Dissecting Machine Learning Return Predictability: A Classification Approach[†]

Yang Bai

November 1, 2023

Abstract

This paper examines machine learning predictability using classification models and investigates the relation between predictability and stock returns. The classifiers show significant and robust out-of-sample precision in placing stocks into the correct deciles, outperforming their counterpart machine learning regressions. The corresponding long-short portfolios deliver significant economic gains. The classifiers invest more resources in return state transitions with lower information shortage—and excel in predicting return deciles in the center and edges of the transition probability matrix. The classifiers extract information from different firm characteristics. Following Easley and O'Hara (2004), I show that prediction success is negatively related to the future returns at the stock level, controlling for information shortage. Information shortage also reduces the probability of prediction success. Portfolios conditional on information shortage show enhanced performance. A mimicking portfolio based on the shock of prediction precision generates significant α benchmarking against popular factor models.

Keywords: Artificial neural network, classification, gradient boosting tree, information entropy, information uncertainty, machine learning, portfolio allocation, out-of-sample prediction, random forest.

JEL Classification: C14, C38, C55, G11, G14

[†] Yang Bai (yangbai@mail.missouri.edu) is from Robert J. Trulaske College of Business, University of Missouri.

1. Introduction

What are the consequences of return predictability? Easley and O'Hara (2004) show that the precision in the signals of return forecasting should be negatively related to the future returns. However, the empirical literature provides a paucity of evidence that supports this theoretical prediction for the stock market. This is potentially because the empirical literature has not been able to produce reliable out-of-sample (OOS) performance in backtesting until the recent development of the machine learning literature. For example, Chen et al. (2023) and Gu et al. (2020) show that the machine learning regressions demonstrate promising out-of-sample performance in stock return predictions. In this paper, I extend the financial machine learning literature to investigate the implications of return predictability.

In particular, I apply classifiers to allocate individual stocks into decile portfolios.[1] My use of classifiers complete the methodological picture in the application of supervised learning in finance following the use of machine learning regressions (Gu et al. 2020). More importantly, since a single prediction of a return decile has an explicit modeling outcome, i.e., success or failure, classification methods provide the convenience in measuring stock-level prediction success and market-level prediction precision. Applying the predicted probabilities associated with decile portfolios for each stock, I propose the use of information entropy as a measure of relative information shortage in the cross-section, which can be seen as a general form of information asymmetry and information uncertainty (Merton 1987; Shannon 1948; Zhang 2006).

In addition, although the return predictability in the financial machine learning literature has been quite successful, it remains economically unclear how the machines capture the forward-looking information. By using the stock-level prediction success and the market-level prediction precision, I provide novel economic insights on sources of the machines' power.

[1] More precisely, I model $Prob(stock\ i\ in\ return\ decile\ j\ at\ t+1|characteristics_{i,t}) = F(characeristics_{i,t})$ instead of $E\left[R^e_{i,t+1}\right] = G(characteristics_{i,t})$, where F and G are machine learning classifiers and machine learning regressions, respectively (see Gu et al. 2020).

To analyze the sources and the consequences of the cross-sectional return predictability, I first establish the validity of classifiers in generating out-of-sample performance. I focus on tree models and neural networks because of their superior performance in predicting returns as reported in the literature (Bali et al. 2023; Chen et al. 2022; Gu et al. 2020; Li and Rossi 2020). The results from these models can provide a solid foundation for the analysis on the consequences of return predictability.

Economically, the long-short portfolios with monthly rebalance based on my predictions can achieve Fama-French 5-factor (FF5F) adjusted alphas as high as 2.1% and Sharpe ratios as high as 2.73. The statistical and economic performance of the models emphasizes the success of the classifiers in capturing the pricing kernel.[2] Meanwhile, the prediction precision is robust. By aggregating the precision to the market level, I show that the time-varying prediction precision is always above the precision from the prediction conditional only on the distribution information (10% cross-sectional prediction precision).

Since I synthesize the information based on a comprehensive set of firm characteristics, I argue that the prediction is representative of the publicly available information (Green et al., 2017). Additionally, the excellent out-of-sample performance also implies that the machine learning classifiers capture the information of the pricing kernel, and thus the classifiers are representative for the market's best anticipation of the future returns.[3] In head-to-head comparisons involving a grid search with a wide range of candidate hyperparameters, I find that the classifiers are better than their corresponding machine learning regressions in allocating individual stocks into correct future deciles. On average, the classifiers achieve a precision of above 15%, while the machine learning regressions achieve a precision of 12%.[4] When stacking the predictions from machine

[2] Holding everything equal, the classifiers deliver performance that is comparable (if not superior to) machine learning regressions both statistically and economically.

[3] Before the application of machine learning methods in finance, no evidence has been documented to support out-of-sample performance in cross-sectional return predictions at the individual stock level. Martin and Nagel (2020) also derive a model to show that individual investors may not be able to fully process the public information, implying that machine learning predictions can be a superior choice for prediction studies.

[4] Both my classifiers and their benchmark machine learning regressions deliver statistically meaningful prediction precision compared to the precision of a naïve classifier with a precision of 10%. The naïve classifier is the raw benchmark in machine learning that predictively assigns each observation to the majority class and provides a benchmark of 10% in my case. Models of higher precision than that of the naïve classifier are considered as producing statistically significant predictability. This is similar to the comparison between the numerical predictions and the historical means in Goyal and Welch (2007).

learning classifications and machine learning regressions together, the economic performance in both equal-weighted and value-weighted long-short portfolios increases substantially to produce Sharpe ratios of 3.4 and 1.5 respectively, which emphasizes the meaningful addition in the information captured by the classification models benchmarking to the machine learning regressions.

To understand the economic sources of the machines' predictability, I analyze the success of machine learning predictions under the classification framework. The transition probability matrix shows that the transition of the return states in deciles is uneven. Compared with a pure random distribution, transitions from extreme deciles to extreme deciles and from middle deciles to middle deciles are more certain. Such transitions are associated with probabilities deviating from the random transition probability.

For example, the transition from the lowest decile to the highest decile has a probability of 1.8%, deviating from the randomly distributed probability of 1%. The classifiers take advantage of such unevenness in the transition probabilities and achieve the highest performance in the middle deciles and the extreme deciles. When I measure the information shortage using Shannon's information entropy calculated based on the predicted probabilities, I find that the information shortage replicates the uneven structure of the return transition matrix.[5]

In general, I confirm that the machines benefit heavily from the least uncertain predictions.[6] The machines achieve higher precisions in the predictions for the center and the tails of the return distribution. This unevenness is more pronounced in the predictions of the lowest return decile, where the aggregated predictions based on the individual classifiers deliver a precision of 49%.

Next, I investigate what characteristics influence the prediction precision. The variable importance shows that different models get information from different firm characteristics. For example, the neural networks focus mainly on industry information,

[5] Because it measures the expected minimum binary questions that need to be answered to completely resolve the prediction uncertainty, the unit of the information shortage is "bit", the standard unit of information. Firms with greater information shortage thus require more information to resolve return prediction uncertainty.
[6] Such heterogeneity in return predictability can signal different levels of market efficiency, i.e., market efficiency level is a function of firm characteristics.

while the tree models, especially Gradient Boosting Trees (GBT), rely on volatilities and bid-ask spreads. However, variable importance does not include a conventional statistical inference on the significance of the contribution. More importantly, variable importance does not reflect the direction of the contribution. Therefore, I analyze the machines' predictability through predictive Fama-MacBeth regressions with prediction success as the response variable (Green et al., 2017). I specify the prediction success of a stock at a certain time with an indicator variable of a value 1 if the decile portfolio allocation is correct.

A number of variables, including change in momentum and return on assets, contribute to the success of the aggregated predictions. When categorizing firm characteristics into Hou et al.'s six types of firm characteristics, I find that all types of firm characteristics contribute to the out-of-sample prediction success at the stock level (Hou et al., 2018).[7] Additionally, I measure information shortage as the machines' assessment of the expected minimum number of binary questions that need to be answered before perfect predictions (Shannon 1948). I show that the information shortage is negatively related to prediction success in the classification setup. A one standard deviation increase in information shortage is related to a 3%-5% reduction in the probability of prediction success, i.e., the chance that the prediction for the return in the next period is correct. An immediate implication is that portfolio performance will increase conditional on low information shortage. As an example, limiting the portfolio construction to stocks in the lowest decile of information shortage, the value-weighted long-short portfolio achieves a Sharpe ratio of 1.6.

When I repeat the regression with information shortage as the response variable, I find that firm characteristics like firm age and change in momentum reduce the information shortage. Consistent with the prior literature, firm characteristics such as analyst forecast dispersion increase the information shortage (Zhang, 2006). However, aside from the traditional proxies of information uncertainty related to return prediction, many other variables also contribute to the information shortage. For example, percentage change in

[7] See Hou, Xue, and Zhang (2018). The six categories of firm characteristics are: Momentum, value vs. growth, investments, profitability, trading frictions, and intangibles.

number of employees, market-cap-scaled operating cashflow, and 6-month momentum positively contribute to the information shortage.

The classifiers perform well in the out-of-sample tests and thus are representative of the pricing kernel, providing a good empirical foundation to study the consequences of return predictability. Easley and O'Hara (2004) derive a rational expectation model with explicit implication on the relation between the prediction precision and the stock returns. They predict that the low prediction precision leads to high stock returns. I document empirical evidence that confirms their theoretical prediction.

Controlling for the information shortage measure in Fama-MacBeth regression, I find that both the past period prediction success and the information shortage measure formed based on predicted probabilities are negatively related to future returns (See Jiang et al. 2006 and Zhang 2006). A correct past month prediction is related to a 0.2% reduction in the monthly return; moreover, a one standard deviation increase in information shortage is related to a 0.41% reduction in the stock return.

I continue by investigating the relation between prediction success and pricing errors to determine the economic sources of predictability following the negative relation between returns and prediction success. I proxy for pricing errors with Fama-French factor models' alphas. The results reveal profound insights of pricing errors as a source of predictability: The fluctuation of pricing errors is the primary contribution from pricing error to the prediction success. In particular, both the past pricing error and the change in pricing error contribute negatively to the prediction success, while only the change in the absolute value of the pricing error contributes positively to the prediction success. This highlights the impersistence of pricing errors are the source of predictability.

With the profound relation between prediction success at the stock level and returns, a natural question is whether a portfolio on prediction precision (if tradable) can generate performance that the benchmark models cannot explain. To investigate this question, I map the innovation of market-level prediction precision in percent from an autoregressive model with 1 lag (AR1) to the return space spanned by 170 basis portfolios that are used to form Fama-French factors (Adrian et al., 2014). I then analyze the performance of a mimicking portfolio obtained from the linear mapping process on the common factor models. My results suggest that the mimicking portfolio can generate an annual alpha of -2.4% benchmarking to the common factor models, including the Fama-French 5 factor model, the Fama-French 6 factor model, the q4 factor model, and the q5 factor model.

The machine learning prediction precision at the market level is consistently meaningful statistically across my out-of-sample data. Meanwhile, the precision promptly reacts to events that seemingly increase macroeconomic uncertainty. Although Gu et al. (2022) mention that the interactions between firm characteristics and macroeconomic variables from Goyal and Welch (2008) seem contributing to the variable importance, what types of macroeconomic uncertainty specifically is associated with prediction precision is unclear. In the final portion of the paper, I investigate the relation between the machine learning prediction precision and the economic uncertainty.

Specifically, I include the war factor from Hirshleifer et al. (2023), geopolitical risk from Caldara et al. (2022), economic policy uncertainty from Baker et al. (2016), and Jurado et al.'s (2015) macroeconomic uncertainty, financial uncertainty, and real activity uncertainty. My results suggest that only macroeconomic uncertainty is robustly related to the prediction precision at the market level, highlighting the specific uncertainty concerns from the stock market.[8]

This paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the empirical modeling and introduces the testing variables. Section 4 reports the modeling performance and the portfolio performance. In Section 5, I analyze the economic sources of prediction precision. I discuss the relation between predictability and returns in Section 6. Section 7 concludes the paper.

2. Related Literature

This paper contributes to the finance literature in two aspects. First, this paper contributes to the financial machine learning literature with an alternative perspective of return prediction and portfolio allocation. I frame the cross-sectional return prediction problem as a machine learning classification problem and contribute to the machine

[8] The financial uncertainty from Jurado et al. (2015) is also marginally related to the market-level prediction precision.

learning application in finance by providing an alternative perspective of return predictions. Specifically, the prior literature in asset pricing focuses on the application of machine learning regressions. For example, Gu et al. (2020) are the pioneers in this field, and they survey a range of popular algorithms in a regression setting to make stock return predictions (See also Chen et al. 2023). Bali et al. (2023) and Bianchi et al. (2021) apply the same research setting to stock options and bond market, respectively. Li and Rossi (2020) adopt the setting to mutual fund selections. Aubry et al. (2023) apply neural network to art auction prices.

These studies adopt conventional modeling of returns that involves the minimization of mean squared errors. These errors are the differences between the predicted returns and the realized returns. For example, Fama and French (1992, 2015) and Hou et al. (2014, 2019) study the returns from portfolios and fit regressions to minimize the mean squared errors between the regression fitted values and the realized returns.

On the opposite side, this paper frames the problem of cross-sectional return prediction as a classification problem. Cross-sectional deciles can be seen as representations of the unobservable future state returns for individual stocks. The expected return is the dot product between the future state returns and their associated state probabilities, and the prediction target of classifiers is the state probabilities. Specifically, this paper quantifies the returns as deciles and directly model the portfolio allocation process, i.e., the probability of a stock belonging to a certain decile portfolio. Due to the dearth in the application of multi-class classification to the return predictions, my results complete the methodological picture empirically and document the impressive performance of classification in out-of-sample predictions.

Second, I study the sources and the consequences of cross-sectional predictability in future returns. Thus, I contribute to the empirical asset pricing literature (Easley and O'Hara 2004). Based on the predictions by the classifiers, I measure the stock-level prediction success and the market-level prediction precision, through which I provide novel intuitions on the return predictability. I provide the first evidence in the literature which confirms that the return predictability is negatively related to future returns at the stock level. This result is robust when controlling for information shortage that accounts for information asymmetry and information uncertainty (Jiang et al. 2006; Zhang et al. 2006). I also document a list of variables that contribute to the prediction success. The machine's predictability heavily relies on the fluctuation of pricing errors instead of the past pricing errors and is mainly influenced by the macroeconomic uncertainty. A mimicking portfolio of the predictability at the market level generates annual benchmark-adjusted return of 2.4%.

3. Empirical Methods

I provide a general description of my methods in this section. First, I explain the basics of my modeling process. I briefly introduce the machine learning classification methods and the training process. I also discuss the metrics I adopt in evaluating modeling performance. Finally, I detail my data construction at the end of this section.

3.1 Introduction to Return Prediction as A Classification Problem

I frame the cross-sectional return prediction as a multi-class classification problem. Given a set of candidate outcomes, the classification process selects the most promising outcome as a prediction. This is the foundation of my information uncertainty measures. Following the convention of the asset pricing literature, I group individual stock returns into ten deciles per month and try to allocate each stock to its correct return decile.[9]

I refer to a strategy that performs the classification prediction as a classifier. A classifier takes the input variables and calibrates the parameters through the modeling architecture. The modeling architecture maps the input variables to the probability space such that a loss function is minimized. Figure 1 illustrates the modeling process. Specifically, when I frame the problem of cross-sectional return prediction as a classification problem, my optimization objective is to create a model such that the predicted probabilities distribute exactly like the observed probabilities. I follow the common practice in multi-class classification problems and adopt cross-entropy loss function to achieve this matching process. The cross-entropy function measures the

[9] The classic asset pricing studies and the recent machine learning prediction studies often focus on the decile portfolios. For example, Fama and French (1992) sort stocks into deciles based β loading, while Gu et al. (2020) sort stocks into deciles based on predicted returns.

difference between two probability distributions. For the real return distribution P relative to the predicted distribution Q over a set of return deciles D, a classifier will minimize the loss function below.

$$L = -E_p\,[\log_2 q] = -\sum_{d_{it} \in D} P(d_{it}) \log_2 Q(d_{it}), \tag{1}$$

where $P(d_{it})$ is proxied empirically by the true outcome, i.e., return decile of a stock i at time t, with a value of 1 or 0.

Then, the classifier selects the return decile with the highest predicted probability as its final prediction. In Table 2, I include the benchmark machine learning regression results, and I adopt the standard mean squared error as the loss function for these benchmark models (See Gu et al. 2020).[10]

[Include Figure 1 Here]

My main models include the standard multilayer perceptron, i.e., Artificial Neuron Network (ANN), the random forest (RF), and the gradient boosting trees (GBT). My choice of models depends on two considerations. First, I want to focus on powerful models only. Second, I do not attempt to search for models with marginal improvement in predictive power benchmark to the existing works in the literature. Instead, I want my models to be replicable and intuitive. Thus, I focus on standard models with strong predictive power.

3.2 Artificial Neural Network

Figure 2 illustrates an example of the ANN architecture in this paper. In a fully connected architecture, the standard ANN processes input through backpropagation, which is a calibration process that adjusts parameters to minimize the loss function. A fully connected feedforward neural network includes an input layer, several hidden layers, and an output layer.

[Include Figure 2 Here]

[10] Mean squared error loss is the loss function in ordinary least square regressions. It takes the following form: $L = \frac{1}{N}\sum_{it}(y_{it} - \hat{y}_{it})^2$ for stocks i and time t.

In my ANN classifiers, the input layers include the firm characteristics. Then, the firm characteristics go through the fully connected hidden layers. Each neuron in a hidden layer takes the input from the prior layer. This input is fed to a linear function wrapped in a nonlinear function, which is again included in another linear function (See Hastie et al. 2009). The results are then fed to another hidden layer. The nonlinear function is referred to as activation function. In the end, the last hidden layer feeds its output to the output layer in my ANN classifiers, and the output layer includes ten neurons representing the return deciles. Each neuron in the output layer employs a SoftMax function that translates the output from last hidden layer into probabilities.[11] In the ANN regressions, the output layer includes only a regression neuron.

More specifically, consider my ANNs with multiple hidden layers. The first hidden layer includes N^1 neurons, and the neuron i^1 includes a weight vector $w^1_{m^1 j} \in W^1_{m^1}$ for the corresponding firm characteristics $x_j \in X_J$ and a bias $b^1_{m^1}$.

$$h^1_{m^1} = \sigma\left(\sum_j w^1_{m^1 j} x_j + b^1_{m^1}\right), \tag{2}$$

where σ is an activation function.[12] In this paper, I have two ANN models, including a model with rectifier activation function $\sigma(a) = \max(0, a)$ and the other model with tanh activation function $\sigma(a) = \frac{\exp(a)-\exp(-a)}{\exp(a)+\exp(-a)}$. Then, $h^1_1, \cdots, h^1_{m^1}, \cdots, h^1_{N^1}$ become the input of the second hidden layer. In general, the neuron m^l in the hidden layer $l \in [1, L]$ transforms all N^{l-1} output from hidden layer $l-1$, i.e., $h^{l-1}_1, \cdots, h^{l-1}_{m^{l-1}}, \cdots, h^{l-1}_{N^{l-1}}$ with a weight vector $w^l_{m^l m^{l-1}} \in W^l_{m^l}$ and a bias b_{m^l} as the following.

$$h^l_{m^l} = \sigma\left(\sum_{m^{l-1}} w^l_{m^l m^{l-1}} h^{l-1}_{m^{l-1}} + b^l_{m^l}\right). \tag{3}$$

[11] The SoftMax function is a popular scaling function in regressions to model categorical response variable. For example, multinomial regression also employs SoftMax function.
[12] I do not change the activation function from layer to layer in this paper.

The output layer takes the vector input H_L from the last hidden layer and makes the final linear transformation $f_d = \sum_{m^L} w_{dm^L} h_{m^L}^L + b_d$ for output neuron of class $d \in D$, and the calculation finishes with the SoftMax function as below. Then, the set of predicted probabilities is compared to the realized outcomes in the cross-entropy loss function.[13]

$$Q(d) = \frac{\exp(f_d)}{\sum_D \exp(f_u)}. \tag{4}$$

3.3 Random Forest and Gradient Boosting

I include two powerful tree models, i.e., random forest and gradient boosting tree. Both models are developed from the simple decision tree. Based on the values of the input variables, a classic binary decision tree finds the best splitting strategies to divide a sample into pieces sequentially such that a loss function is minimized. For each subsample that comes out from the splitting process, the tree will assign a class to it for the classification task and a numeric value to it for the regression task. In other words, the decision tree dissects the response space into subspaces conditional on the input variables and gives each of the subspaces a value.

A random forest model builds on top of the decision trees with bootstrap aggregating (bagging). In each bootstrapping sample, the algorithm grows a tree by recursively sample from the input variables for splitting and picks the best split-point until the prespecified node size is reached. Then, the final prediction is made by aggregating the predictions from the trees in the random forest. Usually, an equal-weighted vote is taken as the prediction for the classification problems, while the average value is taken as the prediction for the regression problems.

Consider a decision tree $T(z; \Theta) = \sum_{j \in [1,J]} \gamma_j I(z \in R_j)$, where z is an observation, γ_j is the assigned value in the region R_j, J is the number of regions. Θ denotes the collection of parameters γ_j and R_j for all the regions, and it also includes J.

[13] My benchmark machine learning regressions have a different single neuron output layer that takes a hidden layer's transformation, and the result is compared to the realized return using the mean squared error loss function.

In my multi-class classification task, a boosted tree will make prediction on the probability of each of the outcome classes $d \in D$ and repetitively update the prediction until the loss function is minimized. Specifically, the algorithm initiates the prediction for class d as $f_{d0} = 0$. The following boosted tree grows.

$$f_d(z) = \sum_{b \in B} T(z; \Theta), \tag{5}$$

where B is the collection of all the bootstrapping subsamples. The output of the tree enters the SoftMax function to produce a set of probability predictions as follows.

$$p_d(z) = \frac{\exp\left[f_d(z)\right]}{\sum_{d \in D} \exp\left(f_u(z)\right)}. \tag{6}$$

The algorithm calculates pseudo residuals $r_{db} = y_d - p_d(z)$ for all regions R_{jb}. Then, it updates γ_j through loss minimization and outputs an updated boosted tree.

$$f_{db}(z) = f_{db-1}(z) + \sum_{j \in [1,J]} \gamma_{jdb}\, I\left(z \in R_j\right). \tag{7}$$

The optimization process solves the parameters in a recursive manner with the bootstrapping samples.

$$\widehat{\Theta}_b = \arg\min_{\Theta_b} \sum_{i \in [1,N]} L\left(y, f_{b-1}(z) + T(z; \Theta_b)\right), \tag{8}$$

where y is the response variable of the observation z, and L is the cross-entropy function with the probabilities as the input or the mean square loss function with the numeric prediction as the input.

3.4 Modeling Strategy: Training, Grid Search, and Aggregation

Conditional on time windows, I separate historical observations into training sets, validation sets, and testing sets. In total, I update the models four times (every ten years), and the out-of-sample prediction period starts in January 1983. Figure 3 demonstrates my modeling strategy.

[Insert Figure 3 Here]

Each update of the models includes two stages. First, using the training data set, I fit the individual models with different architectures and hyperparameters. Then, I make predictions in the following validation set, which includes the observations that the models do not see during the training period. I select the best architecture and hyperparameters for each model, which is applied to the out-of-sample predictions in the corresponding testing set. The specific windows that I adopt in this paper are detailed in Appendix Table A1.

I focus on four models, including ANN with rectifier activation function, ANN with tanh activation function, random forest, and gradient boosting tree. The main architectural hyperparameters for ANN models are the number of hidden layers and the number of neurons in each hidden layer, while the main architectural hyperparameter for tree models is the max number of layers that the tree models can grow. I conduct a wide range search of the architectural hyperparameters, and Table 1 reports my modeling specification.

[Insert Table 1 Here]

I build two ANN models. Each will search for 30 sub-models and take the architectural specification with a shrinkage parameter. I also build two tree models. Each will search for four sub-models with the specified numbers of depths. For the ANN models, I specify the number of epochs to 1000 times. Comparably, I specify the number of trees in the tree models to be 1000 trees. The details of the optimization choices can be found in Appendix Table A2.

In Section 4, I report the individual model's prediction performance. However, for brevity, I report the economic analyses based on the measures formed with the aggregated predictions. I take the simplest route and aggregate the predictions from my four models by averaging them. However, I do not average the prediction directly. Instead, I take the average of each decile's predicted probabilities across the models, based on which I make the return decile predictions by selecting the decile with the highest aggregated probability.

$$\widehat{P_{agg}}(d_{i,t}) = \frac{1}{4} \sum_{c \in \{all\ 4\ classifiers\}} \widehat{P_c}\,(d_{i,t}), \tag{9}$$

where $\widehat{P_{agg}}(d_{i,t})$ is the aggregated predicted decile probability for stock i at time t to be in decile $d \in D$ and 4 represents the total of four classifiers. I then use the aggregated predictions and the aggregated probabilities to calculate my uncertainty measures. I discuss the measures in Section 5.

3.5 Data

My data contains 3,342,486 monthly stock observations of 26,302 distinct common stocks listed on three major exchanges covering 196201:202112. The lagged predictors include the return decile, 102 firm characteristics, 2-digit SIC industry indicator, and 2-digit SIC industry lagged returns.[14] Specifically, I construct the firm characteristics following Green et al. (2017) and Gu et al. (2020) based on CRSP and COMPUSTAT. I start by making the data set to be completely CRSP centric with no data elimination if possible.[15] I only eliminate rows with missing current returns and rows that are not common stocks (SHRCD 10, 11, or 12) listed on the major three exchanges (EXCHCD 1, 2, or 3). For factor model tests and risk-free rate, I obtain Fama and French's (2015) five factors from French’s website. Appendix Table A3 reports the definition and summary statistics of my prediction sample.

4. Modeling Performance

In this section, I demonstrate the statistical and economic performance of my models. The out-of-sample performance is important to the objective of this paper. Only when the models perform well in extracting return information from the comprehensive set of public information, the stock-level prediction success and the market-level prediction precision become meaningful foundations for the analysis on the implications of return predictability.

[14] Following Green et al. (2017) and Gu et al. (2020), I lag the annual firm characteristics by at least 6 months, I lag the quarterly firm characteristics by at least 4 months, and I lag the monthly firm characteristics by at least 1 month.
[15] My data construction avoids the problem of fluctuating number of stocks from month to month.

4.1 Prediction Precision

Panel A and B in Table 2 report the precision of the predictions from my models individually and in aggregate. I define the overall cross-sectional prediction precision as the total number of successful predictions scaled by the total number of predictions.

$$Precision = \frac{1}{\#\ Predictions}\sum_{i,t} I(d_{i,t} = \hat{d}_{i,t}), \tag{10}$$

where I is an indicator of value 1 or 0, $\hat{d}_{i,t}$ is the predicted return decile, and $d_{i,t}$ is the realized decile. This calculation uses only the out-of-sample predictions. I define the cross-sectional prediction precision at time t as the total number of successful predictions scaled by the total number of predictions for the specific period.

$$Precision_t = \frac{1}{\#\ Predictions_t}\sum_{t} I(d_{i,t} = \hat{d}_{i,t}), \tag{11}$$

where I is an indicator of value 1 or 0, $\hat{d}_{i,t}$ is the predicted return decile, and $d_{i,t}$ is the realized decile. This calculation uses only the out-of-sample predictions.

The best in-sample and out-of-sample model is the random forest model, delivering prediction precisions of 17.9% and 16% respectively. The ANN models underperform the tree models in both the training set and the testing set. The ANN models produce prediction precision around 15.5%. In general, the training set precision is higher than testing set precision. But the deterioration is small, except for the RF. In appendix Table A4, I detail the parameters selected from the in-sample training and the validation process for each model. Compared to the ANN model with rectifier activation function, the ANN model with tanh activation function tends to select small models. However, the RF model prefers complex structure.

[Insert Table 2 Here]

The naïve classifier precision is the benchmark that assigns the return decile with the largest discrete decile distribution prevalence to all the observations as the predicted decile.[16] In other words, the naïve classifier's prediction maximizes the prediction precision conditional only on the past return decile distribution assuming that the other characteristics contains no information about future returns. Since I balanced my in-sample data following the common practice of the classification, the naïve precision is 10% for the in-sample prediction. The out-of-sample data has slightly higher naïve precision at 10.1%. The binomial tests indicate that the precisions delivered by the machine learning classifiers are statistically meaningful.[17] In other words, all the models are successful in extracting information of future returns from the input variables through the modeling structure. When I aggregate the predictions, the aggregated classification achieves even higher out-of-sample prediction precision at 16.1%. In the appendix Table A5, I report the 2-digit SIC industry-level prediction precision and information shortage across the out-of-sample data set. Industries like Forestry, Metal, Mining, and Oil & Gas Extraction demonstrate the highest prediction precision, whereas Automotive Dealers & Service Stations, Trucking & Warehousing, and Hotels & Other Lodging Places exhibit the lowest prediction precision. The measurement of information shortage is detailed in the next section.

In Table 2 Panel C, I report the performance of the benchmark regression models with the same parameter and hyperparameter specifications of their classifier counterparts such that a head-to-head comparison is possible. Specifically, these machine learning regression models predict numeric returns first and then prediction is sorted to form the decile predictions. I compare the decile predictions from the classifiers and the regressions and conclude that the classifiers achieve higher precision in allocating the stocks into the correct future deciles.

Figure 4 shows the average precision from the aggregated prediction in the out-of-sample period. The precision is consistently higher than 10%. Since this paper focuses on decile predictions, 10% is prediction precision of the naïve classifier based on historical

---

[16] The comparison between the prediction precision from the classifiers and the prediction precision from the naïve classifier is similar to the comparison between the prediction precision from predictive regressions and the historical mean.

[17] The binomial test is popular in testing whether two probabilities of success is equal. Because of the success is measurable for classification, i.e., a correct prediction is a success, the binomial test is often applied in machine learning to test if the classifier learns something from the data that is meaningful, i.e., systematically different from best guess based on historical distribution.

distribution of majority class. Therefore, Figure 4 highlights the robustness of the overall performance of the classification. I include the notable exogenous events that has profound economic impacts in the figure. Specifically, I include 9-11 attack, Hurricane Katrina, Hurricane Maria, and COVID-19 (Kelly and Ljungqvist 2012). Following these events, the precision drops immediately indicating that prediction precision correctly reflect the changes in the economic uncertainty. Section 5 discusses the specific relation between economic uncertainty and market level return predictability in details.

[Insert Figure 4 Here]

4.2 Economic Performance

In this subsection, I discuss the economic performance. Because I have several models, to keep the performance evaluation concise, I focus on the portfolios constructed with the aggregation of the predictions. I form both the equal-weighted and the value-weighted portfolios. I also include the long-short portfolios, where I short the lowest decile portfolio and hold the top highest decile portfolio with a 50%-to-50% weight ratio. In my calculation, return is adjusted with the risk-free rate.

Table 3 reports the portfolio performance based on the decile predictions. I report several important statistics. First, I report the average excess return across the time periods. The excess return is defined as the portfolio return minus the risk-free rate. Second, I report the cumulative return in the out-of-sample period, i.e., 198301:202112. Third, I report alphas from the standard factor models including the capital asset pricing model (CAPM), the Fama-French 3 factor model (FF3F), and the Fama-French 5 factor model (FF5F) (Fama and French 1992, 2015).[18] I obtain the factor-model alpha from fitting the following regression.

$$R^e_{p,t} = \alpha_{p,t} + \boldsymbol{F}_t\mathbf{B}_p + \varepsilon_{p,t}, \tag{12}$$

where $\boldsymbol{F}_t$ contains the factors at time t and $\mathbf{B}_p$ is the risk loadings for the portfolio p. Lastly, I report important portfolio performance including standard deviation, annualized

[18] I report Newy-West t statistics for the alphas with a lag of 6 (Newey and West 1987).

Sharpe ratio, turnover, maximum drawdown, and the average number of stocks in each portfolio.

[Insert Table 3 Here]

I define monthly Sharpe ratio as a portfolio's excess return scaled by the standard deviation of the portfolio return, and I annualize the Sharpe ratio by multiplying the monthly Sharpe ratio with $\sqrt{12}$:

$$SR_p = \frac{E(R_p - R_f)}{\sigma(R_p)} \times \sqrt{12}. \tag{13}$$

The turnover is defined as

$$Turnover = \frac{1}{n}\sum_{i=t}^{t+n}\left(\sum_{j}\left|w_{j,i+1} - \frac{w_{j,i}(1 + r_{j,i+1})}{\sum_k w_{k,i}(1 + r_{k,i+1})}\right|\right), \tag{14}$$

where $w_{j,i}$ represents the weight of stock j during month i in a portfolio (Gu et al. 2020; Neely et al. 2014). I define the maximum drawdown according to the most recent peak of the cumulative return in the sample coverage.

$$MaxDD_{t:t+n} = \min_{t:t+n}\left(\frac{Y_{i+1} - Y_i^{peak}}{Y_i^{peak}}\right), \tag{15}$$

where i is a trading month during the investment window $t: t + n$. Y_i^{peak} is the highest cumulative return until the month i.

Table 3 Panel A reports the equal-weighted portfolio performance using the classification predictions, while Panel B reports the value-weighted portfolio performance using the classification predictions. In general, the aggregate of the algorithms is good at dissecting future returns. The portfolio returns present a linear pattern with the lowest decile delivering the lowest return and the highest decile delivering the highest return. My portfolios deliver average excess return as high as 2.3% (1.4%) monthly for the equal-weighted (value-weight) scheme. The alphas from CAPM and factor models indicate that the standard risk factors cannot fully explain the returns from the portfolios including stocks of returns that are below or above the market median returns. The long-short portfolios deliver Sharpe ratios significantly higher than the Sharpe ratios from holding the market return. The maximum drawdowns decrease significantly in the long-short portfolios. In Appendix Table A6, I include the performance statistics for the portfolios based on the predictions including only the stocks from the top 50% market capitalization. My findings indicate that the performance of the strategy is robust. My analyses in this section together also point out an investor can beat the market substantially by only correctly predicting stock returns for just more than 16% of the time.

In Panel C and Panel D, I report the portfolio allocation performance based on the stacked models including both aggregated predictions from the classification models and the regression models. Only stocks that are predicted to be in a portfolio by both the classification models and the regression models are included in the related portfolio. Comparing to the Panels A and B, the performance of the portfolios based on the stacked predictions increases substantially and deliver Sharpe ratios of 3.4 and 1.5 for equal-weighted long-short portfolio and value-weighted long-short portfolio, respectively. More importantly, if we compare the performance of the stacked predictions with the portfolio performance from the regression models in appendix Table A7, according to the long-short portfolios' Sharpe ratios, the stacked predictions' performance increases significantly by about 13% in the equal-weighted scheme and 48% in the value-weighted scheme, respectively. This indicates that the classification models provide meaningful additional information about future returns benchmarking against machine learning regressions' predictions. In general, the classification models' performance is on par with the machine learning regressions' performance. Since machine learning regressions provide state-of-the-art performance in the cross-sectional return predictability, I conclude that the classification models' prediction precision is a good representation of the market's best anticipated prediction precision and that it is reasonable to adopt such precision measure as a proxy in the analysis of the return predictability's implications (Gu et al. 2020).

5. Sources of Return Predictability

In this section, I leverage the prediction performance of the classifiers to investigate the sources of cross-sectional return predictability. In Table 4, I report the details of the out-of-sample performance of the aggregated predictions in confusion matrices. Panel A reports the number of observations with the predicted decile $\widehat{d_t}$ in contrast with the realized decile d_t. For example, the first row in the first column shows that the aggregated predictions place 122,627 out-of-sample observations in the predicted decile 1, and these observations also realize in decile 1 in the next period. Panel B reports the scaled version of Panel A by the number of observations in the true class. Panel C reports the scaled version of Panel A by the number of observations in the entire sample.

[Insert Table 4 Here]

My results show that the models on average devote the most resources to the deciles on the two tails and around the center of the return distribution. The models also achieve the highest precisions in these deciles. For example, for the real decile 1, the models spend the most resources and made 546,858 predictions, out of which 112,627 observations realize in decile 1.[19] These 112,627 observations make up 5% in the total precision out of the 100 possible combinations between the predicted deciles and the realized deciles. 49% of these observations that realize in decile 1 are detected correctly by the aggregated predictions from the machines. While the model also gains precision from deciles 6-8 and decile 10.

5.1 Predictability, Return Decile Transitions, and Information Shortage

I proxy the information shortage on a market level for individual stocks using Shannon's information entropy based on the predicted probabilities as the following (Shannon 1948).[20]

[19] I view the number of the observations allocated into a predicted decile as the total resources the machines spend on the predictions. For example, the summation of first row in Table 3 Panel A is 546,858, indicating that the machines predict this many observations as decile 1 observations. The numbers of the observations predicted to be in decile 1 to decile 10 are the following: 546,858, 221,697, 57,031, 72,270, 126,301, 435,344, 261,155, 230,693, 137,730, and 411,396. Therefore, the models spend the least resources on decile 3 and decile 4.

[20] Such measure is conditional on the past public information. Since the models condense the information from a comprehensive list of predictors and deliver significant performance, I argue that this information shortage is representative for the best predictions based on public information.

$$E_{i,t} = -\sum_{d_{i,t} \in D} p\widehat{(d_{i,t})} \log_2 p\widehat{(d_{i,t})}, \tag{16}$$

where D includes [1:10] and $p\widehat{(d_{i,t})}$ stands for the predicted probability of the event that the stock i at time t will be in the decile $d_{i,t}$. Note that my measure of information shortage is concurrent because it is directly from the predictions, while the measure of precision depends on past prediction accuracy.

By definition, my information shortage measures the expected minimum number of binary questions that need to be answered to make 100% correct predictions, and the unit of the information shortage is then in "bits". In other words, if a stock is associated with an entropy or information shortage of three, at least three binary questions about the decile returns must be answered so that the prediction can be made without uncertainty. In other words, the three bits represent the information shortage in making return predictions.

With the measures of prediction precision and the information shortage, I continue to investigate the transition probabilities and their relations with the machine learning predictability. Table 5 presents my analyses. Panel A reports the unconditional transition probabilities of cross-sectional return deciles during the out-of-sample period. Panel B demonstrates the prediction precision from the combined model by the transitions, while Panel C reports the information shortage by the transitions.

[Insert Table 5 Here]

Compared with a random distribution of return transition, which should have a 1% probability, the unconditional transition probabilities are distributed in an uneven way. First, the center of the transition matrix highlights the certainty of the return transitions from deciles 4-7 to the center of the return distribution. Such transitions show a probability of around 1.2%. Transitions from decile 1 to deciles 1 and 10 have probabilities of 1.7% and 1.8%, respectively. Similarly, transitions from decile 10 to deciles 1 and 10 also have greater certainty. In Panel B, the prediction precisions for each transition suggest that the machines take advantage of the uneven distribution of the transition probabilities. The machines achieve the highest precision for the transitions from the center deciles to the center deciles and the transitions from the extreme deciles to other deciles.

The results in Panel C of Table 5 emphasize the machines' choice from the information perspective. The table replicates the uneven distribution of transition probabilities from Panel A and reflects the similarity in the information shortage. The transitions in the center and the extreme transitions are clearly of smaller information shortage, while other transitions have greater information shortage.

5.2 Predictability, Variable Importance, and Tests on Firm Characteristics

Figure 5 reports the average variable importance across the training periods for each variable. I take the average percentage of total sum of squared error reduction to estimate the variable importance for the tree models across all the trees and the splitting nodes related to the predictors of interest. Similarly, I apply Gedeon method to compute the variable importance in the neural networks, and Gedeon's method is based on the summation of the squared normalized weights related to each input predictor in all the layers (Breiman 1984, 2001; Gedeon 1997; Hastie et al. 2009).

My results show that the models draw information from different predictors. The ANN models extract information from a wider range of predictors compared with the tree models. Notably, the gradient boosting tree heavily relies on idiosyncratic volatility (*idiovol*), which contributes 45% of the sum of squared error reduction in the model. The ANN models rely more on past industry information (*sich2*) and return decile distribution (*label10*), which contribute more than 20% and more than 6% to the neural weights, respectively. The selection effect is also obvious. Variables such as annual income (*acc* and *absacc*), industry-adjusted percentage change in capital expenditures (*pchcapx_ia*), and analysts' mean annual earnings forecast (*sfe*) contribute the least to the machines' predictions.

[Insert Figure 5 Here]

Next, I study the return predictability from the machines with respect to the firm characteristics, assuming linear relations. Specifically, I focus on two measures, i.e., the prediction success and information shortage. I define the prediction success as a dummy variable with value of 1 indicating the predicted decile is the same as the realized decile, while the information shortage is defined with the information entropy using the predicted probabilities.

I perform two Fama-MacBeth regressions. Specifically, I regress the prediction success and the information shortage on the firm characteristics, which are included in the machine learning models. Therefore, the coefficients of the Fama-MacBeth regressions indicate the marginal contribution from the firm characteristics to the prediction success and the information shortage.

$$Success_{i,t}\ or\ Info.Shotage_{i,t} = \gamma_0 + \boldsymbol{Char._{i,t-1}}\,\boldsymbol{\Gamma} + \varepsilon_{i,t}. \tag{17}$$

[Insert Table 6 Here]

Table 6 presents the results of my analyses, and I report the significant predictors only. Panel A shows a list of variables that are related to the prediction precision. For example, a one standard deviation increase in the change of the 6-month momentum (*chmom*) is related to a 0.4% increase in the prediction precision, while return on assets (*roa*) is related to a 0.05% decrease in the return predictions. In total, there are 24 (30) firm characteristics that are positively (negatively) related to machine learning prediction precision. These characteristics cover all six types of firm characteristics of Hou et al.'s categorization, including momentum, value vs. growth, investment, profitability, trading frictions, and intangibles (Hou et al. 2018).

The appendix Table A8 reports the results of the analysis on information shortage. 49 firm characteristics are positively related to information shortage, including variables such as analysts' earnings forecast dispersion (*disp*), return on equity (*roeq*), earnings-to-price ratio (*ep*), and beta. In comparison, 35 predictors are negatively related to the information shortage, including firm age (*age*), change in 6-month momentum (*chmom*), dividend yield (*dy*), and bid-ask spread (*baspread*). Specifically, for example, a standard deviation increase in analysts' earnings forecast dispersion (*disp)* is associated with the increase in the information shortage of 0.003 bit, while one-year increase in firm age is related to a reduction of 0.012 bit in the information shortage.

5.3 Predictability and Information Shortage

I explore the relation between prediction precision and information shortage. I adopt the information shortage calculated as the information entropy defined in Subsection 5.1 that captures the additional requirement of information to make fully correct predictions. Table 7 reports the results across the models and the aggregated prediction. The results emphasize the negative relation between the probability of successful prediction and the information shortage. In general, for the aggregated predictions, a one-standard-deviation increase in the information shortage leads to a decrease in the successful rate of prediction by 3.9%.

[Insert Table 7 Here]

An immediate implication of this finding is that the portfolio allocation performance may increase if the portfolio construction only includes low information shortage stocks. Indeed, the performance of portfolios increases substantially conditional on the stocks of the bottom decile information shortage. Table 8 reports the portfolio performance. The Sharpe ratios of the equal-weighted and value-weighted long-short portfolios increases to 2.98 and 1.58, respectively.

[Insert Table 8 Here]

5.4 Predictability and Economic Uncertainty

The theory literature often predicts macroeconomic uncertainties as the main drivers of asset variations (Bansal and Yaron 2004; Barro 2009; Lucas 1978). Therefore, an implication to the predictability literature is that the macroeconomic uncertainties may predict stock returns (Lettau and Ludvigson 2001a; Cochrane 2007). The prior subsections investigate the influence from the stock-level firm characteristics on the prediction performance. In this subsection, I focus on the influence of macro-level events as proxied with common uncertainty indices, including war factor, geopolitical risk, economic policy uncertainty, macro uncertainty, financial uncertainty, and real uncertainty (Baker et al. 2016; Caldara et al. 2022; Hirshleifer et al. 2023; Jurado et al. 2015). Specifically, I perform the following time series regression and report Newey-West testing statistics with the lag of 12.

$$Precision_t = \gamma_0 + \ Uncertianty_t + \varepsilon_{i,t}. \tag{18}$$

Table 9 reports regression results. Among the macroeconomic uncertainty indices, only the Macroeconomic Uncertainty of Jurado et al. (2016) show robust significant influence on the market level prediction precision in percent. A one-standard-deviation increase in the macroeconomic uncertainty is associated with 0.31% reduction in the monthly prediction precision at the market level. The financial uncertainty proxy from Jurado et al. is marginally significant. This finding highlights the influence of macroeconomic factors on the market level prediction precision, consistent with the theoretical predictions that the drivers of the asset prices include macroeconomic variables.

[Insert Table 9 Here]

6. Predictability and Stock Returns

Easley and O'Hara (2004) predict that the precision is negatively related to the stock return. For example, it is believed that the better accounting and governance quality can make investors more confident in their predictions of cash flows. Such confidence reduces risk premium (Bansal and Yaron 2004; Christensen et al. 2010). Section 4 establishes the state-of-the-art performance of machines predictions using classification models. Considering the inclusiveness of the broad set of public information, the predictions provide a sound empirical foundation to analyze the theoretical implication of the return predictability in the investment literature.

Specifically, I focus on providing individual stock-level evidence of the relationship between returns and the cross-sectional prediction success. Following Green et al. (2017), I run the using the Fama-MacBeth predictive regression, controlling 102 firm characteristics, 2-digit SIC fixed effects, and past return deciles.

$$\begin{aligned} R_{i,t} = \gamma_0 + \gamma_1 Prediction\ Success_{i,t-1} + \gamma_2 Info.Shortage_{i,t} \\ + \boldsymbol{Controls_{i,t-1}}\boldsymbol{\Gamma} + \varepsilon_{i,t}, \end{aligned} \tag{19}$$

where all regressors are based on lagged information. The regression controls for the information shortage to mitigate concerns on the information uncertainty and information asymmetry (Kelly et al. 2012; Jiang et al. 2006; Merton 1987; Zhang 2006).

I report the regression results in Table 10 covering the monthly stock data during the out-of-sample period from 198301:202112 on more than 2.5 million observations. Controlling for the firm characteristics and the information shortage, the prediction success from the past month remains significant at the 0.01 level, indicating that the relation between return and the prediction precision is strong and robust.[21] A past prediction success is associated with a 0.2% reduction in the one-month-ahead return.

[Insert Table 10 Here]

Note that the 102 firm characteristics include almost all common proxies of uncertainty directly related to or irrelated to traditional information uncertainty proxies. For example, firm age, monthly average of bid-ask spread (*baspread*), standard deviation of analyst earnings forecasts (*disp*), dollar value volatility (*dolvol*), number of analyst coverage (*nanalyst*), return volatility (*retvol*), earnings surprise (*sue*), among others are all included in the regressions (Green et al. 2007; Gu et al 2020; Jiang et al. 2005; Zhang 2006). The stock-level results for the information shortage are also consistent with the theoretical prediction from Merton (1988) and the empirical evidence from Jiang et al. (2006) and Zhang (2006). The standalone regression of information shortage indicates that 1 bit increase in the additional information necessary to make perfect predictions will lead to 0.35% decrease in the future monthly return.

6.1 Predictability and Pricing Error Fluctuations

The relation between stock returns and prediction success leads to a natural question on the role of price continuation in realizing out-of-sample predictability. The literature has documented the potential situations where price delay can happen (Boehmer and Wu 2012; Cohen et al. 2020; Hou and Moskowitz 2005). Such delay of information

[21] I follow Green et al. (2017) and report the predictive Fama-French regression estimates and statistics. In untabulated results, I show that my results are also robust under the ordinary least square (OLS) estimates with regular clustered errors at the firm level.

incorporation in price will lead to mispricing. For example, Stambaugh et al. (2015) show that short selling overpriced stock can be hard because of less arbitrage capital available. On the one hand, it is thus plausible that the predictability of the machines is dependent on such price delay and persistence in pricing errors. On the other hand, the predictability can also rely on the correction process of the pricing errors. In other words, the machines see the information before the market incorporates it and anticipates the time when the market will incorporate the information.

To investigate this possibility, I proxy the pricing errors at the stock level using the most commonly used factor model, i.e., the Fama French 3 Factor model. I calculate the pricing errors in 60-month rolling windows. In order to study both the influence from the pricing errors' direction and the influence from the pricing errors' magnitude, I create two pricing error measures, i.e., the change in pricing error and the absolute change in pricing error. I regress the prediction success on the changes in pricing errors in Fama-MacBeth regression controlling for the past period pricing error following the equation below.

$$\begin{aligned} Success_{i,t} &= \gamma_0 + \gamma_1 \Delta Pricing\ Error_{i,t} + \gamma_2 Pricing\ Error_{i,t-1} \\ &+ \boldsymbol{Controls_{i,t-1}}\boldsymbol{\Gamma} + \varepsilon_{i,t}, \end{aligned} \tag{20}$$

where $\Delta Pricing\ Error_{i,t}$ is difference between realized pricing error for the 60-month window ending in period t and the 60-month window pricing error ending in period $t-1$. I also perform the similar regression using $|\Delta Pricing\ Error_{i,t}|$ in place of $\Delta Pricing\ Error_{i,t}$. Table 11 reports the results of the pricing error analysis.

[Insert Table 11 Here]

The Fama-MacBeth results confirm the significance of the pricing errors' influence on the out-of-sample predictability. The relation between the pricing errors and the prediction success is profound. Specifically, the past pricing error negatively influences the predictability. Higher pricing errors are related to lower predictability. Similarly, higher contemporaneous pricing error change is also related to lower predictability. However, the absolute level of the pricing error changes, i.e., the scale of the pricing error change instead of the direction of the pricing error change, is positively contributing to the out-of-sample predictability. In other words, the predictability does not come from the pricing errors or the continuation of pricing errors. Instead, the predictability comes from the fluctuation in the pricing errors. It is the impersistence in the pricing errors that contribute positively to the out-of-sample predictability.

6.2 Mimicking Portfolio and Spanning Tests

The profound relation between stock returns and prediction success implies a possibility that if a predictability portfolio is tradable, it may generate substantial benchmark-adjusted returns. To test this hypothesis, I first follow the literature and obtain the residuals from the autoregressive model with 1 lag. Then, I map the residuals to the return space spanned by 170 basis portfolio returns obtained from Kenneth French's data library, including single sort, double sort, and triple sort portfolios for the characteristics included in the factors of Fama-French 6 factor model (Adrain et al. 2014).

$$\Delta Precision_t = \alpha + \boldsymbol{R^b}\boldsymbol{\beta} + \varepsilon_t, \tag{21}$$

where $\Delta Precision_t$ is the AR1 innovation and $\boldsymbol{R^b}$ is the matrix of 170 basis portfolios.

Then, I regress the return from the mimicking portfolio on the common factor models, including Fama-French 5 factor model, Fama-French 6 factor model, q4 factor model, and q5 factor model.

$$PMP_t = \alpha + \boldsymbol{F}\boldsymbol{\beta} + \varepsilon_t, \tag{22}$$

where PMP_t is the mimicking portfolio return generated from the projection in equation (21) and F is the factor returns. Table 12 reports the results of spanning tests. None of the common factor models fully explain the mimicking portfolio return. In general, the mimicking portfolio generates an annual alpha of around 2.4%. The mimicking portfolio return is strongly loaded on the market factor, value factor, profitability factor, momentum factor, and the investment factor. This finding confirms that the prediction precision has profound influence on stock returns (Easley and O'Hara 2004).

[Insert Table 12 Here]

7. Conclusion

In this paper, I provide an alternative perspective of machine learning return predictions, through which I shed light on the economic sources and the consequences of return predictability. Specifically, I first dissect stock returns into deciles and construct classification models to allocate the individual stocks to the future decile portfolios. My models deliver statistically meaningful performance and successfully predict 16% of the return deciles, which translates to significant economic performance. Indeed, my classification-based long-short portfolios can achieve a Fama-French five factor (FF5F) monthly α of 1.1% and 2.1% for the value-weighted and the equal weight portfolios, respectively. Conditional on the bottom decile of information shortage, my long-short portfolios can deliver monthly returns as high as 8%, or annualized Sharpe ratios as high as 3. When stacking my models on top of the machine learning regression models, the economic performance increases substantially compared to the performance from only the machine learning regressions. This finding emphasizes the ability of the classification models in capturing the future return information in addition to the information captured by the machine learning regressions.

Based on the models, I measure prediction precision and information shortage. I document that the market transition probabilities distribute unevenly. The transitions from the center and edges of the transition probability matrix are more certainty. I show that the machines take advantage of such unevenness and achieve exceptional detection rates in the transition from the lowest decile in the past to the lowest decile in the future. My measure of information shortage shows that the machines exploited the advantage of the transition matrix's unevenness. In addition, I show that the market-level prediction precision slumps following exogenous macroeconomic shocks, highlighting the precision's reflection of economic uncertainty.

My results show that all six firm characteristics categories of Hou et al. (2018) contribute to the prediction precision and the information shortage. For example, a one standard deviation increase in the change of 6-month momentum (*chmom*) is related to a 0.4% increase in prediction success, while return on assets (*roa*) is related to a 0.05% decrease in the prediction success. 49 firm characteristics are positively related to information shortage, including variables such as analysts' earnings forecast dispersion (*disp*), return on equity (*roeq*), earnings-to-price ratio (*ep*), and beta (*beta*). In comparison, 35 predictors are negatively related to the information shortage, including firm age (*age*), change in 6-month momentum (*chmom*), dividend yield (*dy*), and bid-ask spread (*baspread*).

I investigate the consequences of the return predictability. Specifically, Easley and O'Hara (2004) show that prediction precision is negatively related to firm value. Using the classification setup, I provide the first empirical evidence that confirms the theoretical prediction in the stock market. This result is robust controlling for the information shortage measured as the information entropy in prediction. A past prediction success is associated with a 0.2% (2.4%) reduction in the one-month-ahead (annual) return at the stock level. The spanning tests with the mimicking portfolio confirm that a portfolio of prediction precision, if tradable, can generate an annual benchmark-adjusted performance of -2.4%. My analysis also reveals the profound relation between predictability and pricing errors. I show that the predictability mainly comes from the fluctuation of pricing errors instead of the persistence in pricing errors.

References

Acharya, V., and L. Pedersen, 2005, Asset Pricing with Liquidity Risk, *Journal of Financial Economics* 77, 375–410.

Adrian, T., E. Etula, and T. Muir, 2014, Financial Intermediaries and the Cross-Section of Asset Returns, *Journal of Finance* 69, 2557–2596.

Ahmed, A. S., M. Neel, and D. Wang, 2013, Does Mandatory Adoption of IFRS Improve Accounting Quality? Preliminary Evidence, *Contemporary Accounting Research* 30, 1344–1372.

Altman, E. I., 1968, Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy, *Journal of Finance* 23, 589–609.

Amihud, Y., and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics* 17, 223–249.

Anderson, E. W., E. Ghysels, and J. L. Juergens, 2009, The Impact of Risk and Uncertainty on Expected Returns, *Journal of Financial Economics* 94, 233–263.

Aubry, M., R. Kraussl, G. Manso, and C. Spaenjers, 2023, Biased Auctioneers, *Journal of Finance* 78, 795–833.

Baker, S. R., N. Bloom, and S. J. Davis, 2016, Measuring Economic Policy Uncertainty, *Quarterly Journal of Economics* 131, 1593–1636.

Bali, T. G., H. Beckmeyer, M. M., and F. Weigert, 2023, Option Return Predictability with Machine Learning and Big Data, *Review of Financial Studies*.

Bali, T. G., and H. Zhou, 2016, Risk, Uncertainty, and Expected Returns, *Journal of Financial and Quantitative Analysis* 51, 707–735.

Bansal, R., and A. Yaron, 2004, Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance* 59, 1481–1509.

Barro, R. J., 2009, Rare Disasters, Asset Prices, and Welfare Costs, *American Economic Review* 99, 243–264.

Bekaert, G., E. Engstrom, and Y. Xing, 2009, Risk, Uncertainty, and Asset Prices, *Journal of Financial Economics* 91, 59–82.

Bianchi, D., M. Büchner, and S. Tamoni, 2021, Bond Risk Premiums with Machine Learning, *Review of Financial Studies* 34, 1046–1089.

Biddle, G. C., and G. Hilary, 2006, Accounting Quality and Firm-Level Capital Investment, *Accounting Review* 81, 963–982.

Bloom, N., 2009, The Impact of Uncertainty Shocks, Econometrica 77, 623–685.

Boehmer, E., and J. Wu, 2012, Short Selling and the Price Discovery Process, *Review of Financial Studies* 26, 287–322.

Bonsall, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp, 2017, A plain English measure of financial reporting readability, *Journal of Accounting and Economics* 63, 329–357.

Breiman, L., 1984, Classification and Regression Trees (Belmont, Calif. Wadsworth International Group).

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

Caldara, D., and M. Iacoviello, 2022, Measuring Geopolitical Risk, *American Economic Review* 112, 1194–1225.

Chen, L., M. Pelger, and J. Zhu, 2023, Deep Learning in Asset Pricing, *Management Science*.

Christensen, P. O., L. E. de la Rosa, and G. A. Feltham, 2010, Information and the Cost of Capital: An Ex-Ante Perspective, *Accounting Review* 85, 817–848.

Clement, M., R. Frankel, and J. Miller, 2003, Confirming Management Earnings Forecasts, Earnings Uncertainty, and Stock Returns, *Journal of Accounting Research* 41, 653–679.

Cochrane, J. H., 2007, The Dog That Did Not Bark: a Defense of Return Predictability, *Review of Financial Studies* 21, 1533–1575.

Cochrane, J. H., 2011, Presidential Address: Discount Rates, *Journal of Finance* 66, 1047–1108.

Cohen, L., C. Malloy, and Q. Nguyen, 2020, Lazy Prices, *Journal of Finance* 75, 1371–1415.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou, 2021, Anomalies and the Expected Market Return, *Journal of Finance* 77, 639–681.

Easley, D., and M. O'Hara, 2004, Information and the Cost of Capital, *Journal of Finance* 59, 1553–1583.

Fama, E. F., and K. R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427–465.

Fama, E. F., and K. R. French, 2015, A Five-Factor Asset Pricing Model, *Journal of Financial Economics* 116, 1–22.

Fama, E. F., and J. D. MacBeth, 1973, Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy* 81, 607–636.

Friedman, J. H., 2001, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics* 29, 1189–1232.

Gedeon, T. D., 1997, Data Mining of Inputs: Analysing Magnitude and Functional Measures, *International Journal of Neural Systems* 08, 209–218.

Green, J., J. R. M. Hand, and X. F. Zhang, 2017, The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns, *Review of Financial Studies* 30, 4389–4436.

Gu, S., B. Kelly, and D. Xiu, 2020, Empirical Asset Pricing via Machine Learning, *Review of Financial Studies* 33.

Hastie, T., R. Tibshirani, and J. Friedman, 2009, The Elements of Statistical Learning Springer Series in Statistics (Springer New York, New York, NY).

Hirshleifer, D. A., D. Mai, and K. Pukthuanthong, 2023, War Discourse and the Cross-Section of Expected Stock Returns, *Social Science Research Network*.

Hou, K., H. Mo, C. Xue, and L. Zhang, 2020, An Augmented q-Factor Model with Expected Growth, *Review of Finance* 25.

Hou, K., and T. J. Moskowitz, 2005, Market Frictions, Price Delay, and the Cross-Section of Expected Returns, *Review of Financial Studies* 18, 981–1020.

Hou, K., C. Xue, and L. Zhang, 2018, Replicating Anomalies,. *Review of Financial Studies* 33, 2019–2133.

Jiang, G., C. M. C. Lee, and Y. Zhang, 2005, Information Uncertainty and Expected Returns, *Review of Accounting Studies* 10, 185–221.

Johnstone, D., 2015, The Effect of Information on Uncertainty and the Cost of Capital, *Contemporary Accounting Research* 33, 752–774.

Jurado, K., S. C. Ludvigson, and S. Ng, 2015, Measuring Uncertainty, American Economic Review 105, 1177–1216.

Kedia, S., and S. Rajgopal, 2011, Do the SEC's enforcement preferences affect corporate misconduct?, *Journal of Accounting and Economics* 51, 259–278.

Kelly, B., and A. Ljungqvist, 2012, Testing Asymmetric-Information Asset Pricing Models, *Review of Financial Studies* 25, 1366–1413.

Lettau, M., and S. Ludvigson, 2001, Consumption, Aggregate Wealth, and Expected Stock Returns, *Journal of Finance* 56, 815–849.

Li, B., and A. G. Rossi, 2020, Selecting Mutual Funds from the Stocks They Hold: A Machine Learning Approach, *SSRN Electronic Journal*.

Lucas, R. E., 1978, Asset Prices in an Exchange Economy, *Econometrica* 46, 1429.

Merton, R. C., 1987, A Simple Model of Capital Market Equilibrium with Incomplete Information, *Journal of Finance* 42, 483–510.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou, 2014, Forecasting the Equity Risk Premium: The Role of Technical Indicators, *Management Science* 60, 1772–1791.

Newey, W. K., and K. D. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703.

Shannon, C. E., 1948, A Mathematical Theory of Communication, *Bell System Technical Journal* 27, 379–423.

Stambaugh, R. F., J. Yu, and Y. Yuan, 2015, Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle, *Journal of Finance* 70, 1903–1948.

Welch, I., and A. Goyal, 2007, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, Review of Financial Studies 21, 1455–1508.

Zhang, X. F., 2006, Information Uncertainty and Stock Returns, *Journal of Finance* 61, 105–137.

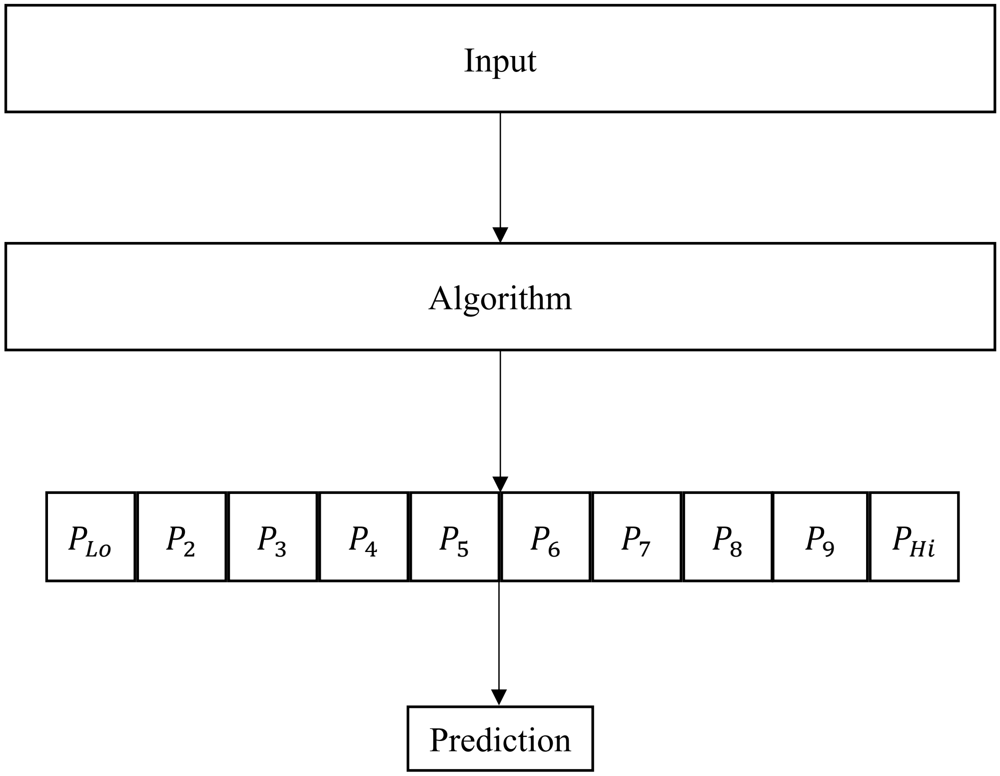

Figure 1 Prediction Process

This figure describes my modeling process. I input the independent variables, i.e., firm characteristics, as the features to a machine learning algorithm. The optimization process uses in-sample training data set to calibrate the parameters such that the predicted probabilities are closely matched to the ground truth distribution of the return deciles conditional on the firm characteristics. Based on the predicted probabilities, such as the probability of the lowest decile P_{Lo}, the algorithm selects the return decile associated with the largest predicted probability as its final prediction. Based on the statistical and economic performance of the classifiers, the classifiers capture the pricing kernel. Thus, the predicted probabilities reflect the state probabilities of the time-varying state returns in deciles.

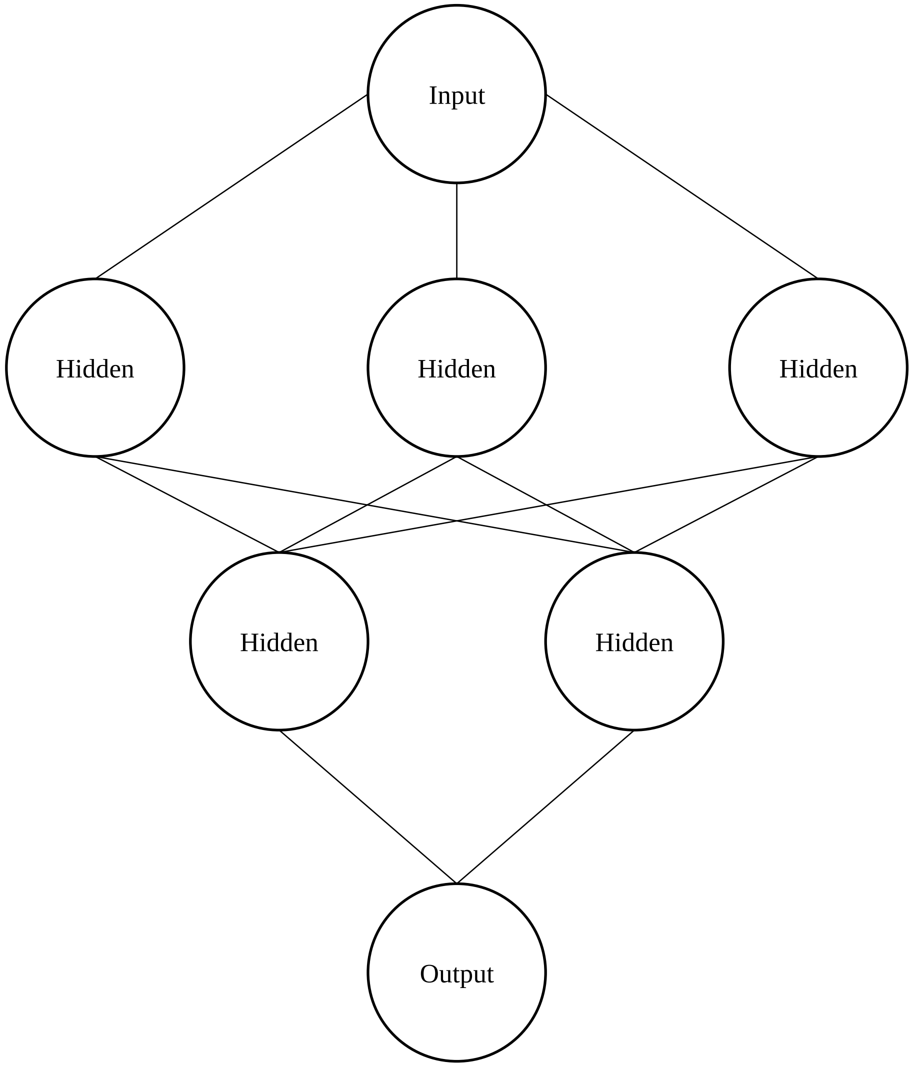

Figure 2 ANN Architecture

This figure illustrates an example structure of ANN with an input layer, two hidden layers of 3 and 2 neurons, and an output layer. The ANN models in this paper take the standard form of the fully connected feed-forward multilayer perceptron. The input layer includes the firm characteristics. The hidden layers make nonlinear transformations. For classification, each neuron in the output layer transforms the input from the hidden layer through fitting a SoftMax function and produces probabilities. I adopt grid search for the combination of layer specifications and lasso shrinkage during the training process. The out-of-sample predictions are made by the best model evaluated with the validation data set. Details of the parameters and hyperparameter search is included in Table 1, appendix TableA1, and Table A2.

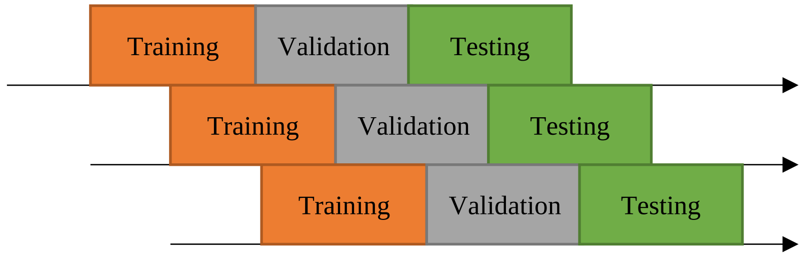

Figure 3 Modeling Strategy

This figure shows my modeling strategy. The models are updated every 10 years in this paper. Each training time uses all the data set available until 5 years before the end of the data. These 5 years are then used in hyperparameter choices. The finalized models are then applied to make predictions for the future observations that the modeling process does not consider. In my tests, I mainly present the aggregated predictions that makes predictions based on the predicted probabilities from all the models.

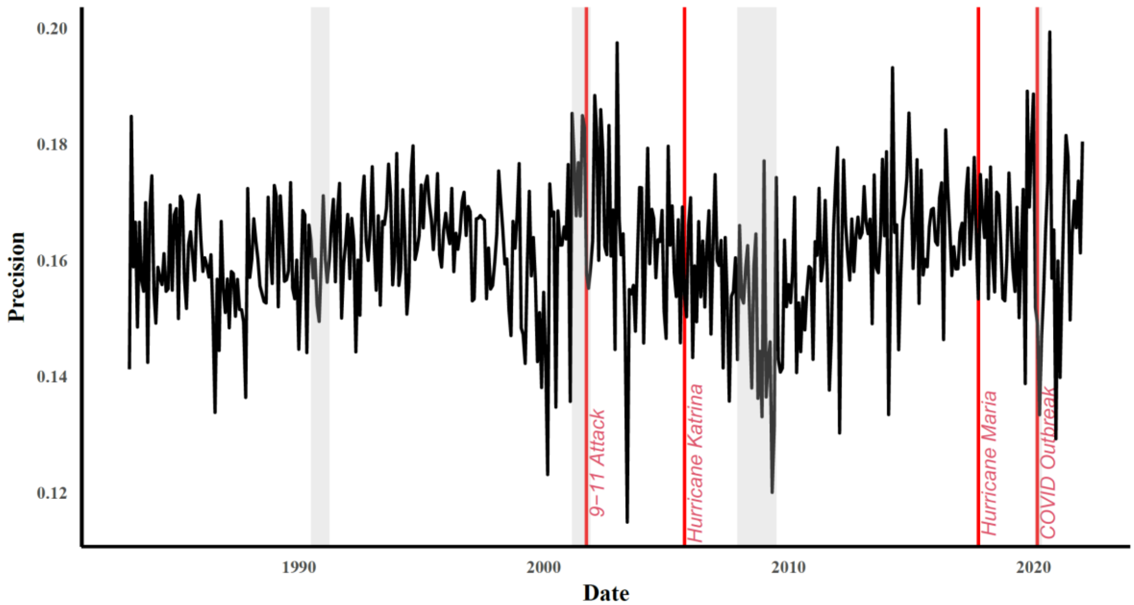

Figure 4 Prediction Precision

This figure demonstrates the precision time series for the aggregated predictions from 198301:202112. The shaded areas denote the NBER-dated recessions. The red lines indicate exogenous shocks to the economy. From the left to the right, the exogenous shocks include 9-11 Attack, Hurricane Katrina, Hurricane Maria, and COVID outbreak. The months after these shocks observe decrease in the prediction precision. However, the precision is always above the best guess based on historical distribution of deciles with a precision around 10%.

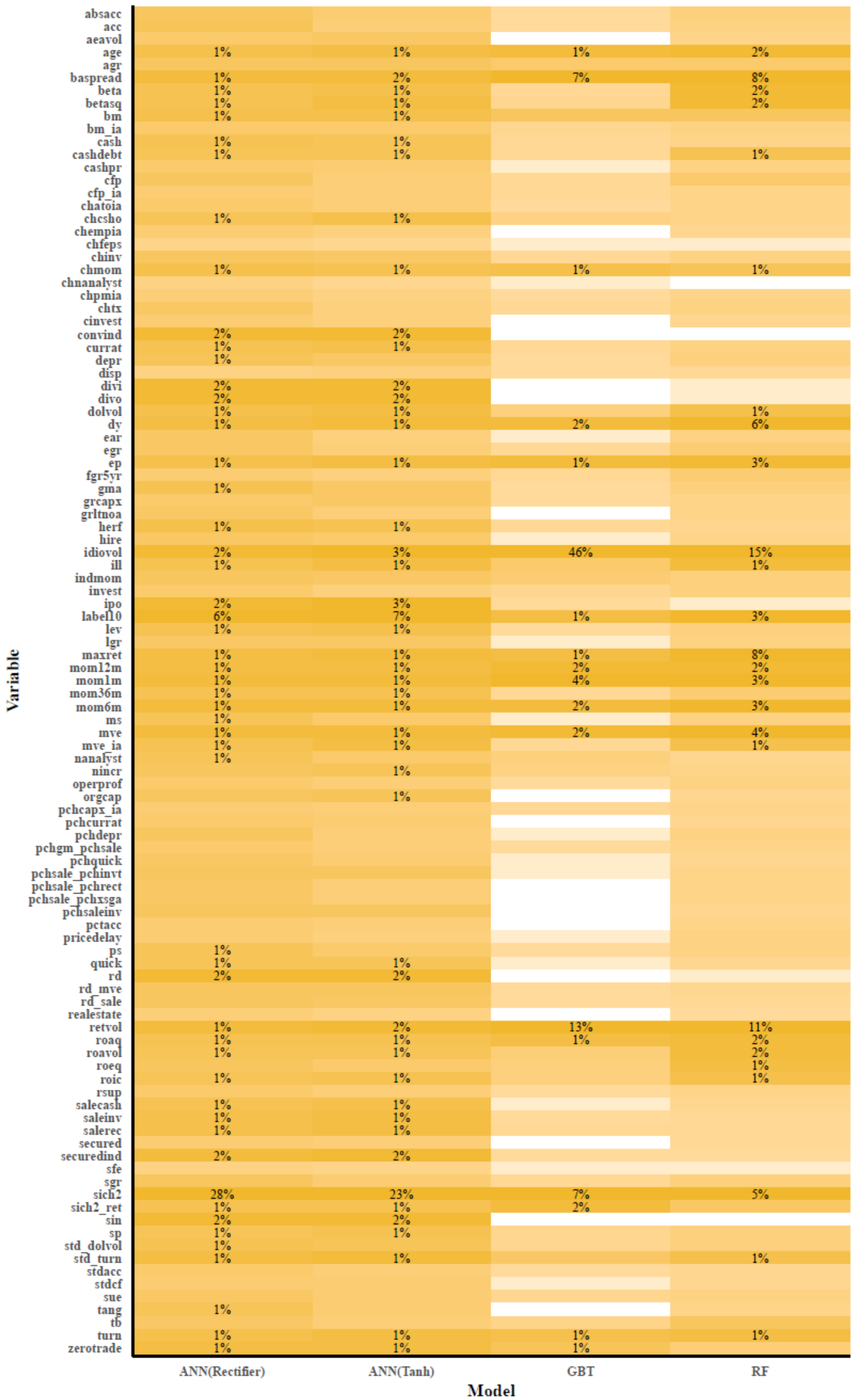

Figure 5 Variable Importance

This figure reports the average variable importance in percentage across all training periods for each individual model.

Table 1 Architectural Search

The table below details the main parameter choices for my models in this paper. Panel A reports the architectural search for the hyperparameters. The hyperparameters are parameters decided through the tuning process happening in the validation data sets instead of the optimization process. For my ANN models, the main architectural choice is about the number of hidden layers and the number of neurons in each hidden layer. For my tree models, the maximum number of depths that the trees can grow is the main architectural parameter. The choice column reports this information. For the ANN models, each pair of parathesis encloses an individual model. Starting from the first hidden layer following the open parathesis until the last hidden layer before the closing parathesis, each number in the parathesis represents the number of neurons in a hidden layer. If a pair of parathesis encloses n numbers, it presents an ANN model with n hidden layers. For the tree models, each number in the search choice represents a separate search of a tree model that specifies the number as the maximum depth of the tree.

Model	Hyperparameter	Search Choice
ANN (ANN Rectifier/Tanh)	1 Layers	(8), (16), (32), (64), (128)
	2 Layers	(128,64), (64,32), (32,16), (16,8)
	3 Layers	(128,64,32), (64,32,16), (32,16,8)
	4 Layers	(128,64,32,16), (64,32,16,8)
	5 Layers	(128,64,32,16,8)
	Shrinkage	L1=0.01 or 0
Tree (RF/GBT)	Depth	2,4,6,8

Table 2 Prediction Precision

This table reports the overall in-sample performance and the overall out-of-sample performance. I pull together the training set predictions, including the predictions in the validation set, to generate the statistics for the in-sample predictions below, and I do the same for the out-of-sample predictions. Panel A reports the model performance from the classifiers, and Panel B reports the out-of-sample precision of the aggregated predictions. Panel C reports the machine learning regressions with the exact same parameters and hyperparameters for a head-to-head comparison with their counterpart classifiers. The decile predictions from the regression models are based on the decile sort of predicted returns (Gu et al. 2020). The two panels are organized in the same way. Column 1 indicates whether the performance is based on in-sample (IS) or out-of-sample (OOS) evaluation. Column 2 reports the precision of the prediction. Columns 3 and 4 report the 5% and 95% bounds of the precision. Column 5 and 6 reports the binomial test results against the naïve classifier's precision. RF indicates random forest, and GBT indicates gradient boosting tree. Aggregation indicates the aggregated predictions based on all the classifiers.

Panel A: Classification Prediction Precision

	(1)	(2)	(3)	(4)	(5)	(6)
	Data Set	Precision	5% Bound	95% Bound	Naïve Classifier Precision	Binomial Test P Value
ANN Rectifier	IS	0.157	0.156	0.157	0.100	0.000
	OOS	0.155	0.155	0.155	0.101	0.000
ANN Tanh	IS	0.154	0.154	0.154	0.100	0.000
	OOS	0.154	0.154	0.155	0.101	0.000
RF	IS	0.179	0.179	0.179	0.100	0.000
	OOS	0.160	0.159	0.160	0.101	0.000
GBT	IS	0.172	0.172	0.172	0.100	0.000
	OOS	0.159	0.159	0.159	0.101	0.000

Panel B: Out-of-Sample Aggregated Prediction Precision

Aggregated	OOS	0.161	0.161	0.162	0.101	0.000

Panel C: Out-of-Sample Regression Prediction Precision

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Data	Precision	5% Bound	95% Bound	Naïve Classifier Precision	Binomial Test P Value
ANN Rectifier	OOS	0.126	0.126	0.127	0.101	0.000
ANN Tanh	OOS	0.129	0.128	0.129	0.101	0.000
RF	OOS	0.124	0.123	0.124	0.101	0.000
GBT	OOS	0.120	0.120	0.121	0.101	0.000

Table 3 Portfolio Performance

This table reports the economic performance of the portfolios constructed based on the aggregated predictions from the individual classifiers. The statistics are based on the out-of-sample period covering 198301:202112. The decile portfolios are sorted based on the predicted deciles monthly, which are the deciles with the highest predicted probabilities. The column "market" reports the performance of the buy-and-hold strategy using all common stocks in the three major exchanges. The cumulative returns are in decimal unit representing gross returns in the sample period. $\alpha's$ are for the corresponding factor models, e.g., CAPM or Fama-French 3 Factor model. The t statistics for the $\alpha's$ are Newey-West t statistics of lag 6. The performance statistics are based on excess return adjusted with risk-free rate, i.e., 30-day US treasury bill. I report annualized Sharpe ratios. Turnover is the average total percentage of holding changes in absolute value. Max drawdown is the max difference between current price and the most recent price peak in percentage across all months in my sample period. Panel A reports the equal-weighted portfolio performance based on the classification, while Panel B reports the value-weighted portfolio performance based on classification. Panel A reports the equal-weighted portfolio performance based on the overlapped portfolio allocation from both classification and machine learning regression, while Panel B reports the value-weighted portfolio performance based on the overlapped portfolio allocation from both classification and machine learning regression. A robustness check of the portfolio performance using only the stocks above the median market capitalization of the market is reported in the Appendix Table A6. The benchmark portfolio performance from machine learning regressions is reported in the Appendix Table A7.

Panel A: Classification Equal-weighted Decile Portfolios

Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.009	-0.003	0.002	0.002	0.005	0.004	0.008	0.011	0.013	0.015	0.023	0.023
CAPM Alpha	0.000	-0.014	-0.007	-0.006	-0.003	-0.001	0.003	0.005	0.005	0.006	0.014	0.025
	(0.049)	(-4.229)	(-3.969)	(-2.911)	(-1.495)	(-0.535)	(1.645)	(3.271)	(2.677)	(3.013)	(4.258)	(10.406)
FF3F Alpha	0.000	-0.013	-0.007	-0.006	-0.004	-0.002	0.002	0.004	0.004	0.006	0.014	0.024
	(0.340)	(-5.333)	(-7.168)	(-4.445)	(-3.458)	(-1.295)	(1.882)	(6.342)	(5.111)	(6.524)	(6.418)	(12.036)
FF5F Alpha	0.002	-0.007	-0.006	-0.006	-0.005	-0.003	0.000	0.003	0.003	0.006	0.017	0.021
	(1.513)	(-3.350)	(-5.439)	(-3.902)	(-4.064)	(-1.818)	(0.170)	(5.254)	(3.877)	(6.274)	(6.643)	(12.483)
Standard Deviation	0.058	0.092	0.066	0.057	0.052	0.036	0.037	0.042	0.053	0.063	0.080	0.029
Sharpe Ratio	0.515	-0.102	0.131	0.144	0.310	0.385	0.766	0.940	0.860	0.848	1.023	2.729
Turnover	0.105	0.163	0.102	0.086	0.074	0.063	0.053	0.060	0.074	0.093	0.142	0.153
Max Drawdown	-0.607	-0.904	-0.666	-0.666	-0.676	-0.663	-0.468	-0.476	-0.539	-0.543	-0.579	-0.154
Mean N	5342	1168	474	122	154	270	930	558	493	294	879	2047

Table 3 (Continues)

Panel B: Classification Value-weighted Decile Portfolios												
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.008	-0.002	0.003	0.007	0.005	0.005	0.007	0.009	0.009	0.015	0.014	0.013
CAPM Alpha	0.000	-0.015	-0.009	-0.003	-0.004	-0.002	0.001	0.002	0.001	0.004	0.002	0.014
	(-1.674)	(-5.016)	(-4.172)	(-1.767)	(-1.803)	(-0.928)	(0.595)	(2.834)	(0.743)	(1.792)	(0.699)	(5.701)
FF3F Alpha	0.000	-0.013	-0.007	-0.003	-0.005	-0.003	0.000	0.001	0.001	0.005	0.004	0.014
	(-1.738	(-5.799)	(-4.537)	(-1.609)	(-3.008)	(-2.180)	(-0.276)	(2.412)	(1.030)	(3.059)	(1.661)	(6.492)
FF5F Alpha	0.000	-0.006	-0.004	-0.001	-0.006	-0.004	-0.002	0.001	0.001	0.006	0.008	0.011
	(-1.003)	(-3.381)	(-2.995)	(-0.741)	(-3.732)	(-3.327)	(-3.576)	(1.126)	(0.733)	(4.043)	(3.380)	(4.809)
Standard Deviation	0.045	0.099	0.078	0.070	0.059	0.047	0.041	0.044	0.054	0.073	0.088	0.044
Sharpe Ratio	0.583	-0.058	0.135	0.336	0.280	0.379	0.602	0.692	0.607	0.691	0.558	1.011
Turnover	0.057	0.132	0.096	0.070	0.064	0.051	0.047	0.048	0.065	0.087	0.118	0.125
Max Drawdown	-0.527	-0.958	-0.824	-0.753	-0.720	-0.625	-0.502	-0.509	-0.616	-0.559	-0.702	-0.414
Mean N	5342	1168	474	122	154	270	930	558	493	294	879	2047

Table 3 (Continues)

Panel C: Classification + Regression Equal-weighted Decile Portfolios												
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.009	-0.013	0.001	0.005	0.006	0.002	0.008	0.011	0.014	0.016	0.039	0.050
CAPM Alpha	0.000	-0.025	-0.009	-0.003	-0.003	-0.003	0.002	0.005	0.006	0.007	0.029	0.051
	(0.049)	(-7.155)	(-4.691)	(-1.342)	(-1.075)	(-1.494)	(1.208)	(2.942)	(3.444)	(2.799)	(6.604)	(12.509)
FF3F Alpha	0.000	-0.023	-0.009	-0.004	-0.003	-0.004	0.001	0.004	0.006	0.007	0.030	0.050
	(0.340)	(-8.985)	(-5.811)	(-1.839)	(-2.195)	(-2.595)	(1.083)	(4.452)	(5.452)	(4.146)	(8.423)	(13.216)
FF5F Alpha	0.002	-0.017	-0.008	-0.004	-0.005	-0.005	-0.001	0.002	0.004	0.007	0.034	0.048
	(1.513)	(-8.146)	(-4.446)	(-1.672)	(-2.915)	(-3.243)	(-0.563)	(2.985)	(4.022)	(4.085)	(8.063)	(12.145)
Standard Deviation	0.058	0.092	0.068	0.067	0.063	0.041	0.039	0.043	0.052	0.065	0.096	0.051
Sharpe Ratio	0.515	-0.501	0.026	0.240	0.315	0.201	0.697	0.895	0.929	0.850	1.406	3.372
Turnover	0.105	0.165	0.100	0.084	0.069	0.059	0.051	0.058	0.071	0.089	0.166	0.166
Max Drawdown	-0.607	-0.957	-0.792	-0.665	-0.771	-0.601	-0.441	-0.493	-0.501	-0.638	-0.499	-0.080
Mean N	5342	413	114	25	30	39	149	97	90	56	326	739

Table 3 (Continues)

Panel D: Classification + Regression Value-weighted Decile Portfolios												
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.008	-0.008	0.002	0.009	0.007	0.002	0.007	0.009	0.010	0.016	0.024	0.029
CAPM Alpha	0.000	-0.022	-0.010	0.000	-0.002	-0.004	0.000	0.002	0.002	0.007	0.012	0.031
	(-1.674)	(-6.670)	(-3.778)	(-0.092)	(-0.605)	(-1.985)	(0.335)	(2.090)	(1.238)	(2.409)	(3.452)	(6.770)
FF3F Alpha	0.000	-0.020	-0.008	0.000	-0.002	-0.005	0.000	0.002	0.002	0.007	0.013	0.030
	(-1.738)	(-7.689)	(-3.758)	(-0.083)	(-1.201)	(-2.729)	(-0.286)	(1.672)	(1.411)	(3.194)	(4.926)	(8.262)
FF5F Alpha	0.000	-0.012	-0.006	0.002	-0.004	-0.006	-0.002	0.000	0.001	0.008	0.015	0.025
	(-1.003)	(-5.043)	(-2.554)	(0.735)	(-2.395)	(-3.119)	(-2.173)	(0.190)	(0.695)	(3.674)	(4.937)	(7.492)
Standard Deviation	0.045	0.103	0.079	0.083	0.072	0.053	0.042	0.048	0.056	0.079	0.093	0.068
Sharpe Ratio	0.583	-0.285	0.069	0.367	0.345	0.161	0.553	0.663	0.618	0.695	0.890	1.486
Turnover	0.057	0.135	0.092	0.063	0.057	0.046	0.044	0.046	0.058	0.080	0.134	0.135
Max Drawdown	-0.527	-0.976	-0.850	-0.765	-0.770	-0.791	-0.427	-0.459	-0.542	-0.609	-0.705	-0.391
Mean N	5342	413	114	25	30	39	149	97	90	56	326	739

Table 4 Out-of-Sample Prediction Confusion Matrices

This table reports the out-of-sample prediction confusion matrix. Panel reports the machines' allocation of number of observations based on the aggregated predictions from all classifiers. The first column indicates the predicted decile, while the first row indicates the realized decile. For example, in the table cell of predicted decile 1 and realized decile 1, the aggregated predictions include 122,627 observations. The row summation of these numbers reflects the resources spent on the deciles by the machines. Panel B reports the scaled version of Panel A by the number of observations in the true class, while Panel C reports the scaled version of Panel A by the number of observations in the entire out-of-sample testing period. The colored blocks indicate the correct predictions. For example, in Panel B, the number 12% on the diagonal means the classifiers correctly predict 12% of real decile 2 observations. In Panel C, the 5% means that out of the entire sample, the classifiers correctly predict 5% of the observations from real decile 1. The summation of the diagonal percentages in Panel C sum up to the total precision of the aggregated predictions ensembled from the individual classifiers.

Panel A: Out-of-sample Prediction Confusion Matrix

$\widehat{d_t}$	d_t: 1	2	3	4	5	6	7	8	9	10
1	122627	73891	49420	35848	32758	29953	31127	36050	47802	87382
2	22132	28897	25348	21123	19137	18465	18843	21039	24204	22509
3	4115	6348	6601	6050	5797	5747	5807	5966	6011	4589
4	3526	7059	8525	8481	8137	8271	8137	8117	7472	4545
5	4010	9549	13739	15869	17885	17476	16173	14466	11189	5945
6	7641	25723	44113	56564	63520	66562	63255	54990	38509	14467
7	6507	17053	25970	31158	34405	36774	36581	33948	26666	12093
8	9745	19485	23467	24557	25248	26727	28121	29081	27974	16348
9	10044	14235	14036	13144	12413	12818	13569	15572	17757	14142
10	59041	47375	37785	30769	31508	29023	30236	34215	42804	68640

Panel B: Out-of-sample Prediction Confusion Matrix
(Scaled by Total Number of Observations in the True Class)

$\widehat{d_t}$	d_t: 1	2	3	4	5	6	7	8	9	10
1	49%	30%	20%	15%	13%	12%	12%	14%	19%	35%
2	9%	12%	10%	9%	8%	7%	7%	8%	10%	9%
3	2%	3%	3%	2%	2%	2%	2%	2%	2%	2%
4	1%	3%	3%	3%	3%	3%	3%	3%	3%	2%
5	2%	4%	6%	7%	7%	7%	6%	6%	4%	2%
6	3%	10%	18%	23%	25%	26%	25%	22%	15%	6%
7	3%	7%	10%	13%	14%	15%	15%	13%	11%	5%
8	4%	8%	9%	10%	10%	11%	11%	11%	11%	7%
9	4%	6%	6%	5%	5%	5%	5%	6%	7%	6%
10	24%	19%	15%	13%	13%	12%	12%	14%	17%	27%

Panel C: Out-of-sample Prediction Confusion Matrix
(Scaled by Total Number of Observations in the Entire Sample)

$\widehat{d_t}$	d_t: 1	2	3	4	5	6	7	8	9	10
1	5%	3%	2%	1%	1%	1%	1%	1%	2%	3%
2	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
5	0%	0%	1%	1%	1%	1%	1%	1%	0%	0%
6	0%	1%	2%	2%	3%	3%	3%	2%	2%	1%
7	0%	1%	1%	1%	1%	1%	1%	1%	1%	0%
8	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%
9	0%	1%	1%	1%	0%	1%	1%	1%	1%	1%
10	2%	2%	2%	1%	1%	1%	1%	1%	2%	3%

Table 5 Out-of-Sample Prediction Precision by Return Decile Transition

This table reports my analysis of machine learning return predictability during my out-of-sample period by transitions. Panel A reports the unconditional transition probabilities. Probabilities deviating from the random distribution probability 1%, regardless of the direction, indicate that the transition has higher certainty. Panel B reports the prediction precision from the aggregated model by return decile transitions. For example, my prediction managed to achieve a precision of 37.4% for the return transition from decile 1 to decile 1. Panel C reports the information shortage created based on the predicted probabilities by return decile transitions.

Panel A: Transition Matrix

d_{t-1} \ d_t	1	2	3	4	5	6	7	8	9	10
1	1.7%	1.1%	0.8%	0.7%	0.7%	0.7%	0.7%	0.8%	1.0%	1.8%
2	1.1%	1.1%	1.0%	0.9%	0.9%	0.9%	0.9%	1.0%	1.1%	1.2%
3	0.9%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	0.9%
4	0.7%	0.9%	1.0%	1.1%	1.1%	1.1%	1.1%	1.1%	1.0%	0.8%
5	0.7%	0.9%	1.0%	1.1%	1.1%	1.2%	1.2%	1.1%	1.0%	0.8%
6	0.7%	0.9%	1.0%	1.1%	1.2%	1.2%	1.2%	1.1%	1.0%	0.8%
7	0.7%	0.9%	1.0%	1.1%	1.1%	1.2%	1.2%	1.1%	1.0%	0.8%
8	0.8%	1.0%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.0%	0.8%
9	1.0%	1.1%	1.1%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	0.9%
10	1.7%	1.2%	1.0%	0.8%	0.8%	0.7%	0.7%	0.8%	0.9%	1.3%

Panel B: Precision

d_{t-1} \ d_t	1	2	3	4	5	6	7	8	9	10
1	37.4%	1.3%	0.1%	0.1%	0.8%	1.9%	4.2%	7.0%	7.8%	65.2%
2	44.4%	5.3%	0.8%	0.8%	4.0%	9.6%	12.8%	16.8%	13.2%	37.6%
3	41.0%	9.3%	1.8%	1.9%	6.2%	20.8%	18.3%	17.1%	10.4%	26.1%
4	39.2%	10.3%	2.2%	2.4%	8.2%	31.0%	18.6%	15.7%	8.6%	20.1%
5	41.0%	10.4%	2.4%	3.1%	8.1%	35.6%	18.3%	12.4%	6.6%	19.1%
6	42.4%	11.4%	2.6%	3.4%	8.3%	35.1%	18.4%	12.0%	5.7%	18.0%
7	40.7%	13.8%	4.2%	4.3%	8.6%	36.7%	18.2%	11.4%	5.1%	15.4%
8	44.5%	15.8%	4.4%	5.0%	9.1%	35.9%	12.9%	9.6%	5.0%	13.6%
9	53.3%	22.2%	4.5%	7.6%	9.4%	25.0%	10.0%	7.0%	4.6%	12.6%
10	82.5%	15.6%	2.7%	4.7%	4.4%	8.6%	4.6%	3.0%	2.9%	7.5%

Panel C: Information Shortage

d_{t-1} \ d_t	1	2	3	4	5	6	7	8	9	10
1	3.16	3.20	3.22	3.22	3.22	3.23	3.23	3.23	3.22	3.17
2	3.23	3.25	3.26	3.26	3.25	3.25	3.26	3.26	3.26	3.25
3	3.24	3.26	3.25	3.25	3.24	3.24	3.24	3.25	3.26	3.26
4	3.24	3.26	3.25	3.24	3.23	3.23	3.23	3.24	3.25	3.26
5	3.24	3.25	3.24	3.23	3.23	3.23	3.23	3.23	3.25	3.25
6	3.24	3.25	3.24	3.23	3.23	3.23	3.23	3.23	3.25	3.25
7	3.24	3.25	3.24	3.23	3.23	3.23	3.23	3.24	3.25	3.25
8	3.23	3.25	3.25	3.24	3.24	3.24	3.24	3.24	3.25	3.25
9	3.22	3.25	3.25	3.25	3.25	3.25	3.25	3.26	3.26	3.24
10	3.11	3.19	3.22	3.23	3.23	3.23	3.23	3.22	3.21	3.16

Table 6 Predictability and Firm Characteristics

This table reports the Fama-MacBeth regression results in the investigation of the relation between prediction success and firm characteristics. Prediction success is a dummy variable of value 1 if the prediction is correct and 0 otherwise. The table reports the results for the regression $Prediction\ Success_{i,t} = \gamma_0 + Characteristics\Gamma + \varepsilon_{i,t}$, where the prediction precision is based on the aggregated predictions from the individual classifiers. I report for only variables that are statistically significant in the linear regressions, and I split the table into the positive column and the negative column, where the positive column reports results for variables that are positively related to the prediction precision and the negative column reports for the variables that are negatively related to the prediction precision. "FM t" represents Fama-MacBeth t statistics with Newey-West correction.

	Positive Relation			Negative Relation	
	Coefficient	FM t		Coefficient	FM t
chmom	0.007	10.820	pchsale_pchinvt	-0.001	-2.053
baspread	0.021	9.020	depr	-0.001	-2.127
age	0.003	8.880	cfp	-0.001	-2.377
mve_ia	0.004	8.270	roic	-0.001	-2.457
turn	0.008	7.625	sue	-0.001	-2.590
idiovol	0.008	7.273	currat	-0.003	-2.761
betasq	0.015	6.880	cashdebt	-0.001	-2.890
mom12m	0.005	6.551	securedind1	-0.002	-3.260
ms	0.003	6.021	gma	-0.002	-3.336
dy	0.003	4.936	salecash	-0.001	-3.488
retvol	0.011	4.766	rd_mve	-0.002	-3.591
pctacc	0.001	4.717	secured	-0.002	-3.710
mom1m	0.005	4.475	divi0	-0.018	-3.935
agr	0.002	4.371	rsup	-0.001	-4.252
nincr	0.001	4.365	divi1	-0.022	-4.457
rd0	0.003	3.295	roeq	-0.002	-4.470
chtx	0.001	2.689	mom36m	-0.002	-4.582
absacc	0.001	2.548	nanalyst	-0.003	-4.607
pchcapx_ia	0.001	2.376	fgr5yr	-0.003	-4.862
sgr	0.001	2.354	sp	-0.002	-5.322
saleinv	0.001	1.800	ep	-0.004	-5.545
pchdepr	0.001	1.743	disp	-0.003	-5.906
lev	0.001	1.726	zerotrade	-0.003	-5.988
ill	0.001	1.670	std_turn	-0.004	-6.390
			cash	-0.004	-7.319
			sfe	-0.004	-7.796
			bm	-0.003	-8.394
			beta	-0.017	-8.960
			roaq	-0.007	-10.570
			mom6m	-0.011	-11.812
Constant	0.175	11.928			
102 Characteristics	Y				
Industry FE	Y				
Past Return Decile	Y				
Mean N	5362				
Mean Adj. R-square	0.028				

Table 7 Predictability and Information Shortage

This table reports the results from the Fama-MacBeth regression examining the relation between the prediction success at the stock-month level and the information shortage for all individual classifiers and the aggregated predictions. The information shortage is computed based on the predicted probabilities using binary information entropy, which measures the expected minimum number of binary questions a forecaster has to answer correctly before reaching 100% correct predictions. In the regressions, I control for the 102 firm characteristics, industry fixed effects, and past return decile. The t statistics are Newey-West t statistics of lag 12 from Fama-MacBeth regression.

Panel A: Summary Stats of Prediction Success and Information Shortage

	Mean	Standard Deviation	Min	Median	Max
Success	0.161	0.368	0.000	0.000	1.000
Info. Shortage	3.234	0.088	2.249	3.258	3.322

Panel B: Prediction Success and Information Shortage

Dependent Variable	Prediction Success				
	(1)	(2)	(3)	(4)	(5)
Models	Aggregate	Ann Tanh	Ann Rectifier	GBT	RF
Info. Shortage$_t$	-0.440	-0.274	-0.334	-0.324	-0.588
	(-24.905)	(-37.920)	(-20.461)	(-31.996)	(-15.175)
Constant	1.571	1.087	1.221	1.255	2.072
	(28.300)	(22.682)	(23.742)	(21.267)	(16.778)
102 Characteristics	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
Past Return Decile	Y	Y	Y	Y	Y
Mean N	5362	5362	5362	5362	5362
Mean Adj. R^2	0.032	0.034	0.034	0.030	0.033

Table 8 Portfolio Performance Conditional on Information Shortage

This table reports the economic performance of the *conditional* portfolios constructed based on the aggregated predictions from the individual classifiers using only the stocks in the highest decile of past 12-month precision and the stocks in the lowest decile of the information shortage. I report the results of the equal-weighted and the value-weighted portfolios in Panel A and B, respectively. The statistics are based on the out-of-sample period covering 198301:202112. The decile portfolios are sorted based on the predicted deciles monthly, which are the deciles with the highest predicted probabilities. The cumulative returns are in decimal unit representing gross returns in the sample period. $\alpha's$ are for the corresponding factor models, e.g., CAPM or Fama-French 3 Factor model. The t statistics for the $\alpha's$ are Newey-West t statistics of lag 12. The performance statistics are based on excess return adjusted with risk-free rate, i.e., 30-day US treasury bill. I report annualized Sharpe ratios. Turnover is the average total percentage of holding changes in absolute value. Max drawdown is the max difference between current price and the most recent price peak in percentage across all months in my sample period. Panel A and C report the equal-weighted portfolio performance, while Panel B and D report the value-weighted portfolio performance.

Panel A: Classification Equal-weighted Decile Portfolios Conditional on the Lowest Prediction Information Shortage

Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.009	-0.014	-0.002	0.005	0.005	0.005	0.010	0.012	0.014	0.019	0.067	0.078
CAPM Alpha	0.000	-0.024	-0.012	-0.004	-0.004	0.000	0.004	0.005	0.006	0.010	0.057	0.078
	(0.340)	(-5.190)	(-5.225)	(-2.025)	(-2.020)	(-0.134)	(3.792)	(5.125)	(4.651)	(3.886)	(7.955)	(12.474)
FF3F Alpha	0.002	-0.015	-0.011	-0.003	-0.005	-0.001	0.003	0.004	0.004	0.009	0.064	0.078
	(1.513)	(-3.164)	(-4.117)	(-1.751)	(-2.339)	(-0.758)	(2.721)	(4.337)	(3.838)	(3.411)	(7.561)	(12.334)
FF5F Alpha	0.000	-0.026	-0.011	-0.003	-0.005	-0.002	0.000	0.002	0.000	0.007	0.032	0.059
	(-1.003)	(-4.693)	(-3.702)	(-0.895)	(-2.522)	(-0.896)	(0.256)	(2.091)	(0.074)	(2.115)	(4.539)	(9.133)
Standard Deviation	0.058	0.133	0.077	0.067	0.058	0.035	0.033	0.038	0.049	0.067	0.153	0.090
Sharpe Ratio	0.515	-0.370	-0.111	0.232	0.303	0.485	1.010	1.054	1.020	0.967	1.512	2.980
Turnover	0.105	0.239	0.111	0.067	0.055	0.042	0.038	0.039	0.055	0.080	0.239	0.239
Max Drawdown	-0.607	-0.967	-0.840	-0.579	-0.751	-0.614	-0.384	-0.374	-0.516	-0.637	-0.581	-0.295
Mean N	5342	117	48	13	16	27	93	56	50	30	88	206

Table 8 (Continues)

Panel B: Classification Value-weighted Decile Portfolios Conditional on the Lowest Prediction Information Shortage												
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.008	-0.026	-0.003	0.004	0.005	0.007	0.008	0.009	0.010	0.018	0.037	0.060
CAPM Alpha	0.000	-0.040	-0.014	-0.005	-0.003	0.001	0.003	0.004	0.003	0.009	0.022	0.059
	(-1.674)	(-6.697)	(-4.415)	(-1.533)	(-1.444)	(0.573)	(1.810)	(2.847)	(1.407)	(2.211)	(3.408)	(8.917)
FF3F Alpha	0.000	-0.038	-0.013	-0.005	-0.004	0.000	0.002	0.003	0.002	0.009	0.024	0.059
	(-1.738)	(-7.742)	(-4.606)	(-1.572)	(-2.234)	(0.041)	(1.416)	(2.464)	(1.145)	(2.375)	(4.266)	(9.133)
FF5F Alpha	0.000	-0.026	-0.011	-0.003	-0.005	-0.002	0.000	0.002	0.000	0.007	0.032	0.055
	(-1.003)	(-4.693)	(-3.702)	(-0.895)	(-2.522)	(-0.896)	(0.256)	(2.091)	(0.074)	(2.115)	(4.539)	(7.666)
Standard Deviation	0.045	0.150	0.091	0.079	0.062	0.048	0.037	0.041	0.052	0.078	0.166	0.131
Sharpe Ratio	0.583	-0.601	-0.097	0.192	0.266	0.484	0.712	0.784	0.645	0.784	0.774	1.582
Turnover	0.057	0.205	0.098	0.053	0.049	0.034	0.036	0.036	0.048	0.069	0.207	0.206
Max Drawdown	-0.527	-0.997	-0.881	-0.662	-0.705	-0.611	-0.374	-0.409	-0.693	-0.630	-0.666	-0.878
Mean N	5342	117	48	13	16	27	93	56	50	30	88	206

Table 9 Predictability and Economic Uncertainty

This table report the results of test on the relation between economic uncertainty and predictability. The tests include the war risk index from Hirshleifer et al. (2023), geopolitical risk index from Caldara et al. (2022), the US economic policy uncertainty index from Baker et al. (2016), and Jurado et al.'s (2015) macroeconomic uncertainty index, financial uncertainty index and real uncertainty index. The tests regress the market level prediction precision in percent on the uncertainty indices during the period 198301:202112. The t statistics are Newey-West t statistics with lag of 12.

Panel A: Summary Stats of Uncertainty Indices

	Mean	Standard Deviation	Min	Median	Max
War	0.103	0.021	0.042	0.102	0.177
Geopolitical Risk	82.571	31.112	28.031	77.604	303.585
Economic Policy Uncertainty	99.037	40.321	37.266	89.000	271.832
Macroeconomic Uncertainty	0.627	0.080	0.530	0.617	1.088
Financial Uncertainty	0.885	0.172	0.633	0.841	1.553
Real Uncertainty	0.597	0.052	0.511	0.590	0.878

Panel B: Prediction Precision and Economic Uncertainty

Dependent	Prediction Precision						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
War	0.011						-0.008
	(0.346)						(-0.230)
Geopolitical Risk		0.000					0.000
		(-0.328)					(0.117)
Economic Policy Uncertainty			0.000				0.000
			(-1.217)				(0.217)
Macroeconomic Uncertainty				-0.040			-0.039
				(-6.414)			(-2.330)
Financial Uncertainty					-0.018		-0.011
					(-3.323)		(-1.810)
Real Uncertainty						-0.048	0.023
						(-3.483)	(0.863)
(Intercept)	0.160	0.161	0.163	0.186	0.177	0.190	0.181
	(43.468)	(80.619)	(79.695)	(50.740)	(40.718)	(23.111)	(24.872)
N	442	442	418	442	442	442	418
Adj. R^2	-0.002	-0.002	0.005	0.076	0.072	0.047	0.087

Table 10 Predictability, Information Shortage, and Stock Returns

This table reports the Fama-MacBeth regression results investigating the influence of past predictability and the prediction information shortage on the monthly stock returns. The information shortage is computed based on the predicted probabilities using binary information entropy, which measures the expected minimum number of binary questions a forecaster has to answer correctly before reaching 100% correct predictions. In the regressions, I control for the 102 firm characteristics, industry fixed effects, and past return decile. The t statistics are Newey-West t statistics of lag 12 from Fama-MacBeth regression.

Dependent		Returns	
	(1)	(2)	(3)
Prediction Success$_{t-1}$	-0.001		-0.002
	(-2.621)		(-4.126)
Info. Shortage$_t$		-0.039	-0.041
		(-4.743)	(-5.122)
Constant	0.058	0.182	0.187
	(1.241)	(3.346)	(3.452)
102 Characteristics	Y	Y	Y
Industry FE	Y	Y	Y
Past Return Decile	Y	Y	Y
Mean N	5362	5362	5362
Mean Adj. R^2	0.094	0.094	0.095

Table 11 Predictability and Pricing Error Fluctuations

This table below reports the relation between the prediction success and the change in pricing error as proxied by the α from Fama-French 3 Factor model for each stock with a 60-month rolling window. The analysis is based on Fama-MacBeth regression performed for the out-of-sample period 198301:202112. The dependent variable is the individual stock-month prediction success, which is an indicator variable equal to 1 if the prediction is correct. The main independent variables area $\Delta\alpha$ and absolute value of $\Delta\alpha$. The t statistics are Newey-West t statistics with a lag of 12.

Dependent Variable	Prediction Success	
	(1)	(2)
$\boldsymbol{\Delta\alpha_{t-1:t}}$	-2.602	
	(-4.013)	
$\boldsymbol{\lvert\Delta\alpha_{t-1:t}\rvert}$		12.603
		(20.588)
α_{t-1}	-0.137	-0.107
	(-3.487)	(-3.956)
Constant	0.186	0.154
	(10.940)	(6.594)
102 Characteristics	Y	Y
Industry FE	Y	Y
Past Return Decile	Y	Y
Mean N	4352	4352
Mean Adj. R^2	0.028	0.035

Table 12 Mimicking Portfolio Spanning Tests

This table below reports the spanning tests for the mimicking portfolio based on the market level prediction precision calculated as the percent of successful predictions for a month. Following the literature, the prediction precision is first regressed on its lagged value such that a shock factor is obtained through the residuals. The contemporaneous prediction precision innovation is then mapped into the return space using all 170 basis portfolio returns, including single sort, double sort, and triple sort portfolios for characteristics included in the factors of Fama-French 6 factor model. The spanning tests regress the mimicking portfolio returns on the common factors, including Fama-French 5 factors, Fama-French 6 factors, q4 factors, and q5 factors. The tests cover the entire OOS period 198302:202112. The t statistics are Newey-West t statistics with lag of 12.

Dependent	Prediction Precision Mimicking Portfolio Return			
	(1)	(2)	(3)	(4)
$\boldsymbol{\alpha}$	-0.001	-0.002	-0.002	-0.002
	(-2.912)	(-4.012)	(-3.734)	(-3.391)
Mkt-RF	0.060	0.074		
	(6.020)	(10.018)		
SMB	0.025	0.021		
	(1.580)	(1.451)		
HML	0.056	0.097		
	(3.221)	(6.921)		
RMW	0.187	0.174		
	(8.310)	(9.684)		
CMA	-0.026	-0.052		
	(-0.845)	(-2.201)		
MOM		0.070		
		(4.410)		
R_MKT			0.070	0.068
			(6.408)	(6.632)
R_ME			0.021	0.018
			(0.967)	(0.765)
R_IA			0.057	0.057
			(2.201)	(2.122)
R_ROE			0.187	0.196
			(9.231)	(8.519)
R_EG				-0.025
				(-0.806)
N	467	467	467	467
Adj. R^2	0.313	0.422	0.329	0.329

Appendix

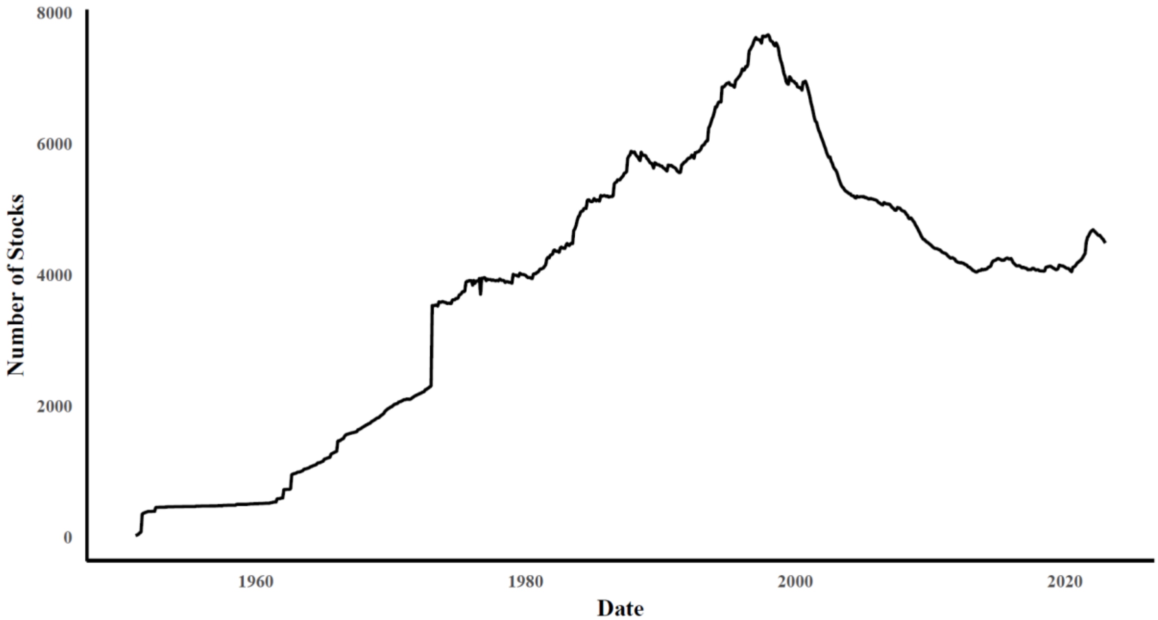

Figure A1 Number of Stocks

This figure demonstrates the monthly number of stocks from CRSP-Compustat database during the period 195101:202112. To minimize the look-ahead bias, the sample adopted in this paper covers 196301:202112 with 4,421 stocks on average per month, since the Compustat database was first released in 1963. The full sample includes 24,136 distinct stocks. The out-of-sample testing period starts in 198301 with 5,362 stocks on average per month. The out-of-sample period includes 22,242 distinct stocks.

Table A1 Modeling Windows

This table reports the specification of the modeling windows. The models are updated every ten years in this paper. The starting date of the training process is January 1962. Every update will train the model using the training data set for in-sample fitting. The fitted models will make predictions for the validation set, and the best combination of architecture and hyperparameters is chosen to make the out-of-sample predictions in the testing periods.

Window	Train Start	Train End	Validation End	Test End
1	01/31/1962	12/31/1977	12/31/1982	12/31/1992
2	01/31/1962	12/31/1987	12/31/1992	12/31/2002
3	01/31/1962	12/31/1997	12/31/2002	12/31/2012
4	01/31/1962	12/31/2007	12/31/2012	12/31/2021

Table A2 Additional Optimization Choices

I conduct a grid search for the best parameters and hyperparameters in training and validation data sets. I train all the sub-models first in the training data set. Then, I select the best performing model in the validation data set for the hyperparameter values. Panel A details the additional optimization choice of my grid search. Panel B reports the selected modeling parameters and hyperparameters after training and hyperparameter tuning.

Model	Parameter	Choice
ANN (ANN Rectifier/Tanh)	Loss Function	Cross entropy for classification/mean squared error for regression
	Learning Rate	Adadelta with rho=0.99 and epsilon=1e-8
	Activation	Rectifier or Tanh for two ANN models separately
	# Epochs	1000
GBT	Loss Function	Cross entropy for classification/mean squared error for regression
	# Trees	1000
	Learning Rate	0.1
RF	Loss Function	Cross entropy for classification/mean squared error for regression
	# Trees	1000

Table A3 Prediction Sample Firm Characteristics and Summary Statistics

The table reports the firm characteristics used in the prediction process and the summary statistics of the firm characteristics following Green et al. (2017). I construct the sample such that the data is CRSP centric, and I attempt to include as many common share stocks listed on three major exchanges (NYSE, AMEX, and NASDAQ) as possible. However, I do not include other securities such as REITS. My data construction avoids issues, including high volatility in the number of stocks from month to month. In my models, I normalize these following predictors monthly. Panel A defines the characteristics following Green et al. (2017). Panel B reports the summary statistics of the raw characteristics.

Panel A: Firm Characteristics

Acronym	Definition
absacc	Absolute value of accrual
acc	Accrual
aeavol	Average daily trading volume change around earnings
age	Firm age
agr	Percentage change in assets
baspread	Bid-ask spread
beta	Market beta
betasq	Market beta squared
bm	Book to Market
bm_ia	Industry adjusted book to market
cash	Cash to asset
cashdebt	Earnings to debt
cashpr	Cash productivity
cfp	Cash to market
cfp_ia	Industry-adjusted cash to market
chatoia	Industry-adjusted sales to assets
chcsho	Annual percentage change in shares outstanding
chempia	Industry-adjusted change in number of employees
chfeps	Change in earnings forecast
chinv	Change in inventory to assets
chmom	Cumulative returns from months t-6 :t-1 minus months t-12:t-7
chnanalyst	Change in number of analyst forecasts
chpmia	Industry-adjusted change in earnings to sales
chtx	Percentage change in total tax
cinvest	Change in capital investment
convind	An indicator equal to 1 if a firm has convertible debt
currat	Current assets to current liabilities
depr	Depreciation to PP&E
disp	Analyst forecast dispersion
divi	An indicator equal to 1 if a firm pays dividend this year but skipped the prior year
divo	An indicator equal to 1 if a firm discontinues the dividend payment this year
dolvol	Dollar value trading volume
dy	Dividend yield
ear	3-day total return around quarterly earnings announcement
egr	Annual percentage change in book value
ep	Earnings to price ratio
fgr5yr	5-year analyst forecast of growth
gma	Novy-Marx (2013) profitability
grcapx	3-year percentage change in capital expenditure
grltnoa	Growth in long-term net operating assets
herf	Sales concentration
hire	Percentage change in number of employees
idiovol	3-year weekly standard deviation of return residuals on equal weighted market returns
ill	Average of daily absolute return over dollar volume
indmom	Equal weighted average industry 12-month returns
invest	Investment to assets
ipo	An indicator equal to 1 if first year in CRSP
lev	Liabilities to market capitalization
lgr	Annual percentage change in liabilities
maxret	Maximum daily return in the past month
mom12m	11-month cumulative returns ending in t-1

Table A3 (Continues)

Panel A: Firm Characteristics	
Acronym	Definition
mom1m	1-month cumulative returns ending in t-1
mom36m	Cumulative returns from months t-36:t-13
mom6m	5-month cumulative returns ending in t-1
ms	Mohanram score of fundamental performance
mve	Market capitalization in t-1
mve_ia	Industry-adjusted market capitalization in fiscal year end
nanalyst	Number of analyst forecasts in I/B/E/S
nincr	Number of consecutive quarters with increasing earnings
operprof	Operating profitability
orgcap	Capitalized SG&A expenses
pchcapx_ia	Industry-adjusted percentage change in capital expenditures
pchcurrat	Percentage change in the ratio of current assets to current liabilities
pchdepr	Percentage change in depreciation
pchgm_pchsale	Percentage change in gross margin minus percentage change in sales
pchquick	Percentage change in quick ratio
pchsale_pchinvt	Annual percentage change in sales minus annual percentage change in inventory
pchsale_pchrect	Annual percentage change in sales minus annual percentage change in receivables
pchsale_pchxsga	Annual percentage change in sales minus annual percentage change in SG&A
pchsaleinv	Percentage change in sales to inventory
pctacc	Accrual in percentage of absolute value of *ib*
pricedelay	The proportion of variation explained by 4 lags of market returns incremental to contemporaneous market return
ps	Fundamental health
quick	(Current assets - inventory)/current liabilities
rd	An indicator equal to 1 if R&D expense to assets increases more than 5%
rd_mve	R&D to fiscal-year-end market capitalization
rd_sale	R&D to sales
realestate	Buildings and capitalized leases to gross PP&E
retvol	Standard deviation of daily returns in t-1
roaq	Quarterly income before extraordinary items to assets
roavol	Standard deviation of 16-quarter income before extraordinary items divided by average quarterly total assets
roeq	Earnings before extraordinary items divided by lagged common shareholders' equity
roic	Annual EBIT minus nonoperating income divided by non-cash enterprise value (*ceq+lt-che*)
rsup	Sales from quarter t minus sales from quarter t-4 divided by fiscal-quarter-end market capitalization
salecash	Annual sales divided by cash and cash equivalents
saleinv	Annual sales divided by total inventory
salerec	Annual sales divided by accounts receivable
secured	Total liability scaled secured debt
securedind	An indicator equal to 1 if a firm has secured debt
sfe	Analysts mean annual earnings forecast for the nearest upcoming fiscal year prior to month of portfolio formation divided by price per share at the fiscal quarter end
sgr	Annual percentage change in sales
sin	An indicator equal to 1 if a firm's industry classification is in smoke or tobacco, beer or alcohol, or gaming
sp	Annual revenue divided by fiscal-year-end market capitalization
std_dolvol	Monthly standard deviation of daily dollar trading volume
std_turn	Monthly standard deviation of daily share turnover
stdacc	16-quarter standard deviation of accruals divided by sales
stdcf	16-quarter standard deviation of cash flows divided by sales
sue	Unexpected earnings
tang	Asset tangibility
tb	Tax income divided by income before extraordinary items
turn	3-month average trading volume ending in month t-1 scaled by the number of shares outstanding in current month
zerotrade	Turnover weighted number of zero trading days for the most recent month

Table A3 (Continues)

Panel B: Summary Statistics of Raw Firm Characteristics					
Variable	Mean	Standard Deviation	Min	Median	Max
absacc	0.098	0.114	0.000	0.066	1.086
acc	-0.023	0.142	-1.039	-0.019	0.582
aeavol	0.853	2.051	-1.000	0.290	21.222
age	15.076	12.893	1.000	11.000	71.000
agr	0.283	1.105	-0.693	0.083	35.398
baspread	0.055	0.069	-0.430	0.036	0.985
beta	1.083	0.651	-1.489	1.014	3.910
betasq	1.602	1.810	0.000	1.032	15.291
bm	0.755	0.726	-2.581	0.585	7.894
bm_ia	23.174	691.727	-2360.690	0.021	16500.928
cash	0.170	0.217	-0.143	0.076	0.980
cashdebt	-0.045	1.670	-382.788	0.127	2.851
cashpr	-0.570	55.119	-656.405	-0.510	594.905
cfp	0.019	0.312	-4.130	0.042	7.626
cfp_ia	12.595	303.092	-310.191	0.016	6795.637
chatoia	-0.005	0.243	-1.380	0.003	1.306
chcsho	0.221	1.005	-0.892	0.008	28.089
chempia	-0.101	0.651	-24.055	-0.061	3.647
chfeps	0.003	0.603	-19.140	0.000	20.950
chinv	0.015	0.059	-0.287	0.001	0.426
chmom	-0.001	0.567	-8.455	-0.006	7.783
chnanalyst	0.026	1.571	-42.000	0.000	38.000
chpmia	0.305	7.505	-93.863	-0.004	111.909
chtx	0.001	0.013	-0.121	0.000	0.145
cinvest	-0.027	6.895	-157.600	-0.002	3390.067
convind	1.130	0.336	1.000	1.000	2.000
currat	3.381	5.994	0.102	1.971	105.898
depr	0.269	0.440	-0.984	0.152	8.147
disp	0.171	0.465	0.000	0.044	12.500
divi	2.006	0.263	1.000	2.000	3.000
divo	1.998	0.246	1.000	2.000	3.000
dolvol	11.129	3.048	-3.060	10.982	19.490
dy	0.018	0.035	-6.122	0.001	0.556
ear	0.003	0.083	-0.458	0.001	0.504
egr	0.215	1.942	-38.569	0.082	43.328
ep	-0.026	0.364	-8.012	0.048	0.683
fgr5yr	16.814	11.617	-74.000	14.830	208.830
gma	0.376	0.389	-1.520	0.313	2.977
grcapx	1.270	4.806	-18.500	0.177	67.915
grltnoa	0.096	0.172	-0.917	0.060	1.256
herf	0.067	0.081	0.003	0.043	1.000
hire	0.091	0.339	-0.700	0.008	3.917
idiovol	0.065	0.037	0.000	0.055	0.266
ill	0.000	0.000	0.000	0.000	0.001
indmom	0.142	0.300	-0.757	0.116	3.102
invest	0.100	0.235	-0.562	0.046	2.990
ipo	1.058	0.234	1.000	1.000	2.000
lev	2.191	4.712	0.000	0.668	73.048
lgr	0.309	1.060	-0.792	0.080	15.515
maxret	0.075	0.072	0.000	0.053	0.846
mom12m	0.129	0.595	-0.972	0.051	11.365

Table A3 (Continues)

Panel B: Summary Statistics of Raw Firm Characteristics					
Variable	Mean	Standard Deviation	Min	Median	Max
mom1m	0.010	0.155	-0.728	0.000	2.000
mom36m	0.315	0.937	-0.986	0.141	14.514
mom6m	0.054	0.368	-0.911	0.020	7.533
ms	3.609	1.688	0.000	4.000	8.000
mve	11.734	2.252	2.357	11.579	18.588
mve_ia	-189.253	7566.268	-26395.790	-364.757	142031.617
nanalyst	4.884	6.657	0.000	2.000	57.000
nincr	0.945	1.299	0.000	1.000	8.000
operprof	0.831	1.603	-10.005	0.615	18.265
orgcap	0.144	0.485	-0.702	0.015	8.223
pchcapx_ia	3.754	54.529	-890.899	-0.561	939.472
pchcurrat	0.194	1.229	-0.915	-0.004	23.397
pchdepr	0.106	0.565	-0.961	0.023	7.789
pchgm_pchsale	-0.096	1.144	-20.502	-0.002	6.174
pchquick	0.243	1.464	-0.938	-0.002	29.768
pchsale_pchinvt	-0.065	0.862	-10.579	0.013	4.163
pchsale_pchrect	-0.061	0.771	-10.015	-0.001	5.431
pchsale_pchxsga	0.029	0.427	-2.897	-0.001	6.642
pchsaleinv	0.154	1.035	-121.036	0.010	30.974
pctacc	-0.647	5.934	-63.600	-0.258	65.444
pricedelay	0.143	0.999	-16.494	0.062	13.838
ps	4.089	1.762	0.000	4.000	9.000
quick	2.667	5.466	0.061	1.294	98.567
rd	2.077	0.367	1.000	2.000	3.000
rd_mve	0.065	0.112	-0.034	0.028	2.228
rd_sale	0.825	6.751	-218.737	0.031	210.899
realestate	0.266	0.200	0.000	0.231	1.000
retvol	0.033	0.026	0.000	0.026	0.262
roaq	-0.009	0.070	-1.047	0.006	0.219
roavol	0.032	0.069	0.000	0.013	1.238
roeq	-0.007	0.196	-4.833	0.022	2.773
roic	-0.128	1.152	-20.737	0.066	1.266
rsup	-0.048	3.987	-2580.272	0.013	6.239
salecash	50.266	161.272	-1230.906	9.833	2942.250
saleinv	26.255	71.165	-106.622	7.549	1203.586
salerec	11.789	50.632	-21796.000	5.918	276.499
secured	0.571	0.517	0.000	0.585	4.013
securedind	1.387	0.487	1.000	1.000	2.000
sfe	-0.596	7.512	-326.471	0.043	4.062
sgr	0.239	0.789	-0.984	0.100	13.743
sin	1.007	0.085	1.000	1.000	2.000
sp	2.222	3.651	-35.942	1.028	55.651
std_dolvol	0.862	0.410	0.000	0.794	3.332
std_turn	4.587	13.885	0.000	1.914	625.712
stdacc	9.588	60.087	0.000	0.141	1138.612
stdcf	17.605	119.120	0.000	0.156	2723.991
sue	-0.006	0.190	-11.824	0.000	3.305
tang	0.541	0.157	0.000	0.550	0.984
tb	-0.118	1.532	-25.942	-0.072	12.172
turn	1.103	2.197	0.000	0.531	76.062
zerotrade	1.369	3.366	0.000	0.000	20.046

Table A4 Selected Models after Training and Hyperparameter Tuning

This table reports the selected hyperparameters for each combination of models and training and validation period. Column "Classification" reports the parameters for the classification models of the corresponding modeling architecture, while column "Regression" reports the parameter for the regression models of the corresponding modeling architecture.

Model	Training Window		Validation Window		Classification		Regression	
	Train Start	Train End	Validation Start	Validation End	Hidden	l1	Hidden	l1
ANN (Tanh)	01/31/1962	12/31/1977	01/01/1978	12/31/1982	16	0	(64, 32, 16)	0
	01/31/1962	12/31/1987	01/01/1988	12/31/1992	16	0	(64, 32, 16, 8)	0
	01/31/1962	12/31/1997	01/01/1998	12/31/2002	(32, 16)	0	(16, 8)	0
	01/31/1962	12/31/2007	01/01/2008	12/31/2012	16	0	8	0
	Train Start	Train End	Validation Start	Validation End	Hidden	l1	Hidden	l1
ANN (Rectifier)	01/31/1962	12/31/1977	01/01/1978	12/31/1982	(128, 64, 32, 16, 8)	0	(32, 16)	0
	01/31/1962	12/31/1987	01/01/1988	12/31/1992	(128, 64, 32)	0	8	0
	01/31/1962	12/31/1997	01/01/1998	12/31/2002	(128, 64, 32)	0	(128, 64, 32)	0
	01/31/1962	12/31/2007	01/01/2008	12/31/2012	(128, 64, 32)	0	(64, 32, 16, 8)	0
	Train Start	Train End	Validation Start	Validation End	Max Depth		Max Depth	
GBT	01/31/1962	12/31/1977	01/01/1978	12/31/1982	2		4	
	01/31/1962	12/31/1987	01/01/1988	12/31/1992	4		4	
	01/31/1962	12/31/1997	01/01/1998	12/31/2002	4		2	
	01/31/1962	12/31/2007	01/01/2008	12/31/2012	4		4	
	Train Start	Train End	Validation Start	Validation End	Max Depth		Max Depth	
RF	01/31/1962	12/31/1977	01/01/1978	12/31/1982	8		8	
	01/31/1962	12/31/1987	01/01/1988	12/31/1992	8		8	
	01/31/1962	12/31/1997	01/01/1998	12/31/2002	8		8	
	01/31/1962	12/31/2007	01/01/2008	12/31/2012	8		8	

Table A5 Average Precision and Average Information Shortage by Industry

This table reports the industry averages of prediction precision and information shortage across the out-of-sample period computed with all common stocks in the three major exchanges (NYSE, AMEX, and NASDAQ). Panel A reports the averages for the prediction precision, while Panel B reports averages for the information shortage. The prediction precision is defined as the ratio between number of successful predictions and the total number of predictions. The information shortage is defined based on the aggregated predicted decile probabilities $E_{i,t} = -\sum_{d_{i,t}\in D} p_{agg}\widehat{(d_{i,t})} \log_2 p_{agg}\widehat{(d_{i,t})}$.The information shortage measures the minimum number of binary questions that need to be answered to completely eliminate the return prediction uncertainty. In other words, it measures the shortage of information.

Panel A: Industry Level Prediction Precision

2-digit SIC Industry	Code	Precision	2-digit SIC Industry	Code	Precision
Forestry	8	0.263	Apparel & Other Textile Products	23	0.154
Membership Organizations	86	0.241	Real Estate	65	0.153
Services, Not Elsewhere Classified	89	0.192	Rubber & Miscellaneous Plastics Products	30	0.152
Metal, Mining	10	0.186	Wholesale Trade – Nondurable Goods	51	0.152
Motion Pictures	78	0.184	Educational Services	82	0.151
Agricultural Production – Livestock	2	0.184	Fabricated Metal Products	34	0.151
Non-Classifiable Establishments	99	0.183	Insurance Carriers	63	0.149
Chemical & Allied Products	28	0.179	Heavy Construction, Except Building	16	0.149
Legal Services	81	0.178	Personal Services	72	0.148
Coal Mining	12	0.176	General Building Contractors	15	0.148
Oil & Gas Extraction	13	0.173	Security & Commodity Brokers	62	0.147
Business Services	73	0.172	Transportation Equipment	37	0.147
Local & Interurban Passenger Transit	41	0.169	Eating & Drinking Places	58	0.147
Holding & Other Investment Offices	67	0.169	Transportation Services	47	0.147
Instruments & Related Products	38	0.166	Auto Repair, Services, & Parking	75	0.147
Electronic & Other Electric Equipment	36	0.166	Apparel & Accessory Stores	56	0.147
Water Transportation	44	0.166	Furniture & Fixtures	25	0.146
Nonmetallic Minerals, Except Fuels	14	0.166	Building Materials & Gardening Supplies	52	0.146
Electric, Gas, & Sanitary Services	49	0.166	Lumber & Wood Products	24	0.146
Communications	48	0.164	Food & Kindred Products	20	0.146
Health Services	80	0.163	General Merchandise Stores	53	0.145
Special Trade Contractors	17	0.163	Paper & Allied Products	26	0.144
Miscellaneous Manufacturing Industries	39	0.163	Tobacco Products	21	0.144
Engineering & Management Services	87	0.162	Petroleum & Coal Products	29	0.143
Industrial Machinery & Equipment	35	0.161	Stone, Clay, & Glass Products	32	0.143
Amusement & Recreation Services	79	0.161	Transportation by Air	45	0.142
Insurance Agents, Brokers, & Service	64	0.160	Primary Metal Industries	33	0.142
Furniture & Home furnishings Stores	57	0.158	Railroad Transportation	40	0.141
Depository Institutions	60	0.157	Social Services	83	0.140
Nondepository Institutions	61	0.157	Hotels & Other Lodging Places	70	0.140
Miscellaneous Retail	59	0.157	Trucking & Warehousing	42	0.139
Wholesale Trade – Durable Goods	50	0.157	Food Stores	54	0.139
Pipelines, Except Natural Gas	46	0.156	Textile Mill Products	22	0.139
Printing & Publishing	27	0.155	Automotive Dealers & Service Stations	55	0.136
Agricultural Production – Crops	1	0.154	Leather & Leather Products	31	0.136
Agricultural Services	7	0.154	Miscellaneous Repair Services	76	0.133

Table A5 (Continues)

Panel B: Industry Level Information Shortage					
2-digit SIC Industry	Code	Info. Shortage	2-digit SIC Industry	Code	Info. Shortage
Electric, Gas, & Sanitary Services	49	3.190	Miscellaneous Manufacturing Industries	39	3.247
Depository Institutions	60	3.196	Nondepository Institutions	61	3.247
Tobacco Products	21	3.197	Health Services	80	3.248
Non-Classifiable Establishments	99	3.198	Furniture & Fixtures	25	3.248
Chemical & Allied Products	28	3.204	Industrial Machinery & Equipment	35	3.248
Pipelines, Except Natural Gas	46	3.211	Wholesale Trade – Nondurable Goods	51	3.248
Railroad Transportation	40	3.212	Stone, Clay, & Glass Products	32	3.249
Forestry	8	3.213	Real Estate	65	3.249
Membership Organizations	86	3.218	General Merchandise Stores	53	3.250
Metal, Mining	10	3.221	Auto Repair, Services, & Parking	75	3.250
Insurance Carriers	63	3.223	Apparel & Other Textile Products	23	3.250
Petroleum & Coal Products	29	3.226	Rubber & Miscellaneous Plastics Products	30	3.250
Motion Pictures	78	3.231	Special Trade Contractors	17	3.251
Business Services	73	3.231	Wholesale Trade – Durable Goods	50	3.251
Nonmetallic Minerals, Except Fuels	14	3.232	Hotels & Other Lodging Places	70	3.251
Insurance Agents, Brokers, & Service	64	3.232	Services, Not Elsewhere Classified	89	3.252
Holding & Other Investment Offices	67	3.232	Miscellaneous Retail	59	3.252
Communications	48	3.233	Agricultural Services	7	3.252
Paper & Allied Products	26	3.233	Local & Interurban Passenger Transit	41	3.253
Food & Kindred Products	20	3.233	Educational Services	82	3.254
Printing & Publishing	27	3.234	Eating & Drinking Places	58	3.255
Oil & Gas Extraction	13	3.235	Building Materials & Gardening Supplies	52	3.257
Water Transportation	44	3.237	Lumber & Wood Products	24	3.257
Engineering & Management Services	87	3.237	Trucking & Warehousing	42	3.257
Instruments & Related Products	38	3.238	Social Services	83	3.260
Personal Services	72	3.240	General Building Contractors	15	3.261
Security & Commodity Brokers	62	3.241	Legal Services	81	3.262
Electronic & Other Electric Equipment	36	3.242	Leather & Leather Products	31	3.262
Amusement & Recreation Services	79	3.244	Heavy Construction, Except Building	16	3.263
Fabricated Metal Products	34	3.244	Primary Metal Industries	33	3.263
Coal Mining	12	3.244	Automotive Dealers & Service Stations	55	3.263
Agricultural Production – Crops	1	3.245	Apparel & Accessory Stores	56	3.263
Food Stores	54	3.245	Furniture & Home furnishings Stores	57	3.264
Agricultural Production – Livestock	2	3.245	Transportation by Air	45	3.266
Transportation Equipment	37	3.245	Textile Mill Products	22	3.268
Transportation Services	47	3.246	Miscellaneous Repair Services	76	3.285

Table A6 Performance of Portfolios Including Stocks with Top 50% Market Capitalization

This table reports the economic performance of the portfolios *using only the stocks with above median capitalization* constructed based on the aggregated predictions from the individual classifiers. The statistics are based on the out-of-sample period covering 198301:202112. The decile portfolios are sorted based on the predicted deciles monthly, which are the deciles with the highest predicted probabilities. The column "market" reports the performance of the buy-and-hold strategy using all common stocks in the three major exchanges. The cumulative returns are in decimal unit representing gross returns in the sample period. $\alpha's$ are for the corresponding factor models, e.g., CAPM or Fama-French 3 Factor model. The t statistics for the $\alpha's$ are Newey-West t statistics of lag 6. The performance statistics are based on excess return adjusted with risk-free rate, i.e., 30-day US treasury bill. I report annualized Sharpe ratios. Turnover is the average total percentage of holding changes in absolute value. Max drawdown is the max difference between current price and the most recent price peak in percentage across all months in my sample period. Panel A reports the equal-weighted portfolio performance, while Panel B reports the value-weighted portfolio performance.

Panel A: Classification Equal-weighted Decile Portfolios

Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.009	-0.002	0.003	0.004	0.004	0.004	0.008	0.010	0.012	0.015	0.014	0.013
Cumulative Return	24.199	-0.962	0.129	1.628	2.438	3.488	27.891	77.556	135.289	310.412	120.312	341.751
CAPM Alpha	0.000	-0.016	-0.008	-0.005	-0.004	-0.002	0.002	0.003	0.004	0.004	0.002	0.015
	(0.049)	(-5.166)	(-4.224)	(-2.636)	(-2.137)	(-1.277)	(1.270)	(2.360)	(1.975)	(2.296)	(0.721)	(5.920)
FF3F Alpha	0.000	-0.013	-0.007	-0.005	-0.005	-0.003	0.001	0.003	0.003	0.005	0.003	0.014
	(0.340)	(-7.339)	(-6.773)	(-3.821)	(-4.378)	(-2.668)	(1.172)	(4.022)	(3.492)	(4.662)	(2.031)	(7.075)
FF5F Alpha	0.002	-0.007	-0.006	-0.005	-0.006	-0.005	-0.001	0.001	0.001	0.005	0.006	0.010
	(1.513)	(-4.781)	(-5.426)	(-3.625)	(-4.683)	(-3.618)	(-1.114)	(2.063)	(1.959)	(4.758)	(3.551)	(5.727)
Standard Deviation	0.058	0.098	0.073	0.062	0.057	0.043	0.040	0.045	0.055	0.067	0.085	0.036
Sharpe Ratio	0.515	-0.073	0.143	0.227	0.263	0.334	0.694	0.807	0.766	0.758	0.569	1.267
Turnover	0.105	0.134	0.100	0.082	0.071	0.059	0.052	0.058	0.073	0.092	0.120	0.127
Max Drawdown	-0.607	-0.936	-0.649	-0.637	-0.667	-0.667	-0.450	-0.480	-0.541	-0.558	-0.659	-0.430
Mean N	5342	283	252	66	99	126	692	407	387	202	158	441

Table A6 (Continues)

Panel B: Classification Value-weighted Decile Portfolios												
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.008	0.000	0.003	0.007	0.005	0.005	0.007	0.009	0.009	0.015	0.013	0.011
Cumulative Return	19.733	-0.934	-0.019	6.261	3.084	5.432	17.497	36.998	39.210	236.704	66.740	91.874
CAPM Alpha	0.000	-0.015	-0.009	-0.003	-0.004	-0.002	0.001	0.002	0.001	0.004	0.001	0.012
	(-1.674)	(-4.644)	(-4.058)	(-1.659)	(-1.785)	(-0.931)	(0.576)	(2.753)	(0.660)	(1.707)	(0.183)	(4.795)
FF3F Alpha	0.000	-0.012	-0.007	-0.003	-0.005	-0.003	0.000	0.001	0.001	0.005	0.003	0.012
	(-1.738)	(-5.239)	(-4.304)	(-1.488)	(-2.939)	(-2.134)	(-0.304)	(2.336)	(0.944)	(2.931)	(1.102)	(5.328)
FF5F Alpha	0.000	-0.005	-0.004	-0.001	-0.006	-0.004	-0.002	0.001	0.001	0.006	0.007	0.009
	(-1.003)	(-2.612)	(-2.745)	(-0.710)	(-3.643)	(-3.208)	(-3.622)	(1.070)	(0.646)	(3.888)	(2.658)	(3.624)
Standard Deviation	0.045	0.102	0.079	0.072	0.060	0.048	0.041	0.044	0.054	0.074	0.093	0.050
Sharpe Ratio	0.583	-0.015	0.140	0.333	0.281	0.374	0.600	0.689	0.602	0.681	0.498	0.753
Turnover	0.057	0.125	0.095	0.068	0.064	0.050	0.047	0.048	0.064	0.087	0.113	0.119
Max Drawdown	-0.527	-0.960	-0.829	-0.753	-0.721	-0.637	-0.502	-0.512	-0.617	-0.559	-0.723	-0.510
Mean N	5342	283	252	66	99	126	692	407	387	202	158	441

Table A7 Portfolio Performance based on Machine Learning Regressions

This table reports the economic performance of the portfolios *using only the machine learning regressions* constructed based on the aggregated predictions from the individual classifiers. The statistics are based on the out-of-sample period covering 198301:202112. The decile portfolios are sorted based on the predicted deciles monthly, which are the deciles with the highest predicted probabilities. The column "market" reports the performance of the buy-and-hold strategy using all common stocks in the three major exchanges. The cumulative returns are in decimal unit representing gross returns in the sample period. $\alpha's$ are for the corresponding factor models, e.g., CAPM or Fama-French 3 Factor model. The t statistics for the $\alpha's$ are Newey-West t statistics of lag 6. The performance statistics are based on excess return adjusted with risk-free rate, i.e., 30-day US treasury bill. I report annualized Sharpe ratios. Turnover is the average total percentage of holding changes in absolute value. Max drawdown is the max difference between current price and the most recent price peak in percentage across all months in my sample period. Panel A reports the equal-weighted portfolio performance, while Panel B reports the value-weighted portfolio performance.

Panel A: Regression Equal-weighted Decile Portfolios

Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.009	-0.011	0.000	0.003	0.007	0.008	0.008	0.010	0.012	0.016	0.033	0.041
CAPM Alpha	0.000	-0.022	-0.010	-0.005	-0.002	0.000	0.001	0.003	0.005	0.008	0.024	0.043
	(0.049)	(-7.222)	(-4.895)	(-3.712)	(-1.253)	(-0.095)	(0.405)	(1.715)	(2.631)	(3.567)	(5.922)	(11.162)
FF3F Alpha	0.000	-0.021	-0.009	-0.005	-0.002	0.000	0.000	0.003	0.004	0.008	0.024	0.042
	(0.340)	(-8.908)	(-5.954)	(-6.141)	(-2.527)	(-0.438)	(0.551)	(3.437)	(5.167)	(6.299)	(7.556)	(11.718)
FF5F Alpha	0.002	-0.015	-0.005	-0.003	-0.001	0.000	0.000	0.002	0.004	0.007	0.027	0.040
	(1.513)	(-7.846)	(-3.889)	(-3.354)	(-1.418)	(-0.528)	(-0.033)	(2.800)	(4.551)	(6.226)	(7.127)	(10.860)
Standard Deviation	0.058	0.085	0.069	0.059	0.054	0.051	0.049	0.049	0.051	0.055	0.090	0.048
Sharpe Ratio	0.515	-0.450	0.000	0.205	0.417	0.525	0.579	0.732	0.849	0.980	1.287	2.967
Turnover	0.105	0.151	0.115	0.099	0.090	0.085	0.081	0.081	0.084	0.097	0.157	0.154
Max Drawdown	-0.607	-0.951	-0.806	-0.678	-0.627	-0.561	-0.560	-0.541	-0.544	-0.529	-0.513	-0.120
Mean N	5342	534	534	534	534	534	534	534	534	534	534	1068

Table A7 (Continues)

Panel B: Regression Value-weighted Decile Portfolios											
Statistic	Market	lo	2	3	4	5	6	7	8	9	hi	hi-lo
Mean Excess Return	0.008	-0.006	0.003	0.006	0.007	0.007	0.007	0.010	0.010	0.012	0.018	0.021
CAPM Alpha	0.000	-0.019	-0.007	-0.003	-0.001	0.000	0.000	0.002	0.002	0.004	0.009	0.024
	(-1.674)	(-6.395)	(-3.912)	(-3.085)	(-1.017)	(-0.023)	(0.269)	(3.648)	(2.498)	(3.214)	(4.718)	(7.611)
FF3F Alpha	0.000	-0.017	-0.006	-0.003	-0.001	0.000	0.000	0.002	0.002	0.004	0.009	0.023
	(-1.738)	(-6.837)	(-3.841)	(-2.518)	(-1.292)	(-0.323)	(-0.026)	(3.602)	(2.808)	(3.123)	(5.506)	(8.122)
FF5F Alpha	0.000	-0.010	-0.002	-0.001	-0.001	-0.001	0.000	0.001	0.001	0.002	0.009	0.016
	(-1.003)	(-4.905)	(-1.971)	(-0.657)	(-0.857)	(-1.390)	(-0.710)	(1.336)	(2.024)	(1.687)	(4.461)	(6.629)
Standard Deviation	0.045	0.089	0.065	0.054	0.047	0.044	0.042	0.044	0.045	0.051	0.070	0.068
Sharpe Ratio	0.583	-0.236	0.142	0.380	0.516	0.579	0.592	0.753	0.748	0.789	0.882	1.071
Turnover	0.057	0.109	0.077	0.064	0.058	0.053	0.051	0.051	0.052	0.055	0.082	0.095
Max Drawdown	-0.527	-0.961	-0.830	-0.684	-0.517	-0.493	-0.449	-0.456	-0.556	-0.618	-0.542	-0.430
Mean N	5342	534	534	534	534	534	534	534	534	534	534	1068

Table A8 Information Shortage and Firm Characteristics

This table reports the Fama-MacBeth regression results in the investigation of the relation between information shortage and firm characteristics. The information shortage is computed based on the predicted probabilities using binary information entropy, which measures the expected minimum number of binary questions a forecaster has to answer correctly before reaching 100% correct predictions. The table reports the results for the regression $Information\ Shortage_{i,t} = \gamma_0 + Characteristics\Gamma + \varepsilon_{i,t}$, where the prediction precision is based on the aggregated predictions from the individual classifiers. I report for only variables that are statistically significant in the linear regressions, and I split the table into the positive column and the negative column, where the positive column reports results for variables that are positively related to the information shortage and the negative column reports for the variables that are negatively related to the information shortage. "FM t" represents Fama-MacBeth t statistics with Newey-West correction.

Table A8 (Continues)

Positive Relation			Negative Relation		
	Coefficient	FM t		Coefficient	FM t
disp	0.006	26.843	dolvol	-0.002	-1.766
cashdebt	0.004	17.093	pchsale_pchxsga	-0.001	-1.801
sp	0.007	16.279	chempia	-0.001	-1.814
roic	0.005	16.218	quick	-0.001	-1.910
fgr5yr	0.008	15.292	invest	0.000	-1.952
roaq	0.014	13.047	rd_sale	0.000	-2.282
roeq	0.002	12.832	lev	-0.001	-2.382
gma	0.004	10.666	sgr	0.000	-2.412
mom6m	0.019	10.415	operprof	0.000	-2.564
hire	0.002	10.347	turn	-0.003	-2.892
egr	0.002	9.953	std_dolvol	-0.002	-3.909
pricedelay	0.002	9.827	absacc	-0.002	-4.761
bm	0.005	9.810	pctacc	-0.002	-5.247
rsup	0.003	9.663	saleinv	-0.001	-5.276
cash	0.007	9.418	mom1m	-0.015	-5.432
securedind1	0.003	9.363	betasq	-0.023	-5.567
cfp	0.004	9.282	maxret	-0.004	-5.570
sue	0.002	9.150	mve_ia	-0.008	-6.597
secured	0.005	8.971	pchdepr	-0.001	-7.057
divi1	0.031	8.605	rd0	-0.004	-7.322
lgr	0.001	8.400	mom12m	-0.005	-7.649
ipo1	0.024	8.188	chtx	-0.001	-8.217
currat	0.004	7.483	chfeps	-0.001	-8.588
beta	0.032	7.463	dy	-0.008	-8.850
divi0	0.026	7.212	baspread	-0.018	-9.194
ep	0.006	6.634	mve	-0.012	-9.774
salerec	0.001	6.121	nincr	-0.002	-10.443
rd_mve	0.002	6.104	ps	-0.002	-10.448
chcsho	0.001	5.894	retvol	-0.019	-11.525
mom36m	0.004	5.600	agr	-0.004	-11.894
convind1	0.005	5.479	divo0	-0.007	-13.310
std_turn	0.002	5.371	idiovol	-0.015	-13.409
nanalyst	0.003	5.090	ms	-0.006	-13.842
salecash	0.001	4.981	chmom	-0.013	-14.319
tang	0.002	4.795	age	-0.012	-23.100
aeavol	0.001	4.656			
depr	0.001	4.375			
sfe	0.003	4.215			
zerotrade	0.002	3.844			
pchgm_pchsale	0.001	3.731			
chinv	0.001	3.701			
chnanalyst	0.000	3.188			
sin1	0.005	3.117			
grcapx	0.000	2.699			
ear	0.000	2.680			
cinvest	0.000	2.347			
roavol	0.001	2.063			
cashpr	0.001	1.914			
herf	0.007	1.774			
Constant	3.172	736.765			
102 Characteristics	Y				
Industry FE	Y				
Past Return Decile	Y				
Mean N	5362				
Mean Adj. R^2	0.589				