Do the Collective Trades of Market Participants Contain Information about Stocks? A Machine Learning Approach

Victor DeMiguel Li Guo Bo Sang Zhe Zhang*

Current draft: June 2023

Abstract

We use machine learning to study whether the joint trading behavior of multiple market participants contains information about future stock returns. Our machine learning approach can capture complex interactions among market participants and the nonlinear effects of their collective trading on stock returns. Our trading-based predictor produces robust out-of-sample performance, with monthly alphas for a long-short portfolio exceeding 1%. Moreover, it forecasts firm fundamentals related to future cash flows and assigns stocks on the right side of anomalies. Overall, our findings suggest that accounting for convoluted interactions between the trades of diverse market participants provides valuable information for price discovery.

JEL classification: G10, G11, G23
Keywords: Return predictability; Trading; Machine learning; Institutions; Retail investors; Short sellers

*DeMiguel: London Business School. avmiguel@london.edu. Guo: School of Economics, Fudan University; Shanghai Institute of International Finance and Economics. guo_li@fudan.edu.cn. Sang: School of Business, University of Bristol. bo.sang@bristol.ac.uk. Zhang: Lee Kong Chian School of Business, Singapore Management University. joezhang@smu.edu.sg.

1 Introduction

How information is impounded into asset prices through trading is a central theme in financial markets. While there is an extensive literature on the informativeness of the trades of various market participants, most studies tend to focus on one particular type of investors, such as mutual funds, hedge funds, short sellers, or retail investors.[1] McLean, Pontiff, and Reilly (2022) conduct a comprehensive analysis on the trades of nine market participants and their relation with return predictability and anomalies, by focusing on the marginal effect of each type of participant.[2]

If market participants trade on value-relevant information, a combined signal reflecting their collective trading should predict future stock returns. Different investors may possess different types of value-relevant information, and they may be informed at different times. For uninformed investors who systematically make wrong bets, their trading can also reveal *negative* information about future stock returns. Prior research has highlighted the importance of studying the interactions between investors (e.g., learning, competition) to explore the information content of their trades (Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981; Goldstein and Yang, 2015). Naturally, a combined signal from various investor types should account for such interactions. In this paper, we examine the combined informational role of multiple market participants.[3] In particular, we construct a composite return predictor by aggregating information from the trading of the same nine market participants considered by McLean, Pontiff, and Reilly (2022) and examine its ability to predict stock returns out of sample.

In constructing the composite return predictor, we adopt a set of nonlinear machine learning methods to allow for possible complex interactions among trading signals and the

[1] E.g., Diether, Lee, and Werner (2009), Aggarwal and Jorion (2010), Boehmer, Huszar, and Jordan (2010), Baker, Litov, Wachter, and Wurgler (2010), Kaniel, Liu, Saar, and Titman (2012), Kelley and Tetlock (2013), Cao, Liang, Lo, and Petrasek (2018), Boehmer, Jones, Zhang, and Zhang (2021).

[2] A few studies examine the pairwise relations between two investor types (e.g., Sias and Whidbee, 2010; Massa, Qian, Xu, and Zhang, 2015; Jiao, Massa, and Zhang, 2016).

[3] Mutual funds, hedge funds, banks, firms, insurance companies, other institutions, wealth management firms, short sellers, and retail investors.

nonlinear relationship between the trading signals and stock returns. Specifically, we use random forest (RF) and gradient boosting regression trees (GBRT), which are based on regression trees, and artificial neural networks (ANN) to form the composite signal.[4] These models are well suited to identify nonlinearities and interactions in the relation between market participant trades and future stock returns. For comparison, we also employ five linear models, including principal components regression (PCR), partial least squares (PLS), adaptive least absolute shrinkage and selection operator (ALasso), ridge regression (Ridge), and elastic net (ENet), in addition to OLS. These algorithms have been widely used in the recent asset pricing literature (e.g., Chinco, Clark-Joseph, and Ye, 2019; Gu, Kelly, and Xiu, 2020; Dong, Li, Rapach, and Zhou, 2022; Leippold, Wang, and Zhou, 2022), and they can help reduce overfitting due to collinearity among predictors.[5]

Every month, for every stock in our sample, we train each of the above models to predict future monthly returns using changes in the holdings of nine market participants (or subsets of them) as inputs, with a five-year rolling window. For each stock, we then take the average predicted returns across the six linear models to construct a composite linear return predictor (LCP), and we take average predicted returns across the three nonlinear models to form a composite nonlinear return predictor (NLCP).

We sort our sample of stocks into decile portfolios based on the LCP (NLCP) every month, and evaluate the out-of-sample monthly factor-adjusted alphas. We consider three different factor models: the Fama and French (2015) five-factor model (FF5), the Hou, Mo, Xue, and Zhang (2021) model (q5), and the Stambaugh and Yuan (2017) mispricing model (MISP).

[4]The literature has documented the presence of nonlinear relationships between predictors and future asset returns (e.g., Kelly, Pruitt, and Su, 2019; Gu, Kelly, and Xiu, 2020; Bianchi, Büchner, and Tamoni, 2021; van Binsbergen, Han, and Lopez-Lira, 2022; Leippold, Wang, and Zhou, 2022). Given the complex interactions among multiple types of market participants, it is reasonable to expect similar nonlinear relation between trading signal-based predictors and future stock returns.

[5]Hastie, Tibshirani, and Friedman (2009) discuss how nonlinear machine learning models, such as neural networks and regression trees, can be used to capture complex, nonlinear relationships in the data. They note that such models are less sensitive to multicollinearity because they do not require linear relationships between the predictors and the response. The authors also explain that nonlinear models can be more flexible in terms of the types of functional relationships they can capture, which can help to alleviate multicollinearity.

For LCP-sorted portfolios, the decile with the highest predicted returns (High) generally outperforms the bottom decile (Low), although the difference is not always statistically significant. The FF5 and MISP alphas for the portfolio that goes long High and short Low (High - Low) are 0.85% and 0.93%, respectively, both statistically significant at the 5% level. The alpha spread for High – Low with respect to the q5 model is 0.34%, albeit statistically insignificant. The linear models seem to have difficulty in predicting returns of poorly performing stocks. For example, although the alphas for the *Low* decile are negative, most of them are statistically insignificant, with the exception of the alpha for the MISP model.

The predictability of the NLCP is much stronger. The return spreads and alphas for High – Low are positive and statistically significant at the 5% level across all factor models. The High – Low alpha spreads span from 1.11% for the q5 model to 1.39% for the FF5 model, substantially higher than those for the LCP-sorted portfolios. Moreover, the NLCP can predict the returns for stocks with both good and bad performance. The alphas for the *Low* decile are consistently negative, ranging from -0.75% for the q5 model to -1.04% for the FF5 model, and the alphas for the *High* decile range from 0.35% to 0.36%, all statistically significant at the 5% level. These results highlight not only the information content of the trading of multiple market participants, but also the nonlinearities and interactions in the relation between these trading signals and future stock returns.

Does the return predictability of the composite signal come from the (nonlinear) interaction of the trading by various types of market participants, or is it driven by the trading of a few types? McLean, Pontiff, and Reilly (2022) show that only short selling and firm trading are on the right side of the anomalies and return predictability. To examine the relative contribution of the trading from various market participants to the composite return predictor, we follow Nembrini, König, and Wright (2018) to compute the importance for each variable. Among all types of market participants in our sample, firms, short sellers, and retail investors contribute the most to the predictive model through their trading, together accounting for more than 50% of the model performance, followed by mutual funds. However, it is worth noting that while individually the other types contribute less to the model's performance, collectively the (nonlinear) combinations of their trading signals contain important information about future stock returns. When we train restrictive models using only firm trading, short selling, and retail trading, the predictive power of their composite predictor is subsumed by that of the full model. When we include the trading of individual types of market participants together with NLCP for the full model in Fama-Macbeth regressions, NLCP remains the only statistically significant predictor of next month returns.

The predictability of NLCP is robust to transaction costs, and lasts for up to nine months. The return predictability of NLCP is stronger for smaller firms and firms with higher idiosyncratic volatility, consistent with the combined information in the trades of market participants being more valuable in more uncertain information environments and for firms with greater information asymmetry, although the effects are similar regardless of firm age or the level of analyst coverage.

If the nonlinear predictor NLCP can predict future realized stock returns, the predicted returns may contain information about firm fundamentals related to future cash flows, or they may identify current mispricing and forecast returns driven by mispricing. We find evidence of both. Prior studies have mixed evidence on whether the trading by various types of institutions and retail investors predict firm fundamentals (e.g., Baker, Litov, Wachter, and Wurgler, 2010; Boehmer, Jones, Wu, and Zhang, 2020; Boehmer, Jones, Zhang, and Zhang, 2021). We show that NLCP predicts both the levels and the changes of return on assets (ROA), cash flow, and earnings surprises (SUE) for the next quarter. One standard deviation increase in NLCP leads to 27% increase in ROA, 25% increase in free cash flow, and 1.56% increase in SUE, all statistically significant at the 5% level.

Prior studies suggest that market participants such as institutional investors trade on the wrong side of anomalies; see, for instance, Hirshleifer, Teoh, and Yu (2011), Edelen, Ince, and Kadlec (2016), Patton and Weller (2020), and McLean, Pontiff, and Reilly (2022). We sort our sample stocks into quintile portfolios based on the NLCP, the quintile with the highest predicted returns being the *High* portfolio, and that with the lowest predicted returns *Low*. We then examine firm characteristic values of these portfolios with respect to each anomaly, as well as the characteristic spreads between *High* and *Low* quintiles.

Out of the 102 stock return anomalies documented in Green, Hand, and Zhang (2017), we select 12 whose returns are statistically significant (t-stat >1.66) during our sample period. Our nonlinear machine learning prediction is on the right side of most anomaly returns, that is, the sign of the firm characteristic value spread between the *High* and *Low* quintile is consistent with the direction of the corresponding anomaly, with the exception of the *market capitalization* and *idiosyncratic volatility* anomalies. We also follow Green, Hand, and Zhang (2017) and sort our sample stocks into quintile portfolios based on their Net anomaly indicator scores (NET). The nonlinear composite predictor (NLCP) for the *Long* quintile is 0.36% higher than that for the *Short* portfolio, and statistically significant at the 1% level. This is also consistent with our predictive signal obtained from market participant trades being on the right side of the return anomalies. Interestingly, when we only include the firm trading and short sales, or firm trading, short sale, and retail trading to construct the nonlinear composite return predictor, the Long - Short return spread decreases to 0.23% and 0.22%, respectively. This is again consistent with earlier results that including the full set the trading signals improves the model's performance, even though some trading signals individually are not strong predictors of future stock returns.

Given that the NLCP-sorted portfolios are on the right side of most anomalies, we expect that a factor model based on the NLCP may help to explain the cross-section of anomaly returns. To test this, we construct a factor-mimicking portfolio by taking a long position in the top decile and a short position in the bottom decile portfolio of stocks sorted by NLCP. We then form a two-factor model with the market factor and the combined trading signal factor. Using the same set of anomaly-sorted portfolios as above, we test the portfolio alphas with respect to our two-factor model, and compare them with the alphas with respect to the FF5 model, the q5 model, and the MISP model. The average absolute value of alpha with respect to FF5, q5, and MISP are 0.64%, 0.52%, and 0.58%, with average absolute t-values of 1.9, 1.5, and 1.9, respectively. The F-test rejects the null hypotheses that the alphas are jointly zero. On the other hand, the average absolute value of alpha with respect to the two-factor trading signal model is only 0.4%, with average absolute t-value of 1.2. The F-test cannot reject the null hypotheses that the anomalies alphas are jointly zero.

Our manuscript contributes to the literature that studies whether market participant trading contains information about future stock returns (Diether, Lee, and Werner, 2009; Aggarwal and Jorion, 2010; Boehmer, Huszar, and Jordan, 2010; Baker et al., 2010; Kaniel et al., 2012; Kelley and Tetlock, 2013; Cao et al., 2018; Boehmer et al., 2021; McLean, Pontiff, and Reilly, 2022). The key distinguishing feature of our manuscript is that we highlight the importance of using machine learning to capture the nonlinearities and interactions in the relation between the trades of multiple market participants and future stock returns.

The manuscript also adds to the growing literature that uses machine learning models in asset pricing and investment. Most of the existing work focuses on using machine learning to study cross-sectional and time-series stock returns, using historical data on returns and firm characteristics (Gu, Kelly, and Xiu, 2020; Kozak, Nagel, and Santosh, 2020; Bryzgalova, Pelger, and Zhu, 2021; Chatigny, Goyenko, and Zhang, 2022; Leippold, Wang, and Zhou, 2022). Several recent papers use machine learning to study mutual fund performance (DeMiguel, Gil-Bazo, Nogales, and Santos, 2022; Kaniel, Lin, Pelger, and Nieuwerburgh, 2022). We show that machine learning models are successful in extracting value-relevant information from the combined trading signals of multiple market participants, and the machine learning predictor based on these trading signals strongly predicts future stock returns beyond the historical returns and firm characteristics. Koijen and Yogo (2019) develop an equilibrium asset pricing model based on the asset demand of market participants, and they use institutional holdings data to demonstrate the model's implications on asset market movements, volatility, and predictability. Da (2022) builds a continuous-time model with many investors who possess information of heterogeneous quality and studies the equilibrium consequences of their collective trading on market efficiency. He estimates the model at the individual institutions level. We empirically construct our trading-based return predictor NLCP using the trades of nine different categories of market participants, without imposing structural restrictions on the interactions among market participants. We highlight the importance of nonlinearity in NLCP to predict stock returns and firm fundamentals, and we show that a NLCP-based aggregate pricing factor helps to explain the cross section of anomaly returns.

The remainder of this manuscript is organized as follows. Section 2 describes the data, machine learning techniques, variable construction, and descriptive statistics. Section 3 evaluates the performance of the trading signals from nine market participants as well as their composite return predictors. Section 4 investigates whether the combined trading signal predicts firm fundamentals and anomalies. Section 5 concludes.

2 Data, sample, and descriptive statistics

The main input variables trained in various machine learning models are trading signals from multiple market participants. We construct the trading variables from mutual funds, hedge funds, banks, firms, insurance companies, other institutions, wealth management firms, short sellers, retail investors in accordance with McLean, Pontiff, and Reilly (2022). We also follow Boehmer et al. (2021) algorithm to compute an alternative retail trading variable. Section 2.1 describes the data sources and trading variables as input. Section 2.2 summarizes the selected machine learning techniques and Section 2.3 explains the method we use to construct the variables of interest. The descriptive statistics of our sample is presented in Section 2.4.

2.1 Data sources and trading variables

We get institutional holdings data from *Thomson/Refinitiv S12* and *13F* to compute institutional trading signals. Following McLean, Pontiff, and Reilly (2022), we classify institutions into six types including mutual funds, insurance companies, banks, hedge funds, wealth management firms, and other institutions. Changes in 13F institutional holdings reflect trading signals from institutions. We also include the level of holdings of each type of these institutions to control for the potential effect of persistent demands on future prices (Gompers and Metrick, 2001).

To construct retail trading signals, we obtain daily off-exchange marketable orders from *TAQ* trade dataset between 2008 and 2020. We follow Boehmer et al. (2021) algorithm to identify retail trades and compute retail order imbalance on a monthly basis. We compute an alternative measure of retail order documented in McLean, Pontiff, and Reilly (2022).

The remaining input variables are trading signals from short sellers and firms. We use short interest to proxy for short seller trading. To obtain short interest ratio, we get the monthly short interest data from *Compustat* and scale it by the number of shares outstanding. Like McLean, Pontiff, and Reilly (2022), we sign the short interest ratio such that increases (decreases) in short interest reflect negative (positive) values of short seller trading signal. As for firm trading, we calculate changes in shares, that is, share issues minus share repurchases, divided by shares outstanding, which are extracted from *CRSP*. We sign the variable such that decreases (increases) in shares outstanding reflect positive (negative) values of firm trading signal. Appendix A.1 gives a more detailed explanation of how we construct our variables.

To construct the sample, we choose stocks from the *CRSP* sample after selecting common stocks (share code equals 10 or 11) listed on the NYSE, AMEX, or Nasdaq (exchange code equals 1, 2, or 3) and excluding stocks with prices under $1. Additionally, we obtain the accounting variables from *Compustat*. We download analyst forecast and recommendation data from the *Institutional Brokers' Estimate System (I/B/E/S)*. We follow Green, Hand, and Zhang (2017) to replicate their 102 stock anomalies. Our full sample period is 2008-2020.

2.2 Machine learning methods

We use the trading signals constructed in Section 2.1 as input variables in multiple machine learning models in order to get our key variables. We follow prior studies that use machine learning techniques for stock analysis (e.g., Gu, Kelly, and Xiu, 2020; Kozak, Nagel, and Santosh, 2020). A detailed explanation of using machine learning models to tune in parameters and create variables of interest can be found in Section 2.3 together with Appendix A.2.

For the linear combination of predictors besides the *ordinary least squares* (OLS), we utilize the adaptive version of *least absolute shrinkage and selection operator* (ALasso) as well as *ridge regression* (Ridge) to improve the predicting estimation (Tibshirani, 1996; Zou, 2006; Rapach, Strauss, Tu, and Zhou, 2019). We also use *elastic net* (ENet) to alleviate the inefficiency issue in the OLS. Additionally, we follow Gu, Kelly, and Xiu (2020) and use two dimension reduction techniques (which are, *principle components regression* (PCR) and *partial least square* (PLS)) that help reduce noise and correlation among predictors.

Furthermore, we take into account the nonlinear interactions among trading variables. According to Gu, Kelly, and Xiu (2020), adding nonlinearities in return predictive models substantially improves accuracy. DeMiguel et al. (2022) also finds that nonlinear machine learning models help select actively managed mutual funds that generate significant out-of-sample alphas. We follow Gu, Kelly, and Xiu (2020) and utilize three sets of nonlinear machine learning models: *gradient boosted regression trees* (GBRT), *random forest* (RF), as well as *artificial neural networks* (ANN). To briefly explain, *gradient boosted regression trees* not only helps incorporate multiway predictor interactions in the model, but also adds a boosting algorithm that recursively combines forecasts from regression trees to improve performance. In a similar vein, *random forest* combines forecasts from trees but with a bootstrap aggregation method, further lowering the correlation among regression trees and increasing the stability in prediction. As a more complicated machine learning model, *artificial neural networks* introduces "feed-forward" networks and increases flexibility and complex interactions among predictors. The hidden layers, of which we set the number to be four (ANN1, ANN2, ANN3, ANN4), are essential links between the inputs and output in the model and nonlinearly transform the predictors as a result.

2.3 Construction of key variables

We build the variables of interest from training samples in a rolling basis and examine their out-of-sample performance. Specifically, we aggregate trading signals from multiple market players considering both linear and nonlinear interactions among the signals using the machine learning tools documented in Section 2.2. The trading signals from nine market participants, as seen in Section 2.1, serve as input variables in predictive models. We conduct uniform transformation on these trading signals by normalizing them to the (-1,1) interval.[6]

[6]The uniform transformation is a common way to process the input variables in the machine learning literature; see, for instance, Gu, Kelly, and Xiu (2020) and Feng, Giglio, and Xiu (2020).

Within the training sample starting from January 2008, we fit trading signals from nine market participants into each of the 12 machine learning models for each stock in a monthly frequency, and predict its one-month-ahead excess return (minus risk-free rate) on a rolling basis. To start with, we use trading signals that cover five years ranging from January 2008 to December 2012 as input variables, and estimate the one-month-ahead excess return via the machine learning models. Next, we employ the estimated parameters and data in December 2012, calculated from the above step, to compute the out-of-sample forecasted return for January 2013. We hence generate a return forecast for each one of the 12 machine learning tools as mentioned, for each stock on January 2013. Keeping the five-year length fixed, we then repeat the above calculation for each month on a rolling basis with one month increment. Overall, our out-of-sample period is from January 2013 to December 2020. To construct the linear composite return predictor (LCP) for each month, we obtain the average out-of-sample excess return computed from the six linear models, that is, OLS, PCR, PLS, ALasso, Ridge, ENet. Similarly, we take the average of out-of-sample excess return calculated from the six nonlinear models, i.e., ANN1, ANN2, ANN3, ANN4, GBRT, RF, to construct the nonlinear composite return predictor (NLCP). As mentioned, a more comprehensive description of the machine learning models we use in this paper can be found in Appendix A.2.

2.4 Descriptive statistics

Table 1 presents summary statistics for our sample. The mean, median, standard deviation, 25th percentile and 75th percentile of the main variables are reported. Besides the two composite return predictors constructed in Section 2.3, we also report the summary statistics of the trading signals from nine market participants, i.e., mutual funds, hedge funds, banks, firms, insurance companies, other institutions, wealth management firms, short sellers, and retail investors.

[Insert Table 1 here]

The linear composite return predictor (LCP) has an average value of 0.8% and the average value of the nonlinear composite return predictor (NLCP) is 1.3%. Both variables has a similar median value compared to the average. NLCP has a larger variation compared to LCP, with a standard deviation twice as that of LCP. Descriptive statistics for the input variables in our training sample, i.e., Bank Trading, Firm Trading, Hedge Fund Trading, Insurance Company Trading, Mutual Fund Trading, Other Institutional Trading, Short Seller Trading , Wealth Management Trading, Retail Trading_MPR, Retail Trading_BJZZ, are also reported. Banks, insurance companies and other institutions tend to trade more relative to other market participants. Also, the trading of mutual funds tends to vary more compared to the others.

We also report statistics of commonly-used stock characteristics. Firm size (SIZE) is the natural logarithm of market capitalization. Book-to-market ratio (BM) is the most recent fiscal year-end book value divided by the market capitalization. Momentum (MOM) is the past cumulative returns from month -12 to month -2. Short-term reversal (STR) is the prior month's return. Asset growth (AG) is the annual asset growth from the previous fiscal year. Gross profitability (GP) is the gross profit divided by the total assets from the last fiscal year. Turnover (TO) is the trading volume over the number of shares outstanding in the last month. Idiosyncratic volatility (IVOL) is the standard deviation of the residuals from the CAPM model of daily stock excess returns over the previous 6 months. All the explanatory variables are winsorized at the 1% and 99% levels.

3 Performance of composite return predictors

In this section, we evaluate the performance of the trading signals from nine market participants as well as their composite return predictors via machine learning techniques. We first report the performance of the trading signals that are constructed following McLean, Pontiff, and Reilly (2022) as in Section 2.1 as benchmarks. The univariate portfolio results are shown in Section 3.1. Next, the performance results of the composite return predictors are reported in Section 3.2. In Section 3.3, we further investigate the importance of each trading signal contributing to the composite return predictor.

3.1 Portfolio performance of univariate trading signals

We first evaluate the performance of each trading signal that serves as input variable to build the composite return predictors, as benchmark cases. To facilitate the comparison, we use the same evaluation period, 2013-2020, which is the out-of-sample period when constructing and evaluating the composite return predictors. We report the results for univariate portfolio sorted by each of the trading variables in Table 2.

[Insert Table 2 here]

At the end of each month, we rank all sample stocks based on each of the trading signals and then sort them into decile portfolios. The portfolios are held for one month. The value-weighted returns on each decile as well as on the long-short portfolio – that goes long stocks in the highest decile and short stocks in the lowest decile – are calculated. To reserve space, we only report results from the top (bottom) three deciles and the high-minus-low portfolios. We report the Fama and French (2015) five-factor alphas (FF5) for each portfolio, with the Newey-West adjusted t-statistics.[7]

The long-short portfolio alphas are statistically insignificant as shown in the last column, except that the FF5 alpha for retail trading signal constructed following McLean, Pontiff, and Reilly (2022) measure is negative (-0.57%, t-stat = -2.09). The results here indicate discrepancies with the performance results shown in McLean, Pontiff, and Reilly (2022), partly due to different sample periods where their sample starts from 2006 and ends in 2017, and due to different methods as they run Fama-MacBeth regressions of monthly returns while we conduct portfolio analysis.

Across all top deciles, retail trading as measured by Boehmer et al. (2021) shows strong return predictability, with monthly return as high as 0.43% and statistically significant. For decile 9 portfolios, FF5 alpha for bank trading is significantly negative at -0.36%, with a

[7] Results regarding the Hou et al. (2021) q5 alphas (q5) and the Stambaugh and Yuan (2017) mispricing-factor alphas (MISP) are presented in Table B.1 and B.2.

t-statistic of -1.99. In contrast, mutual funds and short sellers tend to predict returns in the right direction, where the FF5 alpha for mutual fund trading in decile 9 is 0.26% with a t-statistic of 1.82 and short seller trading in decile 9 has a positive and significant alpha (0.19% and t-stat = 2.19).[8] The FF5 alphas in all specifications do not exhibit a clear monotonic pattern from the lowest decile to the highest decile.

3.2 Performance of composite return predictors

We first evaluate the performance of portfolios sorted by either the linear composite return predictor or the nonlinear composite return predictor with further results accounting for transaction costs. We then examine the performance persistence of the NLCP portfolio. We also conduct the Fama-MacBeth regressions to control for additional firm characteristics.

Portfolio results

Our first set of the main results include the portfolio performance of the composite return predictors we construct in Section 2.3. For each month, the linear composite return predictor is the average value of out-of-sample excess returns, which are computed using six linear (nonlinear) machine learning tools in five-year training samples. Table 3 reports the performance of the value-weighted portfolios sorted by either the linear composite return predictor (LCP) or the nonlinear composite return predictor (NLCP).[9] We rank all sample stocks based on the value of LCP or NLCP and then sort them into decile portfolios at the end of each month. The portfolios are held for one month. The out-of-sample period covers from 2013 to 2020, consistent with benchmark evaluation period in Table 2.

[Insert Table 3 here]

[8]Note that we sign the short interest such that increases (decreases) in short interest reflect negative (positive) values of short seller trading signal. Therefore the top (bottom) decile indicates less (more) short selling activities.

[9]We further report the excess returns as well as the Daniel, Hirshleifer, and Sun (2020) behavioral-factor alphas (DHS) for LCP- and NLCP-sorted portfolios in Table B.3.

Panel A of Table 3 reports the Fama and French (2015) five-factor alphas (FF5), Hou et al. (2021) q5-model alphas (q5) and the Stambaugh and Yuan (2017) mispricing-factor alphas (MISP) for each portfolio, with the Newey-West adjusted t-statistics. For the linear composite return predictor, the abnormal returns in all specifications tend to increase, though not strictly monotonically, from the lowest decile to the highest decile. A long-short strategy (H-L) generates a significant FF5 alpha of 0.85% (t-stat = 2.19) on a monthly basis and the MISP alpha equals 0.93% (t-stat = 3.58). The q5 alphas are however statistically insignificant.

In comparison, the nonlinear composite return predictor (NLCP) preforms better. A long-short strategy based on NLCP generates a significant FF5 alpha of 1.39% (t-stat = 4.88) on a monthly basis. The q5 alpha equals 1.11% with a t-statistic of 3.70, and the alphas with respect to MISP equals 1.30% (t-stat = 5.73). Moreover, both the lowest and the highest decile exhibit significant predictability, with the FF5 alpha as low as -1.04% (t-stat = -4.69) for the lowest decile and 0.35% (t-stat = 2.15) for the highest decile, indicating that a high (low) level of NLCP contains positive (negative) information. In contrast to the insignificant FF5 alpha for portfolios sorted by univariate trading signals, results in this panel suggest nonlinear machine learning methods indeed help extract value-relevant information from the collective trading of various groups of market participants.

In Panel B of Table 3, we further incorporate transaction costs ranging from 1bps to 10 bps, and report the portfolio alphas accordingly. To save space, we only report results for the high-minus-low portfolios. For LCP-sorted portfolios, all alphas decrease monotonically. The FF5 alpha is 0.72% and the MISP alpha equals 0.61%, both statistically significant at the 10% level. The NLCP-sorted portfolios show significant alphas across all factor models, with magnitudes similar to those reported in Panel A.

Performance persistence

We next examine the performance persistence of the portfolios sorted by NLCP. Note that the return premiums examined earlier are out-of-sample that can be as far as five years from the trading signal construction period, so the price pressure hypothesis (e.g., Chordia and

Subrahmanyam, 2004) is unlikely to explain the results. Still, we examine the performance after the portfolio construction month up to one year as a robustness check in Table 4.

[Insert Table 4 here]

To reserve space, we only report FF5 alphas from the top (bottom) three deciles and the high-minus-low portfolios, with the Newey-West adjusted t-statistics. While the top decile portfolio loses statistical significance once we increase the holding period to two months, the bottom decile portfolio remains significantly negative even when we hold the portfolio for one year. Specifically, when the holding period is two months, the Low decile portfolio earns a FF5 alpha of -1.11% (t-stat = -4.27), which drops from -1.04% in Table 3. The FF5 alpha does not change monotonically with the increase in the holding period. When the holding period equals six months, the alpha further decreases to -1.60% and gradually increases afterwards.

As for performance of the long-short portfolio (H-L), we observe a non-monotonic pattern along with the holding period. The corresponding FF5 alpha increases from 1.39% in Table 3 to 1.51% when we hold the portfolio for two months in Table 4. The FF5 alpha for H-L remains significant for up to nine months, at which point the FF5 alpha is 0.59% (t-stat = 2.16). Overall, we find that the out-of-sample performance of NLCP long-minus-short portfolio remains robust even when we increase the holding period more than half-year.

Fama-MacBeth regressions

We further conduct Fama-MacBeth regressions of monthly stock returns on main variables of interest controlling for all the individual trading signal variables simultaneously in Table 5. Again, the two key variables are the linear composite return predictor (LCP) and the nonlinear composite return predictor (NLCP). The trading signals from nine market participants, i.e., banks, firms, hedge funds, insurance companies, mutual funds, wealth management firms, other institutions, short sellers, and retail investors, are constructed following McLean, Pontiff, and Reilly (2022). We include these variables along with the composite return predictors to further evaluate the marginal return predictability of LCP and NLCP beyond the predictability of individual trading signals.[10]

[Insert Table 5 here]

In Table 5, we further include control variables such as firm size (SIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), asset growth (AG), turnover (TO), and idiosyncratic volatility (IVOL) in all specifications. Their definitions could be found in Section 2.4. The explanatory variables are standardized with mean equals zero and standard deviation equals one and t-statistics are Newey-West adjusted.

The first column reports benchmark result for including trading variables alone. Consistent with the portfolio results in Table 3, not many market participants yield significant return predictability. We find that only retail trading constructed following Boehmer et al. (2021) yield significant positive monthly returns. A one standard deviation increase in *Retail Trading_BJZZ* generates 18 bps higher return (t-stat = 6.31). We include LCP and NLCP into regressions in sequence and find the coefficient for NLCP is large and significant while it is small and slightly significant for LCP. Specifically, a one standard deviation increase in NLCP yields 1.20% higher return (t-stat = 3.27). The coefficient of NLCP remains largely unchanged when including both LCP and NLCP in the regression.

Results from Tables 2 to 5 overall highlight the importance to aggregate trading signals from multiple market participants via machine learning nonlinear tools. Information extracted from nonlinear interactions among predictors can best predict stock returns. In contrast, linear machine learning models do not yield such robust performance.

[10]The composite return predictors here face a stricter test criteria as they are examined out-of-sample up to five years while the trading variables are tested upon the one-month-ahead returns.

3.3 Contribution of each trading signal

In this section, we further investigate the contribution of each trading signal to the composite return predictor. We follow Nembrini, König, and Wright (2018) to use the random forest model and measure the *feature importance* for each signal.[11]

Specifically, we include all trading signal variables to predict one-month-ahead returns firstly and then filter out one of the trading signal variable while keep the remaining signal variables in the regression, in order to examine the subsequent effect on the result. We repeat this execution process for each trading signal variable. The marginal contribution of each trading signal is calculated as the normalized sum of the reduced mean squared error from the above process. To be consistent with the construction method to build the composite predictor, we also use a five-year rolling window and obtain the *feature importance* measure from the time-series average value of the marginal contribution.

[Insert Figure 1 here]

Figure 1 presents the *feature importance* of each trading signal's one-month-ahead return forecast. The values are normalized to have a sum value as one, and they indicate the magnitude instead of the direction of return prediction. We find that retail trading, constructed in accordance with either Boehmer et al. (2021) or McLean, Pontiff, and Reilly (2022), together with short seller trading and firm trading, tend to have the largest portion contributing to the return predictability. They exhibit a similar level regarding *feature importance* and is approximately 16%. The *feature importance* of mutual fund trading is slightly lower and is around 13%. Altogether these trading signal variables contribute 60% regarding predictability. The remaining 40% of contribution come from other market participants with relatively small variation. The magnitude of their *feature importance* comes in such order: other insti-

[11] We choose random forest model out of two main reasons. It involves less computation workload compared to other more complicated machine learning tools. Besides, one important feature in the random forest is the *feature importance* in the machine learning literature and are commonly used. Nembrini, König, and Wright (2018) base on the standard random forest model to further improve the *feature importance* measure.

tutional trading, wealth management trading, hedge fund trading, bank trading, insurance company trading. It is important to note that while these trading signals are individually insignificant in predicting returns, collectively they provide significant contribution to the composite return predictor, possibly from their nonlinear interactions.

4 Tests on stock fundamentals and anomalies

If the nonlinear composite return predictor (NLCP) indeed predicts returns by containing value-relevant information, one would expect that it can predict future firm fundamentals, and that its return predictability is stronger for stocks with greater uncertainty, where information is more valuable. We examine the these issues in this section.[12] We first report the Fama-MacBeth regression results for the nonlinear composite return predictor in different levels of information environment in Section 4.1. We then examine the predictability of stock profitability and earnings surprises on the nonlinear composite return predictor in Section 4.2. In Section 4.3, we run a series of tests to check whether the nonlinear composite return predictor predict common stock anomalies and compare with other factor models.

4.1 Varying information environment

Prior studies show that returns are positively related to ex-ante information asymmetry for informed trading (e.g., Diamond and Verrecchia, 1991; Verrecchia, 2001; O'Hara, 2003; Easley and O'hara, 2004). In that sense, if the nonlinear composite return predictor reflects valuable information that is yet absorbed by the market, its return predictability should be stronger for stocks with higher information uncertainty.

[Insert Table 6 here]

[12] For the subsequent tests we only focus on the results for the nonlinear composite return predictor, given the inferior performance of the linear composite return predictor.

Table 6 reports Fama-MacBeth regression results for subsamples of stocks that are constructed based on different proxies of ex-ante information environment level, which are firm age, firm size, idiosyncratic volatility, as well as analyst coverage (e.g., Llorente, Michaely, Saar, and Wang, 2002; Zhang, 2006). At the end of each month, all sample stocks are sorted into halves based on an information environment proxy. We measure firm age as the number of years since the firm was first covered by CRSP, firm size as its market capitalization, idiosyncratic volatility as the standard deviation of the residuals from the CAPM model of daily stock excess returns over the previous 6 months, and analyst coverage as the number of analysts covering the firm. The remaining empiric setting is in line with Table 5.

In all specifications, we find that return of stocks with high level of NLCP value are generally larger for firms with younger age, smaller size, higher idiosyncratic volatility.[13] The coefficient on NLCP remain significant in the remaining sub-samples as well, which suggests that the main findings are unlikely to be driven solely by these firm characteristics. The discrepancy between subsamples is large when we use firm size or idiosyncratic volatility as information environment proxy, where the coefficient difference is threefold for the former case and double for the latter one.

4.2 Predicting stock fundamentals

In this section, we test whether NLCP contains information related to firm fundamentals. The first set of proxies for firm fundamentals is the profitability measure such as returns on assets and cash flows. In addition, we examine the predictability of NLCP on future earnings surprises. Specifically, returns on assets (ROA) is calculated as the summation of income before extraordinary items and interest expenses, divided by the lagged total assets. Cash flow (CF) is measured as the difference between income before extraordinary items and total accruals, divided by total assets. We compute standardized unexpected earnings (SUE) as the market-adjusted returns upon earnings announcements over the three-day window. In Table 7, we report monthly Fama-MacBeth regressions of firm fundamentals where the

[13] Note that subsamples of stocks that are constructed based on analyst coverage do not differ much.

change of values from the last period in the first three columns and the levels in the next three columns serve as dependent variables.

[Insert Table 7 here]

We include firm size, book-to-market ratio, momentum, short-term reversal, asset growth and gross profitability as control variables in all specifications. For the first three columns in Table 7, we find that NLCP significantly predicts increase in ROA, CF, and SUE of firms. Specifically, a one standard deviation increase in NLCP yields 5.97% higher ΔROA (t-stat = 2.23) next month and 3.92% higher ΔCF (t-stat = 2.68) and 0.62% higher ΔSUE (t-stat = 2.62) as well. The effect is much stronger when the dependent variables are the level instead of the change. According to the next three columns, a one standard deviation in NLCP will generate 26.92% higher ROA, 26.92% higher CF, and 1.56% higher SUE, all with a t-statistic larger than six. Overall, these results corroborate the notion that the return predictability of NLCP at least partially comes from its predictability of firm fundamentals.

4.3 Anomaly prediction

Anomalies by NLCP portfolios

To deepen our understanding of the information contained in the composite return predictor, we examine stock anomalies under each quintile of the portfolios sorted by NLCP. We follow Green, Hand, and Zhang (2017) to construct 102 anomalies based on firm characteristics. To be consistent with our sample period, we check the excess return of the long-short portfolio sorted by each of the stock anomalies during 2008-2020, and only select anomalies with significant (t-stat >1.66) excess returns. There are in total 12 anomalies after the selection, which are market capitalization (Banz, 1981), book-to-market (Rosenberg, Reid, and Lanstein, 1985), gross margin (Novy-Marx, 2013), illiquidity (Amihud, 2002), idiosyncratic volatility (Ali, Hwang, and Trombley, 2003), momentum_12m (Jegadeesh, 1990), momentum_1m (Jegadeesh and Titman, 1993), asset growth (Cooper, Gulen, and Schill, 2008), dividend yield (Litzenberger and Ramaswamy, 1981), analyst coverage (Elgers, Lo, and Jr., 2001), price delay (Hou and Moskowitz, 2005), and combined fundamental (Mohanram, 2005). We report the time-series average of the cross-sectional means of firm characteristics corresponding to the above stock anomalies for portfolios sorted by NLCP in Table 8.[14]

[Insert Table 8 here]

The characteristic spreads for high-minus-low portfolios are statistically significant across different anomalies. Stocks in the top quintile portfolio of NLCP tend to have higher values of the book-to-market ratio, stronger profitability (gross margin), larger past one-year returns, slower asset growth rates, more analysts coverage, as well as larger combined values of firm fundamentals, compared to stocks in the bottom quintile. In addition, these stocks tend to be more illiquid and yield more dividends. These patterns are consistent with the return directions of the corresponding anomalies during our sample period.[15] On the contrary, we find that the anomaly spreads in the NLCP-sorted high-minus-low portfolios exhibit opposite signs with the direction of the corresponding anomalies such as market capitalization, idiosyncratic volatility, one-month momentum, and price delay, where the spreads are significant only for the former two anomalies (t-stat >2).

Net anomaly prediction

The literature has been studying whether trades from different market participants are consistent with stock anomalies' predictability. For example, Edelen, Ince, and Kadlec (2016) show that institutional investors tend to trade on the wrong side of anomaly strategies. McLean, Pontiff, and Reilly (2022) conduct a more systematic analysis on trades from nine types of market participants with respect to stock anomalies. They find that retail investors trade against anomalies while firms and short sellers buy more stocks in anomaly-long instead of anomaly-short. The results in Table 8 suggest that our nonlinear composite return predictor (NLCP) assign stocks to the right side of at least some anomalies. We next analyze

[14]The corresponding values of firm characteristics for portfolios sorted by the Linear composite factor (LCP) are presented in Table B.4.

[15]Note that results in Table 8 are descriptive and no causal relation is to be set here.

this issue further by examining the time-series average of the cross-sectional means of NLCP in quintile portfolios sorted by Net anomaly indicator (NET) in Table 9.

[Insert Table 9 here]

To construct Net anomaly indicator, we select the anomalies in Table 8 whose predictability exhibits consistent directions with that of the NLCP and is significant during our sample period, in order to evaluate the combined net effect. Next, We sort stocks into quintile portfolios at the end of each month based on the values of the selected anomalies. The top (bottom) quintile portfolio of each anomaly is treated as the long (short) side. For each stock-month observation, NET is computed as the difference between the number of long-side anomaly portfolios and the number of short-side anomaly portfolios that the stock falls into.

For comparison, we also report the values for alternative nonlinear composite return predictor in Table 9. Specifically, we use two (three) trading signals to construct the composite return predictors, denoted as NLCP_two (NLCP_three). The two trading signals are from two market participants – firms and short sellers, and the three trading signals come from three market participants – firms, short sellers and informed retail investors constructed following Boehmer et al. (2021). The construction method are the same with that for NLCP, as shown in Section 2.3.

Table 9 indicates that the average value of NLCP significantly increases from the bottom quintile portfolio to the top quintile portfolio. The value in its high-minus-low portfolio is 0.36% with a t-statistic equals to 6.74. In addition, the average value of both NLCP_two and NLCP_three exhibit a similar pattern along the NET quintile portfolios, although with a smaller magnitude. Specifically, the value in high-minus-low portfolio is 0.23% for NLCP_two and 0.22% for NLCP_three, both with a significance level at 1%. Such finding indicates that the nonlinear composite return predictor targets stocks that are more likely to be undervalued instead of overvalued. The result is not dominated by the trades from informed investor groups such as firms, short sellers, and smart retail traders (McLean, Pontiff, and Reilly, 2022). Overall, Table 9 suggest that the the nonlinear composite return predictor is in the same direction as predicted by well-known anomalies.

Performance of the factor model constructed from NLCP

If the nonlinear composite return predictor (NLCP) is on average positively related to return anomalies, an aggregate NLCP-based factor may help explain the anomaly returns in the cross-section. We create a two-factor model, an NLCP factor plus the market factor. The NLCP factor is the value-weighted return spreads (High - Low) of the NCLP-sorted portfolios. We examine the model's ability to explain the anomaly returns against other well-known factor models: the Fama and French (2015) five-factor model, Hou, Mo, Xue, and Zhang (2021) q5 model, and Stambaugh and Yuan (2017) mispricing-factor model. We adopt the same anomaly selection criteria shown in Table 8 and 9. Similar to Liu, Zhou, and Zhu (2022), we report the average monthly alpha (in absolute value), the average t-statistic (in absolute value), the aggregate pricing error Delta ($\Delta = \alpha^{\mathrm{T}}\Sigma^{-1}\alpha$), as well as associated Gibbons, Ross, and Shanken (1989) F-statistic and p-value in Table 10.

[Insert Table 10 here]

Table 10 shows that our two-factor model yields the smallest average absolute alpha, which equals 0.40% compared to 0.64% for FF5 model, 0.52% for q5 model, and 0.58% for MISP model. In addition, the average absolute t-statistic of our factor model is lower than those of other models, where the value is 1.2 compered to 1.9 or 1.5 in other models. Furthermore, the-two factor model has the smallest aggregate pricing error, *Delta*. Lastly, results from the GRS test corroborate with the above findings. Factor models FF5 and MISP strongly reject the joint hypothesis that all anomalies included in the sample produce zero alphas. The two-factor model, together with the q5 model, exhibit a p-value larger than 0.2, suggesting their ability to explain the stock return anomalies.

5 Conclusion

We use machine learning to study whether the joint trading behavior of multiple market participants contains information about future stock returns. Our machine learning approach can capture complex interactions among market participants and the nonlinear effects of their collective trading on stock returns. A long-short portfolio based on the nonlinear composite predictor (NLCP) generates monthly alphas from various factor models exceeding 1%. Measuring variable importance for each investor type, we show that the return predictability is not dominated by trades from just a few types. In addition, our results remain significant after accounting for transaction costs and exhibit no short-term reversal.

We find that our nonlinear composite return predictor contains information about firm fundamentals in that it predicts ROA, free cash flows, and earnings surprises. Furthermore, the predictor can forecast returns driven by mispricing, where the NLCP-sorted portfolios are on the right side of the anomalies. Our two-factor model with the market factor and the combined trading signal factor tends to outperform existing factor models in explaining anomalies. Overall, our findings suggest that incorporating the intricate interactions of multiple market participants' trading provides valuable information for price discovery.

References

Aggarwal, Rajesh K. and Philippe Jorion. 2010. "The performance of emerging hedge funds and managers." *Journal of Financial Economics* 96 (2):238–256.

Ali, Ashiq, Lee-Seok Hwang, and Mark A. Trombley. 2003. "Arbitrage risk and the book-to-market anomaly." *Journal of Financial Economics* 69 (2):355–373.

Amihud, Yakov. 2002. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of Financial Markets* 5 (1):31–56.

Baker, Malcolm, Lubomir Litov, Jessica A. Wachter, and Jeffrey Wurgler. 2010. "Can Mutual Fund Managers Pick Stocks? Evidence from Their Trades Prior to Earnings Announcements." *Journal of Financial and Quantitative Analysis* 45 (5):1111–1131.

Banz, Rolf W. 1981. "The relationship between return and market value of common stocks." *Journal of Financial Economics* 9 (1):3–18.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni. 2021. "Bond Risk Premiums with Machine Learning." *The Review of Financial Studies* 34 (2):1046–1089.

Boehmer, Ekkehart, Zsuzsa R. Huszar, and Bradford D. Jordan. 2010. "The good news in short interest." *Journal of Financial Economics* 96 (1):80–97.

Boehmer, Ekkehart, Charles M. Jones, Juan Wu, and Xiaoyan Zhang. 2020. "What Do Short Sellers Know?" *Review of Finance* 24 (6):1203–1235.

Boehmer, Ekkehart, Charles M. Jones, Xiaoyan Zhang, and Xinran Zhang. 2021. "Tracking Retail Investor Activity." *The Journal of Finance* forthcoming.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45:5–32.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu. 2021. "Forest Through the Trees: Building Cross-Sections of Stock Returns." *Available at SSRN 3493458* .

Cao, Charles, Bing Liang, Andrew W Lo, and Lubomir Petrasek. 2018. "Hedge Fund Holdings and Stock Market Efficiency." *The Review of Asset Pricing Studies* 8 (1):77–116.

Chatigny, Philippe, Ruslan Goyenko, and Chengyu Zhang. 2022. "Asset Pricing with Attention Guided Deep Learning." *Available at SSRN 3971876* .

Chinco, Alex, Adam D. Clark-Joseph, and Mao Ye. 2019. "Sparse Signals in the Cross-Section of Returns." *The Journal of Finance* 74 (1):449–492.

Chordia, Tarun and Avanidhar Subrahmanyam. 2004. "Order imbalance and individual stock returns: Theory and evidence." *Journal of Financial Economics* 72 (3):485–518.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill. 2008. "Asset Growth and the Cross-Section of Stock Returns." *The Journal of Finance* 63 (4):1609–1651.

Da, Rui. 2022. "Market Efficiency with Many Investors." *Chicago Booth working paper* .

Daniel, Kent, David Hirshleifer, and Lin Sun. 2020. "Short- and Long-Horizon Behavioral Factors." *The Review of Financial Studies* 33 (4):1673–1736.

DeMiguel, Victor, Javier Gil-Bazo, Francisco J. Nogales, and Andre A. P. Santos. 2022. "Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha." *Available at SSRN 3768753* .

Diamond, Douglas W. and Robert E. Verrecchia. 1981. "Information aggregation in a noisy rational expectations economy." *Journal of Financial Economics* 9 (3):221–235.

———. 1991. "Disclosure, Liquidity, and the Cost of Capital." *The Journal of Finance* 46 (4):1325–1359.

Diether, Karl B., Kuan-Hui Lee, and Ingrid M. Werner. 2009. "Short-Sale Strategies and Return Predictability." *The Review of Financial Studies* 22 (2):575–607.

Dong, Xi, Yan Li, David Rapach, and Guofu Zhou. 2022. "Anomalies and the Expected Market Return." *The Journal of Finance* 77 (1):639–681.

Easley, David and Maureen O'hara. 2004. "Information and the Cost of Capital." *The Journal of Finance* 59 (4):1553–1583.

Edelen, Roger M., Ozgur S. Ince, and Gregory B. Kadlec. 2016. "Institutional investors and stock return anomalies." *Journal of Financial Economics* 119 (3):472–488.

Elgers, Pieter T., May H. Lo, and Ray J. Pfeiffer Jr. 2001. "Delayed Security Price Adjustments to Financial Analysts' Forecasts of Annual Earnings." *The Accounting Review* 76 (4):613–632.

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

Fama, Eugene F. and James D MacBeth. 1973. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81 (3):607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2020. "Taming the Factor Zoo: A Test of New Factors." *The Journal of Finance* 75 (3):1327–1370.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. "A Test of the Efficiency of a Given Portfolio." *Econometrica* 57 (5):1121–1152.

Goldstein, Itay and Liyan Yang. 2015. “Information Diversity and Complementarities in Trading and Information Acquisition.” *The Journal of Finance* 70 (4):1723–1765.

Gompers, Paul A. and Andrew Metrick. 2001. “Institutional Investors and Equity Prices.” *The Quarterly Journal of Economics* 116 (1):229–259.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang. 2017. “The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns.” *The Review of Financial Studies* 30 (12):4389–4436.

Grossman, Sanford J. and Joseph E. Stiglitz. 1980. “On the Impossibility of Informationally Efficient Markets.” *The American Economic Review* 70 (3):393–408.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33 (5):2223–2273.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Hirshleifer, David, Siew Hong Teoh, and Jeff Jiewei Yu. 2011. “Short Arbitrage, Return Asymmetry, and the Accrual Anomaly.” *The Review of Financial Studies* 24 (7):2429–2461.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang. 2021. “An Augmented q-Factor Model with Expected Growth.” *Review of Finance* 25 (1):1–41.

Hou, Kewei and Tobias J. Moskowitz. 2005. “Market Frictions, Price Delay, and the Cross-Section of Expected Returns.” *The Review of Financial Studies* 18 (3):981–1020.

Jegadeesh, Narasimhan. 1990. “Evidence of Predictable Behavior of Security Returns.” *The Journal of Finance* 45 (3):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 (1):65–91.

Jiao, Yawen, Massimo Massa, and Hong Zhang. 2016. “Short selling meets hedge fund 13F: An anatomy of informed demand.” *Journal of Financial Economics* 122 (3):544–567.

Kaniel, Ron, Zihan Lin, Markus Pelger, and Stijn Van Nieuwerburgh. 2022. “Machine-Learning the Skill of Mutual Fund Managers.” *Available at SSRN 3977883* .

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman. 2012. “Individual Investor Trading and Return Patterns around Earnings Announcements.” *The Journal of Finance* 67 (2):639–680.

Kelley, Eric K. and Paul C. Tetlock. 2013. “How Wise Are Crowds? Insights from Retail

Orders and Stock Returns." *The Journal of Finance* 68 (3):1229–1265.

Kelly, Bryan T., Seth Pruitt, and Yinan Su. 2019. "Characteristics are covariances: A unified model of risk and return." *Journal of Financial Economics* 134 (3):501–524.

Koijen, Ralph S. J. and Motohiro Yogo. 2019. "A Demand System Approach to Asset Pricing." *Journal of political economy* 127 (4):1475–1515.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. "Shrinking the cross-section." *Journal of Financial Economics* 135 (2):271–292.

Leippold, Markus, Qian Wang, and Wenyu Zhou. 2022. "Machine learning in the Chinese stock market." *Journal of Financial Economics* 145 (2):64–82.

Litzenberger, Robert H. and Krishna Ramaswamy. 1981. "The Effects of Dividends on Common Stock Prices Tax Effects or Information Effects?" *The Journal of Finance* 37 (2):429–443.

Liu, Yang, Guofu Zhou, and Yingzi Zhu. 2022. "Trend Factor in China: The Role of Large Individual Trading." *Available at SSRN 3402038* .

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang. 2002. "Dynamic Volume-Return Relation of Individual Stocks." *The Review of Financial Studies* 15 (4):1005–1047.

Massa, Massimo, Wenlan Qian, Weibiao Xu, and Hong Zhang. 2015. "Competition of the informed: Does the presence of short sellers affect insider selling?" *Journal of Financial Economics* 118 (2):268–288.

McLean, R. David, Jeffrey Pontiff, and Christopher Reilly. 2022. "Taking Sides on Return Predictability." *Available at SSRN 3637649* .

Mohanram, Partha S. 2005. "Separating Winners from Losers among LowBook-to-Market Stocks using Financial Statement Analysis." *Review of Accounting Studies* 10:133–170.

Nembrini, Stefano, Inke R König, and Marvin N Wright. 2018. "The revival of the Gini importance?" *Bioinformatics* 34 (21):3711–3718.

Novy-Marx, Robert. 2013. "The other side of value: The gross profitability premium." *Journal of Financial Economics* 108 (1):1–28.

O'Hara, Maureen. 2003. "Presidential Address: Liquidity and Price Discovery." *The Journal of Finance* 58 (4):1335–1354.

Patton, Andrew J. and Brian M. Weller. 2020. "What you see is not what you get: The costs of trading market anomalies." *Journal of Financial Economics* 137:515–549.

Rapach, David, Jack Strauss, Jun Tu, and Guofu Zhou. 2019. "Industry Return Predictability: A Machine Learning Approach." *Journal of Financial Data Science* 1 (3):9–28.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. 1985. "Persuasive evidence of market inefficiency." *Journal of Portfolio Management* 11 (3):9–16.

Schapire, Robert E. and Yoav Freund. 2012. *Boosting: Foundations and Algorithms.* MIT Press.

Sias, Richard W. and David A. Whidbee. 2010. "Insider Trades and Demand by Institutional and Individual Investors." *The Review of Financial Studies* 23 (4):1544–1595.

Stambaugh, Robert F. and Yu Yuan. 2017. "Mispricing Factors." *The Review of Financial Studies* 30 (4):1270–1315.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1):267–288.

van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2022. "Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases." *The Review of Financial Studies* .

Verrecchia, Robert E. 2001. "Essays on disclosure." *Journal of Accounting and Economics* 32 (1–3):97–180.

Zhang, X. Frank. 2006. "Information Uncertainty and Stock Returns." *The Journal of Finance* 61 (1):105–137.

Zou, Hui. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the Royal Statistical Society. Series B (Methodological)* 101:1418–1429.

Figure 1. Importance of each trading signal

This figure plots the *feature importance* of each trading signal (Nembrini, König, and Wright, 2018). We follow McLean, Pontiff, and Reilly (2022) to compute trading signals from nine market participants, i.e., mutual fund, hedge fund, bank, firm, insurance company, other institutions, wealth management firm, short seller, retail investors (we also use Boehmer et al. (2021) measure for retail trading). Appendix A.1 contains details of constructing trading signal variables presented in this figure. The full sample period is 2008-2020. We include all trading signal variables to predict one-month-ahead returns and then filter out one of the trading signal variable in turn while keep the remaining signal variables in the regression. The *feature importance* for each trading signal is the normalized sum of the reduced mean squared error accordingly.

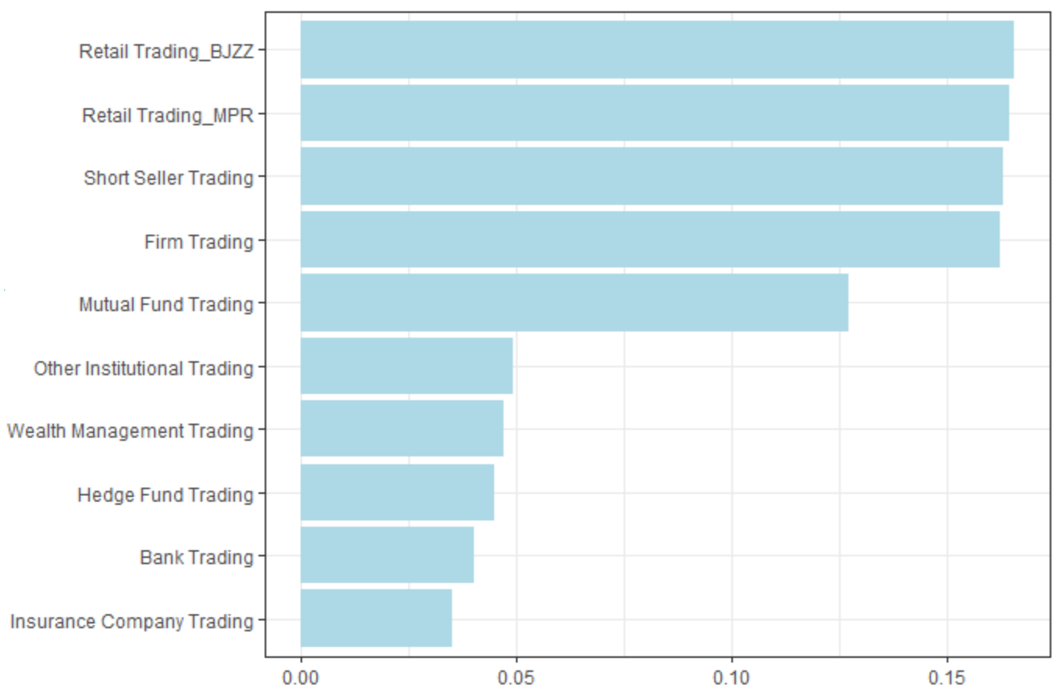

Table 1. Summary statistics

This table reports summary statistics, including mean, median, standard deviation, 10% value and 90% value, of the main variables. We follow McLean, Pontiff, and Reilly (2022) to compute trading signals from nine market participants, i.e., mutual fund, hedge fund, bank, firm, insurance company, other institutions, wealth management firm, short seller, retail investors (we also use Boehmer et al. (2021) measure for retail trading). Appendix A.1 contains details of constructing trading signal variables presented in this table. The full sample period is 2008-2020. In our rolling-based (one-month increment) five-year training sample, trading signals from nine market participants are included in each of the 12 machine learning models mostly documented in Gu, Kelly, and Xiu (2020) to predict out-of-sample return premium. To construct the linear composite return predictor (LCP), we obtain the monthly average of out-of-sample return premiums computed from six linear models, i.e., OLS, PCR, PLS, ALasso, Ridge, ENet. In a similar vein, we use return premiums from six nonlinear models, i.e., ANN1, ANN2, ANN3, ANN4, GBRT, RF, to construct the nonlinear composite return predictor (NLCP). Detailed explanation of the machine learning models can be found in Appendix A.2. Additionally, we report descriptive statistics of commonly-used stock characteristics, with definitions written in Section 2.4. All the explanatory variables are winsorized at the 1% and 99% levels.

Variable	Mean	Median	St.Dev	p10	p90
LCP	0.80%	0.77%	1.35%	-0.62%	2.15%
NLCP	1.34%	1.16%	2.78%	-0.77%	2.88%
Bank Trading	-0.26%	0.00%	2.09%	-0.96%	0.33%
Firm Trading	-7.01%	-0.54%	26.25%	-18.76%	3.47%
Hedge Fund Trading	-0.39%	0.00%	4.85%	-2.27%	1.52%
Insurance Company Trading	-0.07%	0.00%	0.70%	-0.24%	0.03%
Mutual Fund Trading	-0.12%	0.00%	4.23%	-3.26%	2.90%
Other Institutional Trading	-0.05%	0.00%	4.81%	-1.20%	2.60%
Short Seller Trading	-0.02%	0.01%	3.55%	-3.53%	3.48%
Wealth Management Trading	0.00%	0.00%	0.18%	0.00%	0.00%
Retail Trading_MPR	0.58%	-0.54%	10.77%	-8.00%	8.81%
Retail Trading_BJZZ	-2.50%	-1.35%	18.94%	-22.84%	16.01%
SIZE	20.07	20.02	2.13	17.30	22.92
BM	0.760	0.587	0.710	0.159	1.472
MOM	0.062	0.022	0.521	-0.533	0.623
STR	0.006	0.004	0.146	-0.156	0.163
AG	0.108	0.042	0.371	-0.169	0.388
TO	7.17	1.29	33.38	0.20	5.93
IVOL	0.024	0.018	0.019	0.008	0.047

Table 2. Performance of univariate portfolios from trading by nine market participants

This table reports the Fama and French (2015) five-factor alphas (FF5) for the value-weighted portfolios sorted by each of the nine market participants' trading signals. Again, we report two measures of retail order imbalance – Retail Trading_BJZZ as in Boehmer et al. (2021) and Retail Trading_MPR as in McLean, Pontiff, and Reilly (2022). At the end of each month, we rank all sample stocks based on the trading signal and sort them into decile portfolios. The portfolios are held for one month. To reserve space, we only report alphas from the top (bottom) three deciles and the high-minus-low portfolios. To be consistent with the out-of-sample period, we report results in period 2013-2020 here. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	8	9	High	**High - Low**
Bank Trading	0.01%	-0.08%	-0.05%	0.07%	-0.36%*	-0.11%	-0.12%
	(0.04)	(-0.47)	(-0.29)	(0.42)	(-1.99)	(-0.76)	(-0.49)
Firm Trading	-0.28%	0.10%	0.19%	-0.09%	0.04%	-0.05%	0.22%
	(-0.94)	(0.50)	(1.31)	(-0.70)	(0.61)	(-0.43)	(0.66)
Hedge Fund Trading	0.00%	-0.10%	-0.05%	0.06%	-0.13%	0.07%	0.08%
	(-0.02)	(-0.69)	(-0.38)	(0.39)	(-1.03)	(0.46)	(0.29)
Insurance Company Trading	0.10%	-0.03%	-0.16%	-0.10%	-0.27%	-0.16%	-0.26%
	(0.49)	(-0.19)	(-0.85)	(-0.42)	(-1.40)	(-1.20)	(-0.98)
Mutual Fund Trading	-0.02%	0.04%	0.08%	0.02%	0.26%*	0.00%	0.02%
	(-0.15)	(0.33)	(0.27)	(0.06)	(1.82)	(0.04)	(0.12)
Other Institutional Trading	0.00%	-0.15%	0.01%	0.21%	-0.02%	-0.10%	-0.10%
	(0.01)	(-1.18)	(0.08)	(1.57)	(-0.14)	(-0.70)	(-0.38)
Short Seller Trading	0.08%	-0.16%	-0.06%	-0.16%	0.19%**	0.07%	-0.01%
	(0.52)	(-1.45)	(-0.53)	(-1.51)	(2.19)	(0.67)	(-0.04)
Wealth Management Trading	0.06%	-0.13%	-0.16%	-0.11%	0.10%	0.07%	0.01%
	(0.29)	(-1.50)	(-0.90)	(-0.71)	(0.61)	(0.44)	(0.03)
Retail Trading_MPR	0.12%	0.22%*	0.15%	0.01%	-0.14%	-0.45%	-0.57%**
	(0.62)	(1.84)	(0.92)	(0.08)	(-0.48)	(-1.22)	(-2.09)
Retail Trading_BJZZ	0.16%	0.02%	-0.18%**	-0.09%	0.02%	0.43%**	0.27%
	(0.93)	(0.19)	(-2.44)	(-0.52)	(0.14)	(2.02)	(1.20)

Table 3. Performance of portfolio from composite return predictors

This table reports the out-of-sample performance of the value-weighted portfolios sorted by either the linear composite return predictor (LCP) or the nonlinear composite return predictor (NLCP). At the end of each month, we rank all sample stocks based on the value of LCP or NLCP and sort them into decile portfolios. The portfolios are held for one month. Panel A reports the corresponding FF5 alphas, q5 alphas and MISP alphas. In Panel B, we incorporate transaction costs ranging from 1bps to 10 bps, and report the alphas for the high-minus-low portfolios of LCP and NLCP correspondingly. The out-of-sample covers from 2013 to 2020. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Portfolio alphas sorted by LCP or NLCP

	LCP			NLCP		
	FF5	q5	MISP	FF5	q5	MISP
Low	-0.53%	-0.15%	-0.61%**	-1.04%***	-0.75%***	-0.95%***
	(-1.57)	(-0.54)	(-2.53)	(-4.69)	(-3.15)	(-5.04)
2	-0.16%	0.05%	-0.14%	-0.62%***	-0.40%**	-0.57%***
	(-0.85)	(0.28)	(-0.82)	(-3.68)	(-2.05)	(-3.40)
3	-0.08%	0.03%	-0.21%	0.00%	0.14%	0.08%
	(-0.54)	(0.14)	(-1.16)	(0.00)	(1.02)	(0.52)
4	0.21%*	0.38%***	0.08%	-0.16%	-0.11%	-0.20%
	(1.65)	(2.99)	(0.72)	(-1.31)	(-0.83)	(-1.59)
5	-0.20%*	-0.13%	-0.18%*	-0.17%	-0.10%	-0.25%*
	(-1.69)	(-1.06)	(-1.76)	(-1.51)	(-0.77)	(-1.85)
6	-0.08%	-0.05%	-0.10%	-0.04%	0.01%	-0.11%
	(-0.56)	(-0.32)	(-0.73)	(-0.50)	(0.11)	(-1.50)
7	0.10%	0.15%	0.06%	-0.02%	-0.08%	-0.03%
	(1.00)	(1.51)	(0.65)	(-0.18)	(-0.74)	(-0.25)
8	-0.05%	0.00%	-0.03%	0.19%	0.19%	0.21%*
	(-0.58)	(0.04)	(-0.29)	(1.42)	(1.41)	(1.72)
9	-0.09%	-0.12%	-0.04%	0.38%***	0.41%**	0.52%**
	(-0.84)	(-1.11)	(-0.43)	(2.74)	(2.12)	(2.29)
High	0.32%**	0.19%	0.32%**	0.35%**	0.36%**	0.35%**
	(2.26)	(1.23)	(2.56)	(2.15)	(2.19)	(2.35)
H - L	0.85%**	0.34%	0.93%***	1.39%***	1.11%***	1.30%***
	(2.19)	(1.02)	(3.58)	(4.88)	(3.70)	(5.73)

Table 3. (Continued) Performance of portfolio from composite return predictors

Panel B: High-minus-low portfolio alphas after transaction costs

	LCP			NLCP		
	FF5	q5	MISP	FF5	q5	MISP
1 bps	0.84%**	0.33%	0.73%**	1.37%***	1.09%***	1.12%***
2 bps	0.82%**	0.32%	0.71%**	1.35%***	1.07%***	1.11%***
3 bps	0.81%**	0.30%	0.7%*	1.34%***	1.06%***	1.1%***
4 bps	0.8%**	0.29%	0.69%*	1.33%***	1.05%***	1.08%***
5 bps	0.79%**	0.28%	0.67%*	1.31%***	1.03%***	1.07%***
6 bps	0.77%**	0.26%	0.66%*	1.3%***	1.02%***	1.06%***
7 bps	0.76%*	0.25%	0.65%*	1.29%***	1.01%***	1.04%***
8 bps	0.75%*	0.24%	0.63%*	1.27%***	0.99%***	1.03%***
9 bps	0.73%*	0.23%	0.62%*	1.26%***	0.98%**	1.02%***
10 bps	0.72%*	0.21%	0.61%*	1.25%***	0.97%**	1.00%***

Table 4. Portfolio performance over time

This table reports the performance persistence regarding the value-weighted portfolios of the nonlinear composite return predictor (NLCP) constructed in Table 3. At the end of each month, we rank all sample stocks based on NLCP value and sort them into decile portfolios. The holding period of portfolios ranges from two months to twelve months. The corresponding FF5 alphas are reported. To reserve space, we only report alphas from the top (bottom) three deciles and the high-minus-low portfolios. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	8	9	High	**High - Low**
m + 2	-1.11%***	-0.20%	0.19%	0.11%	0.19%	0.41%	1.51%***
	(-4.27)	(-0.93)	(0.75)	(0.95)	(0.99)	(1.53)	(3.93)
m + 3	-0.88%***	-0.05%	-0.17%	0.15%	-0.08%	0.05%	0.93%***
	(-3.18)	(-0.23)	(-1.07)	(0.78)	(-0.65)	(0.25)	(2.65)
m + 4	-1.11%***	-0.35%	-0.28%*	0.25%	0.13%	0.24%	1.35%***
	(-3.27)	(-1.41)	(-1.69)	(1.58)	(1.06)	(1.06)	(3.31)
m + 5	-1.11%***	-0.14%	-0.21%	0.09%	0.33%**	0.33%	1.43%***
	(-3.65)	(-0.59)	(-1.05)	(0.55)	(1.97)	(1.53)	(3.68)
m + 6	-0.61%**	-0.08%	0.05%	0.10%	0.13%	0.10%	0.71%**
	(-2.17)	(-0.52)	(0.31)	(0.90)	(1.09)	(0.71)	(2.08)
m + 7	-1.60%***	-0.43%**	-0.05%	0.07%	0.30%***	0.11%	1.71%***
	(-6.07)	(-1.97)	(-0.29)	(0.57)	(3.15)	(0.71)	(5.25)
m + 8	-0.76%***	-0.03%	-0.04%	-0.14%	0.06%	0.08%	0.85%***
	(-3.49)	(-0.12)	(-0.23)	(-1.63)	(0.44)	(0.41)	(2.78)
m + 9	-0.79%***	0.08%	0.06%	-0.10%	-0.04%	-0.20%	0.59%**
	(-3.48)	(0.36)	(0.33)	(-1.34)	(-0.44)	(-1.45)	(2.16)
m + 10	-0.71%***	0.08%	-0.22%	0.12%	-0.08%	-0.13%	0.58%
	(-2.58)	(0.27)	(-0.82)	(1.28)	(-0.68)	(-0.74)	(1.50)
m + 11	-0.39%	0.02%	0.26%	-0.06%	-0.15%	-0.24%	0.14%
	(-1.34)	(0.07)	(1.56)	(-0.63)	(-1.29)	(-1.56)	(0.41)
m + 12	-0.52%*	0.02%	0.17%	-0.13%	-0.01%	-0.28%	0.24%
	(-1.79)	(0.06)	(0.88)	(-1.04)	(-0.08)	(-1.58)	(0.60)

Table 5. Fama-MacBeth regressions of monthly returns

This table reports the Fama and MacBeth (1973) regressions of monthly stock returns on variables of interest, which include the linear composite return predictor (LCP) and the nonlinear composite return predictor (NLCP), as well as individual trading signals from nine market participants. Again, we report two measures of retail order imbalance – Retail Trading_BJZZ and Retail Trading_MPR. The control variables include firm size (SIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), asset growth (AG), turnover (TO), and idiosyncratic volatility (IVOL). Their coefficients are not reported here for brevity. The out-of-sample period is from 2013 to 2020. The explanatory variables are standardized with mean equals zero and standard deviation equals one. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	Ret	Ret	Ret	Ret
Bank Trading	-0.0164*	-0.0144	-0.0157	-0.0140
	(-1.68)	(-1.54)	(-1.64)	(-1.49)
Firm Trading	0.0011	0.0006	0.0001	-0.0000
	(1.12)	(0.65)	(0.15)	(-0.01)
Hedge Fund Trading	0.0171	0.0169	0.0170	0.0172
	(1.04)	(1.04)	(1.02)	(1.05)
Insurance Company Trading	-0.7398	-0.7425	-0.7345	-0.7234
	(-1.04)	(-1.04)	(-1.04)	(-1.05)
Mutual Fund Trading	-0.0014*	-0.0017*	-0.0014*	-0.0016*
	(-1.83)	(-1.95)	(-1.78)	(-1.86)
Other Institutional Trading	-0.0102	-0.0116	-0.0094	-0.0107
	(-0.49)	(-0.56)	(-0.45)	(-0.52)
Short Seller Trading	-0.0001	-0.0005	-0.0004	-0.0005
	(-0.27)	(-0.91)	(-0.75)	(-0.95)
Wealth Management Trading	0.1282	0.1294	0.1281	0.1263
	(1.10)	(1.11)	(1.11)	(1.11)
Retail Trading_MPR	-0.0009	-0.0003	-0.0003	-0.0001
	(-1.18)	(-0.39)	(-0.35)	(-0.13)
Retail Trading_BJZZ	0.0018***	0.0009**	0.0010***	0.0007
	(6.31)	(2.35)	(2.88)	(1.62)
LCP		0.0068***		0.0035
		(3.53)		(1.49)
NLCP			0.0120***	0.0101**
			(3.27)	(2.50)
Control variables	Y	Y	Y	Y
Obs.	213,262	213,262	213,262	213,262
R-squared	0.05	0.05	0.05	0.05

Table 6. Information environment and return predictability

This table reports the Fama-MacBeth regressions from Table 5 for subsamples of stocks constructed based on different levels of information environment. We use firm age, firm size, idiosyncratic volatility, as well as analyst coverage to proxy for a stock's information environment. At the end of each month, all sample stocks are sorted into halves based on an information environment proxy. We measure firm age as the number of years since the firm was first covered by CRSP, firm size as its market capitalization, idiosyncratic volatility as the standard deviation of the residuals from the CAPM model of daily stock excess returns over the previous 6 months, and analyst coverage as the number of analysts covering the firm. The control variables are the same as in Table 5. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Firm age		Firm size		Idiosyncratic volatility		Analyst coverage	
	Young	Old	Small	Large	High	Low	Low	High
	Ret	Ret	Ret	Ret	Ret	Ret	Ret	Ret
NLCP	0.010***	0.010***	0.012***	0.004*	0.011***	0.004*	0.009*	0.009***
	(3.09)	(2.91)	(3.69)	(1.77)	(4.01)	(1.81)	(1.97)	(3.27)
SIZE	-0.005***	-0.001	-0.015***	0.000	-0.008***	-0.001	-0.009***	0.000
	(-3.63)	(-1.10)	(-3.96)	(0.09)	(-3.70)	(-0.91)	(-3.11)	(0.15)
BM	-0.003*	-0.000	-0.002	-0.004**	-0.002	-0.002	-0.001	-0.003
	(-1.93)	(-0.10)	(-1.61)	(-2.22)	(-1.44)	(-1.47)	(-0.76)	(-1.50)
MOM	-0.003**	-0.001	-0.004***	-0.000	-0.002*	-0.001	-0.003**	0.000
	(-2.08)	(-1.17)	(-2.90)	(-0.29)	(-1.92)	(-0.79)	(-2.57)	(0.04)
STR	-0.002*	-0.004***	-0.003**	-0.001	-0.003***	-0.006***	-0.003**	-0.001
	(-1.73)	(-5.34)	(-2.48)	(-0.54)	(-2.73)	(-3.25)	(-2.35)	(-1.25)
AG	-0.001	-0.000	-0.001	0.000	-0.000	0.001	-0.001	0.000
	(-1.38)	(-0.09)	(-0.68)	(0.27)	(-0.64)	(0.65)	(-0.95)	(0.47)
TO	-0.010***	-0.009***	-0.011***	-0.004	-0.011***	-0.016	-0.011***	-0.001
	(-5.09)	(-5.28)	(-5.47)	(-1.05)	(-5.86)	(-1.35)	(-6.02)	(-0.27)
IVOL	0.005**	0.004	0.005**	0.000	0.007***	0.001	0.004*	0.002
	(2.38)	(1.54)	(2.03)	(0.28)	(3.48)	(0.30)	(1.97)	(0.84)
Obs.	126,865	158,362	132,963	152,264	132,404	152,823	125,277	113,944
R-squared	0.05	0.05	0.05	0.07	0.04	0.05	0.05	0.08

Table 7. Predicting profitability and earnings surprises

This table reports the Fama-MacBeth regressions of subsequent stock profitability and earnings surprises on the nonlinear composite return predictor (NLCP). We use return-on-asset (ROA) and cash flows (CF) to measure profitability. We use standardized unexpected earnings (SUE) to measure earnings surprise. Specifically, ROA is calculated as the sum of income before extraordinary items and interest expenses, divided by the lagged total assets. CF is measured as the difference between income before extraordinary items and total accruals, divided by total assets. SUE is computed as the market-adjusted returns upon earnings announcements over the three-day window. We report the change of values from the last period in the first three columns and the levels in the next three columns, as dependent variables.. We further include firm size (SIZE), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), asset growth (AG), and gross profitability (GP) as control variables. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	ΔROA	ΔCF	ΔSUE	ROA	CF	SUE
NLCP	0.0597**	0.0392***	0.0062**	0.2692***	0.2517***	0.0156***
	(2.23)	(2.68)	(2.62)	(6.95)	(7.94)	(7.70)
SIZE	-0.0007	0.0085	-0.0004	0.1056***	0.1098***	-0.0002
	(-0.12)	(1.61)	(-0.46)	(18.27)	(19.42)	(-0.28)
BM	-0.0463***	-0.0289***	-0.0014	0.0380***	0.0493***	0.0002
	(-4.44)	(-4.09)	(-1.34)	(3.48)	(6.92)	(0.23)
MOM	0.0166***	0.0066*	-0.0026**	0.0338***	0.0242***	0.0005
	(3.11)	(1.98)	(-2.13)	(6.08)	(5.25)	(0.58)
STR	0.0047	0.0046*	-0.0259***	-0.0004	-0.0003	0.0008*
	(1.18)	(1.81)	(-26.17)	(-0.09)	(-0.10)	(1.78)
AG	0.0428***	0.0164**	-0.0002	-0.0371***	-0.0340***	-0.0008
	(3.57)	(2.16)	(-0.20)	(-4.45)	(-5.79)	(-1.41)
GP	-0.0924***	-0.0521***	0.0000	0.0463*	0.0599***	0.0006
	(-3.78)	(-3.26)	(0.02)	(1.73)	(3.22)	(0.76)
Obs.	226,329	222,214	286,434	230,541	223,447	288,550
R-squared	0.04	0.03	0.04	0.17	0.27	0.01

Table 8. Stock anomalies among nonlinear composite return predictor portfolios

This table reports the time-series average of the cross-sectional means of firm characteristics corresponding to the stock anomalies for portfolios sorted by the Nonlinear composite return predictor (NLCP). For each one out of 102 stock anomalies documented in Green, Hand, and Zhang (2017) , we check its long-short portfolio's excess return during our sample period 2008-2020. We only select anomalies with significant (t-stat >1.66) returns, which are market capitalization (Banz, 1981), book-to-market (Rosenberg, Reid, and Lanstein, 1985), gross margin (Novy-Marx, 2013), illiquidity (Amihud, 2002), idiosyncratic volatility (Ali, Hwang, and Trombley, 2003), momentum_12m (Jegadeesh, 1990), momentum_1m (Jegadeesh and Titman, 1993), asset growth (Cooper, Gulen, and Schill, 2008), dividend yield (Litzenberger and Ramaswamy, 1981), analyst coverage (Elgers, Lo, and Jr., 2001), price delay (Hou and Moskowitz, 2005), and combined fundamental (Mohanram, 2005). The t-statistics are Newey-West adjusted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	4	High	High - Low	t-stat
Market capitalization	12.427	13.585	13.937	14.017	13.664	1.24***	10.37
Book-to-market	0.554	0.567	0.561	0.554	0.596	0.04***	2.68
Gross margin	0.187	0.291	0.298	0.301	0.293	0.11***	18.80
Illiquidity	0.810	0.874	0.864	0.765	1.154	0.34**	2.56
Idiosyncratic volatility	0.078	0.052	0.046	0.045	0.047	-0.03***	-16.27
Momentum_12m	-0.010	0.084	0.103	0.120	0.113	0.12***	5.43
Momentum_1m	-0.002	0.009	0.012	0.013	0.015	0.02*	1.97
Asset growth	0.137	0.118	0.107	0.104	0.093	-0.04***	-5.75
Dividend yield	0.011	0.013	0.014	0.014	0.014	0.00***	4.41
Analyst coverage	5.168	7.350	8.069	8.080	7.096	1.93***	6.16
Price delay	0.089	0.064	0.060	0.057	0.077	-0.01*	-1.74
Combined fundamental	3.572	4.163	4.326	4.361	4.237	0.67***	13.77

Table 9. Predicting net anomaly indicator

This table reports the time-series average of the cross-sectional means of the nonlinear composite return predictor (NLCP) in quintile portfolios sorted by *Net* anomaly indicator (NET). For comparison, we report the values for alternative nonlinear composite return predictor using two (three) trading signals, denoted as NLCP_two (NLCP_three). Specifically, the two trading signals are from two market participants – firms and short sellers, and the three trading signals come from three market participants – firms, short sellers and informed retail investors from Boehmer et al. (2021). We use the same construction method as before. To obtain Net anomaly, we select the anomalies in Table 8 whose predictability exhibits consistent directions with that of the NLCP and is significant during our sample period, in order to evaluate the combined net effect. We then sort stocks into quintile portfolios at the end of each month based on the values of the selected anomalies. The top (bottom) quintile portfolio of each anomaly is treated as the long (short) side. For each stock-month observation, NET is computed as the difference between the number of long-side anomaly portfolios and the number of short-side anomaly portfolios that the stock falls into. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	composite return predictors based on past trading		
NET Quintile	NLCP	NLCP_two	NLCP_three
Low	1.16%	0.80%	0.80%
2	0.82%	0.53%	0.53%
3	1.43%	0.97%	0.96%
4	1.49%	1.00%	1.00%
High	1.52%	1.02%	1.02%
High - Low	0.36%***	0.23%***	0.22%***
	(6.74)	(30.64)	(31.32)

Table 10. Performance of factor model based on nonlinear composite return predictor

This table conducts comparison among different factor models regarding anomaly prediction. We create a two-factor model, including an NLCP factor plus the market factor, and examine its ability to explain the anomaly returns against other factor models: Fama and French (2015) five-factor model (FF5), Hou, Mo, Xue, and Zhang (2021) q5 model (q5), and Stambaugh and Yuan (2017) mispricing-factor model (MISP). The dependent variable is anomaly returns. To construct NLCP factor, each month we sort the sample of stocks into decile portfolios based on the NLCP value and then calculate the long-minus-short spread. Each column reports the average monthly alpha in absolute value, the average t-statistic in absolute value, the aggregate pricing error Delta ($\Delta = \alpha^{\mathrm{T}}\Sigma^{-1}\alpha$), as well as associated Gibbons, Ross, and Shanken (1989) F-statistic and p-value. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	FF5	q5	MISP	vw NLCP + market factor
Average $\|\alpha\|$	0.64%	0.52%	0.58%	0.40%
Average $\|t\|$	1.9	1.5	1.9	1.2
Delta	0.44	0.38	0.72	0.34
F-statistic	1.6*	1.3	2.8***	1.2
p-value	0.073	0.234	0.001	0.256

Appendix A Trading variables and machine learning methods

A.1 Trading signals from market participants

The input variables we put into the training sample are the trading signals from nine market participants. Following McLean, Pontiff, and Reilly (2022), we calculate the trading signals as the changes in market participants' holding levels over a one-year horizon. As mentioned in Section 2.1, we look into six types of institutions that report their holdings on SEC 13F form, as well as other types of investors: retail traders, short sellers and firms.

Obtaining available institutional holdings data from *Thomson/Refinitiv S12* and *13f*, we follow McLean, Pontiff, and Reilly (2022) to classify the institutions into six types. 1) We merge mutual fund holdings that are reported in S12 form filings with their corresponding 13F form filings, and classify the number of shares reported by mutual funds as their holdings; 2) In accordance with the type codes created by Brian Bushee (https://accounting-faculty.wharton.upenn.edu/bushee/), we filter out the the institutions as banks or insurance companies, and extract the corresponding holdings; 3) As for wealth management firms and hedge funds, we use McLean, Pontiff, and Reilly (2022) search scheme where we perform case insensitive searches for "Wealth Manag", "Wealth MGNT", "Private", "PRVT" and "advisor." to identify wealth management firms, as well as case insensitive searches for the remaining institutions "LLC", "L.L.C." "L L C", "L. L. C.", "LP", "L.P", "L P", "L. P", or "Partner" to identify hedge funds. We look into the holdings from both types; 4) We treat the remaining institutions as "Other" institutions and obtain the reported holdings. For all six types of the institutions, we scale their holdings by by the number of shares outstanding, and then calculate the corresponding change over a one-year period as trading signals from these institutions: *Bank Trading*, *Hedge Fund Trading*, *Insurance Company Trading*, *Mutual Fund Trading*, *Other Institutional Trading*, *Wealth Management Trading*.

We use two alternative proxies for retail trading signals. First, we obtain daily off-exchange marketable orders from *TAQ* trade dataset. According to Boehmer, Jones, Zhang, and Zhang (2021), retail orders that are internalized or executed by wholesalers are given a small amount of price improvement relative to the National Best Bid or Offer (NBBO), and thus one can separate retail orders based on the sub-penny pricing of the execution from the institutional orders. Specifically, we identify a transaction as a retail buy if the subpenny price is between 60 and 100 basis points, and identify a retail sell if the subpenny price is between 0 and 40 basis points. We then aggregate intraday retail buy volume and retail sell volume for each stock on each trading day. As for the two alternative proxies for retail trading, we first follow Boehmer et al. (2021) to compute the order imbalance as shown in equation (A.1), where $indbvol_{i,t}$ ($indsvol_{i,t}$) is the number of shares of stock i bought (sold) by retail investors on day t.

$$oib1_{i,t} = \frac{indbvol_{i,t} - indsvol_{i,t}}{indbvol_{i,t} + indsvol_{i,t}} \tag{A.1}$$

Alternatively, we follow McLean, Pontiff, and Reilly (2022) by replacing the denominator in the above equation with the number of shares outstanding as reported by *CRSP*, as shown in equation (A.2). To briefly explain, the former method aims to emphasize the "directional" trading signal while the latter method better captures the weight that retail investors put on the underlying stock through buying or selling. Given their weak correlation (smaller than 0.1), we use both methods to capture the retail trading signal for a well-rounded purpose. Similar to the way we construct institutional trading, we aggregate both measures of retail trading to a one-year period: *Retail Trading_MPR* for McLean, Pontiff, and Reilly (2022) approach, *Retail Trading_BJZZ* for Boehmer et al. (2021) approach.

$$oib2_{i,t} = \frac{indbvol_{i,t} - indsvol_{i,t}}{shrout_{i,t}} \tag{A.2}$$

The remaining input variables are trading signals from short sellers and firms. To begin with, we obtain monthly short interest from *Compustat* and scale it by the number of shares outstanding extracted from *Compustat* as well. We sign the scaled short interest ratio such that increases in such ratio result in negative values of short seller trading and decreases in the ratio (net closing of short positions) result in positive values of short seller trading. As for firm trading, we obtain firm's shares outstanding reported from *CRSP* and adjust it for stock splits and stock dividends, following McLean, Pontiff, and Reilly (2022). We then calculate the changes in shares, i.e., share issues minus share repurchases, scaled by the number of shares outstanding. We sign the variable such that decreases (increases) in shares outstanding reflect positive (negative) values of firm trading signal. Again, we aggregate the computed trading signals to a one-year horizon and thus generate *Short Seller Trading* and *Firm Trading* variables.

A.2 Machine learning methods

We employ multiple machine learning models to train the trading signals from nine market participants in order to obtain out-of-sample return premiums. We select the models that are adopted in recent academic papers (Rapach, Strauss, Tu, and Zhou, 2019; Gu, Kelly, and Xiu, 2020; DeMiguel, Gil-Bazo, Nogales, and Santos, 2022), given rapid advances in asset pricing studies using machine learning techniques. The models can be categorized into either linear or nonlinear type.

We use the *sklearn* Python package to train the models. For models that requires tuning the hyper-parameters, we follow the literature by adopting the five-fold cross validation technique (Hastie, Tibshirani, and Friedman, 2009). To briefly summarize, we divide the sample into five *folds* and then remove each *fold* in turn and evaluate the estimation errors associated with different sets of hyper-parameters. We choose the optimal hyper-parameter values that yield the minimum average estimation error.

Linear models

a) Ordinary least squares (OLS)

We start with a basic linear predictive regression model, ordinary least squares, where its wide adoption by research papers and its easy-to-interpret nature make it suitable for a benchmark model. The goal here is set to minimize the objective function that subjects to the parameter vector:

$$\min_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{N} (\alpha_{i,t+1} - \sum_{j=1}^{p} x_{ij,t}\beta_{ij,t})^2 \tag{A.3}$$

where a panel structure of training data regarding both T time points and N stocks are included in the function. Specifically, $\alpha_{i,t+1}$ is the excess return of the ith stock in month t + 1, $x_{ij,t}$ refers to one of the trading signals for the ith stock in month t, and $\beta_{ij,t}$ refers to the parameter variable in the regression. Compared with machine learning tools, a drawback of a multivariate OLS regression is the overfitting issue.

b) Ridge regression (Ridge)

One common way to improve the performance of multivariate linear regression is to employ a shrinkage method. The basic idea is to draw coefficient estimates closer to zero and thus avoid the scenarios when they become too large in magnitude (Gu, Kelly, and Xiu, 2020). Ridge regression tends to improve the forecast accuracy by trading off a small increase in estimation bias for a large reduction in estimation variance. Specifically, ridge regression shrinks the regression coefficients through parameter penalization (Hastie, Tibshirani, and Friedman, 2009), where the goal is to minimize a penalized residual sum of squares:

$$\min_{\beta} [\sum_{t=1}^{T} \sum_{i=1}^{N} (\alpha_{i,t+1} - \sum_{j=1}^{p} x_{ij,t}\beta_{ij,t})^2 + \lambda \sum_{j=1}^{p} \beta_{ij,t}^2] \tag{A.4}$$

where the complexity parameter λ controls the magnitude of shrinkage and is set to be within [0.0001, 0.1]. Again, $\alpha_{i,t+1}$ is the excess return and $\beta_{ij,t}$ refer to one of the parameters. There are p trading signals for each stock i. The panel data cover T time points and N stocks in total.

c) Adaptive least absolute shrinkage and selection operator (ALasso)

Another common solution to alleviate the drawbacks of multivariate OLS regression in the machine learning literature is employing least absolute shrinkage and selection operator

(LASSO) in the predictor estimation (Tibshirani, 1996; Rapach, Strauss, Tu, and Zhou, 2019). In short, LASSO performs both shrinkage and variable selection, by minimizing the function as below:

$$\min_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{N} [||\alpha_{i,t+1} - \sum_{j=1}^{p} x_{ij,t}\beta_{ij,t}||^2 + \lambda \sum_{j=1}^{p} |\beta_{ij,t}|] \tag{A.5}$$

where λ is again the regularization parameter, with the ridge penalty $\sum_{j=1}^{p} \beta_{ij,t}^2$ in Formula A.4 replaced by the lasso penalty $\sum_{j=1}^{p} |\beta_{ij,t}|$ in Formula A.5.

Following Zou (2006), we adopt the adaptive LASSO method to further improves the estimation. Specifically, we assign different weights $w_{ij,t}$ to different parameters and λ again falls within [0.0001, 0.1]:

$$\min_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{N} [||\alpha_{i,t+1} - \sum_{j=1}^{p} x_{ij,t}\beta_{ij,t}||^2 + \lambda \sum_{j=1}^{p} w_{ij,t}|\beta_{ij,t}|] \tag{A.6}$$

where the weight is set to be $\frac{1}{|\hat{\beta}|}$. Each value of $\hat{\beta}$ is obtained from the first-step OLS regression residuals as follows:

$$\underset{\beta}{argmin} \sum_{i=1}^{N} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{A.7}$$

d) Elastic net (ENet)

We also use Elastic net model that helps tune the training sample with the machine adaptively optimizes the hyper-parameters (Gu, Kelly, and Xiu, 2020; DeMiguel et al., 2022). The general model incorporates both shrinkage and selection through the implementation of hyper-parameters, as follows:

$$\min_{\beta} \sum_{t=1}^{T} \sum_{i=1}^{N} [(\alpha_{i,t+1} - \sum_{j=1}^{p} x_{ij,t}\beta_{ij,t})^2 + \lambda\rho \sum_{j=1}^{p} |\beta_{ij,t}| + \lambda(1-\rho) \sum_{j=1}^{p} \beta_{ij,t}^2] \tag{A.8}$$

where $\sum_{j=1}^{p} |\beta_{ij,t}|$ is the 1-norm and $\sum_{j=1}^{p} \beta_{ij,t}^2$ is the 2-norm of the parameter sets. Formula A.8 is a generalized representation built on Formula A.4 when $\rho = 0$ and Formula A.5 when $\rho = 1$. Any intermediate value of the hyper-parameter includes both types of regression. We set the hyper-parameter ρ to be 0.5 and λ within [0.0001,0.1].

e) Principal components regression (PCR)

Besides employing a penalty feature to solve overfitting issue in OLS, we alternatively adjust the baseline model with dimension reduction technique. By averaging across predictors, dimension reduction helps reduce noise and decorrelate among predictors, compared to simple OLS. Each trading signal p for stock i at time t $\sum_{j=1}^{p} x_{ij,t}$ in Formulas A.4 - A.8 can be further decomposed into k linear combinations of predictors as follows:

$$\sum_{j=1}^{p}\sum_{w=1}^{k} x_{ijw,t}W_{ijw,t} \tag{A.9}$$

which serves as a dimension-reduced version of the original predictor set.

We apply one of the dimension reduction tools, principal components regression (PCR), which involves a principal components analysis (PCA) that optimizes the covariance structure and then selecting some leading components in regression. Specifically, PCR chooses the combination weights $\sum_{w=1}^{k} W_{ijw,t}$ recursively, in order to find components that retain the most possible common variation within the predictor set. Although efficient in computation, PCR is unable to predict returns in the dimension reduction step, which leads to applying another method - partial least squares - that directly solves the drawback.

f) Partial least squares (PLS)

As another dimension reduction tool, Partial least squares (PLS) incorporates return prediction when exploiting covariance among predictors in the dimension reduction step (Gu, Kelly, and Xiu, 2020). Unlike PCR, PLS estimates coefficients on return prediction for all univariate predictors and averages across predictors with the highest (lowest) weight on the strongest (weakest) predictors. By choosing components with more potent return predictability, PLS tends to let go the accuracy of weight matrix.

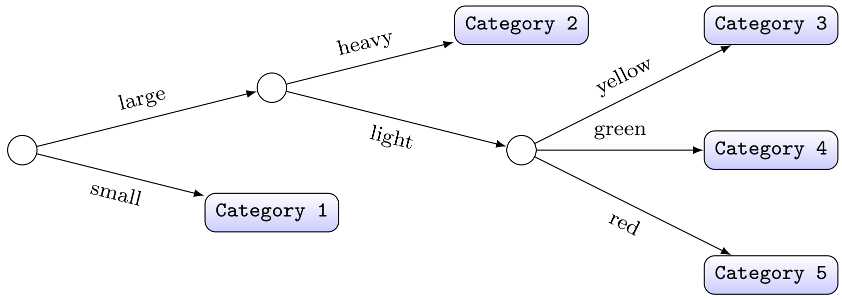

Figure A.1. Decision trees

Nonlinear models

a) Gradient boosted regression trees (GBRT)

To take into account the nonlinear effect of predictors as well as their nonlinear interactions, we first adopt the regression trees model that brings multiway predictor interactions in a nonparametric way. To briefly explain, trees first identify and sort groups of observations with similar behavior and the sorting grows in a sequence. For observations in each group, simple averages of the outcome variable's values are taken to yield forecast. A simplified visualization is shown in Figure A.1, where we sort the observations based on characteristics size, weight, and color in sequence.

Although the trees model can capture interactions, it is subject to the overfitting issue. We tackle this problem by employing the gradient boosted regression trees model (GBRT) (Gu, Kelly, and Xiu, 2020; DeMiguel et al., 2022), where the gradient boosting function combines decision trees in a sequence and it starts from weak decision trees and converges to strong trees. We set the hyper-parameter learning rate, which determines the weight given to the most recent decision tree, to be within 0.01, 0.1. In addition, we let the depth of the tree to fall within 1, 2, and the number of trees in out setting is within [1, 1000]. Overall, GBRT improves the forecasting performance by reducing the prediction variance as well as the prediction bias (Schapire and Freund, 2012).

b) Random forest (RF)

Similar to gradient boosting, a random forest (RF) method combines decision trees altogether. One key distinction between the two methods is that RF aggregates independent decision trees through bootstrap aggregation Breiman (2001) while GBRT aggregates the trees sequentially. Random forest method draws random subsets of the data, where for each sample it utilizes a distinct regression tree and then averages across the estimates to reduce the variance. It tends to reduce the correlation among predictions and thus the estimation variance. The number of trees is set to be 300, in accordance with Gu, Kelly, and Xiu (2020). We set the maximum depth in our setting to be between 1 - 6. The number of features in each split is within 1, 2, 3, 5, 10.

c) Artificial neural networks (ANN)

Lastly, we employ the artificial neural network model that entwines various telescoping layers of nonlinear predictor interactions. The model is regarded as a highly parameterized and complicated machine learning tool. Specifically, we apply a commonly-used "feed-forward" network model (e.g., Gu, Kelly, and Xiu, 2020; Kaniel, Lin, Pelger, and Nieuwerburgh, 2022), which includes *input layers* of predictors, *hidden layers* that capture complex interactions among the predictors, and an *output layer* for an outcome prediction through aggregating *hidden layers*. One illustration example of this "feed-forward" network is shown in Figure A.2.

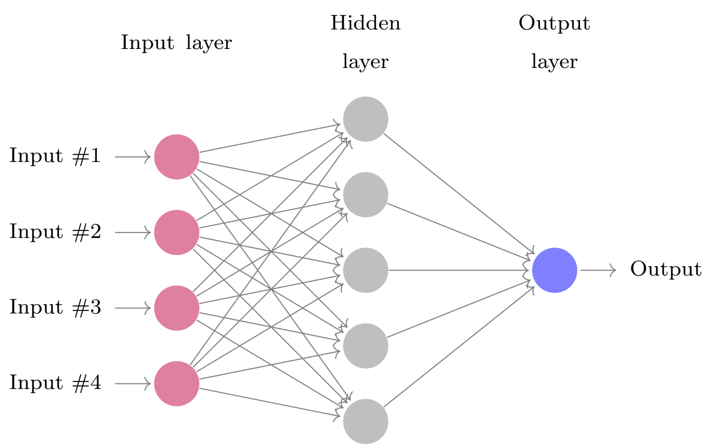

Figure A.2. Neural network

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{A.10}$$

$$ReLU(x) = \max(0, x) \tag{A.11}$$

We consider networks with up to four hidden layers (i.e., ANN1, ANN2, ANN3, ANN4). We follow Gu, Kelly, and Xiu (2020) for the setup of the hyper-parameters in the model, while we set the learning rate to be within 0.00001, 0.0001, 0.001.

As for the nonlinear activation function, we choose two commonly-used functional forms: rectified linear unit (ReLU) and hyperbolic tangent (Tanh), and apply the activation functions at all nodes. To briefly explain, Tanh function in Formula A.10 helps transform the data into zero-centered data in order to make learning for the next layer much easier. ReLu function in Formula A.11 helps deactivate the nodes for which the output of the transformation is less than zero.

Appendix B Robustness tests and additional results

Table B.1. q5 alphas of univariate portfolios from trading by nine market participants

This table reports the Hou et al. (2021) q5 alphas (q5) for the value-weighted portfolios sorted by each of the nine market participants' trading signals. Again, we report two measures of retail order imbalance – Retail Trading_BJZZ as in Boehmer et al. (2021) and Retail Trading_MPR as in McLean, Pontiff, and Reilly (2022). At the end of each month, we rank all sample stocks based on the trading signal and sort them into decile portfolios. The portfolios are held for one month. To reserve space, we only report alphas from the top (bottom) three deciles and the high-minus-low portfolios. To be consistent with the out-of-sample period, we report results in period 2013-2020 here. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	8	9	High	**H - L**
Bank Trading	0.14%	-0.09%	0.02%	-0.02%	-0.20%	0.12%	-0.02%
	(0.68)	(-0.50)	(0.09)	(-0.13)	(-1.13)	(0.62)	(-0.07)
Firm Trading	0.07%	0.47%**	0.26%*	-0.10%	-0.04%	-0.06%	-0.14%
	(0.21)	(2.39)	(1.90)	(-0.64)	(-0.41)	(-0.43)	(-0.42)
Hedge Fund Trading	0.10%	-0.18%	0.02%	0.04%	0.01%	0.25%	0.16%
	(0.37)	(-1.10)	(0.14)	(0.26)	(0.09)	(1.33)	(0.45)
Insurance Company Trading	0.23%	-0.08%	0.00%	-0.15%	-0.13%	0.04%	-0.19%
	(1.12)	(-0.53)	(-0.02)	(-0.71)	(-0.76)	(0.19)	(-0.65)
Mutual Fund Trading	0.09%	-0.07%	0.21%	0.11%	0.29%*	0.07%	-0.02%
	(0.57)	(-0.46)	(0.83)	(0.33)	(1.68)	(0.84)	(-0.08)
Other Institutional Trading	0.08%	-0.18%	0.10%	0.11%	0.16%	0.08%	0.01%
	(0.28)	(-1.13)	(0.59)	(0.96)	(1.04)	(0.47)	(0.03)
Short Seller Trading	0.36%***	-0.01%	-0.05%	-0.23%**	0.18%*	0.22%**	-0.15%
	(2.91)	(-0.08)	(-0.42)	(-2.33)	(1.94)	(2.02)	(-0.94)
Wealth Management Trading	0.15%	-0.22%**	0.11%	-0.15%	0.20%	0.22%	0.07%
	(1.14)	(-2.14)	(0.65)	(-0.90)	(1.29)	(1.25)	(0.59)
Retail Trading_MPR	0.32%	0.32%***	0.19%	-0.07%	-0.24%	-0.30%	-0.62%*
	(1.16)	(2.79)	(1.00)	(-0.36)	(-0.90)	(-0.72)	(-1.76)
Retail Trading_BJZZ	0.34%**	0.06%	-0.12%	-0.08%	0.16%	0.70%***	0.37%*
	(1.99)	(0.51)	(-1.26)	(-0.53)	(1.14)	(3.41)	(1.71)

Table B.2. MISP alphas of univariate portfolios from trading by nine market participants

This table reports the Stambaugh and Yuan (2017) mispricing-factor alphas (MISP) for the value-weighted portfolios sorted by each of the nine market participants' trading signals. Again, we report two measures of retail order imbalance – Retail Trading_BJZZ as in Boehmer et al. (2021) and Retail Trading_MPR as in McLean, Pontiff, and Reilly (2022). At the end of each month, we rank all sample stocks based on the trading signal and sort them into decile portfolios. The portfolios are held for one month. To reserve space, we only report alphas from the top (bottom) three deciles and the high-minus-low portfolios. To be consistent with the out-of-sample period, we report results in period 2013-2020 here. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	8	9	High	**H - L**
Bank Trading	0.11%	-0.10%	0.08%	0.14%	-0.32%*	0.05%	-0.06%
	(0.59)	(-0.54)	(0.46)	(0.97)	(-1.81)	(0.34)	(-0.25)
Firm Trading	-0.34%	0.16%	0.37%**	-0.08%	-0.10%*	-0.27%	0.07%
	(-1.07)	(0.74)	(2.58)	(-0.52)	(-1.67)	(-1.35)	(0.21)
Hedge Fund Trading	0.21%	-0.10%	0.06%	0.15%	-0.05%	0.20%	0.00%
	(0.90)	(-0.63)	(0.39)	(0.97)	(-0.40)	(1.20)	(-0.02)
Insurance Company Trading	0.20%	-0.10%	-0.07%	-0.04%	-0.18%	-0.01%	-0.21%
	(1.04)	(-0.72)	(-0.44)	(-0.19)	(-1.12)	(-0.09)	(-0.80)
Mutual Fund Trading	0.03%	-0.04%	-0.17%	0.08%	0.19%	0.00%	-0.03%
	(0.20)	(-0.29)	(-0.45)	(0.16)	(1.44)	(-0.04)	(-0.17)
Other Institutional Trading	0.21%	-0.11%	0.05%	0.26%*	0.00%	0.09%	-0.11%
	(0.84)	(-0.73)	(0.32)	(1.79)	(0.00)	(0.57)	(-0.42)
Short Seller Trading	-0.16%	-0.36%***	-0.24%**	-0.01%	0.23%***	0.16%	0.32%
	(-0.98)	(-3.26)	(-2.26)	(-0.10)	(3.23)	(1.47)	(1.46)
Wealth Management Trading	0.19%	-0.19%	-0.02%	-0.03%	0.14%	0.12%	-0.07%
	(0.99)	(-1.56)	(-0.14)	(-0.19)	(0.90)	(0.88)	(-0.28)
Retail Trading_MPR	0.08%	0.25%	0.05%	0.22%	0.12%	-0.06%	-0.14%
	(0.47)	(1.30)	(0.30)	(1.23)	(0.35)	(-0.19)	(-0.61)
Retail Trading_BJZZ	-0.20%	-0.27%**	-0.29%***	-0.12%	-0.26%	0.11%	0.31%
	(-0.92)	(-2.21)	(-3.04)	(-0.62)	(-1.38)	(0.42)	(1.28)

Table B.3. Excess returns and DSH alphas of portfolio from composite return predictors

This table reports the excess returns and the Daniel, Hirshleifer, and Sun (2020) behavioral-factor alphas (DHS) of the value-weighted portfolios sorted by either the linear composite return predictor (LCP) or the nonlinear composite return predictor (NLCP). At the end of each month, we rank all sample stocks based on the value of LCP or NLCP and sort them into decile portfolios. The portfolios are held for one month. The out-of-sample covers from 2013 to 2020. The t-statistics are Newey-West adjusted and are shown in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	LCP		NLCP	
	Excess ret	DSH	Excess ret	DSH
Low	0.93% *	-0.35%	0.50%	-0.94% ***
	(1.68)	(-1.18)	(1.03)	(-4.06)
2	1.21% ***	-0.21%	0.78% *	-0.59% **
	(2.59)	(-1.12)	(1.87)	(-2.27)
3	1.25% ***	-0.11%	1.35% ***	0.11%
	(2.87)	(-0.44)	(3.91)	(0.61)
4	1.44% ***	0.11%	1.07% ***	-0.13%
	(3.60)	(0.80)	(2.91)	(-0.89)
5	1.08% ***	-0.13%	1.00% ***	-0.16%
	(3.34)	(-1.02)	(2.65)	(-1.15)
6	1.12%***	-0.05%	1.13% ***	-0.07%
	(3.28)	(-0.30)	(3.94)	(-0.91)
7	1.37% ***	0.15%	1.23% ***	0.01%
	(3.80)	(1.30)	(4.06)	(0.11)
8	1.15% ***	-0.07%	1.38% ***	0.12%
	(3.82)	(-0.82)	(3.91)	(1.24)
9	1.09% ***	-0.03%	1.68% ***	0.24% **
	(4.56)	(-0.28)	(3.36)	(2.02)
High	1.38% ***	0.27% *	1.52% ***	0.23%
	(4.24)	(1.65)	(3.75)	(1.29)
H - L	0.46%	0.61% *	1.01% ***	1.17% ***
	(1.13)	(1.89)	(3.98)	(4.34)

Table B.4. Stock anomalies among linear composite return predictor portfolios

This table reports the time-series average of the cross-sectional means of firm characteristics corresponding to the stock anomalies for portfolios sorted by the Linear composite return predictor (LCP). For each one out of 102 stock anomalies documented in Green, Hand, and Zhang (2017) , we check its long-short portfolio's excess return during our sample period 2008-2020. We only select anomalies with significant (t-stat >1.66) returns, which are market capitalization (Banz, 1981), book-to-market (Rosenberg, Reid, and Lanstein, 1985), gross margin (Novy-Marx, 2013), illiquidity (Amihud, 2002), idiosyncratic volatility (Ali, Hwang, and Trombley, 2003), momentum_12m (Jegadeesh, 1990), momentum_1m (Jegadeesh and Titman, 1993), asset growth (Cooper, Gulen, and Schill, 2008), dividend yield (Litzenberger and Ramaswamy, 1981), analyst coverage (Elgers, Lo, and Jr., 2001), price delay (Hou and Moskowitz, 2005), and combined fundamental (Mohanram, 2005). The t-statistics are Newey-West adjusted. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Low	2	3	4	High	High - Low	t-stat
Market capitalization	12.606	13.410	13.796	13.987	13.832	1.23***	10.09
Book-to-market	0.522	0.556	0.565	0.573	0.614	0.09***	5.71
Gross margin	0.188	0.283	0.301	0.301	0.296	0.11***	17.13
Illiquidity	0.000	0.000	0.000	0.000	0.000	0.00***	5.40
Idiosyncratic volatility	0.079	0.056	0.048	0.044	0.042	-0.04***	-23.23
Momentum_12m	0.011	0.081	0.098	0.110	0.110	0.10***	3.82
Momentum_1m	0.000	0.009	0.011	0.013	0.014	0.01*	1.72
Asset growth	0.155	0.129	0.111	0.094	0.070	-0.08***	-10.83
Dividend yield	0.011	0.013	0.014	0.014	0.014	0.00***	6.20
Analyst coverage	5.684	7.072	7.765	7.965	7.278	1.59***	4.63
Price delay	0.080	0.073	0.062	0.058	0.076	0.00	-0.54
Combined fundamental	3.640	4.089	4.274	4.351	4.303	0.66***	14.19