7,397
Monthly excess return (%): mean / median / SD	0.80 / −0.39 / 18.15
Monthly excess return (%): p5 / p95	−22.44 / 26.03
Lagged market cap ($mm): median / mean	84.0 / 1,861.3
Characteristic coverage (% non-missing): mean / median	74.1 / 74.3
Characteristic coverage: min / max	29.3 (realestate) / 100.0 (indmom)
Characteristics / macro predictors / SIC dummies	94 / 8 / 74

Table 2: Decomposition of forecast disagreement by endowment size

The table separates disagreement in machine-learning return forecasts into feature-endowment, model-choice, and interaction components. Twenty synthetic investors each hold k firm characteristics drawn uniformly from the 94-characteristic set and held fixed over the sample. Each investor estimates five model classes using identical rolling windows from 1987 through 2021. For every stock-month with a complete 20×5 forecast matrix, each investor-model forecast is standardized within the monthly cross-section. The variance of the 100 forecasts is then decomposed into investor main effects (Feature share), model-class main effects (Model share), and their interaction. Feature/Model is the ratio of the two main-effect shares. Cross-inv. SD is the median, across stock-months, of the standard deviation of forecasts across the 20 investors while holding model class fixed. Rank corr. is the average pairwise Spearman correlation between the cross-sectional forecast rankings of two investors while holding model class fixed. Because k is constant within each row, cross-investor dispersion reflects endowment composition rather than the number of characteristics.

k	Feature share	Model share	Interaction	Feature/Model	Cross-inv. SD	Rank corr.
5	47.4%	13.9%	38.7%	3.40	0.770	0.228
10	48.7%	12.6%	38.8%	3.87	0.763	0.233
20	46.1%	15.8%	38.1%	2.93	0.682	0.346

Table 3: Counts of best-performing model classes by endowment size

For each investor, the best-performing model is the class with the highest out-of-sample Sharpe ratio of the monthly value-weighted decile long-short portfolio from 1987 through 2021. Entries report the number of the 20 investors for whom each model class is best under uniform random endowments. The Full column identifies the best-performing model for the single benchmark that uses all 94 characteristics. Under the Sharpe-ratio criterion, the full-feature model is OLS with a Sharpe ratio of 0.82.

Model	$k = 5$	$k = 10$	$k = 20$	Full
OLS	8	5	12	✓
PCR	5	4	1	
PLS	4	2	2	
ENet	1	8	2	
NN3	2	1	3	

Table 4: Forecast disagreement across feature-sharing designs at $k = 10$

The table compares four endowment designs with 20 investors, five model classes, and test years from 1987 through 2021. Uniform assigns all ten characteristics randomly. Core-3 assigns size, book-to-market, and momentum to every investor and adds seven random characteristics. Core-5 also assigns an illiquidity characteristic and a return-volatility characteristic and adds five random characteristics. Vintage limits each annual opportunity set to characteristics published by that date and assigns characteristics according to fixed investor-specific priorities. The variance-share columns are defined in Table 2. Best models reports the number of model classes that attain the highest Sharpe ratio for at least one investor. The bottom panel compares the observed cross-investor dispersion ratio with the orthogonal-characteristic prediction $\sum_j IC_j^2 \pi_j(1-\pi_j)$ evaluated at each design's inclusion rates.

Design	Feature	Model	Interaction	Feat./Mod.	Cross-inv. SD	Rank corr.	Best models
Uniform	48.7%	12.6%	38.8%	3.87	0.763	0.233	5/5
Core-3	32.3%	29.7%	38.0%	1.09	0.502	0.542	4/5
Core-5	18.7%	48.0%	33.3%	0.39	0.375	0.781	2/5
Vintage	45.6%	22.8%	31.6%	2.00	0.685	0.304	4/5

Dispersion ratio relative to uniform: observed and orthogonal-characteristic prediction

	Predicted (IC^2 proxies)	Observed
Core-3	0.97	0.66
Core-5	0.67	0.49

Table 5: Individual, aggregate, and full-feature performance

The table reports performance for value-weighted decile long-short portfolios from 1987 through 2021 under uniform endowments. Panel A reports annualized returns, and Panel B reports annualized Sharpe ratios. Median and best investor rows summarize the 20 NN3 investors. Consensus (NN3) averages the 20 NN3 forecasts before portfolio formation. Consensus (best class) reports the model class whose aggregate forecast has the highest Sharpe ratio at each k, with the class shown in parentheses. Under an annualized-return criterion, the best aggregate class at $k = 10$ is PCR with a return of 9.2%. The full-feature row applies the same estimation and portfolio procedures to all 94 characteristics. ENet has the highest full-feature annualized return, and OLS has the highest full-feature Sharpe ratio.

	$k = 5$	$k = 10$	$k = 20$
Panel A: annualized return			
Median investor (NN3)	0.016	0.014	0.049
Best investor (NN3)	0.055	0.064	0.093
Consensus (NN3)	0.014	0.024	0.070
Consensus (best class)	0.106 (PCR)	0.086 (OLS)	0.126 (OLS)
Full-feature best model		0.158 (ENet)	
Panel B: annualized Sharpe ratio			
Median investor (NN3)	0.106	0.097	0.291
Best investor (NN3)	0.350	0.301	0.641
Consensus (NN3)	0.062	0.121	0.282
Consensus (best class)	0.532 (PCR)	0.507 (OLS)	0.647 (OLS)
Full-feature best model		0.818 (OLS)	

Table 6: Returns to deviation-based portfolios

The table reports returns to monthly value-weighted decile long-short portfolios sorted on forecast deviations. The top panel summarizes, for each model class, the 20 portfolios sorted on $m_i - \bar{m}$ under the uniform $k = 10$ design from 1987 through 2021. Mean annualized return and Mean t are averages across investors. Frac. $|t| > 1.96$ is the fraction of investors with an absolute t-statistic above 1.96. The middle panel applies the same portfolio construction to the full-feature forecast's deviation from the same aggregate forecast. The bottom panel reports the aggregate forecast portfolio for comparison.

	OLS	PCR	PLS	ENet	NN3
Investor deviations $m_i - \bar{m}$ across 20 investors					
Mean ann. return	−0.012	−0.014	−0.011	−0.020	−0.001
Mean t	−0.43	−0.52	−0.39	−0.66	−0.08
Frac. $\lvert t\rvert > 1.96$	0.05	0.25	0.10	0.15	0.05
Full-feature deviation $m_{full} - \bar{m}$					
Ann. return	0.107	0.076	0.048	0.118	0.090
t-statistic	3.67	2.50	1.54	3.53	2.78
Aggregate forecast portfolio					
Ann. return	0.086	0.092	0.089	0.087	0.024
t-statistic	3.00	2.97	2.53	2.62	0.71

Table 7: Performance-based model switching and benchmark strategies

The table compares four model-selection procedures. For each investor, Switch ranks the five model classes at each year-end using the expanding-window Sharpe ratio of realized long-short returns and uses the highest-ranked model in the next year. The equal-weight model forecast is used during a five-year initial period. Pool averages the five model forecasts before portfolio formation. Fixed (median) is the median Sharpe ratio across five strategies that each use one model class throughout the sample. Ex-post best is the best fixed model selected with full-sample hindsight. Entries report the median annualized Sharpe ratio across 20 investors. The next two columns report the percentage of investors for whom switching underperforms Pool and Fixed, respectively. Switches is the median number of model changes per investor.

Design	Switch	Pool	Fixed (median)	Ex-post best	% < pool	% < fixed	Switches
Uniform	0.175	0.285	0.239	0.407	70	75	4.5
Core-3	0.312	0.436	0.428	0.573	70	85	4.0
Core-5	0.290	0.395	0.395	0.541	95	100	7.0
Vintage	0.190	0.293	0.233	0.378	65	65	5.0

Table 8: Time trends in forecast dispersion with fixed inclusion rates

The table reports regressions of the annual median of D_{it} on a linear time trend from 1987 through 2021. D_{it} is the cross-investor standard deviation of within-month standardized forecasts under the uniform $k = 10$ design. Endowments are fixed over the sample. Annual observations more than five median absolute deviations from the median are excluded. The Outliers column reports the number of excluded years. The five excluded ENet years are periods in which validation selects near-total shrinkage for many investors. Start and End report the first and last included annual observations.

Model	Slope per decade	t	Start	End	Outliers
OLS	+0.026	2.48	0.786	0.810	0
PCR	+0.008	1.82	0.871	0.884	0
PLS	−0.002	−0.60	0.840	0.767	0
ENet	+0.035	2.60	0.494	0.722	5
NN3	−0.009	−1.75	0.865	0.812	0

Table 9: Frequency decomposition of the publication-vintage dispersion relation

The table compares observed annual dispersion D_t with the orthogonal-characteristic index $\widehat{D}_t = \sum_j IC_j^2 \pi_{jt}(1 - \pi_{jt})$ under publication-vintage endowments. D_t is the annual median of the cross-investor standard deviation of standardized forecasts. The index uses each characteristic's annual inclusion rate and squared univariate rank information coefficient. The sample contains 1990 through 2021; the common-endowment years from 1987 through 1989 are excluded. Entries report correlations in levels, non-overlapping five-year means, annual changes, and linearly detrended levels.

Levels correlation	+0.648
Five-year block means ($n = 7$)	+0.709
Year-over-year changes	+0.101
Detrended levels	−0.023

Table 10: Aggregate forecast accuracy by dispersion quintile

Within each month, stocks are assigned to quintiles of D_{it}, the cross-investor standard deviation of standardized forecasts under the uniform $k = 10$ design. Within each quintile, we calculate the Spearman rank correlation between the equal-weight aggregate forecast and the realized monthly return. Entries report time-series averages from 1987 through 2021. Q5−Q1 compares the highest- and lowest-dispersion quintiles.

Model	Q1 (low D)	Q2	Q3	Q4	Q5 (high D)	Q5−Q1
OLS	0.0124	0.0144	0.0125	0.0149	0.0082	−0.0042
PCR	0.0291	0.0224	0.0228	0.0205	0.0196	−0.0096
PLS	0.0125	0.0095	0.0097	0.0022	−0.0027	−0.0152
ENet	0.0014	0.0018	0.0026	−0.0035	−0.0083	−0.0098
NN3	0.0131	0.0196	0.0168	0.0128	0.0082	−0.0049

Appendix A. Appendix

A.1. Implementation details

Rolling estimation. For each test year $t \in \{1987, \ldots, 2021\}$, the training sample contains all observations through year $t - 13$. The training window is expanding and contains at least ten years. Hyperparameters are selected by mean squared error over the twelve-year validation period from $t - 12$ through $t - 1$. The selected specification is used to forecast the twelve months of year t. Aggregate predictors are lagged by one month before they are merged with firm-level observations.

Design matrix and standardization. Each characteristic is transformed to its within-month cross-sectional rank and scaled to $[-1, 1]$. Stock-month observations with fewer than 20% of characteristics available are excluded. Remaining missing values are replaced with the monthly cross-sectional median, which is zero after ranking. Investor i uses $[\, x_{S_i} \mid x_{S_i} \otimes z \mid \text{SIC} \,]$, which contains the investor's characteristics, their interactions with eight aggregate predictors, and 74 industry indicators. The design contains $9k + 74$ columns. Before estimation, each design column is standardized using the mean and standard deviation from the training sample. For the disagreement analysis, each investor-model forecast is standardized within the monthly cross-section. This final transformation focuses the dispersion measure on differences in stock rankings rather than forecast scale.

Model classes and hyperparameters. The hyperparameter grids are identical across investors, endowment designs, and estimation windows. Table A.1 reports the grids. OLS is estimated on the complete design without tuning. PCR and PLS select the number of components using the validation sample. The component count is limited by the design width. ENet selects the penalty and mixing parameter jointly. For some endowments, validation selects a specification with nearly complete shrinkage. We retain these forecasts because they are part of the estimated model output. The trend analysis in Section 5.3 identifies the associated annual outliers separately.

Neural network. NN3 contains three hidden layers with 32, 16, and 8 units, rectified linear-unit activations, batch normalization, and a linear output layer. The network is estimated with Adam using a learning rate of 10^{-3}, an ℓ_2 weight-decay parameter of 10^{-5}, and Huber loss with $\delta = 1$. The mini-batch size is 10,000. Estimation continues for at most 100 epochs and stops after five epochs without validation improvement. Each NN3 forecast is the equal-weight average of five independently seeded estimates of the same architecture.

Seed-level variation is therefore averaged within the NN3 forecast and does not enter the model-choice component. Endowment draws, priority orderings, and neural-network seeds are generated from a fixed base seed.

[Table A.1 about here.]

A.2. Baseline feature endowments

Table A.2 lists the ten characteristics assigned to each investor under the uniform $k = 10$ design. Characteristic abbreviations follow Green et al. (2017). The mean pairwise overlap is 1.11 characteristics, with a minimum of zero and a maximum of four. The mean pairwise Jaccard similarity is 0.061. Eighty-one of the 94 characteristics appear in at least one endowment. Thirteen characteristics do not appear, and the most frequently selected characteristic appears in six endowments. The replication files provide the corresponding endowments for the common-core and publication-vintage designs.

[Table A.2 about here.]

A.3. Best-model criterion audit

Table A.3 reports the number of investors for whom each model class has the highest long-short Sharpe ratio under uniform endowments. These are the counts used in Table 3. Under an annualized-return criterion, ENet is the best full-feature model with a return of 15.8%, whereas OLS is the best full-feature model under the Sharpe-ratio criterion with a Sharpe ratio of 0.82. The investor-level counts shift modestly toward ENet when annualized return is used. The conclusion that the best model varies across feature endowments is unchanged.

[Table A.3 about here.]

A.4. Ambiguity decomposition in raw squared-error units

Table A.4 verifies Proposition 2 for the uniform $k = 10$ design. For each model class, mean individual MSE equals aggregate forecast MSE plus the cross-investor dispersion term. The numerical residual is approximately 10^{-17}.

The raw MSE shares require careful interpretation. For PCR, PLS, and ENet, the dispersion term is a small share of total MSE because realized-return variance dominates forecast variance. The aggregate MSE is approximately 0.0386, close to the monthly variance of

realized excess returns. The forecast-variance share provides a scale-free comparison. The dispersion term accounts for 59% to 91% of the pooled variance of individual forecasts. The larger raw shares for OLS and NN3 reflect the unstable magnitude of forecasts from these less regularized specifications in wide designs. Their cross-sectional ranks can remain informative even when forecast magnitudes are poorly calibrated.

[Table A.4 about here.]

A.5. Endowment-quality ordering

Corollary 2 predicts that expected deviation returns vary with endowment quality. We evaluate two empirical proxies. The first is $Q_i = \sum_{j \in S_i} IC_j^2$, where IC_j is the average monthly cross-sectional rank correlation between characteristic j and subsequent returns from 1987 through 2021. The second is Q_i^m, the average monthly cross-sectional R^2 from a joint regression of returns on the ten characteristics in the endowment. The two proxies have a Spearman correlation of 0.61 across the 20 investors.

A useful quality proxy should predict the investor's own out-of-sample performance. Neither proxy satisfies this condition in the present sample. The Spearman correlation between Q_i and the investor's long-short Sharpe ratio ranges from -0.10 to -0.45 across model classes. Its correlation with the investor's forecast information coefficient ranges from -0.28 to 0.33. The correlation between Q_i^m and the investor's Sharpe ratio ranges from -0.20 to 0.01. With only 20 investors and time-varying characteristic premia, the proxies do not provide a validated ordering of out-of-sample endowment quality.

Table A.5 reports regressions of annualized deviation returns on the two proxies. The pooled coefficient is -2.04 with a t-statistic of -3.30 for Q_i and -0.98 with a t-statistic of -1.87 for Q_i^m. Because the proxies do not predict investors' own out-of-sample performance, these coefficients are not interpreted as tests of the theoretical ordering. The main deviation evidence in Section 5.1 relies on the mean across investors and the full-feature comparison.

[Table A.5 about here.]

Table A.1: Hyperparameter grids

The table reports the estimation and validation choices for each model class. The grids are identical across investors, endowment designs, and rolling windows. Hyperparameters are selected by validation-sample mean squared error.

OLS	no tuning (least squares on the full design)
PCR	components $\in \{1, 3, 5, 10, 20, 30\}$ (capped at design width)
PLS	components $\in \{1, 3, 5, 10, 20, 30\}$ (capped below design width)
ENet	ℓ_1 ratio $\in \{0.05, 0.5, 0.95\}$; penalty $\alpha \in \{10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1\}$
NN3	fixed architecture 32−16−8; Adam, lr 10^{-3}, weight decay 10^{-5}; Huber loss ($\delta = 1$); batch 10,000; ≤ 100 epochs, patience 5; five-seed average

Table A.2: Baseline feature endowments under uniform sampling

The table lists the ten characteristics assigned to each of the 20 investors under the uniform $k = 10$ design. Characteristics are drawn without replacement from the 94-characteristic set and remain fixed over the sample.

Investor	Characteristics
inv_00	age, cashdebt, gma, grltnoa, idiovol, lev, pchcapx_ia, pctacc, rsup, tang
inv_01	cfp, mve_ia, operprof, pctacc, rd_mve, retvol, roeq, saleinv, sgr, sin
inv_02	chcsho, chmom, chpmia, currat, grltnoa, lgr, pchsale_pchxsga, quick, rd_sale, sin
inv_03	baspread, chmom, grcapx, hire, pchsale_pchinvt, rd, roavol, std_turn, tang, turn
inv_04	gma, lev, mvel1, pchquick, pchsale_pchinvt, realestate, roaq, roeq, sp, std_turn
inv_05	agr, cashdebt, chempia, cinvest, egr, mom36m, ms, mve_ia, rd, std_dolvol
inv_06	age, cash, chpmia, divo, ep, indmom, mve_ia, roaq, roeq, tang
inv_07	dolvol, indmom, operprof, pchcapx_ia, pchsale_pchxsga, rd_sale, realestate, rsup, salerec, stdacc
inv_08	aeavol, bm, bm_ia, cashdebt, cinvest, depr, maxret, roaq, sgr, stdacc
inv_09	bm, bm_ia, cashdebt, cashpr, chtx, ep, hire, rd_mve, secured, securedind
inv_10	chatoia, dolvol, ep, pricedelay, ps, retvol, salecash, stdacc, tang, zerotrade
inv_11	age, cashdebt, chempia, depr, divi, lgr, operprof, quick, sgr, zerotrade
inv_12	aeavol, cfp, divi, mom6m, mvel1, orgcap, pchquick, pchsaleinv, roic, zerotrade
inv_13	age, divi, hire, pchgm_pchsale, pchsale_pchxsga, rd_mve, roavol, tang, tb, zerotrade
inv_14	acc, cashdebt, divo, dolvol, grcapx, pchquick, pchsale_pchxsga, salecash, turn, zerotrade
inv_15	bm, chcsho, chmom, grltnoa, pchdepr, rd, rsup, sgr, sin, turn
inv_16	chcsho, chempia, divo, hire, indmom, lgr, rd_mve, salerec, securedind, sgr
inv_17	bm_ia, chpmia, convind, divi, maxret, nincr, pchdepr, roaq, roic, stdacc
inv_18	chinv, chtx, convind, ear, gma, mve_ia, nincr, ps, std_turn, stdacc
inv_19	bm_ia, chinv, depr, divi, divo, ear, pchgm_pchsale, roeq, sin, tang

Table A.3: Best-model counts under the Sharpe-ratio criterion

Entries report the number of the 20 investors for whom each model class produces the highest out-of-sample Sharpe ratio under uniform endowments.

Model	$k = 5$	$k = 10$	$k = 20$
OLS	8	5	12
PCR	5	4	1
PLS	4	2	2
ENet	1	8	2
NN3	2	1	3

Table A.4: Ambiguity decomposition in forecast-error space

The table verifies that mean individual MSE equals aggregate forecast MSE plus cross-investor dispersion for the uniform $k = 10$ design. Statistics are pooled over complete stock-month observations from 1987 through 2021. The MSE share divides dispersion by mean individual MSE. The forecast-variance share divides dispersion by the pooled variance of individual forecasts.

Model	Mean indiv. MSE	Aggregate MSE	Dispersion	Share of MSE	Share of forecast var.
OLS	0.0579	0.0406	0.0173	29.9%	89.6%
PCR	0.0386	0.0386	0.0000	0.03%	63.1%
PLS	0.0387	0.0386	0.0001	0.18%	74.2%
ENet	0.0386	0.0386	0.0000	0.03%	59.2%
NN3	0.1880	0.0517	0.1363	72.5%	91.1%

Table A.5: Deviation returns and endowment-quality proxies

The table reports regressions of annualized deviation-portfolio returns on endowment-quality proxies across the 20 investors in the uniform $k = 10$ design. Panel A uses the sum of squared univariate information coefficients. Panel B uses the average joint cross-sectional R^2 of the ten characteristics in each endowment. The pooled specifications include model fixed effects and contain 100 investor-model observations. The proxies do not predict investors' own out-of-sample performance, so the coefficients are reported as descriptive results.

	Panel A: univariate $Q_i = \sum IC_j^2$			Panel B: joint-R^2 Q_i^m		
Model	Slope	t	Spearman	Slope	t	Spearman
OLS	0.22	0.18	0.02	0.97	0.98	0.20
PCR	−2.70	−1.58	−0.38	−1.93	−1.36	−0.36
PLS	−2.54	−1.61	−0.44	−2.07	−1.60	−0.34
ENet	−4.14	−3.04	−0.60	−1.39	−1.04	−0.20
NN3	−1.02	−1.05	−0.21	−0.48	−0.59	−0.15
Pooled (model FE)	−2.04	−3.30	−0.35	−0.98	−1.87	−0.19

References

Bali, T. G., Kelly, B. T., Mörke, M., Rahman, J., 2026. Machine forecast disagreement. *The Review of Financial Studies* p. hhag042.

Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.

Chen, A. Y., Velikov, M., 2023. Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis* 58, 968–1004.

Chordia, T., Goyal, A., Saretto, A., 2020. Anomalies and false rejections. *The Review of Financial Studies* 33, 2134–2179.

DeMiguel, V., Martín-Utrera, A., Nogales, F. J., Uppal, R., 2020. A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies* 33, 2180–2222.

Diether, K. B., Malloy, C. J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. *The Journal of Finance* 57, 2113–2141.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. *The Review of Financial Studies* 33, 2326–2377.

Green, J., Hand, J. R. M., Zhang, X. F., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. *The Review of Financial Studies* 30, 4389–4436.

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

Harvey, C. R., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. *The Review of Financial Studies* 29, 5–68.

Ho, T. K., 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 832–844.

Hong, H., Stein, J. C., 2007. Disagreement and the stock market. *Journal of Economic Perspectives* 21, 109–128.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2023. Is there a replication crisis in finance? *The Journal of Finance* 78, 2465–2518.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. *Journal of Financial Economics* 135, 271–292.

Krogh, A., Vedelsby, J., 1994. Neural network ensembles, cross validation, and active learning. In: *Advances in Neural Information Processing Systems 7*, MIT Press, pp. 231–238.

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

McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? *The Journal of Finance* 71, 5–32.

Miller, E. M., 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32, 1151–1168.

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104, 288–302.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21, 1455–1508.

Yan, X. S., Zheng, L., 2017. Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies* 30, 1382–1423.