Empirical Asset Pricing with Probability Forecasts*

Songrun He, Linying Lv, and Guofu Zhou

December 2024

Abstract

We study probability forecasts in the context of cross-sectional asset pricing with a large number of firm characteristics. Empirically, we find that a simple probability forecast model can surprisingly perform as well as a sophisticated probability forecast model, delivering long-short portfolios whose Sharpe ratios are comparable to those of widely used return forecasts. Moreover, we show that combining probability forecasts with return forecasts yields superior portfolio performance versus using each type of forecast individually. Additionally, probability forecasts augment existing factor models and significantly improve tail risk forecasts. These results underscore the unique and valuable insights probability forecasts offer in understanding the cross-section of stock returns.

JEL Classification: G11, G12, G17

Keywords: Probability Forecast, Machine Learning, Information Ratio, Asset Pricing

*Songrun He is at Washington University in St. Louis (h.songrun@wustl.edu). Linying Lv is at Washington University in St. Louis (llyu@wustl.edu). Guofu Zhou and is at Washington University in St. Louis (zhou@wustl.edu).

We are grateful to Zhiyu Fu, Yuan Liao, Roy Chen Zhang, as well as the conference participants at Wolfe Research 8th Global Quant and Macro Conference and seminar participants at Washington University in St. Louis for their helpful comments and discussions.

1 Introduction

The Fama-MacBeth regression has been a fundamental tool used for years by researchers to construct factors or forecast the cross-section of asset returns. Applications of the approach include Haugen and Baker (1996), Lewellen (2014), and Green, Hand, and Zhang (2017) who find that a large number of firm characteristics have predictive power for the cross-section of asset returns, and they together have greater power than any single one. Recently, there has been an increasing application of sophisticated machine learning (ML) models in finance; examples include Light, Maslov, and Rytchkov (2017); Chinco, Clark-Joseph, and Ye (2019); Feng, Giglio, and Xiu (2020); Freyberger, Neuhierl, and Weber (2020); Gu, Kelly, and Xiu (2020); Kozak, Nagel, and Santosh (2020); Giglio, Liao, and Xiu (2021); Cong et al. (2021); Dong et al. (2022); Avramov et al. (2023) and Chen et al. (2023). A common feature of all these studies is that their results are all based on predicting the expected returns of the stocks.

In this paper, we take a different approach by predicting the probabilities that the stocks will outperform a benchmark, with a large number of firm characteristics as predictors as in those ML studies. The probability associated with a stock's return can be of interest by itself as it captures the information ratio of the stock. In contrast to the expected return, it also takes the risk into account. If we have the ex-ante probability, sorting by the probability is equivalent to sorting by the information ratio, which can potentially be more useful than using the expected return alone if both are estimated accurately. In a given application, however, their performance can vary and depend on individual estimation accuracy of the return and risk.

Empirically, with the same setup as Gu et al. (2020) who predict the expected return in a general framework, we predict the probability, and sort stocks into decile portfolios by their probability forecasts. Since low volatility is often associated with high probability forecast, the decile portfolios have to be adjusted by volatility so that the volatilities are comparable across deciles. Then, the standard long-short portfolio is easily constructed. We find that its performance is comparable with those from the expected return forecasts, with the annualized Sharpe ratios ranging from 1.27 to 1.51.

Moreover, we find that combining probability forecasts with expected return forecasts can generate a better portfolio compared with the one from each type of forecast. The mean-variance combination of the two forecasts yields long-short portfolios with a Sharpe ratio ranging from 1.64 to 1.74 across different prediction models in the out-of-sample testing period. The results suggest that probability forecasts generate incremental and significant information above and beyond the expected return forecasts.

There is another advantage of probability forecasting. We find that simple models in predicting the probabilities can already achieve superb portfolio performance compared to highly complex and nonlinear models. For example, a logistic regression prediction model can generate a long-short portfolio with a Sharpe ratio of 1.46, comparable with the best performance of the sophisticated Neural Network (NN) models used for expected return forecasts by Gu et al. (2020). This finding indicates that the probability target can potentially be a simpler function of underlying characteristics as compared with the expected return on which various studies document complex nonlinearity.

We further demonstrate that probability forecasts can help identify mispriced assets relative to existing factor models. When predicting the probability of a stock outperforming its model-implied benchmark return, we find that the resulting long-short portfolios generate significant alphas against the benchmark models. The enhancement is robust across eight benchmark factor models, including prespecified Fama-French models and recent machine-learning-based latent factor models. The combined portfolio of probability-based strategy with factor portfolios achieves consistent improvements in Sharpe ratios, with information ratios exceeding 1 when compared to factor models alone.

Our probability forecasting framework also exhibits value in managing tail risks. By directly modeling the probability of large price declines, we apply the reduced-form approach to forecast crash risk with a large number of stock characteristics. Our tail risk measure significantly outperforms existing predictors in the literature, leading to portfolios that effectively reduce exposure to market crashes. The low tail risk portfolio not only achieves a 45.6% higher Sharpe

ratio and 33.3% lower maximum drawdown than the market but also shows remarkable resilience during major market downturns, including the burst of the internet bubble, the 2008 financial crisis, and the COVID-19 crash.

Overall, our findings highlight the additional insights and value of modeling the probability instead of just the expected returns alone. By shifting the focus from traditional expected return forecasts only to consider also probability-based predictions, our study suggests a new paradigm for asset pricing and portfolio management.

Our paper is related to two broad strands of literature. There is a large literature on probability forecasts in the time-series context, starting from the early research such as Breen et al. (1989), Pesaran and Timmermann (1995), Leung et al. (2000), and Pesaran and Timmermann (2004). Christoffersen and Diebold (2006) further build a volatility-dependence-based theoretical framework to explain the surprising success of probability forecasting. Such a framework is later used in probability forecasting by Chevapatrakul (2013) and Catania et al. (2019), among others. Moskowitz et al. (2012) and Papailias et al. (2021) provide evidence of sign predictability using each assets' past excess return. However, most earlier studies focus on time-series probability forecasting. In contrast, we study probability forecasts for the cross-section of stock returns, and we use a large number of firm characteristics as predictors.

More recently, Del Viva et al. (2024) examine directional forecasts of stock returns using a simple OLS regression framework. However, their analysis is limited to only eight return-based signals and achieves relatively modest economic gains with a Sharpe ratio of 0.54. In contrast, we study probability forecasts across a broad range of targets using a comprehensive set of 94 firm characteristics. Additionally, we demonstrate incremental value from probability forecasts relative to the widely used expected return forecasts when aggregating information from the same predictors.

There is also a growing literature on using machine learning tools for cross-sectional asset pricing. Gu, Kelly, and Xiu (2021) apply the autoencoder model to consider a non-linear version of IPCA by Kelly, Pruitt, and Su (2019) for latent factor modeling. Chen, Pelger, and Zhu

(2023) extend the idea of GANs or robustness to estimate the SDF. Han, He, Rapach, and Zhou (2024) propose an E-LASSO approach for equity risk premia forecast that augments the traditional Fama-MacBeth regression with machine learning. Moreover, we see applications of machine learning algorithms to other asset classes and contexts, e.g. corporate bond (Bali, Goyal, Huang, Jiang, and Wen (2020)), options (Bali, Beckmeyer, Moerke, and Weigert (2023)), foreign exchanges (Filippou, Rapach, Taylor, and Zhou (2020)), mutual funds (Kaniel, Lin, Pelger, and Van Nieuwerburgh (2023)), and hedge funds (Wu, Chen, Yang, and Tindall (2021)). The existing literature in this strand focuses on modeling only the expected return of the assets. In contrast, our paper seems to be the first one to study probability forecasts with machine learning methods and large number of stock characteristics.

The rest of the paper is organized as follows. Section 2 explores reasons why the probability forecast is of interest. Section 3 presents our methodology, interpretation, and performance evaluation frameworks. Section 4 reports our empirical results. Section 5 presents the application of probability forecasts in tail risk management. Section 6 concludes.

2 Why Probability Forecasts?

In this section, we discuss the rationale behind probability forecasting, why it unifies both risk and return, and the close connection between the probability of outperformance and the information ratio. Furthermore, we demonstrate that under the assumption of a factor model for stock returns, the probability of outperforming the factor is directly related to the t-statistic from the time-series regression of the asset on benchmark factors. Lastly, we present the framework for using characteristics as instruments for the time-varying probability of outperformance.

To start off, assume the returns on a single asset i and a benchmark portfolio b follow a joint normal distribution:

$$\begin{bmatrix} r_i \\ r_b \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_i \\ \mu_b \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & \rho\sigma_i\sigma_b \\ \rho\sigma_i\sigma_b & \sigma_b^2 \end{bmatrix} \right).$$

Under this assumption, the probability that the asset outperforms the benchmark can be written as:

$$\begin{aligned}\mathbb{P}(r_i - r_b > 0) &= 1 - \mathbb{P}(r_i - r_b < 0) \\ &= 1 - \mathbb{P}\left(\frac{(r_i - r_b) - (\mu_i - \mu_b)}{\sigma(r_i - r_b)} < \frac{-(\mu_i - \mu_b)}{\sigma(r_i - r_b)}\right) \\ &= \Phi\left(\frac{E[r_i - r_b]}{\sigma(r_i - r_b)}\right),\end{aligned} \tag{1}$$

where $\Phi(\cdot)$ is the standard normal CDF function and the third equality follows from the normality assumption.

Define the Information Ratio of the asset relative to the benchmark as:

$$IR = \frac{E[r_i - r_b]}{\sigma(r_i - r_b)}. \tag{2}$$

The probability of outperformance in Equation (1) equals applying $\Phi(\cdot)$ on the IR of the asset relative to the benchmark.

Also note that $\Phi(\cdot)$ is an increasing function of its argument. In the setting of cross-sectional asset pricing, sorting based on the probability of outperformance is equivalent to sorting on the information ratio of the asset relative to a benchmark. Therefore, in this paper, we focus on modeling probability for cross-sectional asset pricing.

The information ratio itself is not an observable object in the market. To estimate it, one needs to separately estimate expected returns and volatility, which introduces an additional layer of estimation error that propagates to influence overall model performance.

On the other hand, the realized outcome of whether a stock outperforms a benchmark is directly observable. Therefore, we take the approach by directly modeling the probability and show this can generate distinctive cross-sectional spread patterns compared to forecasting expected returns.

Our framework differs fundamentally from previous work on probability forecasts in time-series settings, particularly Christoffersen and Diebold (2006). While they focus on how volatility dependence can enhance sign predictability in the time series of individual assets, we examine

probability forecasts in the cross-section of asset returns. This cross-sectional approach offers one key advantage: because probability forecasts preserve the ranking of assets' information ratios, they provide a theoretically motivated way to construct long-short portfolios that account for both return and risk. In contrast, time-series probability forecasts primarily aim to predict the directional movements of individual assets without considering their relative performance against other stocks. Our cross-sectional framework thus provides a natural bridge between probability forecasting and portfolio formation in empirical asset pricing.

This information ratio sorting argument extends to factor models. Assume that the realized excess stock returns follow a factor model:

$$r_{i,t+1} - r_{f,t} = \alpha_{it} + \beta_{it}' f_{t+1} + \varepsilon_{i,t+1}, \tag{3}$$

where $r_{f,t}$ is the risk-free rate, α_{it} represents mispricing of the factor model, β_{it} and f_{t+1} are vector of factor loadings and factors respectively, and $\varepsilon_{i,t+1} \sim \mathcal{N}(0, \sigma_{it}^2)$ is the idiosyncratic return.

Following a similar logic, we can write the probability that the asset outperforms the risk-free rate plus $\beta_{it}' f_{t+1}$ as:

$$\mathbb{P}\left(r_{i,t+1} - r_{f,t} - \beta_{it} f_{t+1} > 0\right) = \Phi\left(\frac{\alpha_{it}}{\sigma_{it}}\right). \tag{4}$$

If this probability can be estimated with high accuracy, it provides significant insights into the information ratio of the asset relative to the benchmark factor model.

In finite samples, the information ratio is closely related to the t-statistic derived from the time-series regression of the asset's returns on the corresponding factor models. This relationship, along with the understanding that CDF function is merely an increasing function and that portfolio sorting remains consistent under any increasing transformation, offers a practical framework for testing and evaluating existing factor models.

Researchers may construct a long-short portfolio based on the probability of outperforming the factor model. This portfolio assigns positive weights to stocks with significant positive alphas and negative weights to those with significant negative alphas. These stocks are expected to exhibit

the greatest mispricing relative to the benchmark factor model, thereby offering an opportunity to enhance the baseline model's performance.

This idea is related to Chen, Pelger, and Zhu (2023), which employs a GANs-based approach to adversarially construct the stochastic discount factor (SDF). In each iteration, the SDF is augmented with assets that the previous model struggles to price. In contrast to the deep learning approach, probability modeling is more computationally efficient and provides a transparent and interpretable framework for identifying assets that are most challenging for factor models to price. Moreover, it allows for identifying mispriced assets at the individual stock level rather than the portfolio level.

In the general set up in Equation (4), we allow both α_{it} and idiosyncratic volatility σ_{it} to be time-varying. However, it would be impossible to provide identification for this model because each *i-t* combination only appears once in the panel. To proceed, we model the information ratio as a general function of time-varying stock characteristics:

$$\frac{\alpha_{it}}{\sigma_{it}} = g(z_{i,t};\theta), \tag{5}$$

where $z_{i,t}$ are the stock characteristics, and $g(\cdot)$ is an arbitrary function parameterized by θ. In this setup, $z_{i,t}$ serves as an instrument to capture the dynamics of both α_{it} and σ_{it} over time. Consequently, when we model the probability of outperformance as a function of these instrument characteristics, we implicitly model the information ratio as well.

This framework allows probability forecasts to incorporate not only predictability in expected returns (the α_{it} component) but also predictability in volatility (the σ_{it} component). By adopting an information ratio perspective and accounting for σ predictability, the model spans a richer cross-sectional information set in stock returns than approaches focused solely on $\mathbb{E}[R]$.

3 Methodology

In this section, we introduce the methodology we use for the probability forecast of stock returns, the metrics to evaluate variable importance and performance evaluation measures.

3.1 Prediction Models with Mean Squared Loss

The models we consider can be classified into two categories by their objectives: (1) mean squared error loss with an identity linking function; and (2) cross-entropy loss with a Logit linking function. The linking function transforms the model output into a probability forecast. We introduce the two categories of models in detail in this section.

Similar to the literature on the linear probability model, when the outcome becomes a discrete 0-1 variable, one straightforward way is to just apply the same regression framework to the 0-1 outcome variable. The theory of regression would imply the regression coefficients inherit all the good BLUE properties of simple OLS. In this case, the loss function can be written as:

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t+1} - g(z_{i,t};\theta))^2, \tag{6}$$

where $y_{i,t+1}$ is a 0-1 variable indicating whether return outperforms the benchmark or not, $g(\cdot;\theta)$ is the prediction model, and $z_{i,t}$ are the predictors for stock i at time t.

In this case, although the final forecast may be outside the probability bound of 0 and 1, the model is still capable of learning the probability distribution of the final outcome. We consider two models: the OLS and PLS models.

3.1.1 OLS

For the OLS model, we assume that each predictor has a linear effect on the final probability. Then we can write the prediction model as:

$$g(z_{i,t};\theta) = z'_{i,t}\theta. \tag{7}$$

This model provides a linear benchmark. It is essentially an OLS regression of the discrete outcome y on all the predictors, which can be solved in closed form.

3.1.2 Partial Least Squares

To avoid overfitting, we consider reducing the dimension of the predictors using partial least squares (PLS). The model is first used in finance by Kelly and Pruitt (2013) to construct a market return forecast using many valuation measures. Huang, Jiang, Tu, and Zhou (2015) then apply it to construct an investor sentiment index aligned with predicting market returns.

We write the probability forecast model as follows:

$$g(z_{i,t};\theta) = (z'_{i,t}\Omega)'\theta, \tag{8}$$

where Ω is a $K \times P$ matrix that transforms the K predictors in $z_{i,t}$ into P lower dimensional components.

To extract the j-th component with the PLS, the objective function can be written as:

$$\omega_j = \arg\max_{\omega} Cov(Y, Z\omega), \quad \text{s.t. } \omega'\omega = 1,\ Cov(Z\omega, Z\omega_i) = 0,\ i = 1,2,\cdots j-1. \tag{9}$$

Intuitively, the algorithm extracts predictive components in a sequence where the extracted one should have maximum covariance with the outcome variable while being orthogonal to previous components. Our outcome variable Y here is a 0-1 discrete variable indicating whether the stock

return outperforms its benchmark or not.

3.2 Prediction Models with Cross-Entropy Loss

Another natural loss function in the context of probability forecast is the cross-entropy loss,

$$\mathscr{L}(\theta) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t+1}\log(g(z_{i,t};\theta)) + (1-y_{i,t+1})\log(1-g(z_{i,t};\theta))\right). \tag{10}$$

Minimizing this loss is equivalent to maximizing the log-likelihood of observing all the outcomes. Note that to compute the loss, we require the model output of probability forecast to be strictly within the range of $(0,1)$. One popular choice is to use the sigmoid function as,

$$S(x) = \frac{1}{1+e^{-x}}, \tag{11}$$

which is the link function to map or squash any function output to a value within the range of 0 and 1.

Therefore, the final probability forecast can be written as:

$$g(z_{i,t};\theta) = S(f(z_{i,t};\theta)), \tag{12}$$

where $f(z_{i,t};\theta)$ is output from a prediction model. In this paper, we consider two such prediction models: (1) a linear model and (2) a non-linear neural network following Gu et al. (2020).

3.2.1 Logistic Regression

When $f(z_{i,t};\theta)$ is a simple linear function, the regression becomes a logistic regression. We have

$$f(z_{i,t};\theta) = z'_{i,t}\theta, \tag{13}$$

$$g(z_{i,t};\theta) = \frac{1}{1+e^{-z'_{i,t}\theta}}. \tag{14}$$

Compared to the OLS, the sigmoid linkage function introduces the non-linearity to the model. The property of the linkage function will make extreme input saturate.

3.2.2 Neural Networks

One natural extension to the linear setup in Equation (13) is to consider using a neural network as a prediction model. Consider a three-layer neural network as the prediction model. The entire prediction problem can be written as:

$$f(z_{i,t};\theta) = W_3 \times \sigma\left(W_2 \times \sigma\left(W_1 z_{i,t} + b_1\right) + b_2\right) + b_3\,, \tag{15}$$

$$g(z_{i,t};\theta) = \frac{1}{1+e^{-f(z_{i,t};\theta)}}\,, \tag{16}$$

where $\sigma(\cdot)$ is the ReLU activation function for neural network, W_i and b_i are the weights and bias for layer i. For neural networks with more (less) layers, we are adding (removing) the composition of functions in Equation (15).

For the neural network architecture, we follow Gu et al. (2020) with 32 neurons in the first layer, followed by 16, 8, 4, and 2 in subsequent layers. We consider a neural network with 1 to 5 layers and follow the training and optimization procedure in Gu et al. (2020).[1] This model serves as an extension to the basic logistic regression. We want to examine whether introducing additional non-linearity would improve the prediction performance under the cross-entropy loss specification.

3.3 Model Interpretation

To interpret the prediction model output, we adopt the SHAP (Shapley Additive exPlanations) framework by Lundberg and Lee (2017). The idea originates from Shapley (1953), which solves the problem of distributing the payoff fairly among N players in a game. He shows that the Shapley value is the only metric satisfying three axioms for evenly distributing the payoff based on the

[1] To regularize the model, we add early stop procedures that stop training the neural network once the validation loss stops decreasing. We also retrain the model five times and use an equal-weighted ensemble of the five model outputs as the final prediction.

contributions.

Empirically, the Shapley value can be written as:

$$\phi_i(f,x) = \sum_{z \subseteq x \setminus \{i\}} \frac{|z|!\,(|x|-|z|-1)!}{|x|!} \left[f(z \cup \{i\}) - f(z) \right], \tag{17}$$

where i is the index for predictor, f is the prediction model of interest, x is the total set of predictors considered, and $|\cdot|$ denotes the operator that returns the number of elements in a set. In this paper, we use this measure to account for variable importance across different models. One thing to note is that the Shapley value requires an input for X when it is missing. Following Gu et al. (2020), we use the value of zero as the input given the data has been transformed into cross-sectional rank mapped into the interval of $[-1,1]$.[2]

3.4 Performance Evaluation

To evaluate the performance of our model, we evaluate the probability forecast in our test sample with prediction accuracy and cross-entropy. The accuracy is calculated as:

$$Accuracy_{OOS} = \frac{\sum_{(i,t)\in\mathcal{T}_3} TP_{it} + TN_{it}}{\sum_{(i,t)\in\mathcal{T}_3} TP_{it} + TN_{it} + FP_{it} + FN_{it}}, \tag{18}$$

where $\mathcal{T}_3$ represents the out-of-sample testing period and TP, TN, FP, FN are indicator variables for true positive, true negative, false positive, and false negative, respectively. For model prediction, if the probability output is larger than 50%, we classify the output as $y = 1$. Otherwise, we classify it as $y = 0$.

The out-of-sample cross-entropy is calculated as:

$$CE_{OOS} = -\frac{1}{|\mathcal{T}_3|} \sum_{(i,t)\in\mathcal{T}_3} \left(y_{i,t+1} \log(g(z_{i,t};\hat{\theta})) + (1-y_{i,t+1}) \log(1-g(z_{i,t};\hat{\theta})) \right). \tag{19}$$

[2] After the rank transformation, filling missing value with 0 is equivalent to using the cross-sectional median to impute the missing values.

To gauge the economic significance, we sort stocks based on the probability forecast into value-weighted decile portfolios and examine the out-of-sample decile portfolio performance. We rebalance our portfolio at a monthly frequency. We construct a long-short strategy that longs the decile ten portfolio and shorts the decile one portfolio.

3.5 Variance Adjusted Portfolio Sorting

From our discussion in section 2, the volatility dynamic plays an important role in determining the probability that a stock outperforms the benchmark. In other words, predicting the probability of outperformance is inherently tied to volatility. To alleviate the effect of volatility dependence, we perform a variance adjustment procedure.

To be specific, at the beginning of each month, we calculate real-time volatility using a rolling window of 36 months for each decile portfolio:

$$\sigma_{j,t-1} = \sqrt{\frac{1}{T}\sum_{\tau=t-T}^{t-1}\left(r_{j,\tau} - \bar{r}_j\right)^2}, \qquad \text{for } j = 1, \cdots, D, \tag{20}$$

where $r_{j,\tau}$ is the excess return of decile j portfolio at time τ, $\bar{r}_j$ is the average of excess return of the decile j portfolio over the rolling window, T is the rolling window length, and D is the total number of portfolios in sorting. Next, at each point in time, we scale the decile portfolio excess return to have the same volatility as the first decile portfolio:

$$r_{j,t}^{adj} = r_{j,t}\frac{\sigma_{1,t-1}}{\sigma_{j,t-1}}, \qquad \text{for } j = 1, \cdots, D. \tag{21}$$

We form the variance adjusted zero-cost long short portfolio as $r_{D,t}^{adj} - r_{1,t}^{adj}$ and report variance adjusted decile portfolio performance in empirical anslysis.

4 Data and Empirical Results

In this section, we introduce the data and report the empirical results from the probability forecast. We then compare and contrast the probability forecast with the expected return forecast. Finally, we present evidence on augmenting factor models through probability forecasts.

4.1 Data

Following Gu et al. (2020) paper, our research aggregates a comprehensive dataset comprising monthly individual stock returns sourced from the Center for Research in Security Prices (CRSP). Inspired by Green et al. (2017), we incorporate an array of 94 firm-level predictive characteristics. This dataset captures the market activity of all firms listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ from January 1957 through December 2020.

Our sample features a total of 31,492 unique stocks. The average number of cross-sectional observations per month notably exceeds 5,000, providing a rich basis for in-depth analysis. Additionally, we obtain market returns and risk-free rates from Kenneth French's data library.[3]

Following Gu et al. (2020), we split our data into training, validation, and test sets. For models with no hyperparameters (OLS and Logit), we combine the training and validation set together as the training set. We use an expanding window training scheme and re-train the model each year. For example, for the first model, we train the model using data from January 1957 to December 1974, validate the model on data from January 1975 to December 1986, and test the model from January 1987 to December 1987. For the second model, we expand the training set by one year, keep the validation set length the same for the next 12 years, and test the model prediction in the year 1988. With this setup, our out-of-sample testing period for all models is from January 1987 to December 2020.

[3] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

4.2 Emprirical Results with Probability Forecasts

In this section, we focus on the empirical performance of the models with probability forecasts. For the baseline results, we focus on predicting the probability a stock will outperform the market. Since market returns represent a significant systematic factor affecting all stocks, our approach aims to guide the model toward emphasizing cross-sectional differences. This aligns with the spirit of the Fama and MacBeth (1973) regression framework in the context of expected return forecasting.

In the first section, we report the summary statistics of the model prediction as well as the variable importance. In the second section, we present the portfolio performance from the probability forecasts to gauge the economic value.

4.2.1 Model Prediction and Variable Importance

Table 1 and Table 2 present the summary statistics of the probability forecasts and their corresponding predictive accuracies across various models. The initial observation reveals that the average proportion of stock returns outperforming the market benchmark is 46.4% for the aggregated sample period.

For model-specific performance, we evaluate the performance of eight distinct models: Ordinary Least Squares (OLS), Logistic Regression (Logit), Partial Least Squares (PLS), and a set of Neural Networks with one to five hidden layers (NN1 through NN5). We report accuracy and cross-entropy for each model as forecast performance measures.

This cross-entropy calculates the log loss for each forecast and then averages it over all the forecasts, providing a robust measure of the model's performance in predicting the probability. Higher values of accuracy or lower values of cross-entropy indicate better performance of the forecasting model.

The OLS model exhibits a mean probability forecast of 0.371, along with a comparatively high standard deviation of 0.183. This divergence from the empirical percentage of stocks surpassing

market returns is anticipated, highlighting the potential inadequacies of OLS in capturing the intricacies of non-linear predictor-outcome relationships, as well as its vulnerability to in-sample overfitting phenomena. In contrast, the binary logistic regression model (Logit) manifests a marked enhancement in forecasting precision. The mean prediction soars to 0.459, which closely mirrors the actual observed frequency, and the standard deviation contracts substantially to 0.044. This model's skewness is moderately negative at -0.384, and its kurtosis is nearly negligible at 0.055, indicating a more symmetric and less outlier-prone distribution of forecast errors compared to its linear counterpart.

We further directed towards the PLS model, a technique aimed at dimensionality reduction in regression contexts. This model parallels the improvements seen with Logit, showing predictions that align more closely with actual outcomes than the OLS model. The remaining part of Table 1 focuses on the probability forecasts derived from Neural Networks, ranging from a single-layer network (NN1) to a five-layer network (NN5). The summary statistics show a consistent predictive performance across these neural network models, suggesting a negligible effect of model complexity on probability forecast accuracy.

In Table 2, we adopt a binary classification scheme where forecasted probabilities above 0.5 are considered as positive predictions and those below as negative.[4] Realized stock returns are correspondingly designated as positive if they outperform the market return and negative if they do not. Observations are thus segmented into four distinct categories: stocks with both a positive prediction and actual market outperformance (true positive), those with a positive prediction that underperform the market (false positive), stocks with a negative prediction that do underperform (true negative), and those with a negative prediction but positive actual performance (false negative).

The model's accuracy is measured by how well these predictions match reality. Statistically, the OLS model registers an accuracy of 0.54, accompanied by a cross-entropy score of 0.689. The Logit and PLS models exhibit a comparable level of accuracy and an identical cross-entropy

[4]Notably, This split mainly leads to more negative predictions, as the average forecast falls below the 0.5 threshold.

score, underscoring a consistency in performance. Neural network models, ranging from NN1 to NN5, reveal subtle improvements in accuracy with added complexity—NN1 starts at an accuracy of 0.537, ascending marginally to 0.539 for NN5.

In general, Table 1 and Table 2 highlight the subtle but significant trade-offs involved in model complexity versus predictive accuracy. Non-linear models, such as Logit and neural networks, are effective at navigating the complex, non-linear relationships in the dataset. Yet, the incremental gains in accuracy begin to diminish as model complexity increases, suggesting a threshold beyond which additional complexity does not equate to meaningful increases in predictive performance.

We now examine the relative importance of individual predictors for the performance of each model using the Shapley value method, as described in Section 3. For every observation, the yearly fitted model is used to calculate how each of the 94 anomaly predictors contributes to the forecast in comparison to the dataset's average prediction. The average deviation of each observation in the sample is calculated as a measure of the predictor's significance. Figure 1 displays the importance of the characteristics of the top-20 variables for each model, with the sum of variable importance in each model normalized to one. This normalization permits easy comparison across the models.

Figure 1 indicates that, with the exception of OLS, there is a general consensus among the models regarding the most influential covariates for probability forecasts. The significant variables fall into four categories. The first category includes risk and volatility measures such as return volatility ($retvol$), idiosyncratic return volatility ($idiovol$), market beta ($beta$), volatility of dollar trading volume (std_dolvol), and volatility of share turnover (std_turn), which are consistent with our probability forecast framework.

Momentum-based predictors form the second category, including 1-month short-term momentum ($mom1m$), 12-month long-term momentum ($mom12m$), and 6-month momentum ($mom6m$). The third category is composed of liquidity measures, including share turnover ($turn$), zero trading days ($zerotrade$), bid-ask spread ($baspread$), and Amihud illiquidity (ill), all of which provide insights into the market's ability to facilitate transactions without significantly affecting the stock price.

The final category comprises fundamental characteristics, which include return on invested capital (*roic*), the number of earnings increases (*nincr*), initiation and omission of dividends (*divi* and *divo*), and issuance of secured debt (*securedind*). These reflect the risk factors and financial conditions of firms and are crucial for probability forecasts.

Compared to the variable importance of expected return forecasts reported by Gu et al. (2020) in Figure 3, our approach reveals a different pattern: volatility measures carry greater weight in our probability prediction models. This underscores the unique insights provided by probability forecasts, highlighting their ability to capture distinctive aspects of the cross-section of stock returns.

4.2.2 Portfolio Performance with Probability Forecasts

In this subsection, we focus on the portfolio performance from probability forecasts. At the beginning of every month, we sort stocks into decile portfolios based on the model's predicted probability of outperforming the bench market return, i.e., the market.

Within each portfolio, we weight the stock by its market cap at the end of the previous month to remove the effect of microcaps following Hou, Xue, and Zhang (2020). We then perform the variance adjustment procedure laid out in Section 3.5 to dynamically adjust portfolio weights for each decile portfolio and construct a zero-cost long-short portfolio by longing the decile ten and shorting the decile one portfolio.

To explain the rationale behind the variance adjustment strategy, consider decile one and decile ten excess returns, which both load on common factors in returns:

$$r_{1,t} = \beta_1' f_t + \varepsilon_{1,t}, \tag{22}$$

$$r_{D,t} = \beta_D' f_t + \varepsilon_{D,t}. \tag{23}$$

The long-short return can be written as:

$$r_{LS,t} = (\beta_D - \beta_1)' F_t + \varepsilon_{D,t} - \varepsilon_{1,t}. \tag{24}$$

From our presentation of the analytical framework in Section 2, the prediction based on probability not only sorts stocks based on expected returns but also volatility. As a result, the decile ten portfolio will have lower volatility than the decile one portfolio, which suggests that β_D is of smaller magnitude than β_1.

Therefore, the long-short strategy loads excessively on the factor risk, decreasing its Sharpe ratio. With the adjustment, we have reduced the factor risk while boosting the portfolio's expected return. As a result, we expect to see the Sharpe ratio to improve.[5]

Table 3 presents the portfolio performance from probability forecasts. The average forecast probability ($\widehat{Prob}$) closely mirrors the realized probability ($Prob$), underscoring the efficacy of our models. Except for the OLS model, the returns of the portfolios generally increase in line with the probability forecasts. Notably, the NN1 model performs the best among all models considered, delivering an average monthly return of 2.75% with a Sharpe ratio of 1.51.

It is also worth noting that the simple Logit model generates comparable performance with an average monthly return of 2.61% with a Sharpe ratio of 1.46. The numbers are comparable to the best-expected return forecast model from Gu et al. (2020).[6]

The results further suggest that probability objectives are easier to learn compared to expected return targets. Even the 1-layer Neural Network or the Logit model can beat Neural Networks with

[5] We present the performance of non-variance-adjusted decile portfolios in Online Appendix Table A3. Notably, there are significant differences in volatility and distinct variance sorting patterns among the decile portfolios. However, even without variance adjustment, the probability-based long-short (LS) portfolio demonstrates a significant spread and provides substantial incremental value relative to expected return forecasts. When the two approaches are combined, the resulting portfolio achieves significantly higher Sharpe Ratios. Evidence on combination gains is presented in Table A4.

[6] All conclusions remain robust across a comprehensive set of checks using four alternative benchmarks: (1) r_f, (2) βr_{mkt}, (3) 0, and (4) cross-sectional median. The portfolio sorting results are presented in Online Appendix Tables A5, A7, A9, and A11. Across all these specifications, the simple model such as logistic regressions performs as effectively as the more complex models, producing "H-L" return spreads comparable to those generated by expected return forecasts.

more layers.

4.3 Comparison with Expected Return Forecast

In this section, we focus on discussing the similarities and differences between probability and expected return forecasts.

From our discussion in section 2 equation 1, if returns are jointly normal, the probability that stock i outperforms the market would equal the normal CDF with an argument of the information ratio of stock i relative to the market portfolio. The higher the stock i's expected return, the greater the probability this stock will outperform the market.

However, the probability measure has information beyond just the expected return. It also takes into account the volatility of the difference between stock i and the benchmark, i.e., $\sigma(R_i - R^{mkt})$. As a result, even if expected excess return $\mu_i - \mu^{mkt}$ is constant, as long as the conditional volatility $\sigma(R_i - R^{mkt})$ is predictable, the resulting probability $\Phi\left(\frac{\mu_i - \mu^{mkt}}{\sigma(R_i - R^{mkt})}\right)$ will be predictable.

Therefore, choosing probability as a target to model can improve the signal-to-noise ratio in return forecasting problems. The low signal-to-noise ratio problem in expected return forecast or measurement has plagued empirical asset pricing, as is argued in Black (1986). In many contexts, when there is no need for precise estimates of expected returns, choosing probability as the modeling object can reduce the amount of data needed for the algorithm to converge since there is more signal from the data to learn from. This is especially useful when the market is having structural breaks and time-varying regimes in expected returns (Ang and Timmermann (2012)).

Moreover, another advantage of probability forecasting is that it takes volatility information into account. In Section 2, we show that when returns follow a joint normal distribution, the probability forecast can be exactly mapped into the information ratio of the individual stock return relative to a benchmark. For example, if we choose the risk-free rate as the benchmark, the ratio is exactly the Sharpe ratio. Suppose the learning algorithm is able to uncover the probability forecast perfectly. Sorting on the final output from the forecast would be equivalent to sorting on the Sharpe

ratio/information ratio of the assets. Expected return forecasts, on the other hand, will only sort stocks based on their mean returns regardless of the risks facing investors. This sorting property on Sharpe ratio/information ratios instead of only on returns can potentially generate a richer variation of information in stocks that span the SDF.

Comparing our portfolio performance results in Table 4 to the expected return forecast from Gu et al. (2020), we find that sorting based on probability forecast generate long-short portfolio return and Sharpe ratios comparable to the best neural network models from Gu et al. (2020) that target on expected returns. For a fair comparison, we perform the same variance adjustment procedure to the expected return forecasts from Gu et al. (2020).

Moreover, we find that the probability forecast generates incremental information relative to expected return forecasts. Firstly, the long-short portfolio returns from probability forecasts are only moderately correlated with the ones from expected returns forecasts. Secondly, when we combine the two portfolios using equal-weighted or mean-variance efficient weight, we find that the resulting portfolio has a higher Sharpe ratio than each individual portfolio. The results are shown in Panel C and Panel D of Table 4. For the best-performing model, the Logit, when combining expected return with probability forecast, the long-short portfolio can reach a Sharpe ratio of 1.73, beating each individual one. The resulting improvement is consistent across all the models we considered.[7]

Overall, our empirical results show that the probability forecasts generate incremental information relative to the expected return forecasts, as is evidenced by the increase in the Sharpe ratio when combining the two. Moreover, our finding suggests that the modeling objective is simpler. As a result, even simpler machine learning models such as logistic regression and PLS can generate comparable performance as the complex nonlinear Neural Network models.

[7]In the robustness checks presented in the Online Appendix, we examine four alternative target specifications: (1) r_f, (2) βr_{mkt}, (3) 0, and (4) cross-sectional median. When combined with expected return forecasts, each of these alternative targets significantly improves Sharpe ratios of the $\mathbb{E}[R]$ 'H-L' portfolios. We show the corresponding results in Tables A6, A8, A10, and A12.

4.4 Augmenting Factor Models with Probability Forecasts

In this section, we present how probability can help identify the most mispriced assets and augment existing factor models. Following the argument in Section 2, we focus on predicting the probability that a stock outperforms its real-time β times the factors:[8]

$$\mathbb{P}(r_{i,t+1} \geq r_{f,t} + \beta_{it} f_{t+1}) = S(f(z_{i,t}; \theta)). \tag{25}$$

We consider eight benchmark factor models in the literature:[9]

CAPM: This is the one-factor model with market excess return as the factor, first proposed in Sharpe (1964).

FF3: This is the three-factor Fama-French model from Fama and French (1993).

FF6: This is the Fama-French five-factor model from Fama and French (2015) plus momentum factor, which appeared in Fama and French (2018).

HXZ: This is the five-factor model motivated from the investment CAPM theory. The factors include four factors from Hou, Xue, and Zhang (2015) plus the expected growth factor from Hou, Mo, Xue, and Zhang (2021).

SY: This is the four factor model from Stambaugh and Yuan (2017). In addition to market and size, the model augments with two additional 'mispricing' factors.

DHS: This is the three-factor model from Daniel, Hirshleifer, and Sun (2020), which propose two short- and long- horizon behavioral factors in addition to market.

IPCA: This is the linear latent factor machine learning factor model (Instrumentd Principal Component Analysis) proposed in Kelly, Pruitt, and Su (2019), which extracts factors to explain the most common variations in stock returns.[10]

[8] In this section, we focus on using the logistic regression model to predict the probability but note that our conclusion remains robust to using PLS and NN models.

[9] We thank the authors for providing the factors data on their websites.

[10] We focus on a three-factor IPCA model but note that the conclusions remain robust for alternative specifications.

CA: This is the non-linear neural network latent factor model (Conditional Autoencoder) proposed in Gu et al. (2021), which extends Kelly et al. (2019) to allow for nonlinear factor loadings.[11]

The eight benchmark factor models can be further classified into two categories: (1) the first six are prespecified factor models where factors are defined by researchers using observable characteristics; (2) the last two are latent factor models where the factor loadings change with observable characteristics while factors are latent.

For prespecified factor models, we estimate the βs using rolling windows. To be specific, at the beginning of each month, for each asset, we run time-series regressions of stock returns on factors with a rolling window of 60 months. For stocks with less than 20 observations, we set the regression coefficient as NA. After estimating βs, we fill the all missing values as the cross-sectional median of the corresponding βs. For latent factor models, we use the real-time estimated factor loading as the β_{it} in the probability prediction Equation (25).

Table 5 presents the value-weighted decile portfolio sorted on the probability of outperforming the eight benchmark factor models. We find that across all the models, the probability 'H-L' portfolio generates high Sharpe Ratios with large long-short return spreads, which are comparable to our baseline results in Table 3. The agreement between model-predicted probability versus realized probability of outperforming factor models suggests that overfitting is not a significant concern.

The next important question is whether the probability forecasts add values to the factors, as indicated by our analysis in Section 2. To answer this question, we consider combining the probability 'H-L' portfolios with the real-time tangency portfolios from different factor models.[12]

Table 6 presents the portfolio performance of factors augmented with probability forecasts.

[11] We focus on a three-factor Autoencoder model with three layers of neural nets to model factor loadings but note that the conclusions remain robust for alternative specifications.

[12] In our baseline analysis, we focus on a tangency portfolio of factors, defined as the 1/N average of all factors. Similarly, when combining the probability LS portfolio with factors, we compute the 1/N average of all factors alongside the 'H-L' portfolio. DeMiguel et al. (2009) and Yuan and Zhou (2023) show that '1/N' is not a naive benchmark and can be hard to beat. Our results are robust if we consider a real-time mean-variance combination of 'H-L' portfolios with factors.

First, the H-L' portfolio generally exhibits low or even negative correlations with factor tangency portfolios. Moving to the factor tangency portfolios, some recently proposed factor models demonstrate the ability to generate portfolios with high Sharpe ratios. Among the prespecified models, the HXZ model achieves the highest tangency Sharpe ratio of 1.47. Even in this case, combining the probability-based H-L' portfolio with the factors results in a portfolio with an information ratio exceeding 1, relative to the baseline.

This pattern holds consistently across other factor models, including the recently proposed machine learning-based IPCA and CA models. In most cases, the combined portfolio achieves an information ratio above 1 during the out-of-sample period, leading to higher Sharpe ratios across the board.

From an investment perspective, probability forecasts offer valuable enhancements to existing factor models. From an asset pricing perspective, these forecasts enable the identification of mispricing at the individual stock level relative to baseline factor models. This approach not only improves predictive performance but also facilitates a more interpretable diagnosis of existing asset pricing theories.

5 Forecasting and Managing Tail Risks

This section uses probability forecasts to manage tail risks. Specifically, we focus on predicting the probability that a stock will experience a large decline in the future from 1 month to 1 year under the physical probability measure:

$$\mathbb{P}(R_{i,t+1}^{horizon} < q) = S(f(z_{i,t};\theta)), \tag{26}$$

where horizon represents the different future return time periods, q is a threshold defined by researchers, and $S(\cdot)$ is the sigmoid function mapping the output from $f(\cdot)$ into a probability value between 0 and 1.

Following the literature on default prediction, we adopt the reduced-form approach by utilizing a broad range of individual stock characteristics to directly model the probability of large stock price declines. The trajectory of reduced-form modeling has seen notable milestones. Early frameworks were built on methods like discriminant analysis (Altman (1968); Deakin (1972)) and later incorporated probabilistic models, such as logit (Ohlson (1980)) and probit (Zmijewski (1984)). A pivotal shift occurred with the introduction of Shumway (2001) dynamic hazard model, which established itself as a dominant methodology for forecasting defaults and inspired later literature on default prediction.

Following Shumway (2001), we employ a dynamic logit model that allows for time-varying predictors and accounts for the panel structure of our data. The model produces easily interpretable probability estimates that can be directly used for portfolio formation and risk management.

To benchmark the performance of our tail risk forecasts, we construct a set of 15 characteristics identified in the literature as predictors of stock crashes or financial distress.[13]

For empirical results, we focus on the case where $q = -20\%$, and consider three different horizons: future 1 month, 3 months, and 12 months. We used a logistic regression model to forecast tail risk probabilities.[14]

Table 7 presents the performance of the tail risk forecasts. From Panel A, we find a large spread in model prediction for stocks experiencing more than 20% decline versus those not. To evaluate the economic gains from the tail risk forecasts, we sort stocks based on the out-of-sample tail risk forecast into decile portfolios. We find that stocks in decile 1 have about 1/4 of the volatility of those in decile 10. The portfolio maximum drawdown is -36.2% for the lowest decile versus -99.9% for the highest decile. In the same sample period, the maximum drawdown of the market portfolio is -54.4%

We plot the cumulative log return of the lowest decile tail risk portfolio against the market return in Figure 5. We find that the portfolio avoids the decline after the internet bubble and

[13]We provide detailed descriptions of these characteristics, along with references to the literature in Online Appendix Table A2.

[14]Our conclusions remain robust under alternative q's and for using PLS and NN as forecasting models.

has lower volatility in the 2008 financial crisis and the Covid crisis. Leveraging information in a large number of characteristics, the final long-only portfolio aimed at reducing tail risk exposure outperforms the market across different horizons. For the monthly tail risk forecast, we find the long-only portfolio has a Sharpe ratio that is 45.6% higher than the baseline market Sharpe ratio.

Finally, we compare our tail risk forecast against the baseline in the literature. To evaluate whether the measure contains new information, we run the following regression:

$$\mathbb{I}\left\{R_{i,t+1}^{\text{horizon}} < -0.2\right\} = \alpha_{t+1} + \beta \text{ TailRisk}_{it} + \gamma \text{Controls}_{it} + \varepsilon_{i,t+1}, \tag{27}$$

where TailRisk is the out-of-sample prediction from Equation (26) using all 94 predictor variables. To control against the existing variables in the literature, we consider two approaches. In the first one, we fit the same prediction model as Equation (26) using all variables in the tail risk literature. In the second approach, we include all 15 individual variables in the regression.

Table 8 presents the results of the tail risk forecasting regression. We find that across all specifications, our new measure is the most significant and has an economic magnitude of more than 3 times larger than using only existing variables documented in the literature.

Finally, we probe into the prediction model to evaluate which variables matter for tail risk forecasts and how they differ from our baseline results on the probability of outperforming the market. Using the same Shapley value decomposition framework in Equation (17), we plot the variable importance for tail risk forecast against our baseline results in Figure 4.

Beyond return volatility, we find that the bid-ask spread and idiosyncratic volatility play significant roles in forecasting tail risk among all three horizons. While the bid-ask spread is less critical in predicting the probability of outperforming the market, it proves highly relevant for tail risk. The finding suggests that because of adverse selection, market makers' equilibrium quotes contain valuable information about the future crash risk of a stock. Additionally, the significance of idiosyncratic volatility indicates that a more detailed breakdown of volatility could improve tail risk forecasting accuracy for individual stocks.

6 Conclusion

In this paper, we focus on modeling the probability a stock is going to outperform a benchmark and predict this probability in the cross-section with a large number of firm characteristics. This contrasts with the existing literature that focuses mainly on modeling the expected return dynamics.

Probability forecasting has a clear economic interpretation: it directly captures the information ratio of a stock relative to a benchmark. While expected return forecasts focus solely on return dynamics, our probability target naturally incorporates both return and volatility predictability. This shift in modeling objective reveals distinct cross-sectional patterns in asset returns that are not apparent when examining expected returns alone.

Empirically, we find that forecasting probability can yield significant long-short portfolio return performance that is comparable with those from expected return forecasts. It also generates incremental information relative to the expected return forecasts. Combining probability forecasts with expected return forecasts can generate long-short portfolios with significantly better Sharpe ratios.

We further demonstrate that probability forecasts can enhance existing factor models by identifying individual stocks that are most difficult to price. Through modeling the probability of outperforming factor-implied benchmark returns, our approach enables a transparent assessment of potential mispricing at the individual stock level. Combining probability-based strategies with factor portfolios consistently improves Sharpe ratios relative to baseline factor models.

Our probability forecasting framework also proves valuable for managing tail risks. By directly modeling the probability of large price declines. These probability forecasts, which incorporate a comprehensive set of firm characteristics, significantly outperform existing predictors in the literature. The resulting low tail risk portfolio not only achieves superior risk-adjusted returns but also has significantly lower maximum drawdown and volatility.

Interestingly, we find that, in the context of probability forecast, a simple model, such as Logistic regression, can achieve predictive and portfolio performance similar to a highly complex

neural network model. The results suggest that the probability of outperforming a benchmark is a simpler object to learn.

Given their incremental information value, future research could explore the wide application of probability forecasts in the cross-section. These forecasts can address fundamental questions across multiple asset classes, including stocks, corporate bonds, and currencies. While our analysis focuses on equities, the framework's ability to capture both return and risk dynamics while maintaining model simplicity makes it promising for broader applications in asset pricing and risk management.

References

Altman, Edward I, 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* 23, 589–609.

Ang, Andrew, and Allan Timmermann, 2012, Regime changes and financial markets, *Annu. Rev. Financ. Econ.* 4, 313–337.

Asquith, Paul, Parag A Pathak, and Jay R Ritter, 2005, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Avramov, Doron, Si Cheng, Lior Metzker, and Stefan Voigt, 2023, Integrating factor models, *The Journal of Finance* 78, 1593–1646.

Bali, Turan G, Heiner Beckmeyer, Mathis Moerke, and Florian Weigert, 2023, Option return predictability with machine learning and big data, *The Review of Financial Studies* 36, 3548–3602.

Bali, Turan G, Amit Goyal, Dashan Huang, Fuwei Jiang, and Quan Wen, 2020, Predicting corporate bond returns: Merton meets machine learning, *Georgetown McDonough School of Business Research Paper* 20–110.

Banz, Rolf W, 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *The Journal of Finance* 43, 507–528.

Black, Fischer, 1986, Noise, *The Journal of Finance* 41, 528–543.

Breen, William, Lawrence R Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *The Journal of Finance* 44, 1177–1189.

Campbell, John Y, Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *The Journal of Finance* 63, 2899–2939.

Catania, Leopoldo, Stefano Grassi, and Francesco Ravazzolo, 2019, Forecasting cryptocurrencies under model and parameter instability, *International Journal of Forecasting* 35, 485–501.

Chen, Joseph, Harrison Hong, and Jeremy C Stein, 2001, Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices, *Journal of Financial Economics* 61, 345–381.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2023, Deep learning in asset pricing, *Management Science* .

Chevapatrakul, Thanaset, 2013, Return sign forecasts based on conditional risk: Evidence from the uk stock market index, *Journal of Banking & Finance* 37, 2342–2353.

Chinco, Alex, Adam D Clark-Joseph, and Mao Ye, 2019, Sparse signals in the cross-section of returns, *The Journal of Finance* 74, 449–492.

Christoffersen, Peter F, and Francis X Diebold, 2006, Financial asset returns, direction-of-change forecasting, and volatility dynamics, *Management Science* 52, 1273–1287.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang, 2021, Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai, *Available at SSRN 3554486* .

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short-and long-horizon behavioral factors, *The Review of Financial Studies* 33, 1673–1736.

Deakin, Edward B, 1972, A discriminant analysis of predictors of business failure, *Journal of Accounting Research* 167–179.

Del Viva, Luca, Carlo Sala, and André Souza, 2024, Directional information in equity returns, *Available at SSRN 4575793* .

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The Review of Financial Studies* 22, 1915–1953.

Dong, Xi, Yan Li, David E Rapach, and Guofu Zhou, 2022, Anomalies and the expected market return, *The Journal of Finance* 77, 639–681.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and Kenneth R French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *The Journal of Finance* 75, 1327–1370.

Filippou, Ilias, David Rapach, Mark P Taylor, and Guofu Zhou, 2020, Exchange rate prediction with machine learning and a smart carry trade portfolio .

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *The Review of Financial Studies* 33, 2326–2377.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *The Review of Financial Studies* 34, 3456–3496.

Green, Jeremiah, John RM Hand, and X Frank Zhang, 2017, The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* 30, 4389–4436.

Greenwood, Robin, Andrei Shleifer, and Yang You, 2019, Bubbles for fama, *Journal of Financial Economics* 131, 20–43.

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

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2021, Autoencoder asset pricing models, *Journal of Econometrics* 222, 429–450.

Han, Yufeng, Ai He, David Rapach, and Guofu Zhou, 2024, Cross-sectional expected returns: New fama-macbeth regressions in the era of machine learning, *Review of Finance* 28, 1807–1831.

Haugen, Robert A, and Nardin L Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An augmented q-factor model with expected growth, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *The Review of Financial Studies* 33, 2019–2133.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *The Review of Financial Studies* 28, 791–837.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *The Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65–91.

Kaniel, Ron, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh, 2023, Machine-learning the skill of mutual fund managers, *Journal of Financial Economics* 150, 94–138.

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *The Journal of Finance* 68, 1721–1756.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

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

Leung, Mark T, Hazem Daouk, and An-Sing Chen, 2000, Forecasting stock indices: a comparison of classification and level estimation models, *International Journal of Forecasting* 16, 173–190.

Lewellen, Jonathan, 2014, The cross section of expected stock returns, *Critical Finance Review* .

Light, Nathaniel, Denys Maslov, and Oleg Rytchkov, 2017, Aggregation of information about the cross section of stock returns: A latent variable approach, *The Review of Financial Studies* 30, 1339–1381.

Lundberg, Scott M, and Su-In Lee, 2017, A unified approach to interpreting model predictions, *Advances in neural information processing systems* 30.

Moskowitz, Tobias J, Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Ohlson, JA, 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* .

Papailias, Fotis, Jiadong Liu, and Dimitrios D Thomakos, 2021, Return signal momentum, *Journal of Banking & Finance* 124, 106063.

Pesaran, M Hashem, and Allan Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, *The Journal of Finance* 50, 1201–1228.

Pesaran, M Hashem, and Allan Timmermann, 2004, How costly is it to ignore breaks when forecasting the direction of a time series?, *International Journal of Forecasting* 20, 411–425.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–16.

Shapley, Lloyd S, 1953, A value for n-person games .

Sharpe, William F, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425–442.

Shumway, Tyler, 2001, Forecasting bankruptcy more accurately: A simple hazard model, *The Journal of Business* 74, 101–124.

Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.

Wu, Wenbo, Jiaqi Chen, Zhibin Yang, and Michael L Tindall, 2021, A cross-sectional machine learning approach for hedge fund return prediction and selection, *Management Science* 67, 4577–4601.

Yuan, Ming, and Guofu Zhou, 2023, Why naive diversification is not so naive, and how to beat it?, *Journal of Financial and Quantitative Analysis* 1–32.

Zmijewski, Mark E, 1984, Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research* 59–82.

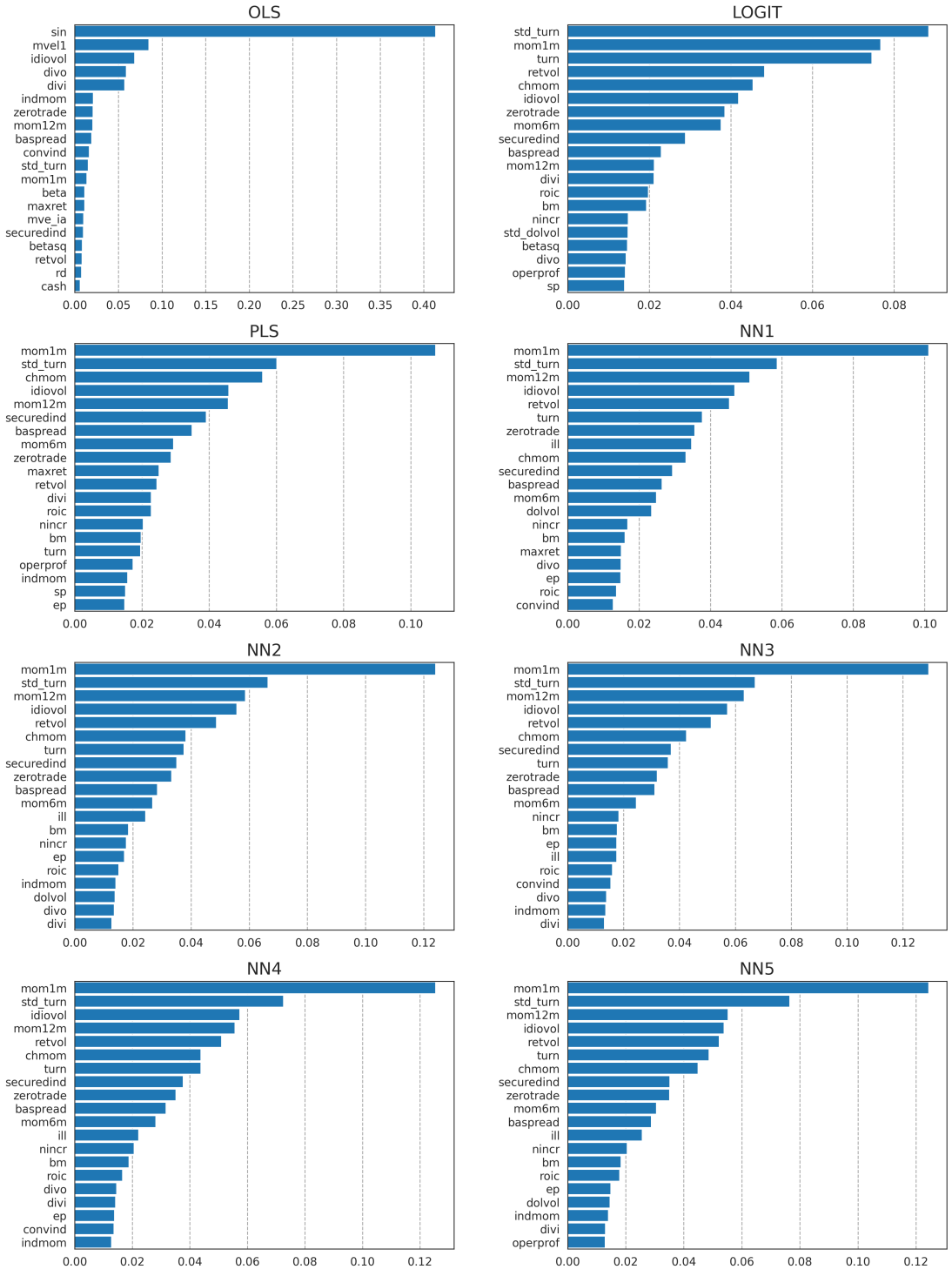

Figure 1 Shapley Values for Variable Importance across Different Models

This figure presents the variable importance for different prediction models, which is calculated following the SHAP framework by Lundberg and Lee (2017). We take an average of the absolute value of the Shapley measure as the variable importance across all observations in our testing sample. We then standardize the variable importance so that they sum up to one.

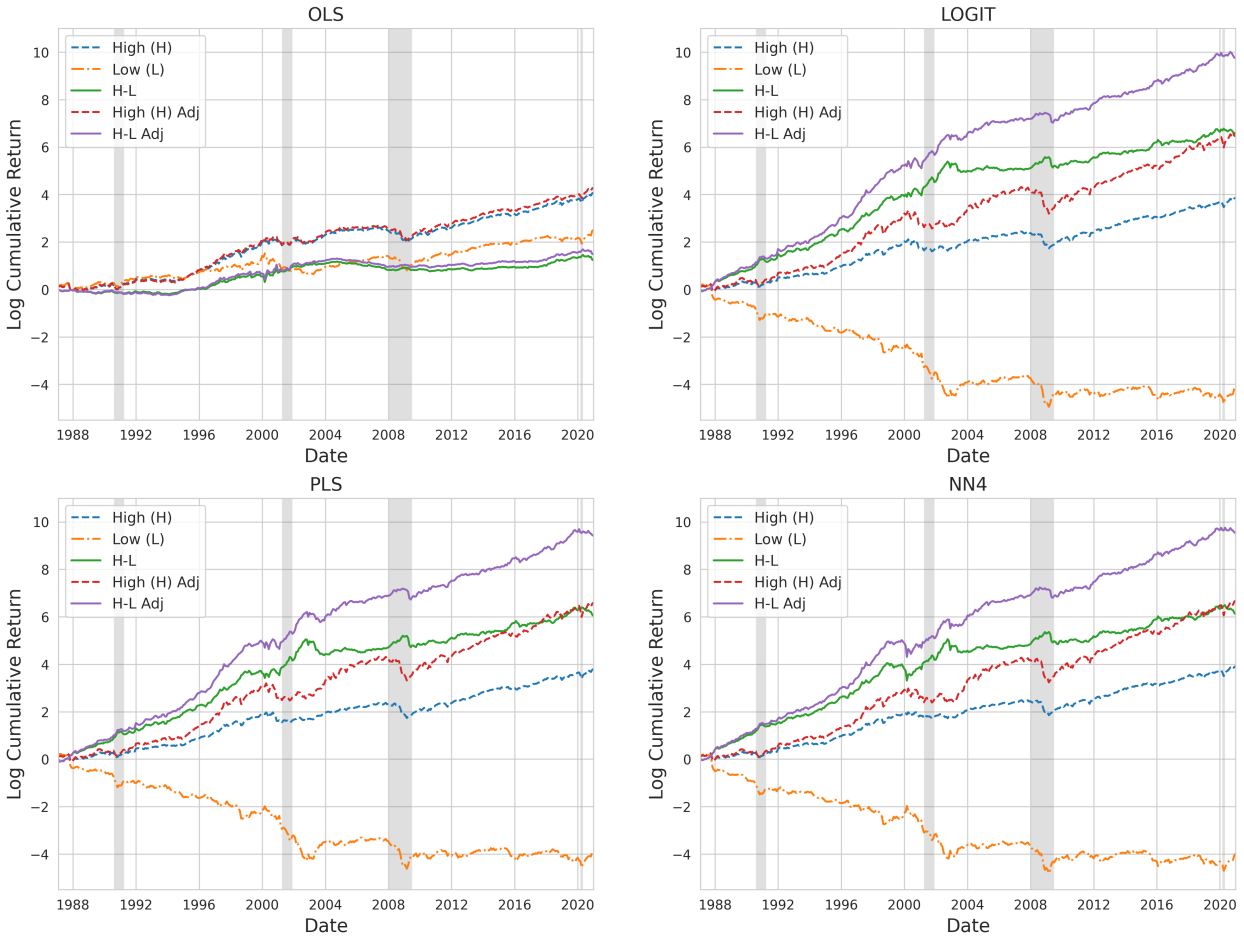

Figure 2 Cumulative Log-return of Portfolios Sorted on Probability Forecasts

The figure displays the cumulative log excess returns of portfolios sorted on probability forecasts for different models. In each figure, we report five time series: (1) the decile ten portfolio return (High (H)); (2) the decile one portfolio return (Low (L)); (3) the long-short portfolio return (H-L); (4) the variance adjusted decile ten portfolio return (High (H) Adj); (5) the variance adjusted long-short portfolio return (H-L Adj). We report the performance of four models: OLS, Logit, PLS, and NN4. The grey bars in all plots indicate NBER-dated recessions. All samples start from January 1987 and end in December 2020.

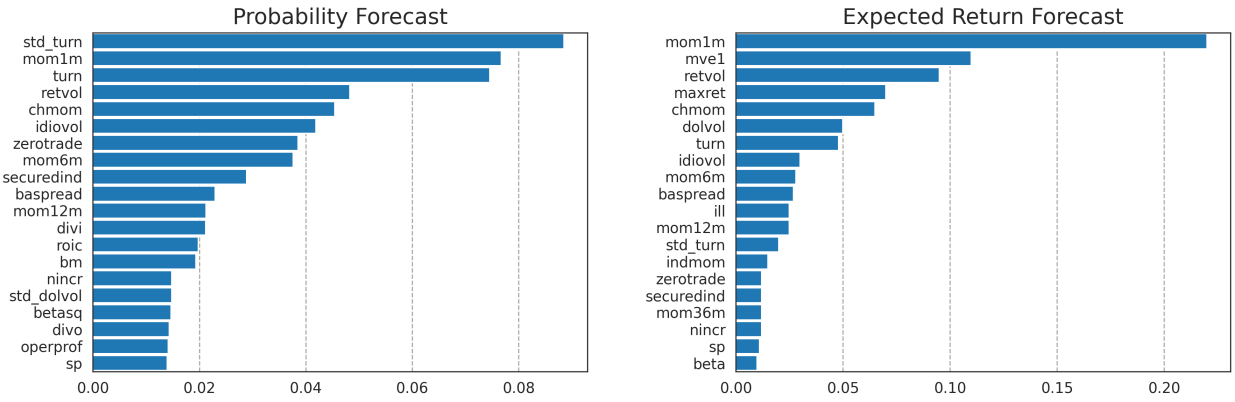

Figure 3 Comparison of Variable Importance: Probability vs. Expected Return Forecasts

This figure compares the variable importance between the probability forecast model and the expected return forecast model. We report the logistic regression model for probability forecast and the 4-layer Neural Net model for expected return forecast. We assess the variable importance using the SHAP framework proposed by Lundberg and Lee (2017). For clarity, the variable importance values are standardized to sum to one within each model.

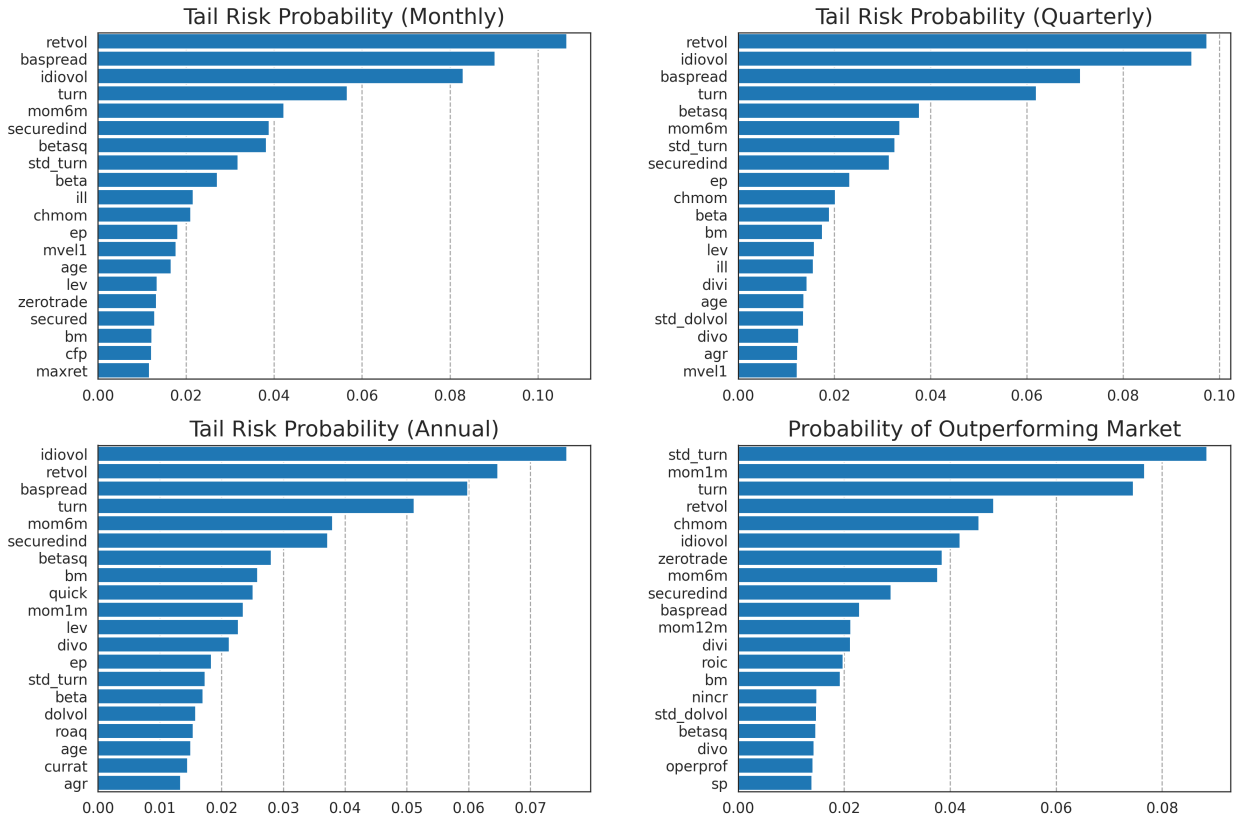

Figure 4 Difference in Variable Importance for Tail Risk Forecasting Models

This figure compares the variable importance between the models forecasting tail risk and the model predicting the probability of outperforming the market. Tail risk probability is defined as the probability that stock returns fall below -20% over three horizons: next month (Monthly), quarter (Quarterly), or year (Annual). We assess the variable importance using the SHAP framework proposed by Lundberg and Lee (2017). Both models employ logistic regression to forecast probabilities. For clarity, the variable importance values are standardized to sum to one within each model.

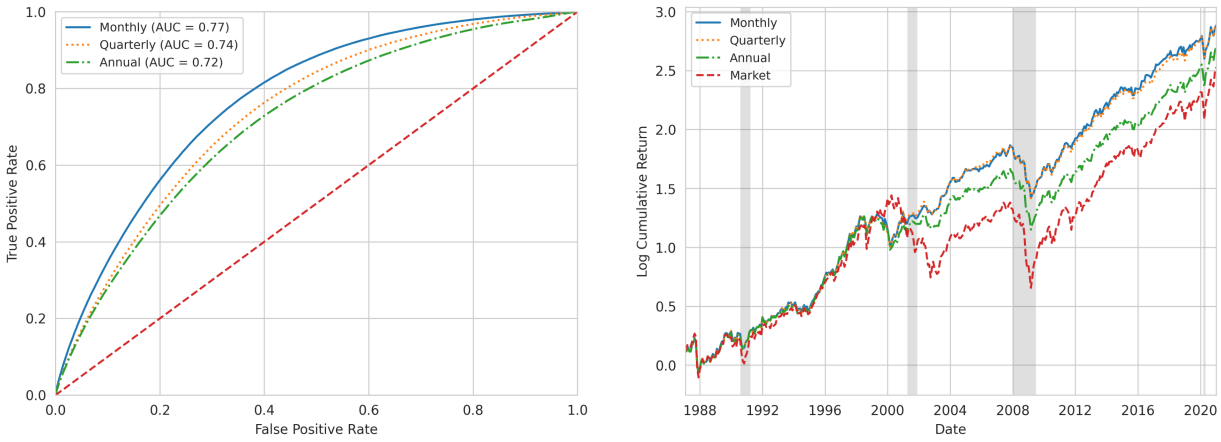

Figure 5 ROC Curves and Cumulative Portfolio Performance for Tail Risk Forecasts

This figure presents the statistical performance and economic gains from tail risk forecasts. Tail risk probability is defined as the likelihood of stock returns falling below -20% over the next month, quarter, or year ('Monthly', 'Quarterly', or 'Annual' lines in the graphs). The left panel displays the ROC curves for the tail risk forecast. The right panel shows the log cumulative returns of a decile portfolio formed by sorting stocks based on their tail risk forecasts, comparing the portfolio with the lowest tail risk to the market benchmark. Both panels use data from the out-of-sample period spanning January 1987 to December 2020. The grey bars in the right panel indicate NBER-dated recessions.

Table 1 Summary Statistics of the Probability Forecasts

This table presents the summary statistics of the probability forecasts and their targets. The first row, '% Beat Mkt', reports the summary statistics for the proportion of stocks beating the market across time. 'β' represents stocks' realtime CAPM βs estimated following Fama and MacBeth (1973). The next eight rows report the probability forecasts from the eight forecasting models (OLS, Logit, PLS, and NN1-5). The statistics are reported for an out-of-sample testing period starting from January 1987 and ending in December 2020.

	Mean	Std	Median	Skew	Kurt
% Beat Mkt	0.464	0.086	0.460	0.308	0.279
β	1.003	0.687	1.000	3.963	164.673
Prob OLS	0.371	0.183	0.431	-1.184	0.661
Prob Logit	0.459	0.044	0.463	-0.384	0.055
Prob PLS	0.459	0.043	0.463	-0.425	0.159
Prob NN1	0.461	0.046	0.466	-0.521	0.237
Prob NN2	0.456	0.043	0.461	-0.585	0.041
Prob NN3	0.449	0.045	0.454	-0.597	0.137
Prob NN4	0.446	0.046	0.451	-0.537	-0.010
Prob NN5	0.445	0.043	0.449	-0.456	-0.014

Table 2 Summary Statistics of the Forecasting Accuracy of the Probability Forecasts

This table presents the prediction accuracy of the eight models considered in this paper (OLS, Logit, PLS, and NN1-5). The 'Pos' ('Neg') row reports the model prediction conditional on stock return outperforming (underperforming) the market returns. Prediction accuracy is computed as the ratio of the number of accurate predictions to the total number of predictions. Additionally, the cross-entropy is calculated using the formula: $CE = -\frac{1}{|\mathcal{T}_3|}\sum_{(i,t)\in\mathcal{T}_3}\left(y_{it}\log(g(z_{i,t};\hat{\theta})) + (1-y_{it})\log(1-g(z_{i,t};\hat{\theta}))\right)$. The statistics are reported for our out-of-sample testing period starting from January 1987 and ending in December 2020.

	OLS		Logit		PLS		NN1	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Pos	0.239	0.761	0.191	0.809	0.182	0.818	0.219	0.781
Neg	0.217	0.783	0.163	0.837	0.155	0.845	0.190	0.810
Accuracy	0.540		0.540		0.540		0.537	
Cross-Entropy	0.689		0.689		0.689		0.689	
	NN2		NN3		NN4		NN5	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Pos	0.165	0.835	0.136	0.864	0.126	0.874	0.101	0.899
Neg	0.143	0.857	0.116	0.884	0.108	0.892	0.086	0.914
Accuracy	0.538		0.539		0.539		0.540	
Cross-Entropy	0.689		0.689		0.690		0.689	

Table 3 Performance of the Probability Forecasts Portfolios

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast. The prediction target is the probability of a stock outperforming the market. For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability (*Prob*), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.45	0.74	5.01	0.51	0.38	0.39	-0.69	7.77	-0.31	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.75	8.11	-0.32
2	0.14	0.44	0.31	5.48	0.19	0.41	0.43	0.14	7.73	0.06	0.41	0.43	0.00	7.77	0.00	0.41	0.43	-0.05	7.95	-0.02
3	0.29	0.44	0.10	5.20	0.07	0.43	0.44	0.48	7.72	0.21	0.43	0.44	0.59	7.76	0.26	0.43	0.44	0.49	8.02	0.21
4	0.38	0.44	0.34	5.25	0.23	0.45	0.45	0.62	7.64	0.28	0.45	0.45	0.63	7.69	0.28	0.45	0.46	0.51	8.08	0.22
5	0.42	0.45	0.53	5.10	0.36	0.46	0.46	0.79	7.63	0.36	0.46	0.46	0.71	7.67	0.32	0.46	0.46	0.76	7.99	0.33
6	0.44	0.46	0.51	5.15	0.35	0.47	0.47	0.93	7.71	0.42	0.47	0.47	0.95	7.62	0.43	0.47	0.47	0.94	7.87	0.42
7	0.47	0.47	0.63	5.04	0.43	0.48	0.48	1.23	7.49	0.57	0.48	0.48	1.12	7.55	0.52	0.48	0.48	1.22	7.91	0.54
8	0.49	0.48	0.73	4.98	0.51	0.49	0.48	1.46	7.40	0.69	0.49	0.48	1.30	7.41	0.61	0.49	0.48	1.35	7.78	0.60
9	0.52	0.48	0.89	4.87	0.63	0.5	0.49	1.57	7.36	0.74	0.5	0.49	1.55	7.42	0.72	0.51	0.49	1.51	7.62	0.69
High (H)	0.58	0.49	1.17	4.86	0.84	0.53	0.51	1.92	7.45	0.89	0.53	0.51	1.91	7.48	0.89	0.53	0.51	2.00	7.70	0.90
H-L	-	-	0.43	3.66	0.41	-	-	2.61	6.19	1.46	-	-	2.54	6.44	1.36	-	-	2.75	6.32	1.51
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.37	0.39	-0.63	8.04	-0.27	0.37	0.39	-0.75	8.16	-0.32	0.37	0.39	-0.64	8.03	-0.27	0.37	0.39	-0.62	8.27	-0.26
2	0.41	0.43	0.24	7.99	0.10	0.4	0.43	0.24	8.14	0.10	0.4	0.43	0.12	7.87	0.05	0.4	0.43	0.32	8.10	0.14
3	0.43	0.45	0.55	8.02	0.24	0.42	0.45	0.52	8.30	0.22	0.42	0.45	0.57	7.98	0.25	0.42	0.45	0.49	8.36	0.20
4	0.44	0.46	0.51	8.06	0.22	0.44	0.46	0.84	8.14	0.36	0.43	0.46	0.48	7.89	0.21	0.43	0.46	0.95	8.19	0.40
5	0.46	0.46	0.78	7.87	0.34	0.45	0.46	0.58	8.10	0.25	0.44	0.46	0.83	7.80	0.37	0.44	0.46	0.66	8.18	0.28
6	0.47	0.47	0.91	7.85	0.40	0.46	0.47	0.82	7.92	0.36	0.46	0.47	0.95	7.89	0.42	0.45	0.47	1.04	8.04	0.45
7	0.48	0.48	1.21	7.80	0.54	0.47	0.47	1.16	8.02	0.50	0.47	0.48	1.02	7.83	0.45	0.46	0.47	1.16	8.02	0.50
8	0.49	0.48	1.27	7.76	0.57	0.48	0.48	1.27	8.02	0.55	0.48	0.48	1.34	7.70	0.60	0.47	0.48	1.23	7.87	0.54
9	0.5	0.49	1.59	7.61	0.72	0.49	0.49	1.51	7.94	0.66	0.49	0.49	1.56	7.66	0.71	0.49	0.49	1.57	7.89	0.69
High (H)	0.52	0.5	1.93	7.76	0.86	0.51	0.5	2.10	7.95	0.91	0.5	0.5	1.95	7.70	0.88	0.5	0.5	2.01	7.92	0.88
H-L	-	-	2.56	6.57	1.35	-	-	2.84	6.86	1.44	-	-	2.59	6.74	1.33	-	-	2.63	7.15	1.27

Table 4 Combining Probability Forecasts with Expected Return Forecasts

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts. The prediction target is the probability of a stock outperforming the market. The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.25	0.34	0.34	0.27	0.32	0.37	0.33	0.30
Panel A: Probability Forecast								
Mean	0.43	2.61	2.54	2.75	2.56	2.84	2.59	2.63
SR	0.41	1.46	1.36	1.51	1.35	1.44	1.33	1.27
α	0.42	1.81	1.73	1.88	1.64	1.85	1.69	1.55
t_α	2.15	5.72	5.60	6.62	5.35	5.49	5.52	5.18
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.71	2.80	2.77	2.50	2.62	2.85	2.79	2.76
SR	1.33	1.76	1.71	1.68	1.58	1.65	1.69	1.62
α	1.57	2.27	2.22	1.82	1.96	2.19	2.21	2.04
t_α	4.36	5.65	5.58	5.99	4.84	4.98	5.73	5.27
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.62	9.62	9.36	9.02	9.21	9.17	9.24	9.09
SR	1.38	1.73	1.68	1.68	1.63	1.65	1.65	1.63
α	6.53	8.29	8.18	6.89	7.61	7.57	7.91	7.53
t_α	4.56	5.37	5.19	5.63	5.00	5.00	5.15	4.94

Table 5 Performance of the Probability Forecasts Portfolios (Probability of Outperforming Factor Models)

This table presents the performance of value-weighted decile portfolios sorted by the predicted probability of outperforming benchmark factor models. We evaluate eight benchmark factor models, including six predefined models (CAPM, FF3, FF6, HXZ, DHS, SY) and two machine-learning-based latent factor models (IPCA and CA). The prediction target is the probability that a stock outperforms its real-time estimated β times the contemporaneous factor return plus the risk-free rate. For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability (*Prob*), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	CAPM					FF3					FF6					HXZ				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.37	0.39	-0.56	8.07	-0.24	0.38	0.40	-0.67	7.96	-0.29	0.39	0.41	-0.60	7.20	-0.29	0.41	0.42	-0.58	6.87	-0.29
2	0.41	0.42	0.14	8.13	0.06	0.41	0.43	0.09	7.99	0.04	0.43	0.45	0.33	7.29	0.15	0.44	0.45	0.19	6.98	0.09
3	0.43	0.44	0.44	8.13	0.19	0.43	0.45	0.30	8.05	0.13	0.45	0.46	0.36	7.15	0.18	0.45	0.46	0.32	6.91	0.16
4	0.45	0.45	0.54	8.01	0.23	0.45	0.46	0.51	7.92	0.22	0.46	0.47	0.59	7.20	0.28	0.46	0.48	0.75	6.86	0.38
5	0.46	0.47	0.87	8.00	0.38	0.46	0.47	0.79	7.95	0.34	0.47	0.48	0.73	7.15	0.35	0.47	0.48	0.77	6.87	0.39
6	0.47	0.48	0.82	8.06	0.35	0.47	0.48	0.97	7.81	0.43	0.48	0.49	1.05	7.11	0.51	0.48	0.49	1.18	6.81	0.60
7	0.49	0.49	1.51	7.74	0.68	0.48	0.48	1.18	7.77	0.53	0.49	0.49	1.18	7.03	0.58	0.49	0.50	1.23	6.70	0.64
8	0.50	0.50	1.62	7.60	0.74	0.49	0.49	1.36	7.69	0.61	0.50	0.50	1.31	6.98	0.65	0.50	0.51	1.48	6.71	0.76
9	0.51	0.51	1.72	7.67	0.78	0.50	0.51	1.72	7.68	0.77	0.51	0.51	1.47	6.95	0.73	0.52	0.51	1.55	6.68	0.80
High (H)	0.54	0.53	2.18	7.76	0.97	0.53	0.52	1.99	7.73	0.89	0.54	0.53	1.72	7.09	0.84	0.54	0.53	1.70	6.83	0.86
H-L	-	-	2.74	6.90	1.37	-	-	2.66	6.50	1.42	-	-	2.32	6.02	1.34	-	-	2.28	5.61	1.41
	DHS					SY					IPCA					CA				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.40	0.42	-0.41	6.81	-0.21	0.39	0.41	-0.64	7.38	-0.30	0.39	0.42	-0.05	9.95	-0.02	0.39	0.41	-0.93	8.06	-0.40
2	0.44	0.45	0.19	7.06	0.09	0.42	0.44	0.18	7.37	0.08	0.42	0.43	0.13	9.87	0.04	0.42	0.43	-0.33	8.18	-0.14
3	0.45	0.47	0.35	7.01	0.17	0.44	0.46	0.56	7.22	0.27	0.43	0.44	0.15	9.83	0.05	0.44	0.45	-0.35	8.46	-0.14
4	0.47	0.48	0.70	6.86	0.35	0.45	0.47	0.38	7.35	0.18	0.44	0.45	0.55	9.73	0.19	0.45	0.46	-0.03	8.38	-0.01
5	0.48	0.49	0.92	6.81	0.47	0.46	0.48	0.71	7.28	0.34	0.45	0.46	0.98	9.62	0.35	0.46	0.47	0.48	8.26	0.20
6	0.49	0.49	1.17	6.75	0.60	0.47	0.48	0.98	7.25	0.47	0.46	0.47	1.02	9.59	0.37	0.47	0.48	0.64	8.05	0.28
7	0.50	0.50	1.28	6.69	0.67	0.48	0.49	1.21	7.12	0.59	0.47	0.48	1.39	9.72	0.49	0.48	0.48	0.79	7.97	0.34
8	0.51	0.51	1.43	6.61	0.75	0.49	0.50	1.41	7.08	0.69	0.48	0.48	1.53	9.34	0.57	0.49	0.49	0.93	8.02	0.40
9	0.52	0.52	1.58	6.62	0.83	0.51	0.51	1.47	7.08	0.72	0.49	0.49	1.53	9.17	0.58	0.50	0.50	1.33	7.79	0.59
High (H)	0.54	0.53	1.79	6.76	0.92	0.53	0.52	1.84	7.18	0.89	0.51	0.51	2.12	9.11	0.81	0.52	0.51	1.65	7.58	0.75
H-L	-	-	2.20	5.67	1.34	-	-	2.48	6.27	1.37	-	-	2.17	9.00	0.84	-	-	2.58	6.42	1.39

Table 6 Combining Probability Forecasts (Probability of Outperforming Factor Models) with Factors

This table presents the incremental information provided by probability-based long-short (LS) portfolios compared to baseline factor models. We evaluate eight benchmark factor models, including six predefined models (CAPM, FF3, FF6, HXZ, DHS, SY) and two machine-learning-based latent factor models (IPCA and CA). Each column corresponds to a specific factor model. The first row reports the correlation between the H-L portfolio of probability forecasts and the tangency portfolios of factors. Panel A reports the value-weighted 'H-L' portfolio's mean, standard deviation (SD), and Sharpe ratio (SR) for the probability-based LS portfolios. Panel B presents the characteristics of the '1/N' tangency portfolios of all the factors. Panel C provides results for the '1/N' combination of probability-based LS portfolios and all the factors. The 'IR' row reports the information ratio of the combined portfolio in Panel C relative to the baseline in Panel B. All samples start from January 1987 and end in December 2020.

	CAPM	FF3	FF6	HXZ	SY	DHS	IPCA	CA
Corr	0.06	-0.11	0.20	0.28	0.20	0.29	-0.28	-0.17
Panel A: Probability 'H-L' Portfolios								
Mean	2.74	2.66	2.32	2.28	2.20	2.48	2.17	2.58
SD	6.90	6.50	6.02	5.61	5.67	6.27	9.00	6.42
SR	1.37	1.42	1.34	1.41	1.34	1.37	0.84	1.39
Panel B: Factor Tangency Portfolios								
Mean	0.73	0.29	0.31	0.44	0.59	0.51	4.42	3.81
SD	4.51	2.05	1.18	1.03	1.48	1.31	9.00	6.42
SR	0.56	0.49	0.92	1.47	1.38	1.35	1.70	2.05
Panel C: Combination of Probability 'H-L' and Factor Tangency Portfolios								
Mean	1.73	0.88	0.60	0.74	0.99	0.91	3.30	3.20
SD	4.24	2.11	1.45	1.44	1.97	1.86	5.39	5.08
SR	1.42	1.45	1.43	1.79	1.74	1.69	2.12	2.18
IR	1.30	1.36	1.09	1.02	1.07	1.01	1.26	0.74

Table 7 Performance and Economic Gains from Tail Risk Forecasts

This table summarizes the statistical accuracy and economic benefits derived from tail risk forecasts. The forecasts are evaluated over three horizons: one month, one quarter, and one year, with tail risk defined as stock returns falling below -20% during the corresponding horizon. Panel A presents the statistical performance of the forecasts. The 'Pos' ('Neg') row reports the probability forecast conditional on stock return is below (above) -20%. Additionally, cross-entropy and AUC (area under the ROC curve) are provided to assess forecast accuracy. Panel B focuses on the economic implications of these forecasts. Stocks are sorted into decile portfolios based on out-of-sample tail risk predictions. For each horizon, we report results for two portfolios: decile 1 ('L'), which contains stocks with the lowest predicted tail risk, and decile 10 ('H'), which contains stocks with the highest predicted tail risk. The statistics include the model-predicted probability ($\widehat{\text{Prob}}$) and realized probability (Prob) of a return drop exceeding 20%, as well as the mean return (%), standard deviation (%), Sharpe ratio (SR), and maximum drawdown (MaxDD) for each portfolio. The analysis covers the out-of-sample testing period from January 1987 to December 2020.

Panel A: Statistical Performance of Tail Risk Forecasts

	$\widehat{\mathbb{P}}\{R^{1m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{3m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{12m}_{i,t+1} < -0.2\}$	
	Mean	Std	Mean	Std	Mean	Std
Pos	0.115	0.078	0.221	0.117	0.337	0.153
Neg	0.053	0.057	0.130	0.102	0.224	0.138
Cross-Entropy	0.212		0.374		0.518	
AUC	0.773		0.737		0.716	

Panel B: Portfolios Sorted on Tail Risk Forecasts

	$\widehat{\mathbb{P}}\{R^{1m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{3m}_{i,t+1} < -0.2\}$		$\widehat{\mathbb{P}}\{R^{12m}_{i,t+1} < -0.2\}$	
	L	H	L	H	L	H
$\widehat{\text{Prob}}$	0.006	0.198	0.028	0.376	0.072	0.554
Prob	0.005	0.204	0.019	0.356	0.062	0.523
Mean	0.763	-0.519	0.765	-0.524	0.723	-0.295
Std	3.257	11.635	3.274	11.297	3.418	10.988
SR	0.811	-0.155	0.809	-0.161	0.733	-0.093
MaxDD	-0.362	-0.999	-0.361	-0.999	-0.402	-0.997

Table 8 Regression Test of Tail Risk Probability Forecasts Relative to Benchmarks

This table presents the regression results of the following model:

$$\mathbb{I}\{R_{i,t+1}^{horizon} < -0.2\} = \alpha_{t+1} + \beta TailRisk_{it} + \gamma \text{Controls}_{it} + \varepsilon_{i,t+1},$$

where the dependent variable is an indicator for whether the stock price declines by more than 20% over three horizons: next month, quarter, or year ($horizon = 1m, 3m, 12m$). TailRisk_{it} denotes the out-of-sample logit prediction using anomaly variables, while BaselineTail_{it} is the same model derived from control variables. Control variables include CAPM Beta, Market Capitalization (MktCap), Book-to-Market Ratio (BM), Gross Profitability (GrossProf), momentum measures (Ret1M, Ret6M, Ret12M), Return Volatility (RetVol), Turnover, Sales Growth (SalesGr), Leverage (Lev), Cash Holdings (CashHold), Stock Price (Price), Cashflow-to-Market Ratio (CFtoMkt), and Short Interest (ShortInt). All control variables are transformed to cross-sectional ranks and rescaled to the interval $[-1, 1]$. The regressions include monthly fixed effects, and standard errors are two-way clustered by permno and date. T-statistics are reported in parentheses. A single asterisk (*) indicates that the absolute value of the t-statistic is greater than 4.

Dependent Variables	$\mathbb{I}\{R_{i,t+1}^{1m} < -0.2\}$			$\mathbb{I}\{R_{i,t+1}^{3m} < -0.2\}$			$\mathbb{I}\{R_{i,t+1}^{12m} < -0.2\}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TailRisk	1.04*	0.83*	0.76*	0.98*	0.78*	0.68*	0.97*	0.76*	0.66*
	(29.54)	(17.28)	(22.10)	(41.80)	(25.62)	(27.72)	(53.16)	(32.88)	(34.32)
BaselineTail		0.25*			0.23*			0.25*	
		(7.11)			(9.79)			(11.46)	
Beta			0.00			0.01			0.01
			(1.98)			(2.85)			(3.88)
MktCap			-0.00			-0.00			-0.00
			(-2.67)			(-0.10)			(-1.31)
BM			-0.00			-0.01			-0.01
			(-3.81)			(-4.03)			(-2.72)
GrossProf			-0.00			-0.00			-0.01*
			(-1.40)			(-2.64)			(-5.70)
Ret1M			-0.01			-0.01			-0.01
			(-3.59)			(-3.20)			(-2.36)
Ret6M			-0.00			-0.00			-0.00
			(-2.16)			(-1.16)			(-1.28)
Ret12M			-0.01			-0.00			0.00
			(-2.97)			(-1.49)			(0.77)
RetVol			0.02*			0.03*			0.05*
			(9.55)			(11.92)			(15.18)
Turnover			0.01*			0.02*			0.02*
			(7.84)			(9.26)			(9.02)
SalesGr			0.00			0.00			0.01
			(0.15)			(2.07)			(3.17)
Lev			0.00			0.00			-0.01
			(3.12)			(0.30)			(-1.62)
CashHold			-0.00			-0.00			-0.01
			(-3.58)			(-1.68)			(-2.63)
Price			-0.01*			-0.03*			-0.04*
			(-6.12)			(-9.90)			(-9.12)
CFtoMkt			-0.01*			-0.01*			-0.01
			(-4.97)			(-4.68)			(-3.14)
ShortInt			-0.00			-0.00			-0.01
			(-1.38)			(-1.42)			(-2.13)
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	14.60%	14.65%	14.82%	20.06%	20.12%	20.28%	20.45%	20.54%	20.75%

Internet Appendix for
"Empirical Asset Pricing with Probability Forecasts"

Not for Publication

Table A1 Probability Prediction Characteristics

This table reports the anomaly variables considered in this paper. We obtain the data from Gu, Kelly, and Xiu (2020) who constructed the dataset following Green et al. (2017). Gu et al. (2020) provide a detailed description of the dataset in their appendix.

Acronym	Description	Acronym	Description
absacc	Absolute accruals	mom36m	36-month momentum
acc	Working capital accruals	mom6m	6-month momentum
aeavol	Abnormal earnings announcement volume	ms	Financial statement score
age	# years since first Compustat coverage	mvel1	Size
agr	Asset growth	mve_ia	Industry-adjusted size
baspread	Bid-ask spread	niincr	Number of earnings increases
beta	Beta	operprof	Operating profitability
betasq	Beta squared	orgcap	Organizational capital
bm	Book-to-market	pchcapx_ia	Industry adjusted % change in capital expenditures
bm_ia	Industry-adjusted book to market	pchcurrat	% change in current ratio
cash	Cash holdings	pchdepr	% change in depreciation
cashdebt	Cash flow to debt	pchgm_pchsale	% change in gross margin - % change in sales
cashpr	Cash productivity	pchquick	% change in quick ratio
cfp	Cash flow to price ratio	pchsale_pchinvt	% change in sales - % change in inventory
cfp_ia	Industry-adjusted cash flow to price ratio	pchsale_pchrect	% change in sales - % change in A/R
chatoia	Industry-adjusted change in asset turnover	pchsale_pchxsga	% change in sales - % change in SG&A
chcsho	Change in shares outstanding	pchsaleinv	% change in sales-to-inventory
chempia	Industry-adjusted change in employees	pctacc	Percent accruals
chinv	Change in inventory	pricedelay	Price delay
chmom	Change in 6-month momentum	ps	Financial statements score
chpmia	Industry-adjusted change in profit margin	quick	Quick ratio
chtx	Change in tax expense	rd	R&D increase
cinvest	Corporate investment	rd_mve	R&D to market capitalization
convind	Convertible debt indicator	rd_sale	R&D to sales
currat	Current ratio	realestate	Real estate holdings
depr	Depreciation / PP&E	retvol	Return volatility
divi	Dividend initiation	roaq	Return on assets
divo	Dividend omission	roavol	Earnings volatility
dolvol	Dollar trading volume	roeq	Return on equity
dy	Dividend to price	roic	Return on invested capital
ear	Earnings announcement return	rsup	Revenue surprise
egr	Growth in common shareholder equity	salecash	Sales to cash
ep	Earnings to price	saleinv	Sales to inventory
gma	Gross profitability	salerec	Sales to receivables
grCAPX	Growth in capital expenditures	secured	Secured debt
grltnoa	Growth in long term net operating assets	securedind	Secured debt indicator
herf	Industry sales concentration	sgr	Sales growth
hire	Employee growth rate	sin	Sin stocks
idiovol	Idiosyncratic return volatility	sp	Sales to price
ill	Illiquidity	std_dolvol	Volatility of liquidity (dollar trading volume)
indmom	Industry momentum	std_turn	Volatility of liquidity (share turnover)
invest	Capital expenditures and inventory	stdacc	Accrual volatility
lev	Leverage	stdcf	Cash flow volatility
lgr	Growth in long-term debt	tang	Debt capacity/firm tangibility
maxret	Maximum daily return	tb	Tax income to book income
mom12m	12-month momentum	turn	Share turnover
mom1m	1-month momentum	zerotrade	Zero trading days

Table A2 Baseline Tail Risk Forecast Characteristics

This table reports the details of the baseline characteristics in literature we use as controls for the tail risk forecasts in Table 8. For each variable, we provide a short description along with the source literature.

Acronym	Description	Tail Risk Literature
Beta	CAPM Beta	Fama and MacBeth (1973)
MktCap	Size	Banz (1981)
BM	Book-to-market	Rosenberg, Reid, and Lanstein (1985)
GrossProf	Gross profitability	Campbell, Hilscher, and Szilagyi (2008)
Ret1M	Short-term reversal	Jegadeesh and Titman (1993)
Ret6M	6-month return momentum	Jegadeesh and Titman (1993)
Ret12M	12-month return momentum	Jegadeesh (1990)
RetVol	Return volatility	Chen, Hong, and Stein (2001)
Turnover	Share turnover dividing by number of shares outstanding	Chen, Hong, and Stein (2001)
SalesGr	Sales growth	Greenwood, Shleifer, and You (2019)
Lev	Leverage	Bhandari (1988)
CashHold	Cash holding to the market value of total assets	Campbell, Hilscher, and Szilagyi (2008)
Price	Log price per share	Campbell, Hilscher, and Szilagyi (2008)
CFtoMkt	Net income to the total market value of assets	Campbell, Hilscher, and Szilagyi (2008)
ShortInt	Short interest scaled by number of shares outstanding	Asquith, Pathak, and Ritter (2005)

Table A3 Performance of the Probability Forecasts Portfolios (Non-variance-adjusted)

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast without variance adjustment. The prediction target is the probability of a stock outperforming the market return. For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability (*Prob*), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.45	0.74	5.01	0.51	0.38	0.39	-0.69	7.77	-0.31	0.38	0.39	-0.62	7.78	-0.28	0.37	0.39	-0.75	8.11	-0.32
2	0.14	0.44	0.40	6.10	0.23	0.41	0.43	0.06	6.60	0.03	0.41	0.43	-0.04	6.81	-0.02	0.41	0.43	-0.08	7.20	-0.04
3	0.29	0.44	0.30	6.76	0.15	0.43	0.44	0.30	5.86	0.18	0.43	0.44	0.38	6.19	0.22	0.43	0.44	0.32	6.30	0.18
4	0.38	0.44	0.55	6.16	0.31	0.45	0.45	0.43	5.44	0.28	0.45	0.45	0.43	5.48	0.27	0.45	0.46	0.33	5.79	0.20
5	0.42	0.45	0.67	5.31	0.44	0.46	0.46	0.46	5.28	0.30	0.46	0.46	0.43	5.24	0.29	0.46	0.46	0.47	5.19	0.31
6	0.44	0.46	0.64	4.92	0.45	0.47	0.47	0.56	4.75	0.41	0.47	0.47	0.59	4.82	0.43	0.47	0.47	0.55	4.95	0.39
7	0.47	0.47	0.66	4.61	0.49	0.48	0.48	0.69	4.55	0.52	0.48	0.48	0.66	4.69	0.48	0.48	0.48	0.66	4.70	0.49
8	0.49	0.48	0.69	4.32	0.55	0.49	0.48	0.78	4.49	0.60	0.49	0.48	0.70	4.45	0.54	0.49	0.48	0.79	4.54	0.60
9	0.52	0.48	0.83	4.07	0.70	0.5	0.49	0.87	4.41	0.68	0.5	0.49	0.87	4.46	0.67	0.51	0.49	0.82	4.35	0.65
High (H)	0.58	0.49	1.10	4.45	0.86	0.53	0.51	1.06	4.65	0.79	0.53	0.51	1.04	4.53	0.79	0.53	0.51	1.09	4.38	0.86
H-L	-	-	0.36	3.38	0.37	-	-	1.75	5.73	1.06	-	-	1.66	5.86	0.98	-	-	1.84	6.00	1.06
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.37	0.39	-0.63	8.04	-0.27	0.37	0.39	-0.75	8.16	-0.32	0.37	0.39	-0.64	8.03	-0.27	0.37	0.39	-0.62	8.27	-0.26
2	0.41	0.43	0.18	7.18	0.09	0.4	0.43	0.19	7.37	0.09	0.4	0.43	0.09	7.45	0.04	0.4	0.43	0.23	7.44	0.11
3	0.43	0.45	0.38	6.43	0.21	0.42	0.45	0.37	6.42	0.20	0.42	0.45	0.42	6.27	0.23	0.42	0.45	0.34	6.33	0.19
4	0.44	0.46	0.37	5.57	0.23	0.44	0.46	0.58	5.88	0.34	0.43	0.46	0.38	5.65	0.23	0.43	0.46	0.62	5.83	0.37
5	0.46	0.46	0.49	5.14	0.33	0.45	0.46	0.40	5.22	0.26	0.44	0.46	0.54	5.20	0.36	0.44	0.46	0.45	5.09	0.31
6	0.47	0.47	0.55	4.93	0.39	0.46	0.47	0.49	5.05	0.34	0.46	0.47	0.58	4.95	0.41	0.45	0.47	0.66	4.85	0.47
7	0.48	0.48	0.68	4.65	0.51	0.47	0.47	0.68	4.67	0.51	0.47	0.48	0.61	4.40	0.48	0.46	0.47	0.68	4.67	0.51
8	0.49	0.48	0.70	4.40	0.55	0.48	0.48	0.73	4.40	0.57	0.48	0.48	0.75	4.53	0.58	0.47	0.48	0.66	4.44	0.52
9	0.5	0.49	0.88	4.51	0.67	0.49	0.49	0.83	4.37	0.66	0.49	0.49	0.86	4.40	0.68	0.49	0.49	0.87	4.41	0.68
High (H)	0.52	0.5	1.03	4.42	0.81	0.51	0.5	1.10	4.31	0.88	0.5	0.5	1.06	4.43	0.83	0.5	0.5	1.08	4.34	0.86
H-L	-	-	1.67	6.06	0.95	-	-	1.84	6.25	1.02	-	-	1.70	6.16	0.96	-	-	1.70	6.50	0.90

Table A4 Combining Probability Forecasts with Expected Return Forecasts (Non-variance-adjusted)

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts without variance adjustments. The prediction target is the probability of a stock outperforming the market return. The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.32	0.46	0.47	0.48	0.52	0.56	0.53	0.58
Panel A: Probability Forecast								
Mean	0.36	1.75	1.66	1.84	1.67	1.84	1.70	1.70
SR	0.37	1.06	0.98	1.06	0.95	1.02	0.96	0.90
α	0.32	1.40	1.31	1.45	1.24	1.35	1.29	1.14
t_α	1.92	5.71	5.58	6.98	5.67	5.84	5.97	5.38
Panel B: Expected Return Forecast								
Mean	2.54	2.54	2.54	1.83	2.28	2.51	2.54	2.33
SR	1.39	1.39	1.39	1.02	1.20	1.31	1.39	1.23
α	2.19	2.19	2.19	1.38	1.84	2.12	2.19	1.96
t_α	5.57	5.57	5.57	3.93	4.34	4.62	5.57	4.96
Panel C: 1/N Combination of Probability and Return Forecasts								
Mean	1.45	2.15	2.10	1.84	1.97	2.18	2.12	2.01
SR	1.24	1.44	1.39	1.21	1.24	1.32	1.34	1.20
α	1.26	1.79	1.75	1.42	1.54	1.73	1.74	1.55
t_α	5.21	6.32	6.26	6.32	5.53	5.64	6.52	5.79
Panel D: Mean-variance Combination of Probability and Return Forecasts								
Mean	3.10	4.15	4.05	3.68	3.96	4.05	4.03	3.72
SR	1.42	1.58	1.56	1.40	1.44	1.49	1.51	1.43
α	2.84	3.64	3.58	3.08	3.39	3.51	3.56	3.23
t_α	5.19	5.49	5.32	4.79	4.78	4.88	5.12	4.90

Table A5 Performance of the Probability Forecasts Portfolios (Probability of Outperforming R_f)

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast. The prediction target is the probability of a stock outperforming the risk-free rate (R_f). For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability ($Prob$), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	-0.01	0.49	0.77	5.13	0.52	0.35	0.38	-0.92	9.15	-0.35	0.35	0.39	-0.98	8.91	-0.38	0.35	0.39	-0.78	9.32	-0.29
2	0.15	0.52	0.26	5.83	0.15	0.4	0.43	-0.62	9.17	-0.23	0.4	0.43	-0.56	9.12	-0.21	0.4	0.43	-0.33	9.40	-0.12
3	0.29	0.46	-0.06	5.42	-0.04	0.42	0.45	0.00	8.98	0.00	0.43	0.45	0.04	8.71	0.02	0.43	0.46	-0.08	9.25	-0.03
4	0.38	0.45	0.23	5.38	0.15	0.45	0.48	0.32	8.92	0.12	0.45	0.48	0.20	8.70	0.08	0.45	0.48	0.29	9.15	0.11
5	0.42	0.47	0.52	5.32	0.34	0.47	0.5	0.55	8.86	0.22	0.47	0.5	0.46	8.61	0.18	0.47	0.5	0.70	9.14	0.26
6	0.45	0.49	0.47	5.27	0.31	0.49	0.51	0.81	8.73	0.32	0.49	0.52	0.82	8.64	0.33	0.49	0.52	0.84	9.12	0.32
7	0.48	0.5	0.58	5.25	0.38	0.5	0.53	1.11	8.78	0.44	0.5	0.53	1.00	8.55	0.41	0.51	0.53	1.07	9.06	0.41
8	0.51	0.52	0.67	5.04	0.46	0.52	0.55	1.62	8.59	0.65	0.52	0.55	1.29	8.46	0.53	0.52	0.55	1.15	8.96	0.45
9	0.54	0.53	0.90	4.89	0.64	0.54	0.56	1.77	8.49	0.72	0.53	0.56	1.79	8.37	0.74	0.54	0.56	1.64	8.77	0.65
High (H)	0.6	0.54	1.18	4.86	0.84	0.56	0.57	2.19	8.50	0.89	0.56	0.57	2.10	8.40	0.87	0.56	0.57	2.24	8.75	0.89
H-L	-	-	0.41	3.64	0.39	-	-	3.10	7.93	1.35	-	-	3.09	7.90	1.35	-	-	3.02	8.53	1.23
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.36	0.39	-0.82	9.33	-0.30	0.36	0.39	-0.84	9.43	-0.31	0.37	0.39	-0.96	9.34	-0.36	0.38	0.39	-0.87	9.18	-0.33
2	0.41	0.43	-0.33	9.14	-0.12	0.4	0.43	-0.25	9.56	-0.09	0.41	0.43	-0.32	9.39	-0.12	0.42	0.43	-0.44	9.16	-0.17
3	0.43	0.46	-0.04	9.19	-0.01	0.43	0.46	-0.09	9.44	-0.03	0.44	0.46	0.05	9.14	0.02	0.44	0.46	0.04	9.17	0.01
4	0.46	0.48	0.30	9.09	0.11	0.46	0.48	0.39	9.32	0.15	0.46	0.48	0.21	9.23	0.08	0.47	0.48	0.38	8.93	0.15
5	0.48	0.5	0.74	9.02	0.29	0.48	0.5	0.59	9.16	0.22	0.48	0.5	0.47	9.11	0.18	0.49	0.5	0.67	8.99	0.26
6	0.49	0.51	1.09	8.97	0.42	0.5	0.51	0.93	9.22	0.35	0.5	0.51	0.88	9.05	0.34	0.5	0.51	0.68	9.03	0.26
7	0.51	0.53	1.07	8.95	0.41	0.51	0.53	0.82	9.28	0.31	0.51	0.53	0.91	9.07	0.35	0.52	0.53	1.07	8.66	0.43
8	0.52	0.55	1.30	8.91	0.51	0.53	0.55	1.36	9.07	0.52	0.53	0.55	1.36	8.81	0.54	0.53	0.55	1.38	8.72	0.55
9	0.54	0.56	1.52	8.68	0.61	0.54	0.56	1.74	8.80	0.68	0.54	0.56	1.80	8.87	0.70	0.54	0.56	1.88	8.61	0.76
High (H)	0.55	0.57	2.20	8.71	0.88	0.55	0.57	2.23	8.89	0.87	0.55	0.57	2.16	8.72	0.86	0.55	0.57	2.09	8.60	0.84
H-L	-	-	3.02	8.43	1.24	-	-	3.07	8.67	1.23	-	-	3.12	8.50	1.27	-	-	2.96	8.13	1.26

Table A6 Combining Probability Forecasts (Probability of Outperforming R_f) with Expected Return Forecasts

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts. The prediction target is the probability of a stock outperforming the risk-free rate (R_f). The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
corr	0.24	0.21	0.18	0.19	0.17	0.15	0.18	0.20
Panel A: Probability Forecast								
Mean	0.41	3.10	3.09	3.02	3.02	3.07	3.12	2.96
SR	0.39	1.35	1.35	1.23	1.24	1.23	1.27	1.26
α	0.27	1.85	1.82	1.67	1.69	1.70	1.77	1.64
t_α	1.78	5.45	5.69	5.20	5.25	5.13	5.67	5.31
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.71	3.05	3.04	2.63	2.85	2.97	3.06	2.93
SR	1.33	1.79	1.81	1.55	1.61	1.66	1.75	1.68
α	1.50	2.29	2.27	1.72	1.99	2.11	2.25	2.09
t_α	4.52	6.27	6.41	5.64	5.17	5.64	6.37	5.61
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.65	10.37	10.71	8.53	11.22	9.76	10.95	10.34
SR	1.37	1.75	1.76	1.55	1.70	1.66	1.73	1.70
α	6.55	8.68	8.99	6.46	9.28	8.02	9.17	8.52
t_α	4.46	5.39	5.32	4.35	4.63	4.83	5.16	4.78

Table A7 Performance of the Probability Forecasts Portfolios (Probability of Outperforming $\beta \times R_{mkt}$)

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast. The prediction target is the probability of a stock outperforming its CAPM β times the market return ($\beta \times R_{mkt}$). We estimate each stock's CAPM β in real-time following Fama and MacBeth (1973). For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability ($Prob$), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	-0.02	0.46	0.76	5.14	0.51	0.37	0.39	-0.55	8.71	-0.22	0.37	0.39	-0.46	8.93	-0.18	0.37	0.39	-0.58	8.84	-0.23
2	0.15	0.49	0.15	5.72	0.09	0.41	0.42	0.09	8.62	0.04	0.41	0.42	0.16	8.85	0.06	0.41	0.42	0.14	8.87	0.06
3	0.29	0.44	0.12	5.41	0.07	0.43	0.44	0.62	8.71	0.25	0.43	0.44	0.60	8.78	0.24	0.43	0.44	0.70	8.74	0.28
4	0.37	0.44	0.38	5.42	0.24	0.45	0.46	0.56	8.59	0.23	0.45	0.46	0.74	8.69	0.29	0.45	0.46	0.68	8.54	0.28
5	0.41	0.45	0.54	5.29	0.35	0.46	0.47	1.00	8.46	0.41	0.47	0.47	0.95	8.74	0.38	0.47	0.47	0.89	8.67	0.35
6	0.45	0.46	0.51	5.30	0.33	0.48	0.48	0.86	8.46	0.35	0.48	0.48	0.82	8.61	0.33	0.48	0.48	0.97	8.61	0.39
7	0.48	0.48	0.61	5.19	0.41	0.49	0.49	1.28	8.40	0.53	0.49	0.49	1.24	8.52	0.50	0.5	0.49	1.04	8.52	0.42
8	0.51	0.49	0.81	5.03	0.56	0.51	0.51	1.58	8.18	0.67	0.51	0.51	1.50	8.49	0.61	0.51	0.51	1.54	8.27	0.65
9	0.54	0.51	0.92	4.90	0.65	0.53	0.53	2.01	8.21	0.85	0.52	0.53	1.86	8.32	0.78	0.53	0.53	1.86	8.34	0.77
High (H)	0.6	0.52	1.25	4.85	0.89	0.55	0.56	2.40	8.18	1.01	0.55	0.55	2.45	8.37	1.01	0.55	0.56	2.42	8.37	1.00
H-L	-	-	0.49	4.13	0.41	-	-	2.94	7.74	1.32	-	-	2.90	8.15	1.23	-	-	3.00	8.34	1.25
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR	$\widehat{Prob}$	$Prob$	Mean	SD	SR
Low (L)	0.37	0.39	-0.53	9.25	-0.20	0.37	0.39	-0.52	9.07	-0.20	0.37	0.39	-0.55	9.05	-0.21	0.38	0.39	-0.43	9.12	-0.16
2	0.4	0.42	0.20	9.14	0.08	0.41	0.42	0.30	8.92	0.12	0.41	0.42	0.25	8.87	0.10	0.41	0.42	0.22	9.09	0.08
3	0.43	0.44	0.81	9.16	0.31	0.43	0.45	0.75	9.00	0.29	0.43	0.45	0.77	8.86	0.30	0.43	0.44	0.75	9.14	0.28
4	0.45	0.46	0.75	9.07	0.29	0.45	0.46	0.79	8.71	0.31	0.45	0.46	0.70	8.70	0.28	0.44	0.46	0.74	8.96	0.28
5	0.46	0.47	0.93	8.96	0.36	0.46	0.47	0.78	8.92	0.30	0.46	0.47	0.95	8.85	0.37	0.46	0.47	0.82	8.98	0.31
6	0.48	0.48	0.83	9.00	0.32	0.48	0.48	0.87	8.71	0.35	0.48	0.48	0.84	8.61	0.34	0.47	0.48	0.76	8.91	0.30
7	0.49	0.49	1.21	8.93	0.47	0.49	0.49	1.08	8.70	0.43	0.49	0.49	1.21	8.64	0.49	0.49	0.49	1.20	8.81	0.47
8	0.51	0.5	1.51	8.62	0.61	0.51	0.51	1.50	8.48	0.61	0.51	0.51	1.61	8.39	0.66	0.5	0.51	1.45	8.59	0.59
9	0.52	0.53	1.94	8.59	0.78	0.52	0.53	1.91	8.55	0.77	0.52	0.53	1.85	8.39	0.77	0.52	0.53	2.08	8.56	0.84
High (H)	0.55	0.56	2.39	8.66	0.95	0.55	0.56	2.35	8.47	0.96	0.54	0.56	2.41	8.48	0.99	0.54	0.56	2.20	8.53	0.89
H-L	-	-	2.92	8.95	1.13	-	-	2.87	8.19	1.21	-	-	2.96	8.59	1.20	-	-	2.63	8.57	1.06

Table A8 Combining Probability Forecasts (Probability of Outperforming $\beta \times R_{mkt}$) with Expected Return Forecasts

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts. The prediction target is the probability of a stock outperforming its CAPM β times the market return ($\beta \times R_{mkt}$). We estimate each stock's CAPM β in real-time following Fama and MacBeth (1973). The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.13	0.29	0.28	0.32	0.33	0.27	0.29	0.32
Panel A: Probability Forecast								
Mean	0.49	2.94	2.90	3.00	2.92	2.87	2.96	2.63
SR	0.41	1.32	1.23	1.25	1.13	1.21	1.20	1.06
α	0.19	1.73	1.66	1.77	1.61	1.64	1.64	1.31
t_α	1.18	5.83	5.35	4.94	4.11	5.23	4.46	3.51
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Expected Return Forecasts								
Mean	1.74	2.97	2.95	2.62	2.80	2.87	2.98	2.76
SR	1.37	1.71	1.65	1.49	1.44	1.57	1.62	1.46
α	1.46	2.23	2.19	1.77	1.95	2.08	2.18	1.92
t_α	4.85	5.91	5.84	5.60	4.58	5.35	5.54	5.00
Panel D: Mean-variance Combination of Probability and Expected Return Forecasts								
Mean	6.66	9.25	8.97	7.97	9.04	8.17	8.92	8.02
SR	1.38	1.70	1.66	1.51	1.53	1.57	1.63	1.52
α	6.62	7.78	7.64	6.04	7.74	6.77	7.60	6.81
t_α	4.52	5.33	5.20	4.66	4.52	4.98	4.99	4.48

Table A9 Performance of the Probability Forecasts Portfolios (Probability of Positive Return)

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast. The prediction target is the probability of a stock having a positive return. For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability (*Prob*), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.51	0.75	5.09	0.51	0.44	0.41	-0.21	8.48	-0.09	0.44	0.41	0.05	8.76	0.02	0.44	0.41	-0.13	8.58	-0.05
2	0.18	0.53	0.13	5.51	0.08	0.48	0.45	0.33	8.55	0.14	0.48	0.45	0.43	8.70	0.17	0.48	0.45	0.65	8.74	0.26
3	0.34	0.48	0.21	5.34	0.14	0.5	0.47	0.66	8.41	0.27	0.5	0.47	0.62	8.68	0.25	0.5	0.47	0.71	8.46	0.29
4	0.43	0.48	0.46	5.32	0.30	0.52	0.49	0.67	8.36	0.28	0.52	0.48	0.73	8.68	0.29	0.52	0.49	0.78	8.49	0.32
5	0.48	0.49	0.54	5.24	0.36	0.54	0.5	1.04	8.28	0.43	0.54	0.5	0.87	8.54	0.35	0.53	0.5	0.88	8.35	0.36
6	0.51	0.49	0.51	5.25	0.34	0.55	0.52	0.93	8.18	0.39	0.55	0.51	0.99	8.48	0.40	0.55	0.52	1.14	8.27	0.48
7	0.55	0.51	0.65	5.04	0.45	0.56	0.53	1.33	8.11	0.57	0.56	0.53	1.27	8.26	0.53	0.56	0.53	1.37	8.13	0.58
8	0.58	0.52	0.83	4.95	0.58	0.58	0.55	1.73	7.98	0.75	0.58	0.55	2.01	8.08	0.86	0.57	0.55	1.59	8.14	0.68
9	0.62	0.52	1.08	4.77	0.78	0.6	0.56	2.02	7.91	0.89	0.6	0.56	1.79	8.14	0.76	0.59	0.56	1.85	8.11	0.79
High (H)	0.7	0.51	1.29	4.83	0.92	0.63	0.56	2.65	8.14	1.13	0.62	0.56	2.60	8.24	1.09	0.62	0.54	2.46	8.35	1.02
H-L	-	-	0.53	4.37	0.42	-	-	2.86	8.45	1.17	-	-	2.55	8.68	1.02	-	-	2.59	8.31	1.08
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.44	0.41	-0.18	8.56	-0.07	0.45	0.41	-0.01	8.60	-0.00	0.45	0.41	0.01	8.59	0.00	0.45	0.41	-0.17	8.68	-0.07
2	0.48	0.45	0.70	8.66	0.28	0.49	0.45	0.66	8.64	0.26	0.49	0.45	0.63	8.75	0.25	0.49	0.45	0.63	8.75	0.25
3	0.5	0.47	0.77	8.49	0.31	0.51	0.47	0.91	8.48	0.37	0.51	0.47	0.71	8.49	0.29	0.51	0.47	0.75	8.63	0.30
4	0.52	0.49	0.92	8.30	0.38	0.52	0.49	0.78	8.43	0.32	0.52	0.49	0.77	8.65	0.31	0.52	0.49	0.75	8.63	0.30
5	0.53	0.5	0.97	8.37	0.40	0.54	0.5	0.92	8.37	0.38	0.54	0.5	1.03	8.21	0.44	0.54	0.5	0.79	8.34	0.33
6	0.54	0.52	1.27	8.21	0.54	0.55	0.52	1.20	8.36	0.50	0.55	0.51	1.08	8.38	0.45	0.55	0.52	1.22	8.42	0.50
7	0.56	0.53	1.41	8.05	0.61	0.56	0.53	1.32	8.19	0.56	0.56	0.53	1.29	8.05	0.55	0.56	0.53	1.37	8.20	0.58
8	0.57	0.55	1.54	7.90	0.68	0.58	0.55	1.59	7.94	0.69	0.57	0.54	1.48	7.96	0.64	0.58	0.55	1.44	8.02	0.62
9	0.58	0.56	1.79	8.03	0.77	0.59	0.56	2.06	8.03	0.89	0.59	0.56	1.95	8.14	0.83	0.59	0.56	2.29	8.07	0.98
High (H)	0.61	0.55	2.51	8.21	1.06	0.62	0.55	2.60	8.27	1.09	0.61	0.56	2.80	8.30	1.17	0.61	0.55	2.63	8.29	1.10
H-L	-	-	2.69	8.18	1.14	-	-	2.61	8.68	1.04	-	-	2.79	8.70	1.11	-	-	2.79	8.78	1.10

Table A10 Combining Probability Forecasts (Probability of Positive Return) with Expected Return Forecasts

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts. The prediction target is the probability of a stock having a positive return. The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.22	0.37	0.32	0.40	0.33	0.35	0.35	0.41
Panel A: Probability Forecast								
Mean	0.53	2.86	2.55	2.59	2.69	2.61	2.79	2.79
SR	0.42	1.17	1.02	1.08	1.14	1.04	1.11	1.10
α	0.28	1.73	1.46	1.33	1.58	1.36	1.52	1.64
t_α	1.47	4.37	3.83	3.84	4.40	3.41	3.84	3.93
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Return Forecasts								
Mean	1.76	2.93	2.77	2.42	2.68	2.74	2.89	2.84
SR	1.32	1.56	1.48	1.34	1.45	1.41	1.52	1.44
α	1.50	2.23	2.09	1.54	1.93	1.94	2.12	2.09
t_α	4.62	5.27	5.16	4.87	4.78	4.45	5.24	4.71
Panel D: Mean-variance Combination of Probability and Return Forecasts								
Mean	6.82	7.85	7.65	5.54	6.92	6.11	7.24	7.20
SR	1.38	1.54	1.51	1.36	1.46	1.33	1.48	1.44
α	6.83	7.13	7.05	4.17	6.10	5.61	6.67	6.70
t_α	4.50	4.73	4.64	4.44	4.59	4.06	4.64	4.38

Table A11 Performance of the Probability Forecasts Portfolios (Probability of Outperforming the Cross-sectional Median)

This table reports the value-weighted decile portfolio performance sorted by each model's probability forecast. The prediction target is the probability of a stock outperforming the cross-sectional median. For each portfolio, we report the following measures: average predicted probability ($\widehat{Prob}$), actual realized probability (*Prob*), mean and standard deviation of excess returns, and the Sharpe ratio. These portfolios are rebalanced at a monthly frequency based on the latest out-of-sample predictions from the respective model. The 'H-L' row represents the strategy of investing in the 10th decile (High) while selling the 1st decile (Low) short. The data spans from January 1987 to December 2020.

	OLS					Logit					PLS					NN1				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	-0.02	0.49	0.76	5.12	0.52	0.4	0.4	-0.71	8.95	-0.27	0.4	0.4	-0.68	9.01	-0.26	0.41	0.41	-0.70	8.90	-0.27
2	0.16	0.5	0.18	5.67	0.11	0.45	0.44	-0.25	8.91	-0.10	0.45	0.44	0.09	9.01	0.03	0.45	0.45	0.09	9.01	0.04
3	0.32	0.47	0.08	5.38	0.05	0.47	0.47	0.53	8.89	0.21	0.47	0.47	0.40	8.84	0.16	0.47	0.47	0.55	8.96	0.21
4	0.41	0.47	0.42	5.38	0.27	0.49	0.49	0.50	8.82	0.20	0.49	0.48	0.65	8.73	0.26	0.49	0.48	0.73	8.72	0.29
5	0.45	0.48	0.49	5.27	0.32	0.51	0.5	0.87	8.81	0.34	0.51	0.5	0.71	8.89	0.28	0.51	0.5	0.84	8.87	0.33
6	0.49	0.49	0.53	5.29	0.35	0.52	0.51	0.84	8.74	0.33	0.52	0.51	1.00	8.73	0.40	0.52	0.51	0.86	8.67	0.34
7	0.52	0.5	0.54	5.17	0.36	0.54	0.52	1.29	8.46	0.53	0.53	0.52	1.12	8.64	0.45	0.54	0.52	1.31	8.54	0.53
8	0.55	0.52	0.81	5.02	0.56	0.55	0.53	1.49	8.45	0.61	0.55	0.53	1.48	8.52	0.60	0.55	0.53	1.47	8.42	0.61
9	0.58	0.53	0.91	4.84	0.66	0.57	0.55	2.04	8.40	0.84	0.56	0.55	1.84	8.37	0.76	0.56	0.54	1.60	8.39	0.66
High (H)	0.65	0.53	1.27	4.85	0.90	0.59	0.56	2.31	8.44	0.95	0.59	0.56	2.36	8.48	0.96	0.58	0.56	2.37	8.44	0.97
H-L	-	-	0.50	3.99	0.44	-	-	3.02	7.90	1.32	-	-	3.04	7.98	1.32	-	-	3.07	8.07	1.32
	NN2					NN3					NN4					NN5				
	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR	$\widehat{Prob}$	*Prob*	Mean	SD	SR
Low (L)	0.4	0.41	-0.64	9.00	-0.25	0.4	0.4	-0.69	9.37	-0.26	0.41	0.41	-0.57	9.15	-0.22	0.41	0.4	-0.65	9.06	-0.25
2	0.44	0.45	0.08	8.94	0.03	0.44	0.45	-0.13	9.37	-0.05	0.45	0.45	-0.06	9.37	-0.02	0.45	0.45	0.04	9.05	0.02
3	0.47	0.47	0.58	9.04	0.22	0.47	0.47	0.59	9.50	0.21	0.47	0.47	0.53	9.28	0.20	0.47	0.47	0.62	9.14	0.24
4	0.49	0.48	0.64	8.86	0.25	0.49	0.49	0.70	9.20	0.26	0.49	0.49	0.69	9.04	0.26	0.49	0.49	0.67	8.76	0.26
5	0.51	0.5	0.94	8.75	0.37	0.51	0.5	0.83	9.25	0.31	0.51	0.5	0.72	9.12	0.28	0.51	0.5	0.85	8.89	0.33
6	0.52	0.51	0.84	8.71	0.33	0.52	0.51	0.75	9.22	0.28	0.52	0.51	0.90	8.97	0.35	0.52	0.51	0.94	8.99	0.36
7	0.54	0.52	1.15	8.63	0.46	0.54	0.52	1.15	9.02	0.44	0.54	0.52	0.93	8.89	0.36	0.54	0.52	1.02	8.58	0.41
8	0.55	0.53	1.32	8.60	0.53	0.55	0.53	1.49	8.81	0.58	0.55	0.53	1.45	8.74	0.58	0.55	0.53	1.71	8.70	0.68
9	0.56	0.54	1.83	8.36	0.76	0.56	0.54	1.88	8.68	0.75	0.56	0.54	1.81	8.57	0.73	0.56	0.55	1.72	8.41	0.71
High (H)	0.58	0.56	2.37	8.56	0.96	0.58	0.56	2.40	8.76	0.95	0.58	0.56	2.32	8.69	0.92	0.58	0.56	2.48	8.58	1.00
H-L	-	-	3.01	7.95	1.31	-	-	3.09	8.60	1.24	-	-	2.89	8.43	1.19	-	-	3.13	8.23	1.32

Table A12 Combining Probability Forecasts (Probability of Outperforming Cross-sectional Median) with Expected Return Forecasts

This table reports the comparison results for the long-short portfolios based on probability forecasts and expected return forecasts. The prediction target is the probability of a stock outperforming the cross-sectional median. The first row reports the correlation between the 'H-L' portfolio of the two approaches. Following this, for each panel, we report the value-weighted 'H-L' portfolio mean, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the t-statistics for α. We consider eight forecasting models in total (OLS, Logit, PLS, and NN1-5). Panel A (B) reports the results based on the probability (expected return) forecast. Panel C (D) reports the equal-weighted (mean-variance efficient) combination of 'H-L' portfolios from probability and expected return forecasts. We estimate the expected return and variance-covariance matrix using real-time data with an expanding window. In constructing the portfolios, we assume a relative risk aversion of 5. All samples start from January 1987 and end in December 2020.

	OLS	Logit	PLS	NN1	NN2	NN3	NN4	NN5
Corr	0.16	0.29	0.28	0.33	0.26	0.25	0.29	0.33
Panel A: Probability Forecast								
Mean	0.50	3.02	3.04	3.07	3.01	3.09	2.89	3.13
SR	0.44	1.32	1.32	1.32	1.31	1.24	1.19	1.32
α	0.23	1.77	1.82	1.75	1.71	1.74	1.61	1.87
t_α	1.35	5.36	5.55	5.13	5.00	5.12	4.76	5.81
Panel B: Expected Return Forecast								
Mean	3.00	3.00	3.00	2.24	2.68	2.87	3.00	2.89
SR	1.43	1.43	1.43	1.18	1.23	1.30	1.43	1.34
α	2.72	2.72	2.72	1.76	2.29	2.52	2.72	2.54
t_α	4.45	4.45	4.45	3.52	3.41	3.70	4.45	4.06
Panel C: 1/N Combination of Probability and Return Forecasts								
Mean	1.75	3.01	3.02	2.66	2.84	2.98	2.94	3.01
SR	1.37	1.71	1.71	1.53	1.60	1.60	1.61	1.63
α	1.48	2.25	2.27	1.76	2.00	2.13	2.16	2.21
t_α	4.79	5.82	6.09	5.75	5.26	5.41	5.54	5.48
Panel D: Mean-variance Combination of Probability and Return Forecasts								
Mean	6.69	9.79	9.50	7.58	9.71	8.79	9.00	9.11
SR	1.37	1.69	1.69	1.55	1.65	1.55	1.59	1.64
α	6.65	8.24	8.10	5.59	8.05	7.32	7.76	7.62
t_α	4.47	5.10	5.20	5.12	5.01	4.81	4.84	4.97