Decoding Mutual Fund Performance: Dynamic Return Patterns via Deep Learning

Norman (Xuxi) Guo*

April 2022

[Link to the latest version]

ABSTRACT

In this paper, we apply a state-of-the-art deep learning model to understand and predict dynamic patterns in mutual fund returns. A long-short portfolio based on the model's prediction generates a 2.8% annualized Carhart 4-factor alpha. This abnormal performance is persistent for up to four years. The model improves the prediction of future fund alphas substantially by increasing the R-squared by more than 25% in a predictive regression that includes other fund skill measures as well as fund and time fixed effects. The model's predictive power derives from its ability in capturing fund skill embedded in dynamic strategies. We construct model-based conditional skill measures that depend on the inferred informativeness of macroeconomic and fundamental variables. Such measures are predictive of fund performance in future periods when the conditioning variables are highly informative. The conditional performance of these measures are also persistent. Overall, our results suggest that mutual fund have various specific skills that generate superior returns when the time is right.

JEL Classification: C45, G11, G17, G23

Key Words: Machine Learning, Mutual Fund Performance, Return Patterns, Investment Strategies, Explainable AI

---

*J. Mack Robinson College of Business, Georgia State University. Email: gxuxi1@gsu.edu. This paper has benefited from discussions with Vikas Agarwal, Sean Cao, Will Cong, Marcin Kacperczyk, Zhen Shi, Baozhong Yang (Dissertation Chair), and Hong Zhang. The usual disclaimer applies.

"History doesn't repeat itself, but it often rhymes."

— Mark Twain

I. Introduction

History has a tendency of reiterating itself, albeit usually in somewhat different forms. For example, there are striking similarities between the 2020s and the "roaring" 1920s, both recovering from a pandemic and experiencing a technology growth burst, notwithstanding important differences between the century-apart eras.[1] In the mutual fund industry, fund managers have also noted recurring patterns in the market and economy that call for certain strategies, which can generate similar future fund returns.[2] These observations relate to the findings in the literature that most mutual funds do not generate superior performance (e.g., Fama and French, 2010, Barras, Scaillet, and Wermers, 2010) and that past fund performance is not a dependable predictor of future performance. Mutual funds may adopt different and dynamic strategies in different economic states, which can be difficult for linear models to capture. In this paper, we aim to answer the following questions: Can we learn dynamic patterns in fund returns as related to macroeconomic and fund conditions? Would such patterns be helpful to predict future fund performance? What can we learn from such patterns about fund skill and strategies?

Given the dynamic and complex nature of potential patterns, traditional econometric models are not well-suited to answer the above questions. In this study, we apply a state-of-the-art deep learning model for time-series predictions to understand and predict dynamic patterns in mutual fund returns. Our model (the Temporal Fusion Transformers model) has several unique features that go beyond the classical out-of-the-box machine learning models, such as decision trees or standard neural networks. First, the model is a

[1] "'Roaring'? Not so fast," Cristina Lindblad, February 1, 2021, Bloomberg Businessweek.

[2] See, for example, *The Most Important Thing: Uncommon Sense for the Thoughtful Investor*, 2011, Howard Marks, Columbia Business School Publishing, "Déjà vu all over again," January 10, 2019, Andrew Pastor, EdgePoint Investment Group, and "ARK Invest's Wood expects market rotation back to growth stocks," September 14, 2021, David Randall, Reuters.

sequence-to-sequence model, i.e., it predicts an entire future time-series of mutual fund returns simultaneously, rather than just a single future return. Second, the model can handle different types of time-series variables well. Specifically, the model has separate treatments for dynamic, deterministic, and static variables[3] that utilize the information contained in these variables efficiently. Third, the model assigns time- and fund-varying informativeness weights to different input variables, unlike traditional machine learning models that assign constant weights to them. This allows the model to adapt dynamically and focus more on the most informative variables for specific time periods and funds. These informativeness weights can further help to interpret the time-varying patterns in mutual fund performance.

The model takes as inputs several classes of variables, including past fund returns, macroeconomic and market variables, and fund characteristics such as fund size, flow, and fees. The dependent or target variables of the model include the time-series of the future 12 monthly (risk-adjusted) returns of funds. The model can capture future return patterns well. Top mutual funds predicted by the model outperform bottom funds by an annualized Fama-French-Carhart four-factor alpha of 2.8%, significantly larger than those generated by OLS or more standard machine learning models. This outperformance is also persistent and remains statistically and economically significant for up to four years.

To investigate whether the model generates a new measure of fund skill, we regress actual fund alphas on the model's predicted alphas and control for historical fund performance, fund characteristics, and other measures of fund skill such as the return gap. We find that the model's predicted alphas improves predictive power even in the most comprehensive regressions, increasing the adjusted R-squared by more than 25%. The prediction power persists with fund and time fixed effects, suggesting that the model can identify time-varying fund skill.

We next try to dig deeper and understand what we can learn from the model. We hy-

[3]Dynamic variables are time-varying variables that are subject to random variations each period. Deterministic variables are variables that follow a determined path (e.g., fund age). Static variables are not time-varying.

pothesize that the model captures dynamic features of mutual fund strategies. For example, mutual funds can adopt "bottom-up" strategies that are based on analyzing company fundamentals and "top-down" approaches that adjust trading strategies with macroeconomic conditions (e.g., Moy and Griffeth, 1995). We first find several pieces of evidence consistent with funds employing such strategies. First, we consider the earnings call cycles of companies, which provide periodic information to the market. We find that the model puts the most weight on mutual fund returns in the month following earnings calls of companies held by the fund. In other words, the model can detect mutual fund returns that are most sensitive to fundamental information. Therefore, funds' use of fundamental information can be important for predicting mutual fund performance and understanding their skill. Second, we also find macroeconomic conditions and past return patterns to be both important determinants of the model's predictive power. For example, historical fund performance and macroeconomic variables are the most important features in the model. Furthermore, the model puts more weight on information from crisis periods, during which fund returns and strategies may be closely related to the abrupt changes in economic conditions.

The interpretability of the model allows us to further analyze funds' specific skills. For each predictive variable (e.g., market return, inflation, or fund's own past return), we construct a model-based conditional skill measure that represent the average abnormal returns of the fund when the predictive variable is most informative. To the extent that the model provides time- and fund-varying informativeness (i.e., variable importance) measures for the conditioning variables, the model-based conditional performance measures capture fund skills that are specific to macroeconomic and firm conditions. We find the conditional skill measures are predictive of fund performance in future periods when the conditioning variables are more informative. For example, funds with high "term spread skill" has an annualized abnormal return of 7.68% in future periods when the term spread variable is more informative, which is 13.87% higher than the abnormal return of funds with low term spread skill.

The conditional variables can be separated into two groups: 1) macroeconomic variables, including market return, inflation, term spread, and default spread, and 2) fundamental variables, including fund past returns and month-of-year.[4] We find conditional skills to predict future performance for these two groups of variables. Furthermore, the performance of these measures are persistent up to four years. The persistence of these measures also provide an explanation of the persistence of our main skill measure, which integrates all specific skills through the model. We also find that the specific skill measures to decline for the largest mutual funds, consistent with the diminishing return-to-scale hypotheses in Berk and Green (2004).

Our results suggest that firms do possess skills that are specific to macroeconomic and firm conditions. We note that such skills are broadly related to the market timing and stock selection skills analyzed in the mutual fund literature but are more specific and include more dimensions. For example, the different conditional skill measures are not highly correlated and some are even weakly negatively correlated. Thanks to our model's unique features, the model can quantify and capture the different specific skills of mutual funds.

This paper contributes to several strands of literature. First, the paper complements a rapidly growing literature that applies machine learning methods in financial economics (e.g., Cong, Tang, Wang, and Zhang, 2020a, Cong, Tang, Wang, and Zhang, 2020b, Feng, Giglio, and Xiu, 2020, Freyberger, Neuhierl, and Weber, 2020, Gu, Kelly, and Xiu, 2020, Gu, Kelly, and Xiu, 2021, and Chinco, Neuhierl, and Weber, 2021). Our paper is the first to introduce a sequence-to-sequence machine learning model that is particularly suitable for capturing dynamic time-series patterns. The model's ability to handle different types of time-varying inputs and flexibility in assigning different weights according to fund and time not only generates superior predictive performance, but also allows intuitive interpretation

[4]We call the second group of variables "fundamental variables" because they are not apparently related to time-varying economic conditions. These fundamental variables may nonetheless provide certain macroeconomic information to the model. However, such information is either orthogonal to those that are already captured by the macroeconomic variables, or represents interactions of fundamental and macroeconomic variables.

of the model's power. We expect this type of models can be used to address more general time-series problems in finance beyond the study of mutual funds.

Second, the paper helps to address questions about the persistence of mutual fund performance. The classical literature (e.g., Jensen, 1968, Elton, Gruber, Das, and Hlavka, 1993, Carhart, 1997, Busse, Goyal, and Wahal, 2010, and Fama and French, 2010) do not find persistence in mutual fund performance. Our paper finds that mutual funds do have predictable performance patterns. However, such patterns can be highly nonlinear and depend on dynamic fund strategies and macroeconomic and information environments. While such patterns are detectable by our model, it may be difficult to identify using in traditional econometric models. This partially answers the lack of performance persistence found in the literature.

The paper is also related to the recent mutual fund skill literature that identifies mutual fund skills through different angles (Carhart, 1997; Kacperczyk, Sialm, and Zheng, 2008; Huang, Sialm, and Zhang, 2011; Amihud and Goyenko, 2013; Hunter, Kandel, Kandel, and Wermers, 2014). Our model contributes to this literature by providing a new measure of time-varying mutual fund skill that can be measured based only on past performance, macroeconomic conditions, and fund characteristics.

Finally, our paper is related to several recent papers on applying machine learning methods to study mutual funds (e.g., Li and Rossi, 2020, Zhang, 2021). Our paper differs from these papers by first introducing a sequence-to-sequence model to capture dynamic fund performance patterns. Furthermore, the rich and flexible features of the model capture both patterns from bottom-up and top-down strategies of mutual funds and offer a more intuitive interpretation of the source of the model's predictive power.

II. Data, Variable Construction, and Sample Overview

II.A. Data sources

We obtain data used in this study from multiple sources. We take mutual fund returns, total net assets (TNA), expense ratio, turnover ratio, investment objective, and other fund characteristics from the Center for Research in Security Prices (CRSP) Survivorship Bias-Free Mutual Fund database. We obtain mutual fund portfolio holdings from the Thomson Reuters Mutual Fund Holdings (s12) database. We merge these two databases via the MFLINKS tables provided by Wharton Research Data Services (WRDS). Finally, macroeconomic data are obtained from Federal Reserve Economic Data (FRED).

Our study is focused on active U.S. equity funds from January 1990 to December 2019.[5] We follow the conventional selection criteria in Kacperczyk, Sialm, and Zheng (2008) to identify domestic equity funds.[6] We further exclude ETFs, fixed income, international, money market, sector, index, target-date, and balanced funds.[7] To mitigate omission bias (Elton, Gruber, and Blake, 2001) and incubation bias (Evans, 2010), we exclude observations prior to the first offer dates of funds, those for which the fund names are missing in the CRSP MF database, and those for which the fund's TNA is below $5 million. Our final sample comprises 3,717 unique funds, and 500,113 fund-month observations.

[5]We set our sample starting from 1990 because some of the macroeconomic variables such as VIX and Oil Price become available after 1990.

[6]Details of the selection criteria are available at Kacperczyk, Sialm, and Zheng (2008), Appendix A, page 2412.

[7]Similar to Jones and Mo (2021), we identify and remove index funds both by CRSP index fund flag and by searching the fundsâ names with the key words *Exchange-traded, Etf, Dfa, Index, Inde, Indx, Inx, Idx, Dow Jones, Ishare, S&P, 500, Wilshire, Russell, Russ, and MSCI*. We exclude target-date funds by searching the fund names with the key words *2055, 2050, 2045, 2040, 2035, 2030, 2025, 2020, 2015, 2010, 2005*, and *Target*.

II.B. Variable Construction

B1. Fund Performance and Characteristics

To measure performance, we compute alphas following based on rolling window estimates of factor betas. Specifically, for each fund-month observation, we use the previous 24 months to estimate the betas on the CRSP value-weighted excess market return (Mktrf), size (SMB), book-to-market (HML), and momentum (UMD) factors from Ken French's website. We then use these betas to risk-adjust the current month's excess return.

Since the CRSP Mutual Fund database lists multiple share classes separately, we aggregate all share classes at the fund level. Specifically, TNA is the aggregate total net assets ($mm) across all share classes of a fund. Cash holdings, turnover ratio (Turnover), expense ratio (Expense), and management fees are the TNA-weighted average across all fund share classes and scaled to percentage points. Manager tenure is the number of years since a portfolio manager is hired; if there are multiple managers for a fund, the longest tenure is used. Load is the dummy variable that equals one if at least one share class has load, and zero otherwise.

We follow the extant literature to identify fund managers' unobservable skill by the return gap measure of . The monthly return gap is the difference between a fund's realized gross return and the hypothetical return on its most recently disclosed portfolio holdings. We define Return Gap as the monthly average return gap over previous 12 months.

B2. Macroeconomic and Market Variables

We obtain a collection of macroeconomic variables that previous studies have shown to be useful for predicting security returns and risks over time. The variables include (1) Industrial Production Index, (2) Consumer Price Index, (3) Crude Oil Price (WTI), (4) 3-month treasury bill rate, (5) Term spread of 10-year treasury and 3-month treasury bill, (6) Default spread of Baa and Aaa corporate bond yields, and 7) NBER recession indicators.

Industry Production Index and Consumer Price Index are measured as the percentage change from a year ago, and the other variables except NBER recession indicators are measured as the percentage change from the previous month.

We also collect a series of market variables that can potentially affect fund manager's investment decisions. We include the percentage change of the CBOE volatility index, the CRSP total-return value-weighted index return, and the S&P500 index return. We also include the factor returns of two widely used factor models: SMB, HML, RMW, CMA from the Fama and French five-factor model and R_ME, R_IA, R_ROE, R_EG from the Hou, Xue, and Zhang (2015) q-factor model factor model.

We list all variables serving as inputs into the machine learning models, their definitions, and sources in Appendix A. The summary statistics are reported in Table 1. All variables are constructed monthly using information available at the previous month-end. All potentially unbounded variables are winsorized at the 1% extremes.

[Insert Table 1 Here]

III. Methodology

III.A. Forecasting in Finance

Traditionally, studies in finance have approached the forecasting problem via predictive regression models. In general, these studies generate various theoretically or intuitively motivated variables that can predict the target variable in the next period. For example, several single-period predictors such as return gap (Kacperczyk, Sialm, and Zheng, 2008), active share (Cremers and Petajisto, 2009), and risk shifting (Huang, Sialm, and Zhang, 2011) are proposed to predict the next-period mutual fund performance. However, such prediction models typically do not utilize the entire paths of history to describe the future, partly because the linear regression models cannot handle a large number of potentially correlated independent variables well. In contrast, time-series models such as the ARIMA models (Box, Jenkins, and MacGregor, 1974) and the exponential smoothing model (Hyndman, Koehler, Ord, and Snyder, 2008) offer a principled framework for modeling and learning time-series patterns such as trend and seasonality. However, such models usually impose structural assumptions and are mainly suitable in the applications where the structure of the time series is well understood.

Deep neural networks (DNNs), or deep learning models, have gained popularity in time-series forecasting and demonstrated strong performance improvements over traditional time-series models (e.g., Rangapuram, Seeger, Gasthaus, Stella, Wang, and Januschowski, 2018, Salinas, Flunkert, Gasthaus, and Januschowski, 2020, Wen, Torkkola, Narayanaswamy, and Madeka, 2017). With their capability to extract higher-order features, deep learning models can identify complex patterns within and across time series, and they usually require little or no structural assumptions about the time series. However, the basic DNN architectures are subject to several limitations when applied to financial data. The biggest challenge is that the financial data have small sizes and weak signal-to-noise ratio (Israel, Kelly, and Moskowitz, 2020). As a result, noisy or irrelevant inputs could dramatically affect the results of machine learning models. In addition, these models often fail to consider the heterogeneity of inputs by simply concatenating static inputs with other time-dependent features in the prediction. Finally, most current architectures are "black-box" models where forecasts are controlled by complex nonlinear interactions between many parameters. This makes it difficult to explain how models arrive at their predictions. A better design of the deep learning models is needed to harness the unique characteristics of financial data and interpret results of the model forecasts.

III.B. Temporal Fusion Transformers Model

In our paper, we adopt the Temporal Fusion Transformer (TFT) model, one of the most recent innovations of neural network architecture introduced by Google in Lim, Arik, Loeff, and Pfister (2019). The TFT model developed several innovative components (shown in

Figure 1) to efficiently build feature representations for different data types while enabling new forms of interpretability. The model uniqueness is fivefold. First, in contrast to one-step-ahead predictions in most prediction models, the TFT model simultaneously generates predictions at multiple future time periods, which allows us access to the evolution of mutual fund performance across the entire desire path. Second, the model includes a gating module (Gated Residual Network) to minimize the contributions of irrelevant inputs. This innovative module is especially helpful in our prediction framework where the precise relationships among historical time-series variables and the target variable are often unknown in advance. For example, some macroeconomic variables may have negligible influences on mutual fund performance, while others may have either linear and non-linear relationships with it. The gating module allows the model to skip over any unused variables and provides the flexibility to apply nonlinear processing only where needed.

[Insert Figure 1 Here]

Third, the TFT model is designed to provide instance-wise variable selection using variable selection networks. For example, the model can endogenously select and laser-focus on the specific variables that are particularly important for each fund-year prediction, removing unnecessary noisy inputs for that instance and improving prediction performance. Fourth, the TFT model employs a sequence-to-sequence neural-network layer, adapted from language translation models, to learn both long- and short-term temporal relationships. This temporal layer allows the model to incorporate information from different types of inputs (targets, dynamic variables, deterministic variables, and statics variables). Finally, to open the “black bo” of forecasts based on complex nonlinear interactions, the TFT model includes a self-attention layer in the neural network to pick up long-range dependencies that may be challenging for standard deep learning architectures to learn. Information from this attention layer can be further exported to enhance interpretability.

III.C. Sample Splitting and Tuning

In preparing the data sample and training the model, we follow the most common approach in the forecast evaluation literature (see, e.g., West, 2006). Specifically, we divide our data into three samples: the training, validation, and testing samples. We first use the training sample to estimate the model subject to a set of hyperparameters. We then use the validation sample to tune the hyperparameters in the following two steps: (1) We construct forecasts using the data from the validation sample based on the estimated model from the training sample; (2) we conduct a grid search of hyperparameters by re-estimating the model from the training sample until the objective function for the validation sample is optimized. The above cross-validation process could help produce reliable performance in out-of-sample tests and avoid overfitting the model to the training sample. Finally, we use the testing sample, which is used for neither estimation nor cross-validation processes, to evaluate a model's predictive performance.

III.D. Model Evaluation

To assess the predictive performance of fund alpha forecasts, we follow the method based on out-of-sample R-squared as Gu, Kelly, Xiu (2020). Specifically, we calculate the out-of-sample R^2 as:

$$R^2_{OOS} = 1 - \frac{\sum_{(i,t)\in\tau_{OOS}}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in\tau_{OOS}} r^2_{i,t+1}}, \tag{1}$$

where i and t indicate the fund and month, and T_{OOS} indicates that R^2 is only assessed on the out-of-sample that never use in model estimation or tuning. R^2_{OOS} pools prediction errors across funds and over time into a panel-level assessment of each model. Since the denominator is the sum of squared excess returns without demeaning, the above measure represents the proportional reduction in mean squared forecast error (MSFE) of the model relative to the benchmark of a naive forecast of zero.

III.E. Interpretable Variance Importance and Attention

The TFT model is designed to provide interpretable variable selection for each data type, including dynamics inputs, deterministic inputs, and static inputs . Below we list several key outputs from the TFT model that will be important for interpreting our results later. Specifically, the variable selection weights of historical inputs are calculated as:

$$\mathbf{w_{i,t}} = Softmax(f(\varepsilon_1, ..., \varepsilon_j, c_s)), \tag{2}$$

where $\mathbf{w_{i,t}}$ is the vector of variable selection weights of each historical variables $j\,(\varepsilon_1, \varepsilon_2, ..., \varepsilon_k)$ for fund i at time t, c_s is the vector of all statics inputs, and f is a gating module to integrate multiple variables.[8] The softmax function is a generalization of the logistic function that rescale inputs into a probability vector with the sum of all the probabilities equal to one. These weights can be exported after the model is estimated, which allows us to understand the importance of each variable j of fund i at time t. In the next step, the inputs are aggregated into the next layer based on the weights of each variable :

$$V_{i,t} = \sum_{j=1}^{k} \varepsilon_{i,j,t} w_{i,j,t} \tag{3}$$

In additional to interpretable variable importance, the TFT model employs a self-attention mechanism to learn short- and long-term relationships across different time steps, the attention is calculated as:

$$A_{i,t} = g(V_{i,t}), \tag{4}$$

where g is a function that encoder aggregated features $V_{i,t}$ into a sequence-to-sequence layer followed by attention architectures. The attention can be exported to provide information that which period will be assigned more attention, hence more important over the prediction history.

[8]For static variable, the context vector c_s is omitted.

IV. Predicting Mutual Fund Performance using Machine Learning Models

We consider the following machine learning model that predicts a sequence of mutual fund alphas for T consecutive periods in the future:

$$\boldsymbol{\alpha}_{i,t} = \boldsymbol{h}(I_{t-1}) + \boldsymbol{\epsilon_{i,t}}, \tag{5}$$

where $\alpha_{i,t} = (\alpha_{i,t+1}, ..., \alpha_{i,t+T})$ is the vector of fund alphas for the future T periods after period t, I_{t-1} is the public information set available at $t-1$, and $\boldsymbol{h}(I_{t-1}) = (h_{i,t+1}(I_{t-1}), ..., h_{i,t+T}(I_{t-1}))$ is a vector-valued function that approximates the expected future fund alphas. The prediction horizon T is the length of the sequence to be predicted. In what follows, we will use "target variable" or "predicted variabl" to refer to the dependent variable in the above estimation, i.e., fund alphas. We use the historical values of a group of variables $\{z_{i,s} : t - T^* \leq s < t\}$ to represent the information set I_{t-1}. The estimation horizon T^* represents the maximum lenth of time we go back and consider historical values of variables. We will refer to these variables as "predictors," "features," "covariates," or "independent variables." The predictors$Z_{i,t}$ consist of three types of variables: (1) dynamic inputs that covary with the target variables over time (e.g., macroeconomic variables), (2) static inputs for which the content is constant over time (e.g., fund style), (3) deterministic inputs that represent characteristics that vary with time with values known in advance (e.g., fund age).

We include a collection of predictive variables that could potentially influence our target variables. For dynamic inputs, we select fund characteristics (fund flow, fund TNA, cash holdings, and equity holdings), macroeconomic variables (industrial production, inflation, oil price, risk-free rate, default spread, and term spread), and market-related variables (SP 500 return, VIX, value-weighted index return, NBER crisis dummy, factor returns from Fama-French five factors model and Hou, Xue, and Zhang (2015) q-factor model, and the momentum factor return). For static inputs, we add fund style, load, and management fee

. For deterministic inputs, we choose the upcoming month of year and manager tenure. To improve the model efficiency and prediction accuracy, all the unbounded inputs and target variables are standardized by month (to have a mean of zero and standard deviation of 1 for each month) before being used in the model. As discussed earlier, due to their different natures, the dynamic, static, and deterministic variables are separately input and treated in our main machine learning model (the TFT model).

Following the machine learning literature, we divide our 30 years of data into 20 years of training and validation sample (1990-2009) and 10 years of testing sample (2010-2019). For the first 20 years, we randomly select 80% of the funds and include their observations in the training sample and the observations of the remaining 20% of funds in the validation sample. To study the return pattern over a prolonged period and reduce the intensive computational costs of the training process, we train and validate a fixed machine learning model for the first 20 years and examine the out-of-sample predictions for the last 10 years. We choose the estimation horizon to be 72 months and the prediction horizon to be 12 months. Hence, we require funds to have at least 84 months of observation in the training, validation, and testing samples. To further reduce the model's noise, we run the same models for 20 times and ensemble the TFT models by taking the mean of their predictions.

IV.A. Performance Comparison

To evaluate the performance of the TFT model, we compare multiple classes of models, which include generalized linear models (OLS regression, Lasso regression, Ridge regression, and Elastic Net), tree-based models (Decision Tree, Ada Boost, and Random Forest), and feed-forward neural networks with two and three layers (NN2, NN3), with the TFT model. Different from the TFT model, the other machine learning models considered here do not treat dynamic, static, and deterministic variables separately. Therefore, we simply concatenate all covariates over the full estimation horizon (72 months) as inputs for these models. The total number of covariates is $72 \times (1 + 33) = 2,448$. In addition, as all models other than TFT do not have multi-horizon features, we use the average of the future 12-month returns as the target variable.

Table 2 presents the comparison of the out-of-sample R^2 among different machine learning techniques, progressing from the simpler to the more sophisticated models. It may not be surprising that the linear OLS model generates close-to-zero prediction performance with an R^2 of 0.02%, because the model cannot handle nonlinear relationships among variables as well as complex intertemporal patterns of variables. The generalized linear models, such as the Ridge model, allows selection of the most important features in the regression. However, the performance of the Ridge model does not improve much over the OLS model, indicating highly nonlinear relationships among different features.

[Insert Table 2 Here]

The decision tree model is designed to capture nonlinear interactions. The single decision tree model, however, only generates an R^2 of 0.004%, potentially due to its large variance, which contributes to poor out-of-sample performance. Ada Boost and Random Forest models are ensembles of decision trees that aggregate information from a number of weak models to form a strong model. Both models produce an improved performance with R^2 of 0.05% and 0.04%, respectively. Neural network models incorporate complex predictor interactions and further improve the R^2 to 0.07% (two-layer feed-forward neural network). Finally, the TFT model, equipped with the unique traits discussed in Section X, produces a far superior R^2 of 0.35%.

Next, we further compare our prediction of the TFT model with other traditional predictors of mutual fund alphas. We define Predicted Alpha as the predicted fund alpha by the TFT model for a given fund and month. We estimate the following panel regression, indexed by fund(i)-month(t), with both year and fund fixed effects, in addition to a host of control variables including *log(TNA)*, *Fund Flow*, *Cash Holdings*, *Expense Ratio*, *Management Fee*, and *Turnover Ratio*:

$$Alpha_{i,t+1} = \beta Alpha\ Predictor_{i,t} + \gamma Control_{i,year} + \alpha_i + \alpha_t + \epsilon_{i,t} \tag{6}$$

Table 3 presents the results. The coefficient of *Predicted Alpha* is statistically significant at the 1% levels in all settings, even after controlling for historical alpha and return gap, suggesting that the information captured from the TFT model is independent of the traditional measures.

In addition, we also calculate the contribution to adjusted R^2 by Predicted Alpha as the ratio of the increase in adjusted R^2 from adding Predicted Alpha (to a regression model without it) to the total adjusted R^2 for the model including Predicted Alpha. The results show that Predicted Alpha consistently contributes 20% of the model's predictive power, even in the most comprehensive model that include all control variables and fixed effects.

[Insert Table 3 Here]

IV.B. Portfolio Performance and Persistence

The results from predictive regressions suggest that the TFT model can help to predict future fund performance. To obtain a more concrete understanding and quantify the value of the model, we next adopt a portfolio approach and identify skilled and unskilled funds. Specifically, we first create the model-predicted future monthly alphas in the next 12 months, $t+1, ..., t+12$, for all fund(i)-year(t) observations. We then sort all funds into deciles based on the predicted alphas for each month $t+i$ for $i = 1, ..., 12$, and construct equally weighted decile portfolios. The decile portfolios are rebalanced each month. We calculate the average monthly Fama-French 4-factor alpha of each decile portfolio over the next 12 months ($t+1, ..., t+12$). Table 4, Column 1 presents the performance of the decile portfolios. The top minus bottom portfolio generates a monthly alpha of 23.24 basis points or an annualized alpha of 2.8%, which is statistically significant at the 1% level.

We next investigate whether the superior performance is persistent. For this purpose, we maintain the monthly portfolio weights so that the same funds are selected in the portfolio for the same month over the next five years. The abnormal returns of the top minus bottom portfolio remain both economically and statistically significant for up to four years (with a monthly alpha of 22.21 basis points in the fourth year), suggesting that the model captures persistent skilled funds.

[Insert Table 4 Here]

V. Variable Importance and Conditional Performance Persistence

V.A. Variable Importance

While the performance of the TFT model is validated in Section IV, in this section, we zoom into the model to understand the source of the predictive power.

The TFT model allows us to open the model black box with its explainable output – variable importance measures from the variable selection network (described in Section III.E). The variable importance $w_{i,j,t}$, from equation (2), represents the weight of each variable j of fund i at time t. This structure gives us a dynamic interpretation of relative variable importance for each fund at different times. We first average $w_{i,j,t}$ across funds and overtime to understand the overall variable importance. Figure 2 reports the rank of overall variable importance in the model. Consistent with the time-series design of the TFT model, the historical alpha is the most prominent variable (28.8%), which suggests the return history itself, and its correlation with the other macroeconomics variables conveys the most information in predicting the future performance. The other most important variables are generally in agreement with the most influential factors mentioned in the shareholder letter, including the size, value, and momentum factors, and important macroeconomic variables such as *SP 500*, *Default Spread*, and *Inflation*.

[Insert Figure 2 Here]

We continue to examine the variable importance of the most important variable: historical alpha (denoted as *Alpha Importance*). Interestingly, we find that *Alpha Importance* appears to have strong seasonality patterns. Within the sample, we aggregate *AlphaImportance* across all funds and calendar months of the year and then aggregate the number of earnings-related announcements of the public companies in each calendar month of the year. Figure 3 plots the relationship between the one-month-ahead *Alpha Importance* and the number of earnings announcements in the month. The figure shows that the pattern of one-month-ahead alpha importance is closely aligned with that of the number of earnings announcements, which implies that our model captures the finding of Pinnuck (2005) that earnings information explains approximately 25% of a mutual fund's average monthly abnormal performance.

[Insert Figure 3 Here]

V.B. Model-Based Conditional Performance Persistence

B1. Performance Persistence

Whether mutual fund performance is persistent is one of the major questions in the mutual funds' literature. The most influential paper on this subject, Carhart (1997) uses the net alpha earned by investors to measure managerial skill. It concludes that there is no evidence of skilled or informed mutual fund managers. Fama and French (2010) use alpha measures to obtain a cross-sectional distribution of managerial talent and find evidence of inferior and superior performance in the extreme tails of the cross-section of mutual fund performance. Berk and van Binsbergen (2015) propose a new value-generated measure for mutual fund skill and find that the performance is persistent for the long-term. The model's interpretability allows us to understand the performance persistence better when the predictive variable is most informative.

B2. V.I. Conditional Alpha

The model's predictive power derives from its ability to capture fund skills embedded in dynamic strategies. To better understand how mutual funds manager adopts the dynamic investment strategy, we construct model-based conditional skill measures that depend on the inferred informativeness of macroeconomic and fundamental variables. In other words, we calculate the abnormal performance when our model believes a variable is most informative. Specifically, for each fund i, we calculate the variable importance conditional alpha (V.I. conditional alpha) as the average Carhart's four-factor alpha during a variable high period over the past five years. Variable high period (low period) is defined as the month when the variable importance of a variable for a fund is higher (lower) than that of 80 percent of the other funds in the same month. The conditional variables can be separated into two groups: 1) macroeconomic variables, including market return, inflation, term spread, and default spread, and 2) fundamental variables, including fund past returns and month-of-year.

Table 5 reports the correlation of V.I. conditional alpha on different macroeconomics and fundamental information. On average, the correlation among the measures is low, which implies that these measures capture information at a different dimension. For example, the measure from month of year, which captures the seasonality of the fundamental information of the funds, has a close-to-zero correlation with the other measures. Some measures are more correlated as the macroeconomics conditions embedded are overlapping - for example, the correlation between the measure from the inflation and market return is 0.27.

[Insert Table 5 Here]

B3. Conditional Performance Persistence

We first investigate the performance persistence without conditional on the model informativeness. Following Carhart (1997), we form portfolios of mutual funds on lagged 12-month Carhart's four-factor alpha and estimate future 12-month Carhart's four-factor alpha on the resulting portfolios. On January 1 of each year, we form five equal-weighted portfolios of mutual funds using the historical alphas. The portfolios of mutual funds sorted on 12-month past alphas demonstrate weak variation in mean alphas, as shown in figure 5, which is consistent with Carhart (1997) that the results of performance persistence are gone by controlling the momentum effect.

[Insert Figure 5 Here]

We then reform five equal-weighted portfolios of mutual funds using the V.I. conditional alphas on different macroeconomics and fundamental information over the past five years. We find that conditional skill measures predict fund performance in future periods when the conditioning variables are more informative. As shown in figure 6, the mean alphas increase monotonically and significantly in the future during the high informative period (when the model recognizes the time in the future is most informative again). Such effects appear on both macroeconomics and fundamental information. However, a similar pattern does appear during the low informative period. The results suggest that funds skills may only be revealed under certain time conditions when the macroeconomics variables are suitable for their investment strategies or when the fundamental information is more available in the market.

[Insert Figure 6 Here]

We conduct a panel regression analysis to confirm the hypothesis. We regress the future 12-month Carhart's four-factor alpha on the V.I. conditional alpha controlling the historical alpha. Specifically, we estimate the following panel regression, indexed by fund(i)-month(t), with both year and style fixed effects, and control variables including *Historical Alpha log(TNA)*and *Fund Flow*:

$$Alpha_{i,t+1} = \beta V.I.\,Conditional\ Alpha_{i,t} + \gamma Control_{i,year} + \alpha_i + \alpha_t + \epsilon_{i,t} \tag{7}$$

Table 6 presents the results. First of all, the coefficient of *Historical Alpha* in column one from Panel A is significant but small in magnitude. In contrast, the coefficients of *V.I. Conditional Alpha* are both economically and statistically significant during a high informative period for fundamental and macroeconomics variables. The *Historical Alpha* becomes insignificant simultaneously. However, such significant effects disappear during the low informative period. The results confirm Figure 6 that funds performance is persistent during a high informative period.

[Insert Table 6 Here]

We next investigate whether conditional persistence is a long-term effect. For this purpose, we use the conditional V.I. conditional alpha to predict the future in multiple horizons, including i) 12-24 months, (ii) 24-36 months, and (iii) 36-48 months. Table 7 presents the coefficients and the t-statistics of the regression of multi-horizon future fund alpha on the V.I. conditional alpha during the high informative period in equation 7. The results show that the effect of conditional persistence is up to at least 48 months. It implies that fund managers will apply a similar investment strategy when the information becomes informative again in the future due to the limited attention.

[Insert Table 7 Here]

B4. Conditional Performance Persistence on Macroeconomics Conditions

Though we find the performance is persistent during the high informative period from the TFT model, similar results could also be found when the actual macroeconomics variables are high. In other words, can we find that the funds' performance is also persistent during the high inflation period? To answer this question, we revisit the regression from equation 7 by adding the *Macro Conditional Alpha* as control. *Macro Conditional Alpha* is calculated as the average Carhart's four-factor alpha when the macroeconomics variable is higher than 80 percent of the time in the sample from 1990m1 to 2019m12. Table 8 presents the results.

Even though *Macro Conditional Alpha* has some limited statistically significant predictive power of the actual alpha, its predictive direction is inconsistent. For example, funds that perform well when market returns are high previous tends to perform worse in the future when the market returns are high again. Overall, the results suggest that funds performance persists when information is informative to funds, but not the information themselves.

[Insert Table 8 Here]

B5. Conditional Performance Persistence under Berk and Green Model

[Insert Table 9 Here]

Berk and Green (2004) introduces a hypothesis that funds' performance is hard to persist, especially when their size becomes too large due to the diminishing return-to-scale in the mutual fund industry. Are similar diminishing return-to-scale effects apply to our conditional alpha measure? To test the hypothesis, we separate the funds into five samples based on funds size and test their conditional performance persist analysis using equation 7. Table 9 presents the results. According to the results, the persistence of model-based conditional skill measures declines for the largest mutual funds, consistent with the diminishing return-to-scale hypotheses.

VI. Conclusion Remarks

In this paper, we apply a state-of-the-art deep learning model to understand and predict dynamic patterns in mutual fund returns. A long-short portfolio based on the model's prediction generates a 2.8% annualized Carhart 4-factor alpha. This abnormal performance is persistent for up to four years. The model improves the prediction of future fund alphas substantially by increasing the R-squared by more than 25% in a predictive regression that includes other fund skill measures as well as fund and time fixed effects. The model's predictive power derives from its ability in capturing fund skill embedded in dynamic strategies.

We construct model-based conditional skill measures that depend on the inferred informativeness of macroeconomic and fundamental variables. Such measures are predictive of fund performance in future periods when the conditioning variables are highly informative. The conditional performance of these measures are also persistent. Overall, our results suggest that mutual fund have various specific skills that generate superior returns when the time is right.

Appendix A: Definitions of Variables

Variable	Definition
Alpha	The Fama-French-Carhart four-factor alpha is the intercept of the rolling window regression of the monthly net return during 24 months on Mktrf, SMB, HML, and UMD factors, expressed in percentage points.
TNA, Log(TNA)	TNA is a a fund's TNA ($mm) prior to month t. Log(TNA) is natural logarithm of a fund's TNA.
Flow	The monthly flow for a fund in month t-1, calculated as $Flow = TNA_{i,t}/TNA_{i,t-1} - 1 - r_{i,t}$, where $r_{i,t}$ is the net return in the prior month, expressed in percentage points.
Expense	The most recent expense ratio prior to month t.
Turnover	The most recent turnover ratio prior to month t.
Load	Dummy variable if funds have load.
Return Gap	The Return Gap measure from Kacperczyk, Sialm, and Zheng (2008).
Cash Holdings	The most recent amount of fund invested in cash prior to month t, expressed in percentage.
Management Fee	The most recent management fees prior to month t.
Manager Tenure	The number of months since a portfolio manager is hired. If there are multiple managers for a fund, the longest tenure is used.
Alpha Mean	Mean of the alpha in a fund-year.
Alpha Std	Standard deviation of the alpha in a fund-year.
Industry production	Percentage change of industry production index (INDPRO) from year ago.
Inflation	Percentage change of consumer price index for all urban consumers (CPIAUCSL) from year ago.
Oil price	Percentage change of crude oil prices:West Taxas Intermediate (WTI) from year ago.
T-Bill yield	Percentage change of 3-month treasury bill (TB3MS).
Term spread	Percentage change of the difference between 10-year treasury (GS10) and 3-month treasury bill (TB3MS).
Default spread	Percentage change of the difference between Baa corporate bond yield (BAA) and Aaa corporate bond yield (AAA).
Crisis Dummy	Crisis dummy defined by NBER.
VIX	Percentage change of The CBOE volatility index.
VWRETD	Return of total return value-weighted index from CRSP.
SP500	S&P 500 index return.
Mkt-RF	Market excess return.
SMB	Size factor return in Fama-French five-factor (FF5) Model.
HML	Value Factor return in FF5 Model.
RMW	Profitablity Factor return in FF5 Model.
CMA	Investment Factor return in FF5 Model.
R_ME	Value Factor return in Hou, Xue, and Zhang (2015) q-factor (q5) Model.
R_IA	Investment factor return in q5 Model.
R_ROE	Equity factor return in q5 Model.
R_EG	Expected growth factor return in q5 Model.

(continued)

Variable	Definition
Announcement Count %	Percentage of the number of the earning announcement of a fund's holding in a month as that number of a year.
Alpha Importance	Variable importance of the historical alpha in the model.
Macro Importance	Sum of the variable importance of macroeconomics in the model. The variables include *Industry Production, Inflation, Oil Price, T-Bill, Term Spread, Default Spread, Crisis Dummy, VIX, VWRETD, SP500, Mkt-RF, SMB, HML, RMW, CMA, R_ME, R_IA, R_ROE, R_EG*

References

Amihud, Yakov, and Ruslan Goyenko, 2013, Mutual Fund's R2 as predictor of performance, *Review of Financial Studies* 26, 667–694.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance* 65, 179–216.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan B., and Jules H. van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.

Box, G. E.P., G. M. Jenkins, and J. F. MacGregor, 1974, Some Recent Advances in Forecasting and Control, *Journal of Applied Statistics* 23, 158–179.

Busse, Jeffrey A., Amit Goyal, and Sunil Wahal, 2010, Performance and persistence in institutional investment management, *Journal of Finance* 65, 765–790.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chinco, Alex, Andreas Neuhierl, and Michael Weber, 2021, Estimating the anomaly base rate, *Journal of Financial Economics* 140, 101–126.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang, 2020a, AlphaPortfolio: Direct Construction Through Reinforcement Learning and Interpretable AI, Working paper.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang, 2020b, Deep Sequence Modeling: Development and Applications in Asset Pricing, *Journal of Financial Data Science*, Forthcoming.

Cremers, K. J.Martijn, and Antti Petajisto, 2009, How Active Is Your Fund Manager A New Measure That Predicts Performance, *Review of Financial Studies* 22, 3329–3365.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2001, A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases, *The Journal of Finance* 56, 2415–2430.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, *Review of Financial Studies* 6, 1–22.

Evans, Richard B., 2010, Mutual fund incubation, *Journal of Finance* 65, 1581–1611.

Fama, Eugene F., and Kenneth R. French, 2010, Luck versus Skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915–1947.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the Factor Zoo: A Test of New Factors, *Journal of Finance* 75, 1327–1370.

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

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical Asset Pricing via Machine Learning, *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.

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

Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011, Risk shifting and mutual fund performance, *Review of Financial Studies* 24, 2575–2616.

Hunter, David, Eugene Kandel, Shmuel Kandel, and Russ Wermers, 2014, Mutual fund performance evaluation with active peer benchmarks, *Journal of Financial Economics* 112, 1–29.

Hyndman, Rob J, Anne B. Koehler, J. Keith Ord, and Ralph D Snyder, 2008, *Springer Series in Statistics Forecasting with Exponential Smoothing*.

Israel, Ronen, Bryan T. Kelly, and Tobias J. Moskowitz, 2020, Can Machines 'Learn' Finance?, *Journal of Investment Management* 1–21.

Jensen, Michael C., 1968, The Performance of Mutual Funds in the Period 1945-1964, *The Journal of Finance* 23, 389.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2014, Time-varying fund manager skill, *Journal of Finance* 69, 1455–1484.

Jones, Christopher S., and Haitao Mo, 2021, Out-of-Sample Performance of Mutual Fund Predictors, *Review of Financial Studies* 34, 149–193.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved actions of mutual funds, *Review of Financial Studies* 21, 2379–2416.

Li, Bin, and Alberto G. Rossi, 2020, Selecting Mutual Funds from the Stocks They Hold: A Machine Learning Approach, *Working paper* .

Li, Shiyang, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu Xiang Wang, and Xifeng Yan, 2019, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, *arXiv* .

Lim, Bryan, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister, 2019, Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, *arXiv* 1–27.

Moy, Ronald L., and Bill Griffeth, 1995, The Mutual Fund Masters., *The Journal of Finance* 50, 1345.

Pinnuck, Matt, 2005, What is the abnormal return performance of mutual funds due to private earnings information?, *Journal of Contemporary Accounting Economics* 1, 193–215.

Rangapuram, Syama Sundar, Matthias Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski, 2018, Deep state space models for time series forecasting, *Advances in Neural Information Processing Systems* 2018-Decem, 7785–7794.

Salinas, David, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski, 2020, DeepAR: Probabilistic forecasting with autoregressive recurrent networks, *International Journal of Forecasting* 36, 1181–1191.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017, Attention is all you need, *Advances in Neural Information Processing Systems* 2017-December, 5999–6009.

Wen, Ruofeng, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka, 2017, A Multi-Horizon Quantile Recurrent Forecaster .

West, Kenneth D., 2006, Forecast Evaluation, in *Handbook of Economic Forecasting*, volume 1, 99–134.

Zhang, Alan, 2021, Uncovering Mutual Fund Private Information with Machine Learning, *Working paper* .

Figure 1: Basic Structure of Temporal Fusion Transformer (TFT) Model

This figure plots the basic structure of the TFT model. The model is multi-horizon forecasting with dynamic, deterministic, and static variables with exportable variable importance and attention outputs.

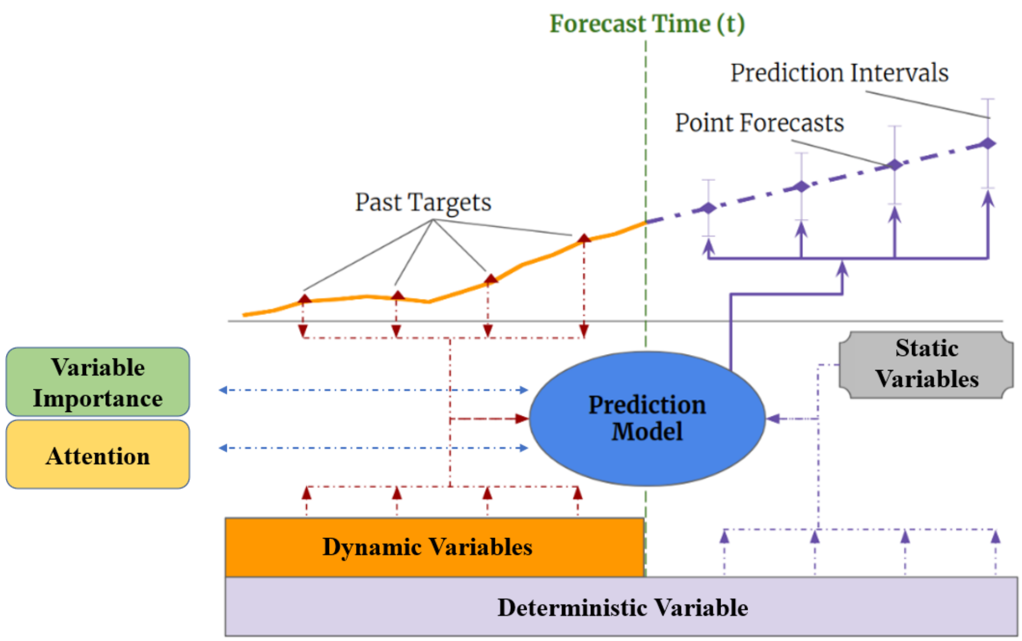

Figure 2: The Relative Importance of Variables in the Model

The figure plots the relative importance of all variables of the TFT model from 1990m1 to 2009m12. The relative importance of each variable is averaged first across funds and then for all months and measured in percentage points. All attributes are defined in Appendix A.

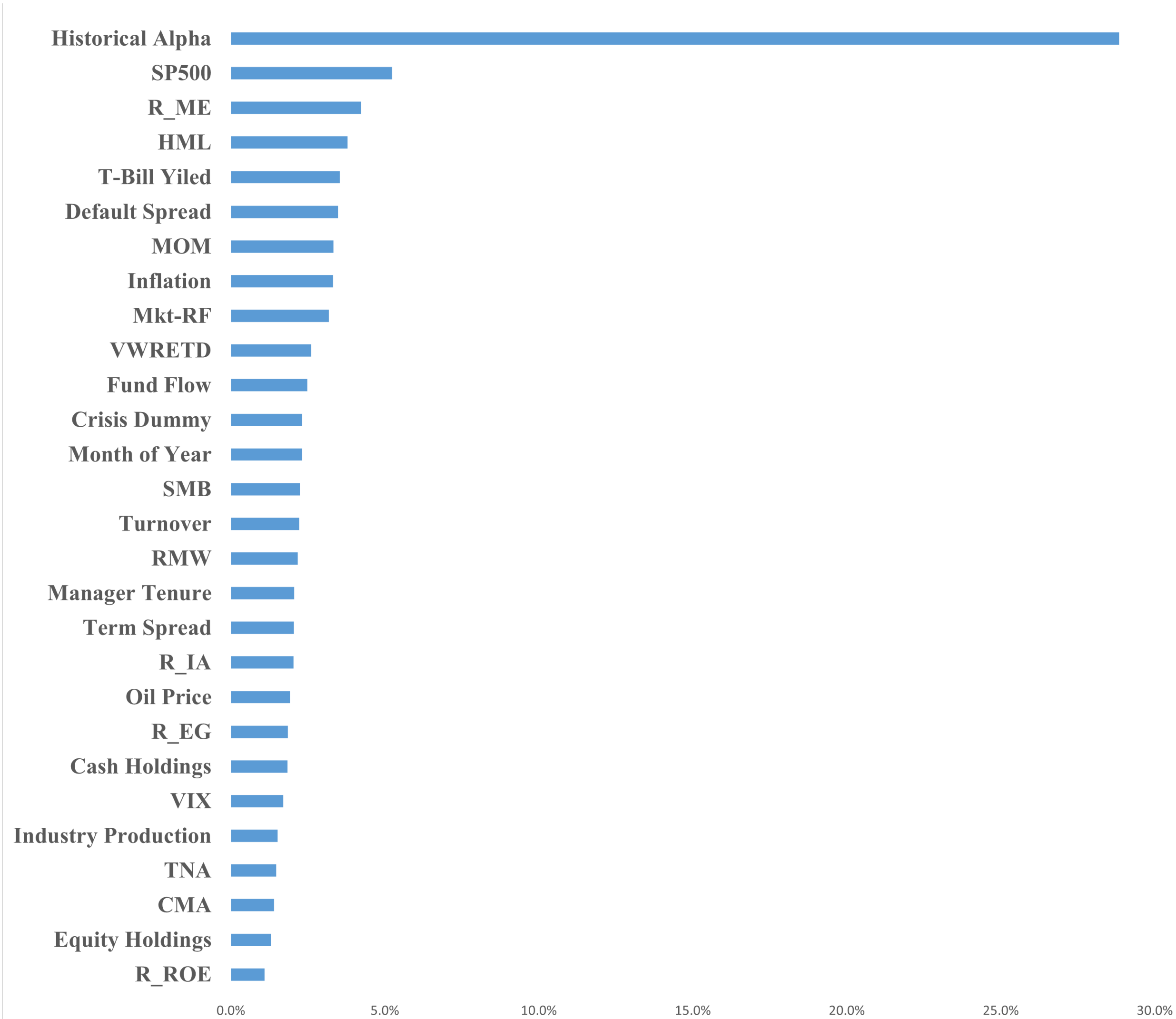

Figure 3: Variable Importance of Returns and Frequency of Earnings Announcements

This figure plots the relationship between the one-month ahead variable importance of mutual fund risk-adjusted returns, or alphas, and the number of earnings announcements in the month. The variable importance of mutual fund alphas is first averaged across all funds and then average across all years in the sample for each calendar month of the year. The frequency of earnings announcements is the total number of public companies announcing earnings in each calendar month of the year in our sample.

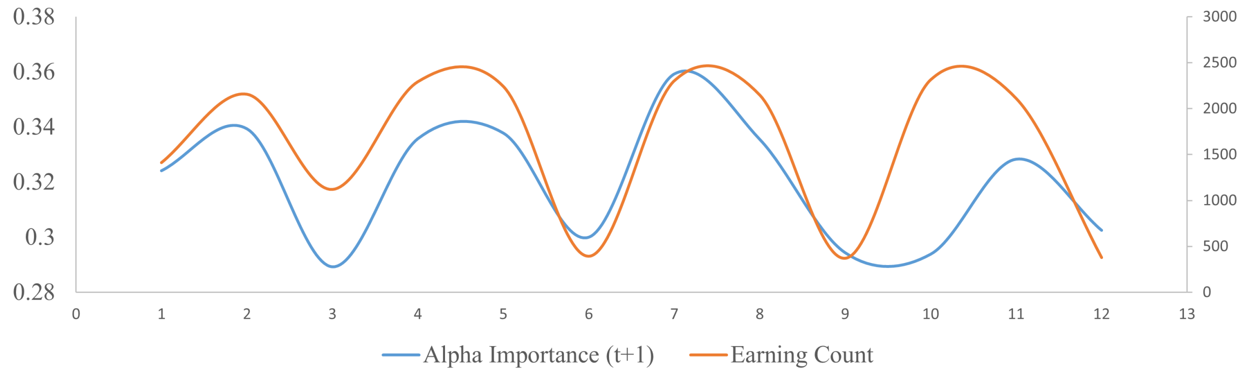

Figure 4: The Times Series of Attention in the Model

This figure plots the average attention of the model during our sample period. Attention for each calendar month is defined as the average attention the model assigned to that month across all funds and forecast horizons. Sections marked as blue denote the crisis periods defined by NBER.

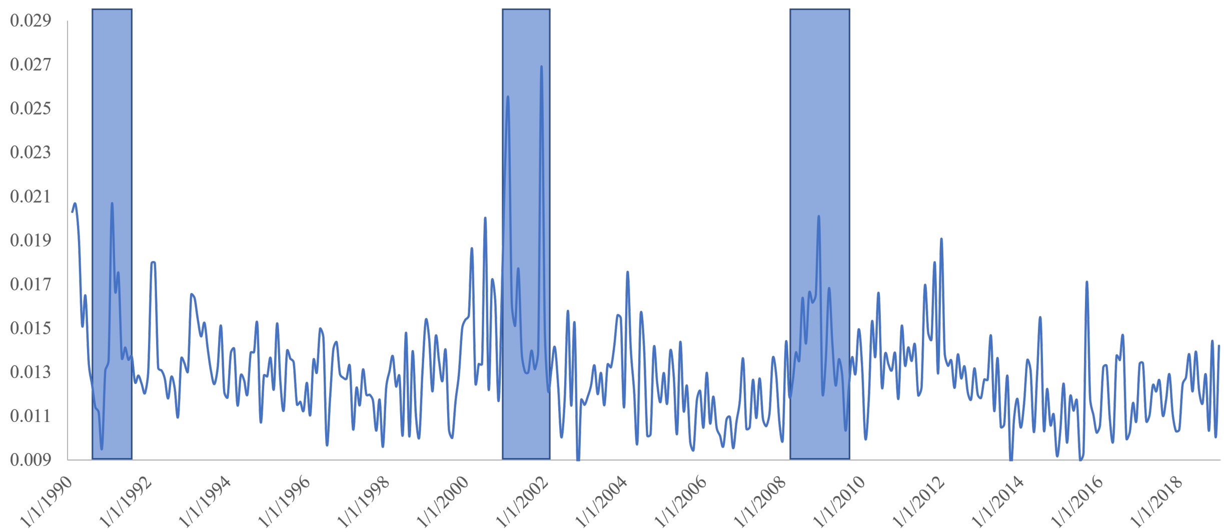

Figure 5: Unconditional Performance Persistence

This figure plots the confidence intervals of subsequent one-year performance by the rank average historical alpha Carhart's four-factor alpha over the past 12 months. In each calendar year from 2010 to 2019, funds are ranked into quintile portfolios based on historical alpha.

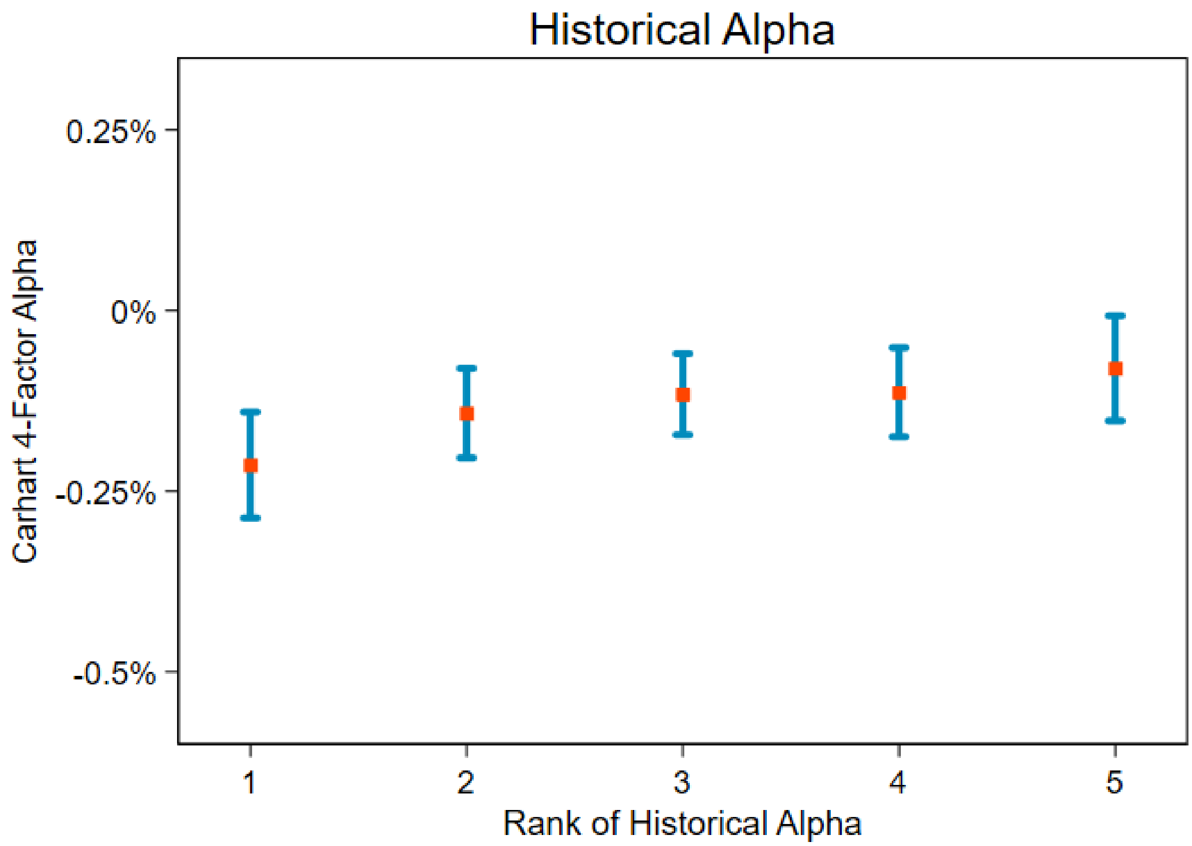

Figure 6: Conditional Performance Persistence

This figure plots the confidence intervals of subsequent one-year performance by the rank V.I. conditional alpha under high and low informative period of macroeconomics and fundamental information. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than 80 percent of the other funds in the same month. In each calendar year from 2010 to 2019, funds are ranked into quintile portfolios based on a variable V.I. conditional alpha. The future time period is separated into the high and low informative period. The x-axis represents the rank of the V.I. conditional alpha. The y-axis represents the Carhart's four-factor alpha.

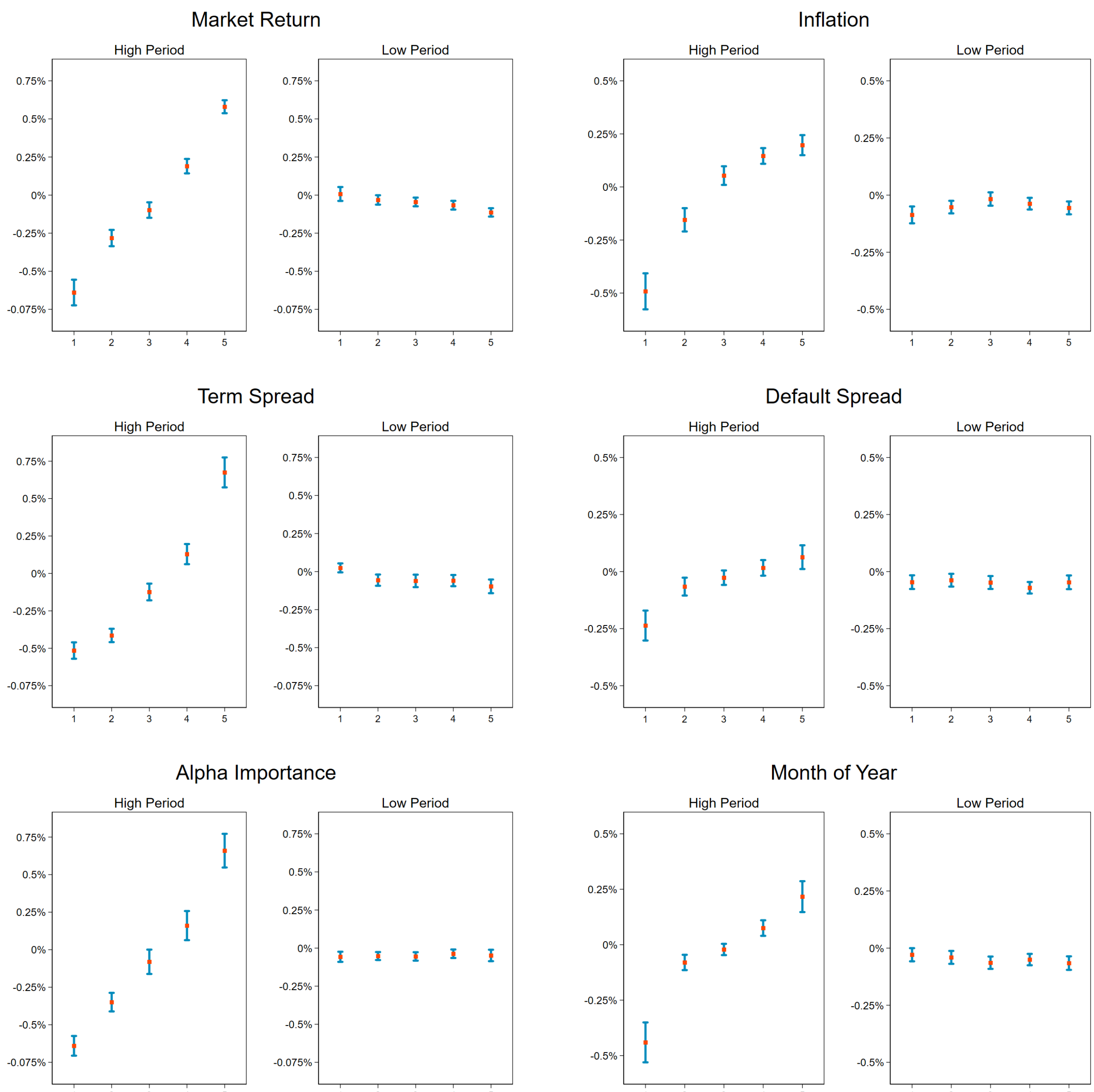

Table 1: Summary Statistics

This table provides summary statistics. Fund returns and characteristics are based on the sample of active US domestic equity mutual funds from 1990 to 2019. Macro-level variables are calculated monthly based on information available in the previous month. Variables are defined in Appendix A.

Variables	Mean	Median	Std	P25	P75
Fund Return & Characteristics					
Alpha	-0.12	-0.11	-0.11	-0.97	0.74
Flow	1.75	-0.20	-0.20	-1.34	1.49
TNA	1,185.01	195.90	195.90	47.30	815.90
Load	0.56	1.00	1.00	0.00	1.00
Cash	5.00	2.42	2.42	0.70	5.49
Expense	1.20	1.15	1.15	0.90	1.46
Management fee	0.58	0.72	0.72	0.50	0.88
Turnover	0.92	0.62	0.62	0.33	1.07
Total number of funds	3717				
Macro Variables					
Industry production	1.90	2.66	2.66	0.71	4.03
Inflation	2.45	2.49	2.49	1.70	3.07
Oil price	8.73	5.20	5.20	-12.28	26.67
T-Bill yield	3.74	0.00	0.00	-3.33	3.85
Term spread	0.01	-2.00	-2.00	-15.00	14.00
Default spread	-0.04	-1.00	-1.00	-4.00	4.00
Crisis Dummy	0.09	0.00	0.00	0.00	0.00
VIX	1.10	-1.38	-1.38	-8.89	6.79
Value-weighted return	0.87	1.34	1.34	-1.69	3.53
S&P 500 return	0.71	1.11	1.11	-1.74	3.25
Market risk premium	0.67	1.18	1.18	-1.90	3.37

Table 2: Comparisons of Performances of Machine Learning Models

This table reports the out-of-sample R^2_{OOS} based on out-of-sample predictions of different models: (i) TFT model, (ii) OLS model, (iii) ridge OLS regression model, (iv) decision tree model, (v) AdaBoost model , (vi) random forest model, (vii) Neural network with two hidden layers (32 and 16 neurons), and (viii) Neural network with three hidden layers (32, 16, and 8 neurons).

TFT	OLS	Ridge	Decision Tree	AdaBoost	Random Forest	NN2	NN3
0.3584	0.0272	0.0052	0.0039	0.0486	0.0363	0.0749	0.0392

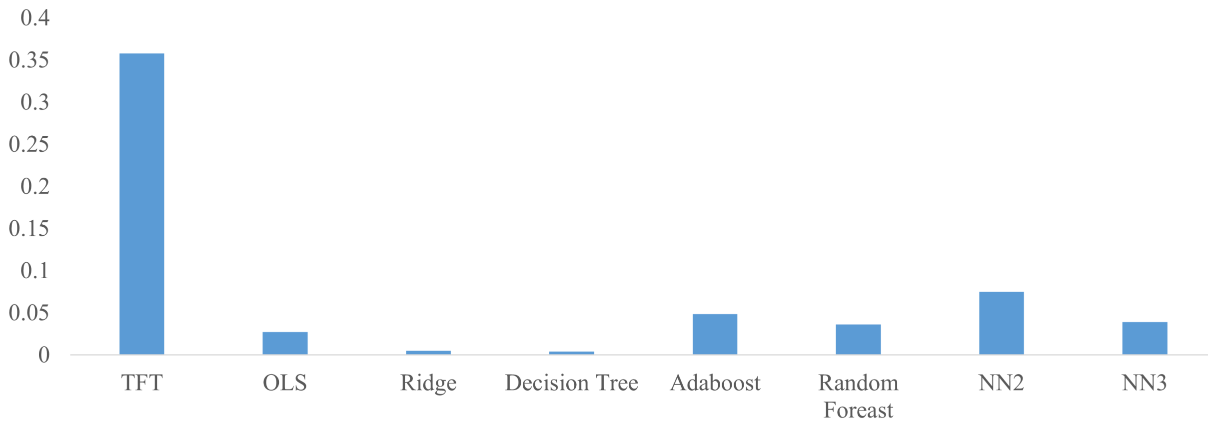

Table 3: Regression Analysis of Model-Predicted Fund Performance

This table reports the regression of future fund alpha on the predicted alpha from the TFT model, historical alpha over the previous 12 months, return gap, and other fund characteristics. Variables are defined in Appendix A. The *t*-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables			Alpha			
Predicted Alpha	0.510***	0.427***	0.434***	0.451***	0.439***	0.454***
	(11.25)	(9.48)	(9.60)	(9.78)	(10.13)	(10.36)
Historical Alpha				-0.014***		-0.012***
				(-3.46)		(-3.04)
Return Gap					-3.633	-3.663
					(-0.89)	(-0.90)
Log (TNA)	-0.000	-0.123***	-0.131***	-0.132***	-0.138***	-0.139***
	(-0.12)	(-7.98)	(-8.52)	(-8.55)	(-9.31)	(-9.29)
Fund Flow	0.395*	-1.154***	-1.249***	-1.231***	-1.298***	-1.281***
	(1.87)	(-4.22)	(-4.58)	(-4.53)	(-4.97)	(-4.90)
Cash Holdings	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001
	(-1.03)	(-0.51)	(-0.31)	(-0.33)	(-0.51)	(-0.52)
Expense Ratio	-8.534***	-8.697	1.025	1.180	-5.322	-5.126
	(-4.36)	(-0.81)	(0.09)	(0.11)	(-0.52)	(-0.49)
Management Fee	0.021	-0.083	-0.056	-0.058	0.000	-0.001
	(0.95)	(-1.21)	(-0.86)	(-0.88)	(0.01)	(-0.02)
Turnover Ratio	-0.019**	-0.013*	-0.009	-0.010	-0.010	-0.010
	(-2.50)	(-1.81)	(-1.33)	(-1.34)	(-0.59)	(-0.58)
Fund Fixed Effect	No	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	No	Yes	Yes	Yes	Yes
Observations	101,076	101,076	101,076	101,076	96,240	96,240
Adjust R-squared	0.007	0.008	0.010	0.010	0.011	0.011
Adjust R-squared Contribution by Predicted Alpha	45.04%	26.58%	20.10%	20.48%	19.72%	20.28%

Table 4: Persistence of Fund Performance

This table reports the results of the persistence analysis of the TFT model. It presents the different prediction horizons of post-ranking monthly basis points of alphas from net fund returns for funds sorted into deciles portfolios based on TFT models. The prediction horizon include: (i) 0-12 months, (ii) 12-24 months, (iii) 24-36 months, (iv) 36-48 months, and (v) 48-60 months. The results reflect 84 individual monthly observations over the 2012m1-2019m12 out-of-sample period. The *t*-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

	Prediction Horizon				
	0-12 months	12-24 months	24-36 months	36-48 months	48-60 months
Bottom	-25.29	-19.47	-22.79	-22.44	-9.17
2	-20.81	-17.90	-19.34	-19.13	-15.94
3	-16.79	-16.46	-19.94	-20.84	-13.51
4	-17.97	-14.14	-20.53	-18.39	-13.94
5	-10.84	-12.82	-14.70	-15.54	-8.55
6	-12.88	-12.74	-11.81	-13.16	-13.77
7	-11.53	-12.64	-9.94	-11.99	-13.80
8	-7.84	-11.01	-8.02	-8.42	-8.02
9	-6.32	-6.78	-5.14	-4.67	-2.68
Top	-2.05	-2.19	2.00	-0.23	-4.35
Top-Bottom	23.24**	17.27*	24.79**	22.21*	4.82
t-Statistic	(2.51)	(1.81)	(2.37)	(1.91)	(0.34)

Table 5: Correlation of V.I. Conditional Alpha

This table reports the correlation of V.I. conditional alpha on different macroeconomics and fundamental information. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than that of 80 percent of the other funds in the same month.

	Market Return	Inflation	Term Spread	Default Spread	Alpha Importance	Month of Year
Market Return	1.00					
Inflation	0.27	1.00				
Term Spread	-0.44	-0.15	1.00			
Default Spread	0.10	0.21	-0.05	1.00		
Alpha Importance	-0.10	-0.09	0.24	-0.01	1.00	
Month of Year	0.09	0.02	0.03	-0.04	-0.04	1.00

Table 6: Conditional Performance Persistence

This table reports the regression of future fund alpha on the V.I. conditional alpha. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period. Variable high (low) informative period is defined as the month during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than that of 80 percent of the other funds in the same month. The regressions are grouped into the high and low informative period. Control variables include historical alpha, size, and flow. The t-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Panel A: Fundamental Conditions

	(1)	(2)	(3)	(4)	(5)
		Target		Month of the Year	
Variables	Unconditional	High	Low	High	Low
V.I. Conditional Alpha		0.321***	-0.020***	0.138***	0.010
		(13.88)	(-2.60)	(5.72)	(1.22)
Historical Alpha	0.066***	0.036	0.045***	0.022	0.065***
	(4.56)	(0.97)	(2.69)	(0.84)	(3.28)
Log (TNA)	0.000***	0.000***	0.000***	-0.000	-0.000
	(6.94)	(12.01)	(13.24)	(-0.24)	(-1.20)
Flow	0.001	-0.031***	-0.002**	0.003	0.000
	(1.12)	(-9.73)	(-2.36)	(1.26)	(0.42)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes
Style Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	96,088	18,784	65,063	18,804	66,859
adjust R-squared	0.0433	0.206	0.0797	0.168	0.0484

Panel B: Macroeconomics Conditions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Market Return		Term Spread		Inflation		Default Spread	
Variables	High	Low	High	Low	High	Low	High	Low
V.I. Conditional Alpha	0.519***	-0.075***	0.409***	-0.065***	0.277***	0.007	0.126***	0.001
	(23.92)	(-6.96)	(13.64)	(-6.76)	(12.90)	(0.85)	(4.03)	(0.08)
Historical Alpha	-0.008	0.047**	0.021	0.097***	-0.018	0.056***	0.013	0.069***
	(-0.25)	(2.53)	(0.67)	(4.39)	(-0.65)	(3.20)	(0.48)	(3.58)
Log (TNA)	0.000***	0.000***	-0.000	0.000***	0.000**	0.000***	0.000***	0.000
	(10.71)	(12.19)	(-0.61)	(6.48)	(2.53)	(3.79)	(8.55)	(1.25)
Flow	-0.015***	-0.009***	-0.004**	0.009***	-0.011***	-0.004***	-0.005	0.002*
	(-4.67)	(-9.53)	(-2.38)	(7.82)	(-4.30)	(-5.07)	(-1.47)	(1.84)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,924	73,262	18,816	67,228	18,703	69,489	18,846	69,741
adjust R-squared	0.240	0.0624	0.211	0.0929	0.242	0.0543	0.172	0.0512

Table 7: Conditional Performance Persistence in Long Term

This table reports the coefficients and the *t*-statistics of the regression of future fund alpha on the V.I. conditional alpha during the high informative period in multiple prediction horizon, including i) 12-24 months, (ii) 24-36 months, and (iii) 36-48 months. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than 80 percent of the other funds in the same month. Control variables include historical alpha, size, and flow. The *t*-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Market Return	Term Spread	Inflation	Default Spread	Target	Month of the Year
High	High	High	High	High	High
12-24 Months					
0.430***	0.337***	0.196***	0.097***	0.242***	0.105***
(20.29)	(12.31)	(10.26)	(3.33)	(10.23)	(5.02)
24-36 Months					
0.339***	0.293***	0.169***	0.101***	0.201***	0.098***
(13.50)	(10.54)	(8.96)	(3.81)	(8.36)	(5.21)
36-48 Months					
0.300***	0.261***	0.127***	0.091***	0.152***	0.073***
(15.04)	(10.19)	(8.23)	(4.59)	(6.71)	(4.01)

Table 8: Conditional Performance Persistence on Macro Conditional Alpha

This table reports the regression of future fund alpha on the V.I. conditional alpha controlling the macro conditional alpha. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than that of 80 percent of the other funds in the same month. The macro conditional alpha is calculated as the average Carhart's four-factor alpha when the macro variable is higher than 80 percent of the time in the sample from 1990m1 to 2019m12. The regressions are grouped into the high informative period and the low informative period. Control variables include historical alpha, size, and flow. The *t*-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Market Return		Term Spread		Inflation		Default Spread	
Variables	High	Low	High	Low	High	Low	High	Low
V.I. Conditional Alpha	0.518***	-0.077***	0.422***	-0.063***	0.274***	-0.006	0.110***	-0.006
	(23.76)	(-7.24)	(13.65)	(-6.45)	(12.49)	(-0.75)	(3.40)	(-0.83)
Macro Conditional Alpha	-0.196*	-0.052	-0.253**	-0.083	0.037	0.420***	0.132**	0.149***
	(-1.84)	(-1.00)	(-2.48)	(-1.46)	(0.35)	(5.42)	(2.13)	(2.94)
Historical Alpha	0.060**	0.109***	0.062*	0.143***	-0.000	0.128***	0.051*	0.119***
	(2.00)	(4.92)	(1.74)	(6.19)	(-0.00)	(6.65)	(1.88)	(6.16)
Log (TNA)	0.000***	0.000***	-0.000	0.000***	0.000**	0.000***	0.000***	0.000
	(10.58)	(12.22)	(-0.14)	(6.19)	(2.48)	(2.88)	(8.34)	(0.57)
Flow	-0.015***	-0.009***	-0.004**	0.009***	-0.012***	-0.005***	-0.006*	0.001
	(-4.79)	(-9.84)	(-2.50)	(7.95)	(-4.38)	(-6.03)	(-1.67)	(1.16)
Time Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Style Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,890	73,180	18,787	67,164	18,688	69,422	18,829	69,665
adjust R-squared	0.241	0.0632	0.212	0.0935	0.242	0.0564	0.172	0.0523

Table 9: Conditional Performance Persistence by Size

This table reports the regression of future fund alpha during the high informative period on the V.I. conditional alpha grouped by fund size. V.I. conditional alpha is calculated as the average Carhart's four-factor alpha during a variable high informative period over the past five years. Variable high (low) informative period is defined as the month when the variable importance of a variable for a fund is higher (lower) than that of 80 percent of the other funds in the same month. Control variables include historical alpha, size, and flow. The *t*-statistics, in parentheses, are based on standard errors clustered by funds. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Group	Market Return	Inflation	Term Spread	Default Spread	Alpha Importance	Month of Year
	High	High	High	High	High	High
5	0.468***	0.308***	0.324***	0.105**	0.244***	0.242***
	(9.83)	(6.46)	(5.34)	(2.32)	(5.06)	(4.82)
4	0.511***	0.244***	0.468***	0.137	0.275***	0.117***
	(10.93)	(5.29)	(9.56)	(1.60)	(4.55)	(2.88)
3	0.517***	0.347***	0.460***	0.103*	0.359***	0.158***
	(14.85)	(6.90)	(11.00)	(1.90)	(7.46)	(4.00)
2	0.496***	0.236***	0.315***	0.088*	0.355***	0.069
	(10.36)	(5.37)	(4.40)	(1.81)	(8.07)	(1.60)
1	0.524***	0.213***	0.386***	0.072	0.315***	-0.016
	(12.59)	(5.28)	(6.83)	(1.64)	(7.31)	(-0.29)