Machine Forecast Disagreement*

Ran Chang†

Current version: June 2022

Abstract

This paper documents a significantly negative cross-sectional relation between machine forecast disagreement (MFD) and future stock returns. MFD measures divergence of opinion among investors about stock value based on the dispersion in machines' expected return forecasts and provides an objective measure of investor disagreement, which is free from behavioral biases and conflicts of interest. A long-short portfolio of stocks sorted by MFD provides a six-factor Fama-French (2018) alpha of 0.72% (1.06%) per month for the value-weighted (equal-weighted) portfolio. The return predictability is driven by mispricing rather than compensation for risk. The disagreement premium is also stronger for stocks that are largely held by retail investors, that receive less investor attention and that are costlier to arbitrage.

JEL Classification: G10; G11; G12; G14.

Keywords: machine forecast disagreement, analyst forecast dispersion, stock returns, costly arbitrage, investor inattention.

*Chang acknowledges financial support from the Shanghai Institute of International Finance and Economics and the Shanghai Pujiang Program.

†Assistant Professor of Finance, Antai College of Economics & Management, Shanghai Jiao Tong University, China, 200030. Phone: (86) 6293-3671, Email: r.zhang@sjtu.edu.cn.

Machine Forecast Disagreement

Abstract

This paper documents a significantly negative cross-sectional relation between machine forecast disagreement (MFD) and future stock returns. MFD measures divergence of opinion among investors about stock value based on the dispersion in machines' expected return forecasts and provides an objective measure of investor disagreement, which is free from behavioral biases and conflicts of interest. A long-short portfolio of stocks sorted by MFD provides a six-factor Fama-French (2018) alpha of 0.72% (1.06%) per month for the value-weighted (equal-weighted) portfolio. The return predictability is driven by mispricing rather than compensation for risk. The disagreement premium is also stronger for stocks that are largely held by retail investors, that receive less investor attention and that are costlier to arbitrage.

JEL Classification: G10, G11, G12, G14.

Keywords: machine forecast disagreement, analyst forecast dispersion, stock returns, costly arbitrage, investor inattention.

1 Introduction

Miller [1977] hypothesizes that stock prices reflect an upward bias as long as there exists divergence of opinion among investors about stock value and pessimistic investors do not hold sufficient short positions (i.e., short-sale constraints). In Miller's model, overvaluation of securities is observed because pessimists are restricted to hold zero shares although they prefer to hold a negative quantity, and the prices of securities are mainly determined by the beliefs of optimistic investors. Since divergence of opinion is likely to increase with firm-specific uncertainty, Miller [1977] predicts a negative relation between uncertainty and expected return; that is, stocks with higher firm-specific uncertainty or greater investor disagreement are anticipated to have higher prices and lower subsequent return.[1]

As an important information intermediary, financial analysts predict companies' future performance and release research reports, including their earnings forecasts and stock recommendations. While financial analysts play a significant role in guiding crucial investment decisions, they are known to be subject to behavioral and psychological biases. Thus, on one hand, accurate and objective analyst forecasts can potentially help correct investor misperceptions, but on the other hand, biased analyst forecasts can reinforce investor misperceptions. The popular perception is that instead of being impartial providers of unbiased opinions, financial analysts can be cheerleaders for the firms they cover. Their impartiality may be compromised because they are also expected to secure underwriting and other investment banking business, so they have an incentive to accommodate firm managers and generate positive earnings surprises (see, e.g., Dugar and Nathan [1995], Michaely and Womack [1999], and Chan et al. [2007]).

In this paper, I propose a novel measure of divergence of opinion among investors about stock value based on the dispersion in machines' expected return forecasts. Compared to financial analysts, machines have neither behavioral biases nor conflicts of interest, thus I argue that machine forecast disagreement (MFD) provides an unbiased estimate of investor disagreement. After introducing a new measure of firm-specific uncertainty proxied by the degree of disagreement of machines' future return forecasts, I show that this newly proposed, objective measure of uncertainty (or investor disagreement) does have a significant impact on the cross-sectional pricing of individual stocks. I find that stocks with higher MFD earn significantly lower future returns than otherwise similar stocks. In particular, a value-weighted (equal-weighted) portfolio of stocks in the

[1] Diether et al. [2002] show that higher dispersion in analysts' earnings estimates, proxying for greater investor disagreement, predicts lower future return in the cross section of individual stocks.

highest MFD decile underperforms a portfolio of stocks in the lowest MFD decile by 5.4% (7.2%) per annum. This return predictability is also stronger for stocks that are largely held by retail investors, that receive less investor attention, and that are costlier to arbitrage. Using the mispricing measure of Stambaugh et al. [2015] and alternative measures of risk, I further show that high-MFD stocks are significantly overvalued, and they have large negative alphas as well as high total, systematic, and idiosyncratic risks, rejecting a risk-based explanation in favor of a mispricing-based explanation of the cross-sectional relation between MFD and equity returns.

I rely on 12 widely used machine learning models, including three dimension reduction models: Principal Components Analysis (PCA), Scaled PCA (SPCA), and Partial Least Squares (PLS); three penalized linear regressions: LASSO, Ridge, and Elastic Net (E-Net);[2] three regression trees: Random Forests (RF), Gradient Boosted Regression Tree (GBRT), and Extreme Tree (ET); and three neural networks: Feed Forward Neural Network (FNN), Recurrent Neural Network (RNN), and Long Short-Term Memory Neural Network (LSTM).[3] I use 10 years (first six years as the training sample and subsequent four years as the validation sample) as the rolling window to estimate parameters of each machine learning model to obtain return forecasts of each stock in month t using information in the 310 stock characteristics available in month $t-1$. The machine forecast disagreement (MFD) is defined as the cross-sectional standard deviation of machines' return forecasts scaled by the absolute value of the mean machines' return forecasts obtained from the 12 machine learning models.[4] If the cross-sectional mean of machines' return forecasts is zero, then the stock is assigned to the highest disagreement category. Excluding observations with a mean return forecast of zero does not change the main findings.

I start my empirical analysis by investigating the predictive power of MFD in forecasting the cross-sectional variation in future stock returns. Specifically, I sort stocks into 10 decile portfolios by machine forecast disagreement during the previous month and examine the one-month- to 12-month-ahead value-weighted and equal-weighted average returns on the resulting portfolios from January 1976 to December 2019. I find that stocks with a higher (lower) MFD earn lower (higher) average return in subsequent months. Furthermore, the value-weighted arbitrage portfolio that

[2]I do not use the OLS model, since it has the overfitting problem in high-dimensional characteristics and is known to perform poorly in out-of-sample predictions (see, e.g., Gu et al. [2020]).

[3]These machine learning models are frequently used in the return forecast literature; e.g., Freyberger et al. [2020], Gu et al. [2020], Kozak et al. [2020], Bali et al. [2020], Bianchi et al. [2021], and Huang et al. [2021]. Details of these models are given in Appendix B2.

[4]I propose two additional measures of investor disagreement based on the cross-sectional dispersion in machines' return forecast errors and the cross-sectional spread between the maximum and minimum of machines' return forecasts. As will be discussed in Section 5.6, my main findings from these two alternative measures of disagreement are very similar to those obtained from the MFD.

takes a short position in 10% of the stocks with the highest MFD (decile 10) and takes a long position in 10% of the stocks with the lowest MFD (decile 1) yields the one-month-ahead abnormal returns (alphas) of 0.72%, 0.55%, 0.71%, and 0.67% per month, estimated, respectively, with the six-factor model of Fama and French [2018], the mispricing factor model of Stambaugh and Yuan [2017], the Q-factor model of Hou et al. [2015], and the behavioral factor model of Daniel et al. [2020]. All alphas are significant at the 1% level, except for the mispricing factor model with the 5% level of significance. I also examine the long-term predictive power of MFD and find that the negative relation between the machine forecast disagreement and future equity returns is not just a one-month affair. The MFD predicts cross-sectional variation in equity returns six months into the future. Finally, I find corroborating evidence on the significance of MFD from bivariate portfolio-level analyses and multivariate Fama–MacBeth regressions when I control for a number of firm characteristics and risk factors, including firm's size, book-to-market ratio, gross profitability, asset growth, momentum, short-term reversal, earnings surprise, illiquidity, turnover ratio, idiosyncratic volatility, lottery demand, dispersion in analysts' earnings estimates, and dispersion in analysts' long-term growth forecasts.

I investigate the source of the significant alpha spread between the high-MFD and low-MFD portfolios and find that the machine forecast disagreement premium is driven by underperformance of high-MFD stocks, but not due to outperformance of low-MFD stocks, as the alphas on high-MFD stocks (decile 10) are negative, economically large and highly significant, whereas the alphas on low-MFD stocks (decile 1) are positive, but economically small and statistically insignificant. Consistent with Miller [1977] model, my results support a mispricing-based explanation of the disagreement premium. Using on the stock-level mispricing (MISP) measure of Stambaugh et al. [2015], I find that high-MFD stocks, on average, have a significantly higher mispricing score than low-MFD stocks, implying that stocks with high MFD are subject to significant overpricing. Moreover, I form value-weighted bivariate portfolios of stocks sorted by MFD and MISP, and find that the negative alpha spread on MFD-sorted portfolios is much stronger, both economically and statistically, for overpriced stocks, compared to underpriced and fairly priced stocks.

I also examine the predictive power of the MFD during the periods with and without earnings announcements. I find that the long-short excess returns and alphas on MFD-sorted portfolios in earnings announcement periods are almost three times higher than the long-short excess returns and alphas in non-earnings announcement periods. Thus, the evidence supports the mispricing argument that investors do not fully incorporate the MFD-driven return predictability information into their earnings forecasts and are therefore surprised when earnings are realized.

In addition to the mispricing-based explanation, I provide complementary economic mechanisms along the lines of informational frictions, limits to arbitrage, and investor inattention. Hong and Stein [1999] propose a theoretical model in which gradual diffusion of information among investors explains the observed predictability of stock returns. In their model, at least some investors can process only a subset of publicly available information because they have either limited data-processing capabilities or limited computing resources. Moreover, using publicly available data, processing all possible complex forecasting models (e.g., machine learning algorithms) and implementing a suggested investment strategy may be costly (Hirshleifer and Teoh [2003]), and there are limits to arbitrage (Shleifer and Vishny [1997]). Due to informational frictions, costly arbitrage, and/or investors' limited attention, new informative signals are incorporated into stock prices partially because at least some investors do not adjust their demand by recovering informative signals from firm fundamentals or observed prices. As a result of under- or delayed-reaction to information by some investors, stock returns exhibit predictability.

Thus, I hypothesize that the return predictability is concentrated in stocks that are largely held by retail investors, that receive less investor attention, and that are costlier to arbitrage. Earlier studies show that less sophisticated individual investors have more limited attention and hence I argue that the informative signals provided by the MFD for stocks largely held by retail investors are not incorporated into prices quickly. However, sophisticated institutional investors, who are able to detect and process information generated by the MFD, can take advantage of mispricing in these stocks so that the information produced by the MFD will be promptly incorporated into stock prices. Since the information is integrated into the prices much faster in the presence of informed investors, there is little room for predictability among stocks with high institutional ownership and low arbitrage costs. Thus, the slow diffusion of information and the resulting return predictability should be more pronounced for stocks with low institutional ownership, low investor attention characteristics, and high arbitrage costs.

To provide a better understanding of the economic mechanisms behind the return predictability, I test whether the predictive power of the MFD is explained by investors' sophistication, informational frictions, investors' limited attention, and/or limits to arbitrage. I find that the disagreement premium is much stronger, both economically and statistically, for stocks predominantly held by retail investors, whereas the return predictability is much weaker for stocks largely held by institutional investors, confirming that investors' sophistication and informational frictions play a significant role in the predictive power of the machine forecast disagreement. When testing the significance of attention mechanism, I find that the abnormal returns on stocks with low investor attention are indeed larger than the abnormal returns on stocks with high investor attention, where the proxies of investor attention are the institutional ownership, analyst coverage, and absolute earnings surprise.[5] I also investigate the impact of costly arbitrage and find that the abnormal returns on stocks with high arbitrage costs are greater than the abnormal returns on stocks with low arbitrage costs, where the proxies for limits-to-arbitrage include the idiosyncratic volatility, illiquidity, and market capitalization. Overall, my results indicate that the MFD-based return predictability is likely due to informational frictions, investors' limited attention, and limits to arbitrage.

The remainder of the paper is organized as follows. Section 2 provides a brief literature review and outlines the contribution. Section 3 describes the data and variables. Section 4 evaluates the empirical performance of machine learning models. Section 5 presents the main empirical results on the cross-sectional equity return predictability. Section 6 investigates the sources of return predictability. Section 7 further distinguishes between risk and mispricing based explanations. Section 8 concludes the paper.

2 Related Literature

This paper is related to the extensive literature on heterogeneous beliefs. Since early contributions of Miller [1977] and Harrison and Kreps [1978], heterogeneous belief models have been frequently used to investigate a number of crucial issues in financial markets, including financial bubbles and differences in the beliefs of equity market investors (e.g., Diether et al. [2002], Chen et al. [2002], Hong and Stein [2003], and Scheinkman and Xiong [2003]). This study contributes to the literature on heterogeneous beliefs by introducing a novel, objective measure of investor disagreement based on machines' expected return forecasts and show that this newly proposed measure, MFD, significantly predicts the cross-sectional variation in future equity returns. It also demonstrates that the predictive power of MFD is distinct from the existing measures of disagreement and is not explained by established return predictors, such as size, value, momentum, investment, profitability, liquidity, volatility, and dispersion in analysts' earnings estimates.

[5] See, e.g., Hou and Moskowitz [2005], Peng and Xiong [2006], Hong et al. [2007], Cohen and Frazzini [2008], Fang and Peress [2009], Hirshleifer et al. [2009], Da et al. [2011], Gervais et al. [2001], Hirshleifer et al. [2013], Bali et al. [2014], Hirshleifer et al. [2018], and Han et al. [2021].

Chen et al. [2002] show that when breadth of equity ownership is low (i.e., when few investors have long positions), the short-sale constraint is binding tightly, so prices are high relative to fundamentals and lower breadth should predict lower future returns. Diether et al. [2002] use dispersion in analysts' earnings forecasts as a proxy for divergence of opinion and find that stocks with higher dispersion in analysts' earnings forecasts generate significantly lower future returns than those with lower dispersion. Johnson [2004] questions the interpretation of Diether et al. [2002] results, and argues that dispersion in analysts' earnings estimates can proxy for higher firm-specific idiosyncratic risk and attributes it to higher leverage. Anderson et al. [2005] find that dispersion in analyst long-term growth forecasts shows a negative and statistically significant predictive power on future stock returns. Boehme et al. [2006] examine the significance of simultaneous effects of differences of opinion and short-sale constraints in the cross-sectional pricing of individual stocks and provide evidence of significant overvaluation for stocks that are subject to both conditions simultaneously.

Atmaz and Basak [2018] show that, in equilibrium, disagreement affects stock returns via two channels. The first channel is a direct effect: disagreement represents uncertainty and investors require a higher expected return to hold a stock when disagreement on the stock increases, suggesting a positive disagreement-return relation. The second channel is an indirect effect: investor disagreement affects stock returns via an amplification effect on the average bias. That is, higher disagreement leads to higher average bias and more overvaluation, thereby suggesting a negative disagreement-return relation. With these two channels, Atmaz and Basak [2018] reconcile the mixed disagreement-return relation documented in the literature.[6] Since investors, regardless of whether they are sophisticated or not, are generally upward biased,[7] the second channel is more likely to dominate the first channel, thereby explaining why my newly proposed measure of disagreement (MFD) negatively predicts the cross-section of future equity returns.

This paper also extends the literature on the usage of machine learning techniques in empirical asset pricing. Kelly et al. [2019] apply instrumented principal component analysis to model the cross-section of returns which allows for latent factors and time-varying loadings. Gu et al. [2020] perform a comparative analysis of machine learning methods to measure equity risk premium based on a large set of stock characteristics. Neuhierl et al. [2021] use a large number of firm and option characteristics to predict the cross-section of future stock returns. Kozak et al. [2020] impose an

[6] See, e.g., Chen et al. [2002]; Diether et al. [2002]; Yu [2011]; Carlin et al. [2014]

[7] See, e.g., Barber and Odean [2008]; Edelen et al. [2016]; DeVault et al. [2019]; Engelberg et al. [2020]

economically motivated prior on stochastic discount factor coefficients that shrinks contributions of low-variance principal components for the cross-section of stock returns and Chen et al. [2020] add to these insights using deep neural networks to estimate an asset pricing model for individual stock returns. Martin and Nagel [2021] show that asset returns may appear predictable in-sample when analyzing the economy ex-post and stress the importance of out-of-sample tests. Feng et al. [2020] propose a new model selection method which accounts for model selection mistakes that produce a bias due to omitted variables, and Lettau and Pelger [2020] construct a new estimator that generalizes principle component analysis by including a penalty on the pricing error in expected returns. A nonparametric method to dissect characteristics based on the adaptive group Lasso is proposed by Freyberger et al. [2020]. Giglio et al. [2021] perform "thousands of alpha tests" to develop a new framework to rigorously perform multiple hypothesis testing in linear asset pricing models.[8]

Although the majority of articles applies machine learning models to predict the cross-section of individual stock returns, they have so far focused on the economic and statistical significance of expected return forecasts. This study contributes to the literature by focusing on the *dispersion* in expected return forecasts generated by alternative machine learning models. I propose a novel measure of divergence of opinion among investors based on the cross-sectional dispersion in machines' return forecasts, which is free from behavioral biases and conflicts of interest that can be observed in the existing measures of disagreement (e.g., standard deviation of analysts' earnings estimates). This paper not only shows that machine learning models significantly disagree on their future return forecasts, but the degree of disagreement also varies across small vs. big stocks, liquid vs. illiquid stocks, stocks with high vs. low institutional ownership, analyst coverage, and so on. More importantly, I present the first empirical evidence that the machine forecast disagreement predicts future equity returns and I provide the economic underpinnings of this disagreement premium as well.

[8]Recent research also expands the application of machine learning models for the prediction of other asset classes. Kelly et al. [2020] propose a conditional factor model for corporate bond returns resting on instrumented principal component analysis. Bali et al. [2020] find that machine learning models substantially improve the out-of-sample performance of stock and bond characteristics when predicting the cross-section of corporate bond returns. Bianchi et al. [2021] apply similar techniques to Treasury securities, whereas Filippou et al. [2020] employ them in the context of exchange rates. DeMiguel et al. [2021] show that machine learning helps to select future outperforming mutual funds and Wu et al. [2021] establish similar conclusions for hedge funds. Finally, Li and Rossi [2020] apply machine learning to select mutual funds on the basis of their exposure to a large set of various stock characteristics. Goyenko and Zhang [2020] and Bali et al. [2021] use machine learning models to predict the cross-section of individual option returns.

3 Data and Variables

My sample includes common stocks trading at the NYSE, AMEX, and NASDAQ with a market value recorded at the Center for Research in Security Prices (CRSP) in the previous month and a non-missing value for common equity in the firm's annual financial statement. I obtain monthly stock returns from CRSP and accounting information from Compustat. I exclude financial and utilities firms. To reduce the effect of small and illiquid stocks, I also exclude the low-priced stocks trading below $5 per share.[9] I follow Shumway [1997] to adjust stock returns for delisting. Specifically, if a delisting return is missing and the delisting event is performance-related, I set the delisting return at −30%. Using the data available at CRSP, Compustat, I/B/E/S, and researchers' websites, I construct a total of 310 stock characteristics. The stock characteristics are obtained from Hou et al. [2015], Haugen and Baker [1996], Lewellen [2015], Harvey et al. [2016], McLean and Pontiff [2016], Green et al. [2017], Freyberger et al. [2020], and Han et al. [2020], among others. Similarly to Hou et al. [2015], Harvey et al. [2016], McLean and Pontiff [2016], Freyberger et al. [2020], and Han et al. [2020], I categorize the individual stock characteristics into six broader subgroups: value, momentum, investment, profitability, intangibles, and frictions. The characteristics and subgroups are listed in Section B of the online appendix.

My sample covers the period from July 1966 to December 2019. I use 10 years (first six years as the training sample and subsequent four years as the validation sample) as the rolling window for estimating the parameters and tuning the hyperparameters of machine learning models, so that I compute the month-t MFD – using the month $t-1$ characteristics – as the cross-sectional standard deviation of the machines' return forecasts scaled by the absolute value of the cross-sectional mean of the machines' return forecasts, and then I conduct out-of-sample cross-sectional asset pricing tests for the period July 1976 – December 2019. Since the analysts' earnings forecast (long-term growth forecast) data are available from the beginning of 1976 (1982), the asset pricing tests that include dispersion in analysts' earnings forecasts (long-term growth forecasts) also start in July 1976 (July 1982). The asset pricing tests are conducted using a total of 2,085,442 firm-level return observations spanning the period from July 1976 to December 2019.

For each month, I winsorize all stock characteristics at the 1st and 99th cross-sectional percentiles. Some characteristics have missing values for some months. I require at least 20% of the

[9] To reduce the impact of micro-cap firms, as a further robustness check, I exclude firms with market capitalization below the 20th percentile of the NYSE size breakpoints. As will be discussed later in the paper, my main findings remain intact from alternative stock samples.

characteristic observations for that month to be non-missing. When a characteristic is included, I fill in any missing values for the month with the cross-sectional mean (median for indicator variables) of the available observations. Without loss of generality, before estimating the forecasting models, I standardize each characteristic to have a monthly cross-sectional mean of zero and standard deviation of one.

In the cross-sectional regression analysis, I control for other firm characteristics that have been shown to predict future returns. Specifically, SIZE is the firm's market capitalization computed as the logarithm of the market value of the firm's outstanding equity at the end of month $t-1$. BM is the logarithm of the firm's book value of equity divided by its market capitalization, where the BM ratio is computed following Fama and French [2008]. Firms with negative book values are excluded from the analysis. Short-term reversal (STR) is the stock's one-month lagged return. MOM is the stock's cumulative return from month $t-12$ to month $t-2$ (skipping the STR month), following Jegadeesh and Titman [1993]. Gross Profitability (GP) is the firm's gross profitability, defined as revenue minus cost of goods sold scaled by total assets, following Novy-Marx [2013]. Asset Growth (AG) is defined as the percent growth rate of total assets between two consecutive fiscal years, following Cooper et al. [2008]. The monthly turnover ratio (TURN) is calculated as the number of shares traded divided by the total number of shares outstanding in month $t-1$. The monthly illiquidity (ILLIQ) measure is computed as the absolute daily return divided by daily dollar trading volume, averaged in month $t-1$, following Amihud [2002]. The idiosyncratic volatility (IVOL) is defined as the standard deviation of daily residuals estimated from the regression of daily excess stock returns on the daily market, size, and value factors of Fama and French (1993) in month $t-1$, following Ang et al. [2006]. The standardized unexpected earnings (SUE) is defined as the actual earnings in the current quarter minus earnings four quarters ago, scaled by stock price in the current quarter, following Livnat and Mendenhall [2006]. The lottery payoff is proxied by the average of the five highest daily returns (MAX) in month $t-1$, following Bali et al. [2011]. DAE is the standard deviation of analysts' earnings forecasts scaled by the absolute value of the mean analysts' earnings forecasts, following Diether et al. [2002]. DALG is the standard deviation of the analysts' long-term earnings growth rate forecasts, following Anderson et al. [2005].

Panel A of Table 1 presents descriptive statistics for the main cross-sectional variables. Concerning my key variable of interest, the sample mean of the MFD is 2.49 with a standard deviation of 1.28. The median of the MFD is 2.04 – somewhat lower than the mean – indicating a positively skewed MFD distribution with the minimum and maximum values of 0.07 and 5.71, respectively.

Panel B of Table 1 includes the panel-level Pearson (Spearman) correlations below (above) the diagonal. The first column and the first row report a negative relation between the MFD and one-month-ahead returns, $R_{i,t+1}$. Consistent with the literature, I find a negative relation between one-month-ahead returns and size, asset growth, past one-month return, idiosyncratic volatility, lottery payoff, and analyst dispersion, whereas there is a positive relation between one-month-ahead returns and gross profitability, book-to-market ratio, past 12-month return, illiquidity, and earnings surprise. These correlation statistics confirm the presence of size, value, momentum, investment, profitability, short-term reversal, earnings momentum, illiquidity, investor disagreement, idiosyncratic volatility, and lottery demand effects in my sample.

The second column and the second row in Panel B of Table 1 show that smaller and less liquid stocks with higher idiosyncratic volatility and higher analyst disagreement also have higher MFD. The positive correlations between the MFD and the existing measures of analysts' earnings forecasts (DAE and DALG) indicate that the machine forecast disagreement is a sensible measure of divergence of opinion among investors about stock value. The positive (negative) correlation of the MFD with idiosyncratic volatility (size) also suggests that the machine forecast disagreement is a reasonably proxy for information uncertainty (see, e.g., Zhang [2006]).

Based on these results, I conjecture that the MFD can be viewed as a measure of "firm-specific uncertainty" as higher MFD implies higher information uncertainty or higher investor disagreement about the firm value. Thus, when I aggregate a firm-level uncertainty measure, I'm supposed to obtain a proxy for market-level uncertainty if the MFD truly captures some form of uncertainty, ambiguity, or disagreement about the firm value. To test this hypothesis, I construct an aggregate measure of the MFD by calculating the value-weighted and equal-weighted average of the stock-level MFD measures. Figure 1 displays the monthly time-series plot of the aggregate MFD measures along with the NBER recession dates. The MFD indices in Figure 1 are standardized to have a zero mean and unit standard deviation. A notable point in Figure 1 is that the value-weighted and equal-weighted MFD measures are highly correlated; the correlation is 95% over the sample period July 1976–December 2019. Another notable point in Figure 1 is that the aggregate MFD indices are generally higher during bad states of the economy and financial market downturns. I also calculate the correlations between the aggregate MFD and the existing measures of market-level or economy-wide uncertainty; the financial and macroeconomic uncertainty measures of Jurado et al. [2015], the economic policy uncertainty measure of Baker et al. [2016], and the S&P100 option implied volatility index (VXO). As reported in Panel C of Table 1, the correlation between the macroeconomic uncertainty measure of Jurado et al. [2015] and the equal-weighted (value-weighted) MFD index is 0.52 (0.35). The correlation between the financial uncertainty measure of Jurado et al. [2015] and the equal-weighted (value-weighted) MFD index is 0.53 (0.48). The correlations between the aggregate MFD and the economic policy uncertainty measure of Baker et al. [2016] and the VXO are roughly 0.50. These results indicate that the aggregate MFD index is a sound measure of market-level uncertainty.

4 Empirical Performance of Machine Learning Models

Following Gu et al. [2020], I assess the empirical performance of a variety of machine learning models, including principal components analysis (PCA), Scaled PCA (SPCA), Partial Least squares (PLS), LASSO, Ridge, Elastic Net (E-Net), random forests (RF), Gradient Boosted Regression Tree (GBRT), Extreme Tree (ET), Feed Forward Neural Network (FNN), Recurrent Neural Network (RNN), and Long Short-Term Memory Neural Network (LSTM).

4.1 Evaluating the Out-of-sample Performance of Machine Learning Models

Following Gu et al. [2020], I use the out-of-sample R-squared as the performance metric to assess the predictive power of individual stock return predictors,

$$R^2_{OS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} \left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}} \tag{1}$$

The R^2_{OS} statistic pools prediction errors across stocks and over time into a grand panel-level assessment of each model, and it measures the proportional reduction in mean squared forecast error (MSFE) for each model relative to a naive forecast of zero excess return benchmark, which assumes that the one-month-ahead expected return on stocks equals the time $t+1$ risk-free rate (see Gu et al. [2020]). To estimate the out-of-sample R^2_{OS}, I follow the commonly used approach in the literature and use 10 years (first six years as the training sample $\mathcal{T}_1$ and subsequent four years as the validation sample $\mathcal{T}_2$) as the rolling window for estimating the parameters and tuning the hyperparameters of machine learning forecasting models. The "test" subsample (from July 1976 to December 2019, $\mathcal{T}_3$) is used to evaluate a model's out-of-sample forecasting performance.

I use the mean squared forecast error (MSFE)-adjusted statistic of Clark and West (2007) to test the statistical significance of R^2_{OS}. Considering the potentially strong cross-sectional dependence among individual excess stock returns, I employ the modified MSFE-adjusted statistic based on the cross-sectional average of prediction errors from each model instead of prediction errors among individual returns. The p-value from the MSFE-adjusted statistic tests the null hypothesis that the MSFE of a naive forecast of zero excess return is less than or equal to the MSFE of a machine learning model against the one-sided (upper-tail) alternative hypothesis that the MSFE of a naive forecast of zero excess return is greater than the MSFE of a machine learning model (H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$).

Table OA1 presents the monthly R^2_{OS} (in percentage) for the entire pooled sample of stocks using all the 310 stock characteristics as covariates. Compared to the dimension reduction models, regression trees, and neural network models, the penalized linear regressions – LASSO, Ridge, and E-Net – perform relatively poorly with the respective R^2_{OS} values of 0.13%, 0.08%, and 0.13%. By forming a few linear combinations of predictors via dimension reduction, PCA, SPCA, and PLS improve the R^2_{OS} to 0.37%, 0.48%, and 0.46%, respectively. Unlike the penalized linear models, regression trees are fully nonparametric and can reduce overfitting in individual bootstrap samples, and make the predictive performance more stable. Consistent with this prediction, RF, GBRT, and ET show a significant increase in R^2_{OS} to 0.51%, 0,54%, and 0.56% per month, respectively. In addition to nonparametric regressions, I investigate the performance of different neural network models, including feed forward neural network (FFN), recurrent neural network (RNN), and long short-term memory neural network (LSTM). As a typical neural network, the feed forward neural network (FFN) produces more flexible prediction approach by adding hidden layers between the inputs and output layer that aggregates hidden layers into the outcome prediction. The recurrent neural network (RNN) processes sequential data such as stock prices and returns. The long short-term memory neural network (LSTM) captures long-term dependencies as a flexible hidden state space model for a large dimensional system. FFN, RNN, and LSTM model produce the largest R^2_{OS} values of 0.57%, 0.58%, and 0.59% per month, compared to the other machine learning models.

4.2 Diebold and Mariano (1995) Test

To compare the out-of-sample predictive power of two methods, I use the modified Diebold and Mariano (1995) test, which accounts for the potentially strong cross-sectional dependence among individual returns. Specifically, the modified Diebold-Mariano statistic is defined as:

$$DM_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}} \tag{2}$$

where $\bar{d}_{12}$ and $\hat{\sigma}_{\bar{d}}$ are, respectively, the time-series mean and Newey-West standard error of $d_{12,t+1}$ over the testing sample. $d_{12,t+1}$ is the forecast error differential between the two methods, (1) and (2), calculated as the cross-sectional average of forecast error differentials from each model over each period $t + 1$,

$$d_{12,t+1} = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_3} \left(\left(\hat{e}_{i,t+1}^{(1)} \right)^2 - \left(\hat{e}_{i,t+1}^{(2)} \right)^2 \right) \tag{3}$$

where $\hat{e}_{i,t+1}^{(1)}$ and $\hat{e}_{i,t+1}^{(2)}$ are the return forecast errors for individual asset i at time t+1 generated by two methods, and $n_{3,t+1}$ is the number of assets in the testing sample.

Table OA2 reports the Diebold-Mariano test statistics for pairwise comparisons of a column model versus a row model. A positive statistic indicates that the column model outperforms the row model. Table OA2 shows that regression trees and neural networks outperform penalized linear regressions, showing positive and statistically significant test statistic with the Diebold-Mariano test statistics ranging from 2.08 to 3.29. Comparing the performance of machine learning methods in the same category, Table OA2 shows that there is little difference in the performance of dimension reduction models (PCA, SPCA, and PLS), penalized linear regressions (LASSO, Ridge, and E-Net), regression trees (RF, GBRT, and ET), and neural networks (FFN, RNN, and LSTM), as the test statistics are not significant. Finally, the last column of Table OA2 shows that the LSTM model produces large and significant statistical improvements over most individual machine learning models.

4.3 Machine Learning Portfolios Constructed Using 310 Stock Characteristics

To further investigate the economic significance of the machine learning models, I form long-short portfolios based on the machine learning expected return forecasts using the 310 stock characteristics. Specifically, I sort stocks into decile portfolios based on each model's forecasts of the one-month-ahead returns and then calculate the one-month-ahead equal-weighted and value-weighted average realized returns of the decile portfolios. Table OA3 of the online appendix reports the monthly performance results. "Low" is the decile portfolio with the lowest one-month-ahead expected return forecast (decile 1), "High" is the decile portfolio with the highest one-month-ahead expected return forecast (decile 10), and "High-Low" denotes the long-short portfolio that buys the highest expected return stocks in decile 10 and sells the lowest expected return stocks in decile 1.

Table OA3 presents the equal-weighted and value-weighted one-month-ahead average realized returns on the long-short portfolios. Consistent with the findings of Gu et al. [2020], Panel A of Table OA3 shows that the equal-weighted portfolios produce positive, economically large and highly significant average return spreads from all machine learning models; in the range of 1.35% and 3.49% per month. The top two best performing hedge portfolios are generated by the Long Short-Term Memory Neural Network (LSTM) and the Recurrent Neural Network (RNN), with the monthly return spread of 3.49% (t-stat.=7.10) and 3.48% (t-stat.=6.72), respectively. Consistent with my earlier findings using R^2_{OS} as the performance metric, the penalized linear regression models perform least well, although the equal-weighted average return spreads obtained from LASSO, Ridge, and E-Net are still very large, ranging from 1.35% (t-stat.=2.62) to 1.60% (t-stat.=3.09).

Panel B of Table OA3 shows that the relative performance of machine learning models remains the same in the value-weighted portfolios which produce economically smaller return spreads than the equal-weighted portfolios, but the return spreads are still very large in the value-weighted portfolios. Again, the best performing hedge portfolios are generated by the neural network models – FFN, RNN, and LSTM – with the respective value-weighted monthly return spreads of 1.86% (t-stat.=3.08), 2.05% (t-stat.=3.39), and 1.89% (t-stat.=3.70). Consistent with my findings from the equal-weighted portfolios, the penalized linear regression models – LASSO, Ridge, and E-Net – perform least well, with the respective value-weighted monthly return spreads of 0.86% (t-stat.=1.47), 0.69% (t-stat.=1.33), and 0.85% (t-stat.=1.40).

Table OA4 of the online appendix reports the Fama and French [2018] six-factor alphas on the long-short portfolios of stocks sorted by the machine learning models' one-month-ahead expected return forecasts. Panel A of Table OA4 shows that controlling for the market, size, value, investment, profitability, and momentum factors of Fama and French [2018] does not reduce much the economic and statistical significance of the return spreads on the equal-weighted portfolios. However, as presented in Panel B of Table OA4, the value-weighted portfolios paint a different picture. Although the relative performance of the 12 machine learning models remain almost identical, the six-factor monthly alpha spreads on the value-weighted portfolios generated by the PLS, LASSO, Ridge, and E-Net models are not statistically significant; ranging from 0.10% (t-stat.=0.75) to 0.31% (t-stat.=0.97). Similar to my earlier findings, the best performing hedge portfolios are again generated by the neural network models – FFN, RNN, and LSTM – with the respective value-weighted monthly alpha spreads of 1.37% (t-stat.=4.90), 1.40% (t-stat.=5.04), and 1.42% (t-stat.=4.70).

5 Empirical Results

In this section, I conduct parametric and nonparametric tests to assess the predictive power of the machine forecast disagreement (MFD) over future stock returns. First, I present results from univariate portfolio-level analysis. Second, I investigate the mispricing-based explanation of the disagreement premium. Third, I report average stock characteristics of the MFD-sorted decile portfolios. Fourth, I conduct bivariate portfolio-level analyses to assess the predictive power of the MFD after controlling for well-known stock characteristics and risk factors. Fifth, I present firm-level Fama-MacBeth cross-sectional regression results. Finally, I run a battery of robustness checks using alternative measures of the MFD.

5.1 Univariate portfolio-level analysis

To construct the long-short portfolio, for each month from July 1976 to December 2019, individual stocks are sorted by the MFD into decile portfolios. I then compute the one-month-ahead value- and equal-weighted average excess return of each decile portfolio. To examine the cross-sectional relation between the MFD and future stock returns, I form a long-short portfolio that takes a long position in the lowest decile of MFD and a short position in the highest decile of MFD.

Table 2 reports the average monthly excess returns of each decile portfolio, and the long-short portfolio (in excess of the one-month Treasury bill rate). I also report the abnormal returns (alphas) estimated with various factor models, including the capital asset pricing model (CAPM) with the market (MKT) factor, the six-factor model (FF6) of Fama and French [2018] with the MKT, size (SMB), value (HML), investment (CMA), profitability (RMW), and momentum (MOM) factors, the q4-factor model (HXZ) of Hou et al. [2015] with the MKT, size (SMB_Q), investment (I/A), and profitability (ROE) factors, the mispricing factor model (SY) of Stambaugh and Yuan [2017] with the MKT, SMB, management (MGMT), and performance (PERF) factors, and the behavioral factor model (DHS) of Daniel et al. [2020] with the MKT, post-earnings-announcement drift (PEAD), and financing (FIN) factors. Controlling for these risk and mispricing factors helps to ensure that the MFD indeed contains incremental predictive power beyond these well-known equity return predictors.

In general, the excess returns and alphas of the MFD-sorted portfolios decrease from decile 1 to decile 10. The long-short portfolio that short-sells 10% of the stocks with the highest MFD (decile 10) and buys 10% of the stocks with the lowest MFD (decile 1) earns a value-weighted (equal-weighted) average return of 0.45% (0.60%) per month with a t-statistic of 2.94 (4.04), translating into an annualized return of 5.40% (7.20%).[10] Controlling for the robust risk and mispricing factors does not change the magnitude and statistical significance of the return spreads on the MFD-sorted portfolios for most of the factor models. The only exception is the alpha of the long-short portfolio under the mispricing factor model, where the alpha decreases from 0.69% (CAPM) to 0.55% (SY model) per month and the corresponding t-statistic decreases from 3.36 to 2.54 for the value-weighted portfolio, suggesting that the return predictability is potentially driven by mispricing rather than compensation for risk. Finally, the significant relation between the MFD and future returns is largely coming from the short leg of the arbitrage portfolio as the economic magnitude and statistical significance of the abnormal returns (alphas) are larger among the stocks in the short leg than those in the long leg. This implies that high-MFD firms are overvalued relative to firms with lower MFD, perhaps due to higher firm-specific uncertainty or higher investor disagreement about stock value of high-MFD firms as well as higher arbitrage costs.

Next, I examine the persistence of the rank of MFD and the persistence of the MFD-based return predictability. If the rank of MFD is persistent, investors would be able to learn from the past and I would not be able to detect mispricing over a long period of time; i.e., mispricing and arbitrage opportunity decay over time. Table 3 presents the probability of staying in the same MFD group or moving to any of the other nine MFD groups next year. Specifically, I present the average probability that a stock in decile i (defined by the rows) in month t will be in decile j (defined by the columns) in month $t + 12$. All the probabilities in the transition matrix should be approximately 10% (ten portfolios) if the evolution for MFD for each stock is random and the relative magnitude of MFD in one period has no implication about the relative MFD values next year. However, Table 3 shows that 63% of stocks in the lowest MFD decile (decile 1) in month t continue to be in the same decile in month $t + 12$. Similarly, 66% of the stocks in the highest MFD decile (decile 10) in month t continue to be in the same decile in month $t + 12$. These results indicate that firm-specific uncertainty (or investor disagreement) proxied by the MFD is a highly persistent stock characteristic.

The results from transition matrix reported in Table 3 suggest that investors underprice (overprice) securities with the lowest (highest) MFD in the past with the expectation that this behavior will persist in the future. If the machine forecast disagreement was a characteristic that evolved

[10] The t-statistics reported in my tables are Newey and West [1987] adjusted with six lags to control for heteroskedasticity and autocorrelation.

randomly over the months, I would expect to see no relation between the MFD and future stock returns. The fact that the MFD is persistent and it has a robust, incremental predictive power for the cross-section of future stock returns suggests the possibility that investors underestimate the magnitude of the cross-sectional persistence uncovered in Table 3. I delve further into this possibility in the test of long-term portfolio returns.

I investigate the long-term predictive power of MFD by calculating the Fama-French (2018) six-factor alphas of the MFD-sorted portfolios from two to 12 months after portfolio formation. The results are presented in Table 4. The six-factor alpha spread remains economically large and highly significant during the second, third, and fourth month after portfolio formation, with the respective monthly alpha spreads of -0.56% (t-stat.$=-2.83$), -0.51% (t-stat.$=-2.29$), and -0.47% (t-stat.$=-2.08$). The predictive power of MFD on future returns diminishes as one moves further away from the portfolio formation month and becomes insignificant after the six month. These results show that the negative cross-sectional relation between the MFD and future returns is not just a one-month affair and the underreaction to firm-specific uncertainty (or MFD-driven information) persists several months into the future, which is consistent with the theoretical evidence of Hong and Stein [1999] as a consequence of the gradual diffusion of information into stock prices.

5.2 Testing the mispricing hypothesis

The results so far suggest that high-MFD stocks tend to be overvalued relative to low-MFD stocks, but I have not yet provided any direct empirical evidence that high-MFD stocks are indeed overvalued. I test this hypothesis by computing the mispricing score of the stocks in MFD-sorted portfolios. Specifically, I use the stock-level mispricing measure (MISP) of Stambaugh et al. [2015] to identify if high-MFD stocks are indeed overvalued.[11] I also conduct independent double sorts based on individual stocks' MISP and MFD; that is, stocks are grouped into 10 decile (5 quintile) portfolios based on independent ascending sorts of MFD (MISP) and the intersections of each of the decile MFD and the quintile MISP portfolios are used to form bivariate portfolios. I then compute the Fama and French [2018] six-factor alpha of each of the 50 (5×10) resulting MISP&MFD sorted intersection portfolios.

Table 5, Panel A, shows that the high-MFD stocks indeed have a higher average mispricing

[11] As discussed in Stambaugh et al. [2015], each month, individual stocks are ranked independently based on 11 prominent equity return predictors (net stock issues, composite equity issues, accruals, net operating assets, asset growth, investment-to-assets, distress, O-score, momentum, gross profitability, and return on assets) in such an order that a higher rank is associated with lower one-month-ahead stock returns. The mispricing measure (MISP) is defined as the arithmetic average of the ranks of the 11 return predictors.

score than the low-MFD stocks. Furthermore, as reported in the last column of Panel A, Table 5, the 10-1 difference in the average mispricing scores between the high- and low-MFD stocks is statistically significant at the 1% level with a t-statistic of 3.50. Consistent with Miller [1977], this evidence supports the mispricing argument that the high-MFD stocks with higher firm-specific uncertainty (or greater investor disagreement) are indeed overvalued.

Next, I investigate whether the cross-sectional relation between the MFD and future returns is stronger for overvalued vs. undervalued stocks. Specifically, I calculate the Fama and French [2018] six-factor alphas of each of the 50 (5×10) MISP&MFD sorted value-weighted portfolios. The last two columns in Panel B of Table 5 present the six-factor alpha spreads between the high-MFD and low-MFD decile portfolios within each MISP quintile along with the Newey-West t-statistics in parentheses. A notable point in Table 5, Panel B, is that the alpha spread is highest at -1.30% per month (t-stat.$=-4.89$) for overvalued stocks (high-MISP quintile). Moreover, the alpha spread on MFD-sorted portfolios of overvalued stocks is economically and statistically greater than the alpha spreads on MFD-sorted portfolios of all other stocks – undervalued or (relatively) fairly valued stocks in the MISP quintiles 1 through 4.[12] Overall, the results in Table 5 confirm Miller [1977] hypothesis that the disagreement premium is stronger for overpriced stocks, i.e., stocks with higher investor disagreement are subject to a higher degree of mispricing and lower subsequent return.

5.3 Average portfolio characteristics

I investigate which firm characteristics can potentially explain the negative relation between the MFD and future stock returns. To do so, I sort stocks by the MFD into decile portfolios each month and report the time-series averages of the cross-sectional medians of various firm-specific characteristics for each decile. Table 6 presents the average stock characteristics of each MFD-sorted decile portfolio and the long-short portfolio. The characteristics include the machine forecast disagreement (MFD), log market capitalization (SIZE), log book-to-market ratio (BM), asset growth (AG), gross profitability (GP), medium-term stock momentum (MOM), short-term reversal (STR), illiquidity (ILLIQ), turnover (TURN), standardized unexpected earnings (SUE), idiosyncratic volatility (IVOL), lottery demand (MAX), dispersion in analysts' earnings forecasts (DAE), and dispersion in analysts' long-term growth forecasts (DALG).

[12] The differences between the six-factor alphas of the zero-cost MFD-sorted portfolios for MISP5 vs. the other MISP quintiles turn out to be economically large and statistically significant. As shown in the lower panel of Table 5, Panel B, the t-statistics between the six-factor alpha of MISP5 and the six-factor alphas of MISP1 to MISP4 are -1.92, -3.29, -2.73, and -1.71, respectively.

By construction, the average MFD increases monotonically from decile 1 to decile 10 and the 10-1 difference of the average MFD is economically large at 0.34 and highly significant (t-stat.=247.01), indicating significant cross-sectional variations in the machine forecast disagreement of individual stocks. The last two rows in Table 6 show that as the MFD increases across the deciles, the established measures of investor disagreement – DAE and DALG – increase as well, indicating a positive correlation between the MFD and the existing measures of divergence of opinion. Earlier studies also find that small, illiquid, lottery-like stocks with high idiosyncratic volatility exhibit high information uncertainty (e.g., Zhang [2006], Kumar [2009], and Bali et al. [2011]). Consistent with the literature, Table 6 shows that the stocks with higher MFD (or higher firm-specific uncertainty) are indeed smaller, less liquid, and have higher idiosyncratic volatility and stronger lottery features.

The literature provides clear evidence that the firm-specific attributes considered in Table 6 are instrumental in explaining the cross-section of expected stock returns. Stocks with higher MFD, higher asset growth, lower profitability, lower momentum returns, lower earnings surprise, higher idiosyncratic volatility, and higher MAX tend to have lower future returns. Considering the prior findings in the literature and the patterns that the firm-specific attributes exhibit across the MFD deciles, one may think that investment, profitability, momentum, post-earnings-announcement drift, idiosyncratic volatility, and/or the lottery demand effect drive the significantly negative relation between the MFD and future stock returns. Thus, in the next two sub-sections, I control for these well-known return predictors in bivariate portfolio sorts and firm-level cross-sectional regressions to further test whether the significant relation between the MFD and future stock returns remains intact after accounting for these established, robust firm characteristics.[13]

5.4 Bivariate portfolio-level analysis

In this section, I control for the established equity return predictors using 10x10 dependent and independent double sorts based on various firm characteristics and the MFD. For dependent bivariate sorts, each month, I first sort stocks into decile portfolios based on firm characteristics (control variables). Then, I further sort stocks by the MFD into decile portfolios within each control variable decile. For each first-stage sorting (or control) variable, this bivariate portfolio analysis provides

[13] In Table 2, I have already controlled for the market, size, value, momentum, investment, and profitability factors of Fama and French [2018] and Hou et al. [2015] as well as the mispricing and behavioral factors of Stambaugh and Yuan [2017] and Daniel et al. [2020] constructed based on earnings surprise (post-earnings-announcement drift) and a number of other well-known return predictors. As discussed in Section 5.1, the alpha spreads on MFD-sorted portfolios remain negative and highly significant in both value-weighted and equal-weighted portfolios after controlling for this large set of equity market factors.

100 conditionally double-sorted portfolios. Portfolio 1 (10) is the combined portfolio of stocks with the lowest (highest) MFD in each control variable decile. For bivariate independent sorts, each month, all stocks are grouped into decile portfolios based on independent ascending sorts of both a control variable and the MFD. The intersections of each of the decile portfolios are used to form 100 unconditionally double-sorted portfolios.

Table 7 presents the Fama and French [2018] six-factor alphas on the value-weighted bivariate portfolios. For brevity, I do not report the alphas for all 100 (10x10) portfolios. Instead, I report the abnormal returns on the value-weighted portfolios of MFD averaged across the 10 control variable deciles to produce the MFD-sorted decile portfolios while accounting for the impact of control variables. Panel A (Panel B) reports the six-factor alphas from the dependent (independent) bivariate portfolios. The last row in Panels A and B of Table 7 shows that the cross-sectional relation between the MFD and future returns remains economically large and highly significant after controlling for a large set of well-known return predictors. The six-factor alpha spreads on the value-weighted MFD-sorted portfolios are in the range of -0.62% per month (t-stat.$=-4.14$) and -0.92% per month (t-stat.$=-4.28$) from dependent bivariate sorts and ranging from -0.57% per month (t-stat.$=-3.08$) to -0.78% per month (t-stat.$=-3.74$) from independent bivariate sorts. These results indicate that even after controlling for various firm characteristics and risk factors in bivariate portfolios, there is a strong negative relation between the MFD and future equity returns. In other words, the predictive power of MFD is not explained by other cross-sectional return predictors, including the existing measures of investor disagreement.

Another notable point in Table 7 is that, even after controlling for these robust, most prominent return predictors, the significant alpha spread on MFD-sorted portfolios remains to be driven by underperformance of high-MFD stocks, but not due to outperformance of low-MFD stocks, as the six-factor alphas on the high-MFD portfolio are negative, economically large and highly significant, whereas the six-factor alphas on the low-MFD portfolio are statistically insignificant. This finding is observed for all control variables without exception and provides further support for the mispricing-based explanation of the disagreement premium.

5.5 Fama-MacBeth cross-sectional regressions

In this section, I conduct firm-level Fama-MacBeth regression analysis to test if the MFD predicts the cross-section of future stock returns while controlling for other known predictors simultaneously. Each month, I run a cross-sectional regression of stock returns in that month on the past MFD as well as a number of control variables, including the lagged size, book-to-market, momentum, gross profitability, asset growth, earnings surprise, short-term return reversal, illiquidity, turnover ratio, idiosyncratic volatility, and lottery demand. I also control for the dispersion in analysts' earnings forecasts and the dispersion in analysts' long-term growth forecasts, as the predictive power of MFD may be correlated with two exsiting measures of analysts' disagreement. I also control for the industry fixed effects following the 48-industry classification scheme of Fama and French [1997]. The stock-level cross-sectional regressions are run each month and the standard errors of the average slope coefficients are corrected for heteroskedasticity and autocorrelation following Newey and West [1987].

Table 8 reports the results of stock-level Fama-MacBeth regressions. I control for the industry fixed effect using Fama-French 48-industry classifications in all columns, except column (3). In column (1), I include the MFD as well as other well-known return predictors in the cross-sectional regressions. Consistent with the portfolio results, I find a negative and significant relation between the MFD and one-month-ahead returns controlling for a large number of predictors. The average slope coefficient on the MFD is -1.45 with a t-statistic of -2.78. The spread in the average standardized MFD between deciles 10 and 1 is approximately 0.30, and multiplying this spread by the average slope of -1.45 yields an economically significant return difference of -0.44% per month, controlling for all else. In most cases, the slope coefficients on the control variables are consistent with prior literature; the short term reversal (STR), asset growth (AG), and MAX are negatively correlated with the future return, and momentum (MOM), gross profitability, and earnings surprise (SUE) are positively related to the next month's return. In addition, the coefficient on the dispersion in analysts' earnings forecasts is negative but insignificant, indicating that the MFD subsumes the cross-sectional predictive power of the dispersion in analysts' earnings forecasts. When I independently test the predictive power of the dispersion in analysts' earnings forecast measure, I find that its coefficient is negative and significant at the 5% level, consistent with earlier studies. In column (2), I further control for the dispersion in analysts' long-term growth forecasts. The MFD retains its significant predictive power, although the magnitude of the average slope coefficient somewhat reduces to -1.40, corresponding to the disagreement premium of -0.42% per month.

In column (3), I include $INDRET_{t+1}$, which is computed as the value-weighted Fama-French 48-industry portfolio returns, as a control variable in my main regression to account for the industry effect. Specifically, I adjust the dependent variable by subtracting the firm's value-weighted Fama-

French 48-industry return $INDRET_{t+1}$ from the firm's current month return. Doing so allows us to tease out the return predictive power from the MFD rather than the one-month industry momentum effect (Moskowitz and Grinblatt [1999]). The coefficient of the MFD remains similar after controlling for the industry return directly. In column (4), I further control for the common characteristics that are shown to affect stock returns systematically. Specifically, I follow Daniel et al. [1997] and compute the characteristics-adjusted returns as the difference between the firm's return and the corresponding DGTW benchmark portfolio returns. I replace the firm's raw return with this characteristics-adjusted return as the dependent variable and run the same set of monthly cross-sectional regressions. Again, the magnitude of the slope coefficient on MFD becomes slightly weaker, but it remains highly significant.

Overall, these results indicate that the MFD provides incrementally value-relevant information. The predictive power of the MFD is distinct and robust to the inclusion of other well-known return predictors and the existing measures of investor disagreement.

5.6 Robustness Check

My key variable of interest, the machine forecast disagreement (MFD), is defined as the cross-sectional standard deviation of machines' return forecasts scaled by the absolute value of the mean machines' return forecasts obtained from the 12 machine learning models described in Section 4. In this section, I introduce two auxiliary measures of investor disagreement and investigate their performance in predicting future equity returns.

Specifically, I propose an alternative measure of divergence of opinion as the ratio of the cross-sectional standard deviation of machines' return *forecast errors* to the absolute value of the mean machines' return forecasts, denoted by MFED. Given that *Realized Return* = *Predicted Return* + *Forecast Error*, the MFED is constructed based on the cross-sectional dispersion in forecast errors obtained from the 12 machine learning models. Once I compute the MFED for each stock and month in my sample, I form the long-short equity portfolios of stocks sorted by the MFED and present the excess returns and alphas on the MFED-sorted portfolios in Table OA5 of the online appendix. Similar to my original findings from the MFD (reported in Table 2), Table OA5 shows an economically and statistically significant relation between the MFED and future stock returns. Controlling for the robust risk and behavioral factors does not change the magnitude and statistical significance of the return spreads on the MFED-sorted portfolios for most of the factor models. The only exception is the alpha of the long-short portfolio under the mispricing factor model, where the alpha decreases from 0.55% (CAPM) to 0.43% (SY model) per month and the corresponding t-statistic decreases from 2.64 to 1.91 for the value-weighted portfolio, suggesting that the return predictability is largely driven by mispricing rather than compensation for risk.

As a further robustness check, I develop another measure of investor disagreement (or firm-specific uncertainty) as the ratio of the cross-sectional difference between the maximum and minimum of machines' return forecasts to the absolute value of the mean machines' return forecasts. This alternative measure, denoted by MAX–MIN, captures the largest spread in the cross-sectional distribution of expected return forecasts obtained from the 12 machine learning models and can be viewed as a proxy for firm-specific uncertainty. Since a larger spread in the cross-sectional distribution of return forecasts signals higher information uncertainty about the firm, stocks with a higher MAX–MIN spread are anticipated to have lower future returns. To test this hypothesis, I construct the long-short portfolios of stocks sorted by the MAX–MIN and report the excess returns and alphas in Table OA6 of the online appendix. Similar to my earlier findings from the MFD and MFED, Table OA6 presents a highly significant relation between the MAX–MIN and future stock returns. Controlling for the established factor models does not change the significance of the return spreads on the MAX–MIN sorted portfolios. Again, the only exception is the alpha of the long-short portfolio under the mispricing factor model, where the alpha decreases from 0.49% (CAPM) to 0.38% (SY model) per month and the corresponding t-statistic decreases from 2.31 to 1.78 for the value-weighted portfolio, confirming that the predictive power of the MAX–MIN spread, proxying for information uncertainty, is mainly driven by mispricing.

6 Sources of return predictability

Having established that the predictive power of the MFD may be driven by slow dissemination of disagreement-related information due to investors' underreaction, I seek to understand the economic underpinnings of my main finding based on investors' sophistication, informational frictions, investors' limited attention, and limits to arbitrage.

6.1 Informational frictions

Market reactions to firm-specific uncertainty or information uncertainty about individual stocks can generate important insights on how the market processes divergence of opinion that may influence the information efficiency of the equity market. I conjecture that time-series and cross-sectional variations in firm-specific uncertainty, information uncertainty, and investor disagreement that generate low vs. high MFD are harder to interpret by average investors, compared to the direct and well-defined information events studied in the literature. Thus, consistent with Hirshleifer et al. [2013] who emphasize that investors would have more difficulty in processing information that is less tangible, I conjecture that the elusive nature of the MFD thus makes investors face more severe informational frictions. As a result, the stock market can underreact to my uncertainty/disagreement proxy, and the informative signals provided by the MFD for stocks largely held by retail investors are not incorporated into prices quickly. On the other hand, sophisticated institutional investors, who are able to detect and process information generated by the MFD, can take advantage of mispricing in these stocks so that the information produced by the MFD will be promptly incorporated into stock prices. Since the information is integrated into the prices much faster in the presence of informed investors, there is little room for predictability among stocks with high institutional ownership. Thus, the slow diffusion of information and the resulting return predictability should be more pronounced for stocks with low institutional ownership.

In this section, I investigate the differing levels of institutional ownership among stocks with high and low MFD.[14] I test two hypotheses. First, I investigate whether the level of institutional ownership is lower for high-MFD stocks and are more likely to earn negative returns in the next month. This would be true if institutional investors were better able to capture the persistence in MFD and shied away from those stocks which have experienced recent negative returns. Second, I test whether the magnitude of the negative relation between the MFD and future returns is larger for those stocks in which retail investors are more active compared to those stocks in which institutional investors are more active.

In Panel A of Table 9, I present the time-series averages of cross-sectional means for percentage institutional ownership (INST) for equity deciles formed via a univariate sort based on the MFD. The results show that equities with higher MFD are more likely to be held by individual investors. The percentage institutional ownership is equal to 34% for decile 1. In contrast, for decile 10 which includes the equities with the highest MFD, the percentage institutional ownership drops to 26%. The difference in institutional holdings between the extreme MFD deciles is highly significant with a t-statistic of 6.18.

[14] Institutional holdings data are obtained from Thompson-Reuters' Institutional Holdings (13F) database. To measure a stock's institutional holdings (INST), I define month-t INST to be the fraction of total shares outstanding that are owned by institutional investors as of the end of the last fiscal quarter during or before month t. Values of INST are available for the period from January 1980 to December 2019.

Next, I analyze the strength of the disagreement premium across institutional ownership portfolios using a dependent double sort analysis. Specifically, I first sort stocks into quintile portfolios every month based on the level of institutional ownership. Then, I divide each institutional ownership quintile into deciles based on the MFD. In Panel B of Table 9, I present the Fama-French (2018) six-factor alpha for each of the 50 (5×10) resulting INST&MFD sorted portfolios as well as the six-factor alpha spread between the extreme MFD deciles, and associated t-statistics. A notable point in Table 9, Panel B, is that the magnitude of the abnormal return (six-factor alpha) to the zero-cost portfolio that buys stocks with the highest MFD and sells stocks with the lowest MFD increases monotonically in absolute value as one moves towards the stocks for which the level of institutional holdings is lowest (INST1). For stocks in which institutional investors are most active (INST5), the six-factor alpha to the zero-cost portfolio is negative but economically and statistically insignificant; -0.19% per month (t-stat.$=-1.56$), whereas the corresponding alpha spread on MFD-sorted portfolios is much higher at -1.23% (t-stat.$=-4.72$) for stocks in which retail investors are most active (INST1). The diff-and-diff analysis of the six-factor alpha spreads of the stocks with high vs. low institutional holdings also generates an economically and statistically significant difference. Specifically, the difference between the six-factor alphas of the zero-cost MFD-sorted portfolios among the extreme institutional ownership quintiles (INST5 – INST1) is 1.04% with a t-statistic of 3.27. Collectively, these results indicate that the disagreement premium is much stronger for equities with high informational frictions or equities that are held by less informed, retail investors.

6.2 Investors' limited attention

Barber and Odean [2008] argue that individual investors can only process limited investment choices due to limited time and resources they have. Hirshleifer et al. [2009] find that investors' underreaction to earning surprises and post-earnings-announcement drift are stronger for firms that announce earnings on days that many other firms announce earnings due to investors' limited attention. Cohen and Frazzini [2008] provide evidence that suppliers' have delayed responses to the information disclosure of their customers. Hirshleifer et al. [2013] emphasize that investors would have more difficulty in processing information that is less tangible, so the elusive nature of the MFD makes the investors' attention constraints more likely to be binding. These constraints would be even more binding for retail investors who are more active in high-MFD stocks, compared to institutional investors. Moreover, as indicated in the model of Peng and Xiong [2006], an investor who optimizes the amount of attention would allocate more attention to systematic shocks and less to firm-specific shocks. Thus, a case can be made for under- or delayed-reaction to firm-specific uncertainty based on theories of investor attention.

Following the aforementioned articles, I argue that investors may pay limited attention to firm-specific uncertainty or investor disagreement proxied by the MFD. Following earlier studies, I use three proxies of investor attention; (*i*) institutional ownership, INST, (*ii*) analyst coverage, CVRG, and (*iii*) absolute earnings surprise, abs(SUE). Institutional ownership and analyst coverage are commonly used in the literature as proxies for investor attention.[15] Bali et al. [2018] show that firms with greater absolute earnings surprises are more likely to attract investor attention, increasing investor awareness of firms' specific characteristics.[16] Therefore, firms with lower institutional ownership, lower analyst coverage, or lower absolute SUE receive less attention from investors and should exhibit more sluggish stock price reactions to the information contained in MFD and greater predictability of stock returns.

I test the investor attention hypothesis by dividing my sample into subgroups based on an investor attention proxy and investigate whether the predictive power of MFD is stronger (weaker) for stocks that receive less (more) investor attention. Specifically, I first sort stocks into tercile portfolios every month based on an attention proxy – INST, CVRG, or abs(SUE). Then, I divide each attention tercile into deciles based on the MFD. I test whether the strength of the disagreement premium exhibits a pattern across the attention terciles. The investor inattention theory predicts that the return/alpha to the zero-cost portfolio that buys (sells) stocks with the highest (lowest) MFD should be more negative for those stocks with low attention characteristics.

Table 10 reports the average return and alpha differences between the extreme MFD deciles in each attention group. Consistent with the attention hypothesis, I find that the return and alpha spreads on MFD-sorted portfolios are negative and larger in absolute magnitude, and statistically more significant for stocks in low-attention terciles; low-INST, low-CVRG, and low-abs(SUE), compared to the return and alpha spreads on MFD-sorted portfolios for stocks in high-attention terciles. The diff-and-diff analysis of the return and alpha spreads of the stocks with high vs. low attention characteristics also generates an economically and statistically significant difference. Specifi-

[15] Analyst coverage data come from the Institutional Brokers' Estimate System (I/B/E/S) database and cover the period from 1976 to 2019. Analyst coverage (CVRG) is defined as the number of analysts following a firm in the portfolio formation month.

[16] The standardized unexpected earnings (SUE) is defined as the actual earnings in the current quarter minus earnings four quarters ago, scaled by stock price in the current quarter, following Livnat and Mendenhall [2006]. For abs(SUE), I use the last non-missing SUE value that is released prior to the June of each year during the past 12 months.

cally, the differences between the return and alphas of the zero-cost MFD-sorted portfolios among the extreme attention terciles (high-INST–low-INST; high-CVRG–low-CVRG, high-abs(SUE)–low-abs(SUE)) are economically and statistically significant. Overall, these results indicate that the disagreement premium is stronger for stocks that receive less investor attention.

6.3 Limits to arbitrage

My results suggest that informational frictions and investors' inattention contribute to the cross-sectional relation between the MFD and future equity returns, but I do not fully understand what sustains this return predictability. In this section, I further explore the role of limits-to-arbitrage. If the predictive power of the MFD is driven by mispricing to some extent, then I should expect the return predictability to be more pronounced for stocks with high arbitrage costs. In my next test, I use three proxies of limits-to-arbitrage that are prevalent in the literature.

The prior literature singles out idiosyncratic risk as the primary arbitrage cost (e.g., Pontiff [2006]). I rely on Ang et al. [2006] and measure the monthly IVOL as the standard deviation of the daily residuals from the regression of daily excess stock returns on the three factors of Fama and French [1993] over the past one month. Moreover, following Amihud [2002], I use the monthly illiquidity measure as my second proxy, computed as the absolute daily return divided by daily dollar trading volume, averaged in month $t-1$. Finally, I rely on the market capitalization (size) as my third proxy, which is another widely used measure to capture costly arbitrage (e.g., Cohen and Lou [2012]; Lee et al. [2019]). I test the limits-to-arbitrage hypothesis using bivariate portfolios. Specifically, I first sort stocks into tercile portfolios every month based on a limits-to-arbitrage proxy – IVOL, ILLIQ, or SIZE. Then, I divide each limits-to-arbitrage tercile into deciles based on the MFD. I test whether the strength of the disagreement premium exhibits a pattern across the arbitrage cost terciles. The limits-to-arbitrage hypothesis predicts that the return/alpha to the zero-cost portfolio that buys (sells) stocks with the highest (lowest) MFD should be more negative for those stocks with high arbitrage costs.

I conduct the dependent double sorting. Specifically, I first sort stocks into tercile portfolios every month based on an arbitrage cost proxy – IVOL, ILLIQ, or SIZE. Then, I divide each arbitrage cost tercile into deciles based on the MFD. Consistent with the limits-to-arbitrage hypothesis, Table 11 shows that the return and alpha spreads on MFD-sorted portfolios are negative and larger in absolute magnitude, and statistically more significant for stocks with high arbitrage costs; high-IVOL, high-ILLIQ, and low-SIZE, compared to the return and alpha spreads on MFD-sorted portfolios for stocks with low arbitrage costs. The diff-and-diff analysis of the return and alpha spreads of the stocks with high vs. low arbitrage costs also generates a highly significant difference in the machines' forecast disagreement premium. Specifically, the differences between the return and alphas of the zero-cost MFD-sorted portfolios among the extreme arbitrage cost terciles (high-IVOL–low-IVOL; high-ILLIQ–low-ILLIQ, high-SIZE–low-SIZE) are economically large and statistically significant. Thus, I conclude that the slow diffusion of information into stock prices due to limits-to-arbitrage provides a complementary explanation to the predictive power of the MFD.

7 Risk versus mispricing explanation

The results so far suggest that the established asset pricing models of risk do not explain the cross-sectional variation in equity returns associated with the MFD. However, there is still the possibility of a risk-based mechanism that leads to the return predictability. For example, the MFD can predict the future change in risk, which would lead to a change in the firm's expected return. Thus, in this section, I conduct additional tests to explore whether alternative measures of risk could plausibly explain my results.

7.1 Earnings prediction

If investors could not fully capture the impact of firm-specific uncertainty or investor disagreement – proxied by the MFD – on firm's profitability, they would be surprised by the earnings realizations in the future. Thus, I investigate whether the MFD can predict the future earnings controlling for the past earnings. I use standardized unexpected earnings (SUE), defined as actual earnings in the current quarter minus earnings four quarters ago, scaled by stock price in the current quarter, following Livnat and Mendenhall [2006], to proxy for earnings surprise. I conduct Fama-MacBeth regressions of the SUE from quarter $q+3$ in year $y+1$ to quarter $q+2$ in year $y+2$ on the MFD and other accounting variables at the end of year y as well as other priced-based controls in last month prior to each quarter. I also control for the industry effects following the 48-industry classification of Fama and French [1997]. Finally, I examine the future SUEs over longer time periods, while keeping all independent variables the same. If the MFD contains information about future earnings, I should expect the slope coefficient to be negative and significant.

Consistent with my expectation, the first column of Table 12 shows that the coefficient on the MFD is significantly negative at -0.27 with a t-statistic of -3.05, after accounting for past SUE, control variables, and the industry effects. Moreover, consistent with Bernard and Thomas [1989], the lagged SUE at quarter q is strongly positively correlated with the future SUE. In columns (2) to (4), I repeat the Fama-MacBeth regressions with the same independent variables, but I replace the dependent variable (SUEs) in subsequent quarters. The absolute values of the slope coefficients on the MFD decrease monotonically from column (2) to column (4), and they become statistically insignificant in columns (3) and (4), indicating that the earnings predictability of the MFD decays quickly after two quarters. These results are consistent with the underreaction and mispricing hypotheses that the MFD reflects slow diffusion of cash flow news into stock prices rather than a change in the future discount rate or compensation for risk.

7.2 Portfolio returns during earnings vs. non-earnings announcement periods

To further differentiate the risk vs. mispricing explanations, I examine stock price reactions around earnings announcements. If the return predictability were explained by underlying risk, I would expect the returns to be evenly affected in the subsequent periods. In contrast, if the effect is consistent with mispricing, then the returns must be disproportionately affected around earnings announcements, meaning that the long-short portfolio returns and alphas around earnings announcements should be higher than those around non-earnings announcement periods, if investors are surprised by the good or bad news during those periods.

I test these two distinct hypotheses by examining the long-short portfolio returns and alphas during the periods with and without earnings announcements, defined as in Engelberg et al. [2018]. Specifically, I divide the entire sample period into months with and without earnings announcements, and report the value-weighted average return and alpha spreads on the MFD-sorted portfolios for these two different periods.

Table 13 presents the results. In earnings announcement periods, the long-short value-weighted MFD-sorted portfolio generates an average return and Fama and French [2018] six-factor alpha of 0.86% and 1.03% per month with the respective t-statistics of 3.68 and 3.60. In non-earnings announcement periods, the same long-short portfolio produces much smaller average return and six-factor alpha; 0.32% per month (t-stat.=2.01) and 0.40% (t-stat.=1.73), respectively. The long-short excess returns (alphas) on MFD-sorted portfolios in earnings announcement periods are 2.69 (2.58) times higher than the long-short excess returns (alphas) in non-earnings announcement periods. These results are consistent with the findings of Engelberg et al. [2018] that the equity market anomaly returns on average are three times higher during the periods of earnings announcements.

Thus, the evidence supports the mispricing argument that investors do not fully incorporate the MFD-driven return predictability information into their earnings forecasts and are therefore surprised when earnings are realized.

7.3 Testing potential risk-based explanations

The results have so far shown that the standard factor models or traditional measures of risk do not explain the cross-sectional variation in stock returns associated with the machine forecast disagreement. In this section, I provide comprehensive evidence from testing alternative risk-based explanations. Specifically, I rely on the established rational asset pricing models and investigate whether these models' implied measures of risk can be the driving force of the MFD-return relation.

I first test whether the CAPM explains the disagreement premium. Specifically, I report total volatility, idiosyncratic volatility, and market beta for each MFD-sorted decile portfolio. The CAPM implied measures of market beta, total volatility, and idiosyncratic volatility are estimated for each month using the past 60-month individual stock returns, following Fama and French [1992]. Table 13 shows that the CAPM does not explain the disagreement premium as the low-MFD stocks with high average returns have lower total volatility, lower idiosyncratic volatility, and lower market beta than the high-MFD stocks with low average returns.

Next, I investigate if the MFD effect can be explained by the intertemporal CAPM (ICAPM) of Merton (1973) and/or the consumption CAPM (CCAPM) of Breeden [1979]. Following Ang et al. [2006] and Campbell et al. [2018], I use the change in VIX – S&P500 index option implied volatility – as the second factor of the two-factor ICAPM model.[17] Specifically, I estimate the VIX beta for each stock and each month by running the time-series regressions of excess stock returns on the excess market returns and the change in VIX in the past 60 months. To test for the CCAPM explanation, I compute the consumption beta for each stock and each month by regressing the excess stock returns on the consumption growth rate in the past 60 months.[18] I convert the quarterly consumption data to monthly frequency using linear and cubic spline interpolation methods and the consumption beta estimates turn out to be similar from both methods.[19] Results in Table 14

[17] Campbell et al. [2018] extend Merton's original model by proposing a two-factor ICAPM with stochastic volatility in which an unexpected increase in future market volatility represents deterioration in the investment opportunity set.

[18] The central implication of the CCAPM is that the expected return on an asset is related to "consumption risk," that is, how much uncertainty in consumption would come from holding the asset. Assets that lead to a large amount of uncertainty offer large expected returns, as investors want to be compensated for bearing consumption risk. Thus, the expected excess return on a risky asset is proportional to the covariance of its return and consumption in the period of the return.

[19] The quarterly consumption data (CAY) are obtained from Martin Lettau's online data library:

show that neither the ICAPM nor the CCAPM explains the MFD effect. Specifically, the low-MFD stocks tend to have a higher VIX beta than the high-MFD stocks, implying lower future return for the low-MFD stocks in the ICAPM framework. Also, as presented in Table 14, the VIX beta difference between the high-MFD and low-MFD groups is statistically insignificant. In addition, the low-MFD stocks have a lower consumption beta than the high-MFD stocks, rejecting the CCAPM explanation for the disagreement premium.

Finally, I investigate the magnitude of the factor exposures to see if the MFD-driven return spread is negatively loaded on these factors. Specifically, I estimate stock exposure to each factor (ex-ante factor beta) for each month by regressing the excess stock returns on each of these well-established factors in the past 60 months. Generally, the stocks in the lowest MFD decile have lower factor exposures than those in the highest MFD decile. The only exception is the mispricing factor of Stambaugh and Yuan [2017]; the difference between the high-MFD and low-MFD stock exposures to the performance (PERF) factor of Stambaugh and Yuan [2017] is negative and statistically significant, which provides further supporting evidence for the mispricing explanation. Overall, these results indicate that the predictive power of the MFD is not explained by alternative measures of risk.

8 Conclusion

This paper introduces a novel measure of divergence of opinion among investors about stock value based on the dispersion in machines' expected return forecasts (MFD), which is free from behavioral biases and conflicts of interest that can be observed in the existing measures of disagreement. I document a significantly negative cross-sectional relation between this newly proposed, objective measure of uncertainty (or investor disagreement) and future stock returns. In particular, the value-weighted arbitrage portfolio that takes a short position in 10% of the stocks with the highest MFD and takes a long position in 10% of the stocks with the lowest MFD yields the one-month-ahead abnormal returns (alphas) of 0.55-0.72% per month, estimated with established factor models. I also examine its long-term predictive power and find that the negative relation between the MFD and future equity returns is not just a one-month affair. The MFD predicts cross-sectional variation in equity returns six months into the future. Finally, I find corroborating evidence on the significance of MFD from bivariate portfolios and multivariate Fama–MacBeth regressions when I control for a

https://sites.google.com/view/martinlettau/data?authuser=0.

large number of firm characteristics and risk factors.

I investigate the source of the significant alpha spread between the high-MFD and low-MFD portfolios and find that the machine forecast disagreement premium is driven by underperformance of high-MFD stocks, but not due to outperformance of low-MFD stocks, as the alphas on high-MFD stocks are negative, economically large and highly significant, whereas the alphas on low-MFD stocks are economically small and statistically insignificant. I also show that the high-MFD stocks are subject to significant overpricing, and the negative alpha spread on MFD-sorted portfolios is much stronger for overpriced stocks, compared to underpriced and fairly priced stocks. Thus, my findings support the mispricing explanation of the disagreement premium, consistent with Miller [1977]. I conduct comprehensive analyses to differentiate the risk vs. mispricing explanations. First, I examine the market reactions around earnings announcements and find that the long-short average returns/alphas on MFD-sorted portfolios in earnings announcement periods are about three times higher than the long-short average returns/alphas in non-earnings announcement periods. Second, the stocks in the lowest MFD decile portfolio have lower average beta, total and idiosyncratic volatility, and their exposures to the established risk factors are lower than those in the highest MFD decile portfolio. These results suggest that the return predictability is driven by mispricing rather than compensation for risk.

To provide a better understanding of the economic mechanisms behind the return predictability, I test if the predictive power of the MFD is explained by investors' sophistication, informational frictions, investors' limited attention, and/or limits to arbitrage. I find that institutional (individual) investors are less (more) likely to be active in equities with high MFD and the disagreement premium is more pronounced for equities with high ownership of retail investors. I also show that the disagreement premium is stronger for stocks that are more likely to be held by retail investors and that receive less investor attention. Another potential explanation is limited arbitrage since the negative relation between the MFD and future returns is found to be most pronounced for stocks with high arbitrage costs. Thus, I conclude that the MFD-driven return predictability is likely due to informational frictions, investors' limited attention, and limits to arbitrage.

References

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

Anderson, E. W., Ghysels, E., and Juergens, J. L. (2005). Do heterogeneous beliefs matter for asset pricing? *Review of Financial Studies*, 18:875–924.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 61:259–299.

Atmaz, A. and Basak, S. (2018). Belief dispersion in the stock market. *Journal of Finance*, 73:1225–1279.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131:1593–1636.

Bali, T. G., Beckmeyer, H., Moerke, M., and Weigert, F. (2021). Option return predictability with machine learning and big data. *Available at SSRN 3895984*.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99:427–446.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., and Wen, Q. (2020). Different strokes: Return predictability across stocks and bonds with machine learning and big data. *Available at SSRN 3686164*.

Bali, T. G., Hirshleifer, D. A., Peng, L., and Tang, Y. (2018). Attention, social interaction, and investor attraction to lottery stocks. *Availabe at SSRN 3343769*.

Bali, T. G., Peng, L., Shen, Y., and Tang, Y. (2014). Liquidity shocks and stock market reactions. *Review of Financial Studies*, 27:1434–1485.

Barber, B. M. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, 21:785–818.

Bernard, V. L. and Thomas, J. K. (1989). Post-earnings-announcement drift: delayed price response or risk premium? *Journal of Accounting Research*, 27:1–36.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *Review of Financial Studies*, 34:1046–1089.

Boehme, R. D., Danielsen, B. R., and Sorescu, S. M. (2006). Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis*, 41:455–487.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7:265–296.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). Classification and regression trees. *CRC press.*

Bushee, B. J. (2001). Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research*, 18:207–246.

Campbell, J. Y., Giglio, S., Polk, C., and Turley, R. (2018). An intertemporal capm with stochastic volatility. *Journal of Financial Economics*, 128:207–233.

Carlin, B. I., Longstaff, F. A., and Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, 114:226–238.

Chan, L. K., Karceski, J., and Lakonishok, J. (2007). Analysts' conflicts of interest and biases in earnings forecasts. *Journal of Financial and Quantitative Analysis*, 42:893–913.

Chen, J., Hong, H., and Stein, J. C. (2002). Breadth of ownership and stock returns. *Journal of Financial Economics*, 66:171–205.

Chen, L., Pelger, M., and Zhu, J. (2020). Deep learning in asset pricing. *Available at SSRN 3350138*.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *Journal of Finance*, 63:1977–2011.

Cohen, L. and Lou, D. (2012). Complicated firms. *Journal of Financial Economics*, 104:383–400.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *Journal of Finance*, 63:1609–1651.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance*, 66:1461–1499.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance*, 52:1035–1058.

Daniel, K., Hirshleifer, D., and Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, 33:1673–1736.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., and AP Santos, A. (2021). Can machine learning help to select portfolios of mutual funds? *Can Machine Learning Help to Select Portfolios of Mutual Funds.*

DeVault, L., Sias, R. W., and Starks, L. T. (2019). Sentiment metrics and investor demand. *Journal of Finance*, 74:985–1024.

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

Dugar, A. and Nathan, S. (1995). The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations. *Contemporary Accounting Research*, 12:131–160.

Edelen, R. M., Ince, O. S., and Kadlec, G. B. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics*, 119:472–488.

Engelberg, J., McLean, R. D., and Pontiff, J. (2018). Anomalies and news. *Journal of Finance*, 73:1971–2001.

Engelberg, J., McLean, R. D., and Pontiff, J. (2020). Analysts and anomalies. *Journal of Accounting and Economics*, 69:1–13.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47:427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43:153–193.

Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, 63:1653–1678.

Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, 128:234–252.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance*, 64:2023–2052.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, 75:1327–1370.

Filippou, I., Rapach, D., Taylor, M. P., and Zhou, G. (2020). Exchange rate prediction with machine learning and a smart carry portfolio. *Available at SSRN 3455713*.

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

Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning. *Springer series in statistics New York, NY, USA*, 1.

Gervais, S., Kaniel, R., and Mingelgrin, D. H. (2001). The high-volume return premium. *Journal* *of Finance*, 56:877–919.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63:3–42.

Giglio, S., Liao, Y., and Xiu, D. (2021). Thousands of alpha tests. *Review of Financial Studies*, 34:3456–3496.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. *MIT Press*.

Goyenko, R. and Zhang, C. (2020). The joint cross section of option and stock returns predictability with big data and machine learning. *Available at SSRN 3747238*.

Green, J., Hand, J. R. M., and Zhang, X. F. (2017). The characteristics that provide inde- pendent information about average us. monthly stock returns. *Review of Financial Studies*, 30:4389–4436.

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

Han, B., Hirshleifer, D., and Walden, J. (2021). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, pages 1–42.

Han, Y., He, A., Rapach, D., and Zhou, G. (2020). Firm characteristics and expected stock returns. *Available at SSRN 3185335*.

Harrison, J. M. and Kreps, D. M. (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *Quarterly Journal of Economics*, 92:323–336.

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

Hastie, T., Tibshirani, R., and Friedman, J. (2009). Elements of statistical learning. *Springer Publisher*.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41:401–439.

Hirshleifer, D., Hsu, P.-H., and Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics*, 107:632–654.

Hirshleifer, D., Hsu, P.-H., and Li, D. (2018). Innovative originality, profitability, and stock returns. *Review of Financial Studies*, 31:2553–2605.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. (2009). Driven to distraction: Ex- traneous events and underreaction to earnings news. *Journal of Finance*, 64:2289–2325.

Hirshleifer, D. and Teoh, S. H. (2003). Limited attention, information disclosure, and fi- nancial reporting. *Journal of Accounting and Economics*, 36:337–386.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9:1735–1780.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54:2143–2184.

Hong, H. and Stein, J. C. (2003). Differences of opinion, short-sales constraints, and market crashes. *The Review of Financial Studies*, 16:487–525.

Hong, H., Torous, W., and Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, 83:367–396.

Hou, K. and Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies*, 18:981–1020.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28:650–705.

Huang, D., Jiang, F., Li, K., Tong, G., and Zhou, G. (2021). Scaled pca: A new approach to dimension reduction. *Management Science*.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91.

Johnson, T. C. (2004). Forecast dispersion and the cross section of expected returns. *Journal of Finance*, 59:1957–1978.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105:1177–1216.

Kelly, B. T., Palhares, D., and Pruitt, S. (2020). Modeling corporate bond returns. *Available at SSRN 3720789*.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134:501–524.

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

Kumar, A. (2009). Who gambles in the stock market? *Journal of Finance*, 64:1889–1933.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.

Lee, C. M. C., Sun, S. T., Wang, R., and Zhang, R. (2019). Technological links and predictable returns. *Journal of Financial Economics*, 132:76–96.

Lettau, M. and Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies*, 33:2274–2325.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review*, 4:1–44.

Li, B. and Rossi, A. G. (2020). Selecting mutual funds from the stocks they hold: A machine learning approach. *Available at SSRN 3737667*.

Livnat, J. and Mendenhall, R. R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research*, 44:177–205.

Martin, I. and Nagel, S. (2021). Market efficiency in the age of big data. *Journal of Financial Economics*.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy return predictability? *Journal of Finance*, 71:5–32.

Michaely, R. and Womack, K. L. (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies*, 12:653–686.

Miller, E. M. (1977). Risk, uncertainty and divergence of opinion. *Journal of finance*, 32:1151–1168.

Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *Journal of Finance*, 54:1249–1290.

Neuhierl, A., Tang, X., Varneskov, R. T., and Zhou, G. (2021). Stock option predictability for the cross-section. *Available at SSRN 3795486*.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

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

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80:563–602.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, 42:35–52.

Scheinkman, J. A. and Xiong, W. (2003). Overconfidence and speculative bubbles. *Journal of Political Economy*, 111:1183–1220.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*, 52:35–55.

Shumway, T. (1997). The delisting bias in crsp data. *Journal of Finance*, 52:327–340.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance*, 70:1903–1948.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, 30:1270–

1315.

Wu, W., Chen, J., Yang, Z., and Tindall, M. L. (2021). A cross-sectional machine learning approach for hedge fund return prediction and selection. *Management Science*, 67:4577–4601.

Yu, J. (2011). Disagreement and return predictability of stock portfolios. *Journal of Financial Economics*, 99:162–183.

Zhang, X. F. (2006). Information uncertainty and analyst forecast behavior. *Contemporary Accounting Research*, 23:565–590.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B.*, 67:301–320.

Figure 1: Aggregate MFD Index

This figure presents the monthly time-series plot of the aggregate MFD indices. The blue line depicts the value-weighted average of the stock-level MFD measures. The red line depicts the equal-weighted average of the stock-level MFD measures. Both MFD indices are standardized to have a zero mean and unit standard deviation. The vertical bars correspond to NBER-dated recessions. The sample period is from July 1976 to December 2019.

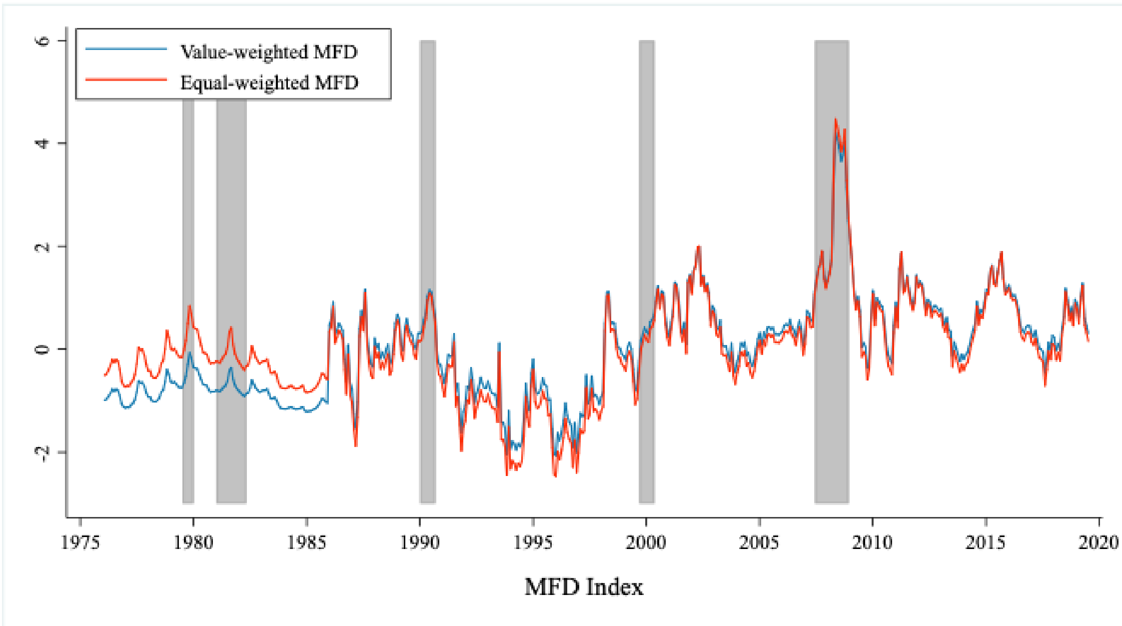

Table 1: Descriptive Statistics

Panel A reports the summary statistics for the cross-sectional variables. The sample consists of all common stocks (share codes equal to 10 or 11) that are listed on NYSE, Amex, and Nasdaq. Financial firms (with one-digit SIC = 6), utility firms (with two-digit SIC = 49), and stocks trading below $5/share are excluded from the analysis. RET_{t+1} is the one-month-ahead return of individual stocks. MFD is defined as the cross-sectional standard deviation of return forecasts scaled by the absolute value of the cross-sectional mean of return forecasts. SIZE is the firm's market capitalization computed as the logarithm of the market value of the firm's outstanding equity at the end of month $t-1$. BM is the logarithm of the firm's book value of equity divided by its market capitalization, where the BM ratio is computed, following Fama and French (2008). Firms with negative book values are excluded from the analysis. Asset Growth (AG) is a percentage of total asset growth between two consecutive fiscal years, following Cooper et al. (2008). Gross Profitability (GP) is the firm's gross profitability calculated as revenue minus cost of goods sold scaled by total assets, following Novy-Marx (2013). MOM is the stock's cumulative return from the start of month $t-12$ to the end of month $t-2$, following Jegadeesh and Titman (1993). Short-term reversal (STR) is the stock's one-month lagged return, following Jegadeesh (1990). ILLIQ is the monthly illiquidity measure computed using daily data in month $t-1$, following Amihud (2002). TURN is the monthly turnover computed as the number of trading shares divided by the total number of shares outstanding in month $t-1$. SUE is the standardized unexpected earnings defined as actual earnings in the current quarter minus earnings 4 quarters ago, scaled by stock price in the current quarter, following Livnat and Mendenhall (2006). IVOL is the idiosyncratic volatility over month $t-1$, following Ang et al. (2006). MAX is average of the five highest daily returns of each stock in month $t-1$, following Bali et al. (2011). DAE is the dispersion in analysts' earnings forecasts, following Diether et al. (2002). DALG is the dispersion in analysts' long-term growth forecasts, following Anderson et al. (2005). All variables are winsorized at the 1% level for both tails to mitigate the effect of outliers. The mean, standard deviation (SD), minimum, median, and maximum of each variable are shown in Panel A. The sample period is from July 1976 to December 2019. Panel B presents the panel Pearson (Spearman) correlations of the variables below (above) diagonal. Panel C presents the correlations between the aggregate MFD indices and the established measures of uncertainty. The value-weighted (equal-weighted) MFD index is the value-weighted (equal-weighted) average of the stock-level MFD measures. The established measures of uncertainty include Jurado, Ludvigson, and Ng (2015) economic and financial uncertainty measures, Baker, Bloom, and Davis (2016) economic policy uncertainty measures, and the VXO index. The MFD indices and Jurado, Ludvigson, and Ng (2015) economic and financial uncertainty indices start from July 1976. The Baker, Bloom, and Davis (2016) economic policy uncertainty index starts from January 1985. The VXO index starts from January 1986.

Panel A: Summary statistics

	Mean	Stdev	Min	Median	Max
RET_{t+1}	0.01	0.14	-0.41	0.00	1.58
MFD	2.49	1.28	0.07	2.04	5.71
SIZE	4.90	1.91	0.62	4.78	10.44
BM	-0.61	0.87	-3.40	-0.53	1.43
AG	1.32	0.88	0.87	1.25	5.68
GP	0.38	0.29	-0.54	0.36	1.27
MOM	0.13	0.51	-0.82	0.05	2.61
STR	0.01	0.14	-0.40	0.00	0.57
ILLIQ	0.22	0.43	0.00	0.05	23.83
TURN	1.19	1.28	0.02	0.79	8.33
SUE	0.10	1.90	-54.21	0.09	34.63
IVOL	0.04	0.02	0.02	0.04	0.20
MAX	0.03	0.02	0.00	0.03	0.19
DAE	2.22	0.83	0.10	1.98	4.31
DALG	1.12	0.42	0.15	0.92	3.75

Panel B: Pearson (Spearman) correlations below (above) the diagonal

	RET_{t+1}	MFD	SIZE	BM	AG	GP	MOM	STR	ILLIQ	TURN	SUE	IVOL	MAX	DAE	DALG
RET_{t+1}		-0.006	-0.003	0.022	-0.026	0.018	0.024	-0.005	0.005	-0.014	0.051	-0.029	-0.042	-0.102	-0.030
MFD	-0.005		-0.038	0.066	-0.018	0.132	-0.011	0.005	0.037	0.041	0.019	0.021	-0.021	0.033	0.020
SIZE	-0.003	-0.045		-0.279	0.046	0.021	0.194	0.071	-0.408	0.261	0.033	-0.330	-0.164	0.024	0.017
BM	0.023	0.065	-0.260		-0.175	-0.129	-0.022	0.017	0.112	-0.227	-0.021	0.007	0.080	-0.068	-0.012
AG	-0.030	-0.015	0.046	-0.259		-0.113	-0.015	-0.021	-0.059	0.106	-0.037	0.061	0.041	-0.127	-0.005
GP	0.024	0.133	0.013	-0.100	-0.088		0.055	0.023	0.017	-0.024	-0.036	-0.086	-0.064	-0.035	-0.093
MOM	0.030	-0.013	0.200	-0.018	-0.011	0.092		0.023	-0.141	0.128	0.078	-0.130	-0.025	-0.006	-0.046
STR	-0.006	0.006	0.050	0.019	-0.021	0.025	0.024		-0.066	0.052	0.059	-0.083	0.154	-0.004	-0.016
ILLIQ	0.003	0.040	-0.344	0.165	-0.072	0.017	-0.107	-0.050		-0.153	-0.017	0.361	0.465	0.001	0.069
TURN	-0.010	0.035	0.157	-0.130	0.156	-0.026	0.137	0.052	-0.171		-0.007	0.295	0.189	-0.039	-0.040
SUE	0.063	0.015	0.035	-0.018	-0.032	-0.026	0.075	0.061	-0.010	-0.013		-0.029	-0.043	-0.014	-0.010
IVOL	-0.028	0.019	-0.406	0.004	0.037	-0.088	-0.173	-0.091	0.437	0.225	-0.033		0.692	0.055	0.009
MAX	-0.038	-0.024	-0.168	0.068	0.043	-0.073	-0.024	0.144	0.347	0.226	-0.039	0.743		0.060	0.013
DAE	-0.111	0.032	0.031	-0.059	-0.093	-0.042	-0.009	-0.004	0.001	-0.048	-0.015	0.041	0.067		0.034
DALG	-0.026	0.021	0.017	-0.015	-0.004	-0.067	-0.036	-0.018	0.080	-0.040	-0.016	0.009	0.010	0.041	

Panel C: Correlations between the MFD index and existing measures of uncertainty

Correlations	Value-weighted MFD index	Economic uncertainty	Financial uncertainty	Economic policy uncertainty	VXO index
Equal-weighted MFD index	0.950	0.520	0.526	0.509	0.512
Value-weighted MFD index		0.344	0.482	0.500	0.507
Economic uncertainty			0.524	0.220	0.551
Financial uncertainty				0.360	0.813
Economic policy uncertainty					0.344

Table 2: Univariate Portfolio Analysis

Panel A reports the average monthly excess returns and alphas on the value-weighted portfolios of stocks sorted by the MFD. Panel B reports the average monthly excess returns and alphas on the equal-weighted portfolios of stocks sorted by the MFD. For each month t from July 1976 to December 2019, individual stocks are sorted into decile portfolios by the MFD in month $t-1$, and are held for the next one month. P1 is the portfolio of stocks with the lowest MFD and P10 is the portfolio of stocks with the highest MFD. L/S is a zero-cost portfolio that buys stocks in decile 10 (highest MFD) and sells stocks in decile 1 (lowest MFD). All returns and alphas are expressed in percentage. Excess return is the raw return of the portfolio over the risk-free rate. Alpha is the intercept from a time-series regression of monthly excess returns on the factors of alternative models: the CAPM, Fama and French [2018] six-factor model (FF6), Stambaugh and Yuan [2017] mispricing factor model (SY), Hou et al. [2015] q-factor model (HXZ), Hou et al. [2015] q4-factor model (HXZ), and Daniel et al. [2020] behavioral factor model (DHS). Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Panel A: Value-weighted MFD-sorted decile portfolios

Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.72***	0.19	0.20	0.01	0.18	0.15
	(2.90)	(0.92)	(0.87)	(0.38)	(0.91)	(0.95)
P2	0.61**	0.17	0.17	-0.02	0.16	0.13
	(2.26)	(0.55)	(0.51)	(-0.08)	(0.89)	(0.49)
P3	0.55**	0.04	0.03	-0.16	-0.03	-0.11
	(2.20)	(0.84)	(0.23)	(-0.06)	(-0.36)	(-0.33)
P4	0.57**	0.16	0.10	-0.07	0.13	-0.03
	(2.22)	(0.42)	(0.51)	(-0.07)	(0.42)	(-0.36)
P5	0.45**	-0.06	-0.12	-0.16	-0.18	-0.34
	(2.16)	(-1.07)	(-0.33)	(-0.86)	(-0.21)	(-0.97)
P6	0.43	-0.11	-0.13	-0.20	-0.36	-0.38*
	(1.35)	(-1.13)	(-0.60)	(-1.00)	(-0.51)	(-1.89)
P7	0.35	-0.43*	-0.37*	-0.42	-0.49**	-0.49**
	(0.59)	(-1.93)	(-1.73)	(-1.45)	(-2.03)	(-2.17)
P8	0.36	-0.29*	-0.32	-0.25	-0.47**	-0.39*
	(0.89)	(-1.81)	(-0.78)	(-1.01)	(-1.99)	(-1.93)
P9	0.33	-0.50**	-0.45*	-0.44**	-0.49**	-0.51***
	(0.33)	(-2.11)	(-1.90)	(-2.06)	(-2.20)	(-2.81)
P10	0.27	-0.50***	-0.52*	-0.54**	-0.53***	-0.52***
	(0.05)	(-2.63)	(-1.92)	(-2.44)	(-2.65)	(-2.82)
L/S	-0.45***	-0.69***	-0.72***	-0.55**	-0.71***	-0.67***
	(-2.94)	(-3.36)	(-3.45)	(-2.54)	(-2.91)	(-2.83)

Panel B: Equal-weighted MFD-sorted decile portfolios						
Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.84***	0.26	0.33	-0.01	0.21	0.16
	(3.28)	(0.99)	(0.37)	(-0.58)	(0.45)	(0.99)
P2	0.78***	0.12	0.17	-0.34	0.08	-0.04
	(2.78)	(0.36)	(0.04)	(-0.83)	(0.41)	(-1.37)
P3	0.58**	0.08	0.12	-0.37	0.04	-0.21*
	(2.09)	(0.33)	(-0.39)	(-1.26)	(-0.05)	(-1.76)
P4	0.58*	0.05	-0.01	-0.37	-0.10	-0.30*
	(1.85)	(0.10)	(-0.82)	(-1.50)	(-0.81)	(-1.80)
P5	0.56*	-0.10	-0.06	-0.46*	-0.11*	-0.38*
	(1.73)	(-0.04)	(-1.28)	(-1.93)	(-1.89)	(-1.84)
P6	0.51	-0.22	-0.14	-0.48***	-0.14**	-0.38**
	(1.63)	(-0.30)	(-1.35)	(-2.64)	(-2.17)	(-2.04)
P7	0.41	-0.24	-0.31	-0.54***	-0.38**	-0.40**
	(1.61)	(-1.45)	(-1.38)	(-2.67)	(-2.49)	(-2.16)
P8	0.38	-0.28	-0.38*	-0.55***	-0.56***	-0.47***
	(1.60)	(-1.52)	(-1.94)	(-2.84)	(-2.87)	(-2.65)
P9	0.32	-0.43**	-0.44**	-0.57***	-0.61***	-0.54***
	(1.54)	(-2.14)	(-2.53)	(-3.15)	(-2.96)	(-3.88)
P10	0.24	-0.59***	-0.73***	-0.67***	-0.72***	-0.63***
	(1.28)	(-2.72)	(-3.60)	(-3.47)	(-3.09)	(-3.98)
L/S	-0.60***	-0.86***	-1.06***	-0.66***	-0.93***	-0.79***
	(-4.04)	(-4.15)	(-4.41)	(-3.58)	(-4.36)	(-4.14)

Table 3: Transition Matrix

This table presents transition probabilities for MFD at a lag of 12 months from July 1976 to December 2019. For each month t, all stocks are sorted into deciles based on an ascending ordering of the MFD. The procedure is repeated in month t+12. P1 is the portfolio of stocks with the lowest MFD and P10 is the portfolio of stocks with the highest MFD. For each MFD decile in month t, the percentage of stocks that fall into each of the month $t + 12$ MFD decile is calculated. Table presents the time-series averages of these transition probabilities. Each row corresponds to a different month t MFD portfolio and each column corresponds to a different month $t + 12$ MFD portfolio.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	63	16	10	6	2	2	1	0	0	0
P2	17	48	15	8	4	3	2	1	1	1
P3	7	18	40	14	9	5	3	2	1	1
P4	5	7	15	37	16	9	6	2	2	1
P5	4	4	9	16	35	15	8	3	3	3
P6	2	2	5	8	13	38	14	8	6	4
P7	2	2	2	4	9	14	40	14	8	5
P8	0	1	2	3	6	6	14	52	10	6
P9	0	1	1	2	3	5	9	10	56	13
P10	0	1	1	2	3	3	3	8	13	66

Table 4: Long-Term Predictive Power

This table presents the long-term predictive power of the MFD. P1 is the value-weighted portfolio of stocks with the lowest MFD and P10 is the value-weighted portfolio of stocks with the highest MFD. L/S is a zero-cost value-weighted portfolio that buys stocks in decile 10 (highest MFD) and sells stocks in decile 1 (lowest MFD). The table reports Fama and French [2018] six-factor alphas for each of the MFD-sorted decile portfolios from two to 12 months after portfolio formation. The last column shows the differences of monthly Fama and French [2018] six-factor alphas between deciles 1 and 10. Newey and West [1987] adjusted t-statistics are presented in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively.

	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$	$t+9$	$t+10$	$t+11$	$t+12$
P1	0.11	0.07	0.04	0.05	-0.01	-0.07	-0.15	-0.13	-0.10	-0.11	-0.15
	(0.71)	(0.44)	(0.34)	(0.52)	(-0.53)	(-0.48)	(-0.36)	(-0.16)	(-0.14)	(-0.17)	(-0.07)
P2	-0.05	0.01	0.00	0.04	-0.02	-0.08	-0.18	-0.17	-0.17	-0.18	-0.17
	(-0.61)	(0.29)	(0.16)	(0.05)	(-0.39)	(-0.23)	(-0.04)	(-0.15)	(-0.12)	(-0.07)	(-0.04)
P3	-0.14	-0.03	-0.06	-0.13	-0.19	-0.13	-0.20	-0.19	-0.20	-0.18	-0.18
	(-0.26)	(-0.47)	(-0.65)	(-0.40)	(-0.16)	(-0.29)	(-0.46)	(-0.04)	(-0.03)	(-0.04)	(-0.18)
P4	-0.12	-0.01	-0.05	-0.08	-0.18	-0.12	-0.19	-0.18	-0.18	-0.18	-0.17
	(-0.31)	(-0.05)	(-0.42)	(-0.03)	(-0.26)	(-0.06)	(-0.24)	(-0.10)	(-0.07)	(-0.03)	(-0.14)
P5	-0.15	-0.16	-0.27	-0.14	-0.23	-0.13	-0.25	-0.21	-0.20	-0.21	-0.18
	(-1.10)	(-1.35)	(-0.84)	(-0.51)	(-0.72)	(-0.54)	(-0.46)	(-0.12)	(-0.02)	(-0.22)	(-0.20)
P6	-0.16	-0.20	-0.27	-0.16	-0.25	-0.13	-0.30	-0.21	-0.21	-0.23	-0.18
	(-1.17)	(-1.39)	(-0.90)	(-0.54)	(-0.89)	(-0.78)	(-0.47)	(-0.31)	(-0.03)	(-0.27)	(-0.23)
P7	-0.39**	-0.36*	-0.35*	-0.29	-0.29	-0.17	-0.33	-0.27	-0.25	-0.27	-0.20
	(-2.30)	(-1.92)	(-1.69)	(-1.14)	(-1.26)	(-1.08)	(-0.76)	(-0.49)	(-0.15)	(-0.41)	(-0.25)
P8	-0.25**	-0.34*	-0.27	-0.27	-0.28	-0.15	-0.32	-0.25	-0.22	-0.26	-0.20
	(-2.30)	(-1.82)	(-1.36)	(-0.96)	(-1.24)	(-0.80)	(-0.67)	(-0.48)	(-0.09)	(-0.30)	(-0.23)
P9	-0.41**	-0.39*	-0.36*	-0.40	-0.33	-0.20	-0.37	-0.30	-0.27	-0.27	-0.21
	(-2.41)	(-1.95)	(-1.80)	(-1.61)	(-1.26)	(-1.24)	(-0.87)	(-0.54)	(-0.22)	(-0.54)	(-0.25)
P10	-0.45**	-0.44**	-0.43**	-0.42*	-0.33	-0.35	-0.40	-0.32	-0.28	-0.27	-0.23
	(-2.41)	(-2.04)	(-2.02)	(-1.75)	(-1.53)	(-1.36)	(-1.19)	(-0.83)	(-0.60)	(-0.57)	(-0.27)
L/S	-0.56***	-0.51**	-0.47**	-0.47*	-0.32*	-0.28	-0.25	-0.19	-0.18	-0.16	-0.08
	(-2.83)	(-2.29)	(-2.08)	(-1.94)	(-1.87)	(-1.58)	(-1.25)	(-1.03)	(-0.91)	(-0.89)	(-0.52)

Table 5: Testing Mispricing Explanation

Panel A reports the average mispricing (MISP) score of the MFD-sorted univariate decile portfolios. High (low) MISP indicates a higher degree of overvaluation (undervaluation). Panel B presents the Fama and French [2018] six-factor alphas on the 5x10 bivariate portfolios of stocks independently sorted into quintile portfolios by MISP and decile portfolios by MFD. Panel B also reports the differences of monthly six-factor alphas on MFD-sorted portfolios within each MISP quintile. The last four rows present the differences between the six-factor alphas of the zero-cost MFD-sorted portfolios for MISP5 vs. the other MISP quintiles. Newey and West [1987] adjusted t-statistics are given in parentheses. The sample period is from July 1976 to December 2019.

Panel A: Average mispricing score of MFD-sorted decile portfolios

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
MISP	48.63	49.11	49.25	49.03	49.09	49.49	49.37	49.29	49.54	49.57	0.94	3.50

Panel B: Six-factor alphas on MISP&MFD-sorted portfolios

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
MISP1	0.20	0.17	0.03	0.10	-0.11	-0.13	-0.35	-0.32	-0.46	-0.55	-0.75	-3.07
MISP2	0.10	0.08	0.01	0.05	-0.06	-0.06	-0.17	-0.15	-0.22	-0.27	-0.37	-1.76
MISP3	0.13	0.11	0.02	0.07	-0.08	-0.09	-0.27	-0.21	-0.31	-0.35	-0.48	-2.29
MISP4	0.18	0.15	0.03	0.09	-0.11	-0.12	-0.34	-0.29	-0.41	-0.48	-0.67	-3.26
MISP5	0.36	0.31	0.05	0.18	-0.22	-0.24	-0.66	-0.57	-0.79	-0.94	-1.30	-4.89
MISP5 – MISP1	0.15	0.14	0.02	0.08	-0.11	-0.11	-0.31	-0.26	-0.33	-0.40	-0.55	-1.92
MISP5 – MISP2	0.26	0.23	0.04	0.13	-0.16	-0.18	-0.50	-0.43	-0.57	-0.67	-0.93	-3.29
MISP5 – MISP3	0.22	0.20	0.03	0.11	-0.14	-0.14	-0.39	-0.36	-0.48	-0.59	-0.82	-2.73
MISP5 – MISP4	0.17	0.16	0.03	0.09	-0.11	-0.12	-0.33	-0.29	-0.38	-0.46	-0.63	-1.71

Table 6: Average Stock Characteristics of MFD-sorted Portfolios

This table presents the average stock characteristics of the univariate decile portfolios formed based on the MFD. P1 is the portfolio of stocks with the lowest MFD and P10 is the portfolio of stocks with the highest MFD. Table reports the time-series averages of the monthly cross-sectional medians for MFD and various firm-specific characteristics for each MFD-sorted portfolio. The last two columns show the differences between P1 and P10 and the associated Newey and West [1987] adjusted t-statistics. The MFD and other firm-specific characteristics are defined in Table 1. The sample period is from July 1976 to December 2019.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
MFD	2.37	2.39	2.47	2.50	2.53	2.56	2.60	2.62	2.63	2.71	0.34	247.02
SIZE	4.68	4.41	4.28	4.89	4.74	4.36	4.42	4.32	4.04	4.24	-0.44	-13.13
BM	-0.67	-0.78	-0.54	-0.52	-0.57	-0.64	-0.68	-0.55	-0.55	-0.47	0.20	12.72
AG	1.15	1.42	1.20	1.24	1.23	1.35	1.57	1.41	1.37	1.53	0.38	5.38
GP	0.46	0.38	0.37	0.42	0.42	0.39	0.38	0.39	0.33	0.35	-0.12	-62.70
MOM	0.14	0.14	0.15	0.13	0.12	0.13	0.12	0.11	0.12	0.12	-0.02	-5.94
STR	-0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	5.41
ILLIQ	0.13	0.22	0.15	0.18	0.20	0.21	0.26	0.29	0.28	0.24	0.11	5.34
TURN	1.00	1.01	1.10	1.16	1.22	1.19	1.16	1.17	1.20	1.23	0.23	17.38
SUE	0.15	0.12	0.09	0.09	0.06	0.06	0.03	0.03	0.00	-0.03	-0.18	-9.42
IVOL	0.01	0.02	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03	13.27
MAX	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.03	8.30
DAE	2.21	2.16	2.23	2.25	2.26	2.23	2.27	2.22	2.27	2.28	0.07	10.27
DALG	1.09	1.08	1.12	1.10	1.09	1.12	1.12	1.12	1.14	1.13	0.04	9.76

Table 7: Bivariate Portfolio Analysis

This table presents results from the value-weighted bivariate portfolios based on dependent and independent double sorts of various firm-specific characteristics and MFD. Panel A reports results from dependent double sorts. First, decile portfolios are formed every month based on a firm-specific characteristic. Next, additional decile portfolios are formed based on MFD within each firm-specific characteristic decile. Panel B reports results from independent double sorts. All stocks are grouped into decile portfolios based on independent ascending sorts of a firm-specific attribute and MFD each month. The intersections of each of the decile groups are used to form the bivariate portfolios. P1 is the portfolio of stocks with the lowest MFD averaged across each firm-specific characteristic decile. P10 is the portfolio of stocks with the highest MFD averaged across each firm-specific characteristic decile. The table reports one-month-ahead six-factor alphas associated with each MFD decile. L/S shows the differences of monthly alphas between MFD deciles 1 and 10 after controlling for each firm-specific characteristic. Alphas are calculated after adjusting for the market, size, value, investment, profitability, and momentum factors of Fama and French [2018]. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Panel A: Dependent bivariate sorts

	SIZE	BM	AG	GP	MOM	STR	ILLIQ	TURN	SUE	IVOL	MAX	DAE	DALG
P1	0.00	0.18	0.17	0.11	0.28	0.20	0.20	0.04	0.17	0.30	0.24	0.17	0.20
	(0.61)	(0.60)	(0.82)	(0.49)	(0.34)	(0.19)	(0.65)	(0.52)	(0.31)	(0.82)	(0.84)	(0.62)	(0.56)
P2	0.09	-0.15	0.06	0.26	-0.07	0.18	0.13	0.29	0.18	0.08	-0.03	0.10	0.12
	(0.88)	(-0.47)	(0.08)	(0.98)	(-0.59)	(0.04)	(0.38)	(0.61)	(0.85)	(0.03)	(-0.40)	(0.10)	(0.56)
P3	-0.01	-0.20	-0.18	0.09	-0.28**	0.03	-0.15	0.04	0.16	-0.34	0.05	-0.07	-0.52
	(-0.44)	(-0.97)	(-0.90)	(0.27)	(-2.25)	(0.76)	(-1.05)	(1.33)	(0.09)	(-0.58)	(0.41)	(-1.33)	(-0.53)
P4	-0.04	-0.22	-0.20	0.03	-0.33**	-0.07*	-0.26	0.03	0.16	-0.44*	-0.12	-0.13	-0.56
	(-0.29)	(-1.21)	(-1.21)	(0.00)	(-2.57)	(-1.85)	(-1.06)	(1.44)	(0.87)	(-1.69)	(-0.37)	(-1.48)	(-1.14)
P5	-0.17	-0.19	-0.15	-0.11	-0.09	0.12	0.04	-0.32	0.09	-0.11	-0.13	0.08	0.02
	(-0.56)	(-0.12)	(-0.18)	(-0.25)	(-0.70)	(0.22)	(0.00)	(-1.63)	(0.93)	(-0.41)	(-1.00)	(0.80)	(0.25)
P6	-0.26	-0.28	-0.22	-0.16	-0.37***	-0.12**	-0.09	-0.33**	0.03	-0.13	-0.15*	0.02	-0.50
	(-1.20)	(-1.28)	(-1.37)	(-0.77)	(-2.73)	(-2.00)	(-0.70)	(-2.27)	(0.95)	(-0.58)	(-1.80)	(-0.89)	(-0.22)
P7	-0.42	-0.30**	-0.25**	-0.19	-0.42***	-0.60**	-0.39*	-0.39**	-0.24	-0.46**	-0.20*	-0.25	-0.60
	(-1.55)	(-2.52)	(-2.00)	(-0.94)	(-3.74)	(-2.12)	(-1.77)	(-2.35)	(-1.17)	(-1.96)	(-1.93)	(-1.54)	(-1.52)
P8	-0.62**	-0.39***	-0.61**	-0.48***	-0.54***	-0.60**	-0.55**	-0.53***	-0.41**	-0.47**	-0.32***	-0.36**	-0.60***
	(-2.53)	(-3.65)	(-2.15)	(-3.33)	(-3.76)	(-2.49)	(-2.11)	(-3.61)	(-2.32)	(-2.36)	(-3.35)	(-2.00)	(-2.90)
P9	-0.50**	-0.29**	-0.22*	-0.34***	-0.41***	-0.39**	-0.62***	-0.41***	-0.26**	-0.53***	-0.23***	-0.45***	-0.61***
	(-2.04)	(-2.42)	(-1.65)	(-2.74)	(-2.90)	(-2.06)	(-2.62)	(-2.94)	(-2.06)	(-3.52)	(-2.90)	(-2.75)	(-2.94)
P10	-0.63***	-0.54***	-0.69**	-0.64***	-0.65***	-0.70***	-0.65***	-0.66***	-0.56***	-0.55***	-0.60***	-0.60***	-0.62***
	(-3.76)	(-4.00)	(-2.50)	(-3.81)	(-3.89)	(-3.52)	(-2.82)	(-3.94)	(-2.73)	(-3.69)	(-3.83)	(-3.16)	(-3.79)
L/S	-0.62***	-0.72***	-0.86***	-0.74***	-0.92***	-0.90***	-0.85***	-0.70***	-0.73***	-0.84***	-0.84***	-0.76***	-0.82***
t-stat	(-4.14)	(-4.40)	(-2.75)	(-4.19)	(-4.28)	(-3.87)	(-4.20)	(-4.33)	(-3.00)	(-4.06)	(-4.21)	(-3.47)	(-4.17)

Panel B: Independent bivariate sorts													
	SIZE	BM	AG	GP	MOM	STR	ILLIQ	TURN	SUE	IVOL	MAX	DAE	DALG
P1	0.20	0.15	0.16	0.20	0.16	0.16	0.19	0.18	0.14	0.20	0.13	0.19	0.10
	(0.78)	(0.59)	(0.05)	(0.62)	(0.20)	(0.85)	(0.43)	(0.74)	(0.81)	(0.98)	(0.81)	(0.43)	(0.35)
P2	0.15	0.09	0.02	-0.09	-0.04	0.01	0.12	0.05	0.06	0.03	0.04	0.04	0.01
	(0.14)	(0.06)	(0.08)	(-0.94)	(-0.18)	(0.55)	(0.12)	(0.53)	(0.61)	(0.05)	(0.43)	(0.03)	(0.82)
P3	0.16	0.11	-0.10	0.02	0.10	0.05	0.13	0.05	0.13	0.12	0.06	0.08	0.05
	(0.17)	(0.46)	(-0.07)	(0.60)	(0.06)	(0.67)	(0.02)	(0.54)	(0.78)	(0.63)	(0.64)	(0.07)	(0.45)
P4	0.11	0.03	-0.06	-0.14	-0.07	-0.07	0.04	0.03	-0.02	-0.07	-0.02	0.02	-0.05
	(0.09)	(0.18)	(-0.63)	(-1.25)	(-0.22)	(-0.32)	(0.23)	(0.36)	(-0.12)	(-0.83)	(-0.29)	(0.51)	(-1.17)
P5	0.09	-0.05	-0.09	-0.23	-0.12	-0.13	-0.01	-0.19	-0.03	-0.18	-0.14	-0.08	-0.16
	(0.51)	(-0.64)	(-1.01)	(-1.30)	(-1.19)	(-0.09)	(-0.70)	(-0.35)	(-0.29)	(-0.92)	(-0.02)	(-0.54)	(-1.49)
P6	-0.02	-0.13	-0.38*	-0.25	-0.16	-0.15	-0.01	-0.25	-0.16	-0.39	-0.25	-0.24	-0.23*
	(-1.19)	(-1.04)	(-1.80)	(-1.54)	(-1.32)	(-0.45)	(-0.79)	(-0.15)	(-0.33)	(-1.39)	(-0.32)	(-0.74)	(-1.83)
P7	-0.32	-0.19*	-0.42**	-0.35**	-0.18*	-0.37	-0.38	-0.28	-0.41	-0.44	-0.28	-0.24	-0.27**
	(-1.56)	(-1.91)	(-2.24)	(-2.09)	(-1.65)	(-0.46)	(-0.84)	(-0.14)	(-0.34)	(-1.41)	(-1.41)	(-0.88)	(-2.29)
P8	-0.49*	-0.39**	-0.47***	-0.42**	-0.41**	-0.42	-0.56**	-0.43	-0.48	-0.51**	-0.52**	-0.42***	-0.54***
	(-1.81)	(-2.43)	(-2.62)	(-2.23)	(-2.18)	(-1.56)	(-2.57)	(-0.81)	(-0.70)	(-2.27)	(-2.41)	(-2.66)	(-2.83)
P9	-0.48	-0.27*	-0.45**	-0.38**	-0.31**	-0.38	-0.46*	-0.42	-0.44	-0.46	-0.30	-0.35	-0.37***
	(-1.58)	(-1.95)	(-2.32)	(-2.10)	(-2.12)	(-1.49)	(-1.84)	(-0.27)	(-0.67)	(-1.63)	(-1.55)	(-1.37)	(-2.79)
P10	-0.52*	-0.42**	-0.51***	-0.52***	-0.56***	-0.52**	-0.59***	-0.54	-0.59**	-0.57***	-0.58***	-0.51***	-0.54***
	(-1.89)	(-2.43)	(-2.81)	(-2.60)	(-2.83)	(-2.40)	(-2.95)	(-1.58)	(-2.50)	(-2.75)	(-2.91)	(-2.71)	(-2.85)
L/S	-0.72**	-0.57***	-0.68***	-0.72***	-0.72***	-0.68***	-0.78***	-0.71**	-0.74***	-0.77***	-0.71***	-0.70***	-0.64***
t-stat	(-2.39)	(-3.08)	(-3.55)	(-3.28)	(-3.58)	(-3.04)	(-3.74)	(-2.33)	(-3.16)	(-3.48)	(-3.68)	(-3.43)	(-3.61)

Table 8: Fama-MacBeth Cross-Sectional Regressions

This table reports the Fama and MacBeth [1973] cross-sectional regression results. The first column is from July 1976 to December 2019. The other three columns are from July 1982 to December 2019. The MFD and the control variables in month $t-1$ are matched to stock returns in month t. The monthly price-based variables are calculated using the last non-missing observations prior to each month. The dependent variable is the firm's future raw return in the first two columns, the firm's future excess return over its value-weighted industry peers' return (Column 3), or the firm's DGTW adjusted return (Column 4). I include industry dummies and classify each firm's industry peers based on the Fama-French 48-industry classifications. All returns are expressed in percentage. The MFD and firm-specific characteristics (i.e., control variables) are defined in Table 1. Cross-sectional regressions are run every calendar month, and the time-series standard errors are corrected for heteroskedasticity and autocorrelation. Newey and West [1987] adjusted t-statistics are reported in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively.

Independent Variables	RET	RET	RET-INDRET	DGTW-adj. RET
MFD	-1.45***	-1.40***	-1.27***	-1.11***
	(-2.78)	(-2.65)	(-3.06)	(-3.57)
SIZE	-0.06**	-0.07**	-0.06*	-0.07**
	(-2.21)	(-2.02)	(-1.81)	(-2.17)
BM	0.12	0.08	0.10	0.06
	(1.59)	(1.12)	(1.13)	(0.83)
AG	-0.06	-0.05	-0.09	-0.09
	(-0.35)	(-0.38)	(-0.62)	(-0.61)
GP	0.18	0.23	0.29	0.38
	(0.48)	(0.73)	(0.66)	(0.99)
MOM	0.45	0.54	0.38	0.45
	(0.79)	(1.02)	(0.66)	(0.73)
STR	-1.20***	-1.24***	-1.18**	-1.17***
	(-2.95)	(-3.30)	(-2.44)	(-3.54)
ILLIQ	0.98	4.91	5.35	11.22
	(0.06)	(0.20)	(0.28)	(0.55)
TURN	-0.38	-0.05	-0.43	-0.09
	(-0.98)	(-0.19)	(-1.01)	(-0.29)
SUE	0.08**	0.08**	0.05*	0.05**
	(2.15)	(2.50)	(1.74)	(1.99)
IVOL	-0.73	-0.80	-0.58	-0.79
	(-1.23)	(-1.35)	(-0.82)	(-1.13)
MAX	-1.50***	-1.41**	-1.23**	-1.11**
	(-2.81)	(-2.56)	(-2.38)	(-2.04)
DAE	-0.71	-0.58	-0.62	-0.68
	(-1.41)	(-1.26)	(-1.17)	(-1.02)
DALG		-0.31	-0.32	-0.34
		(-0.56)	(-0.95)	(-0.73)
Intercept	1.06*	0.98*	0.91	0.89
	(1.72)	(1.70)	(1.45)	(1.57)
Industry FEs	Yes	Yes	No	Yes
N	2,085,442	1,662,360	1,662,360	1,654,261
Adj. R2	0.178	0.195	0.104	0.185

Table 9: Institutional Ownership and MFD

Panel A reports the average institutional ownership (INST) of the MFD-sorted univariate decile portfolios. High (low) INST indicates a higher institutional (retail) ownership. Panel B presents the Fama and French [2018] six-factor alphas on the 5x10 bivariate portfolios of stocks dependently sorted into quintile portfolios by INST and decile portfolios by MFD in each INST quintile. Panel B also reports the differences of monthly six-factor alphas on MFD-sorted portfolios within each INST quintile. The last row presents the difference between the six-factor alphas of the zero-cost MFD-sorted portfolios for INST5 vs. INST1 quintiles. Newey and West [1987] adjusted t-statistics are given in parentheses. The sample period is from July 1976 to December 2019.

Panel A: Average INST of MFD-sorted decile portfolios

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
INST	0.34	0.32	0.31	0.30	0.29	0.29	0.28	0.27	0.26	0.26	-0.08	-6.18

Panel B: Six-factor alphas on INST&MFD-sorted portfolios

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
INST1	0.30	0.26	0.06	0.16	-0.20	-0.22	-0.60	-0.56	-0.79	-0.92	-1.23	-4.72
INST2	0.24	0.23	0.05	0.13	-0.16	-0.18	-0.50	-0.41	-0.59	-0.71	-0.95	-3.96
INST3	0.21	0.16	0.03	0.10	-0.11	-0.13	-0.37	-0.33	-0.46	-0.51	-0.72	-3.47
INST4	0.14	0.12	0.02	0.07	-0.08	-0.09	-0.24	-0.20	-0.29	-0.31	-0.46	-2.00
INST5	0.09	0.07	0.01	0.04	-0.05	-0.05	-0.13	-0.11	-0.12	-0.10	-0.19	-1.56
INST5 – INST1	-0.21	-0.20	-0.04	-0.12	0.15	0.17	0.47	0.45	0.66	0.82	1.04	3.27

Table 10: Investor Attention and MFD

This table splits the stock sample into three tercile subsamples based on proxies of investor attention; (*i*) institutional ownership (INST), (*ii*) analyst coverage (CVRG), and (*iii*) Abs(SUE). CVRG is the number of analysts covering the firm at the end of the previous month. INST is the percentage of institutional ownership at the end of the previous fiscal-year end. Abs(SUE) is defined as the absolute value of SUE based on the last non-missing SUE during the 12 months preceding June. I conduct the dependent double sorting. I first sort stocks into tercile portfolios every month based on an attention proxy – INST, CVRG, or abs(SUE). Then, I divide each attention tercile into deciles based on the MFD. Table reports the excess returns and alphas of long-short decile portfolios of stocks sorted by MFD within INST, CVRG, and Abs(SUE) terciles. Portfolio sorts are conducted by every calendar month, and the time-series standard errors are corrected for heteroskedasticity and autocorrelation. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

INST	Excess Return	CAPM	FF6	SY	HXZ	DHS
Low	-0.60***	-0.92***	-0.98***	-0.75***	-0.96***	-0.90***
	(-3.90)	(-4.59)	(-4.57)	(-3.43)	(-3.74)	(-3.84)
Medium	-0.49***	-0.81***	-0.71***	-0.59***	-0.69***	-0.58***
	(-2.71)	(-2.87)	(-3.59)	(-2.86)	(-2.78)	(-2.88)
High	-0.25*	-0.40**	-0.39*	-0.34	-0.43*	-0.39*
	(-1.82)	(-2.06)	(-1.81)	(-1.56)	(-1.69)	(-1.70)
High – Low	0.36**	0.52***	0.59***	0.41**	0.53**	0.51**
	(2.18)	(2.65)	(2.89)	(1.97)	(2.14)	(2.24)
CVRG	**Excess Return**	**CAPM**	**FF6**	**SY**	**HXZ**	**DHS**
Low	-0.58***	-0.88***	-0.91***	-0.73***	-0.90***	-0.87***
	(-3.71)	(-4.39)	(-4.56)	(-3.40)	(-3.90)	(-3.69)
Medium	-0.39***	-0.62***	-0.67***	-0.49***	-0.62***	-0.80***
	(-2.70)	(-3.87)	(-4.08)	(-2.58)	(-2.67)	(-2.89)
High	-0.28*	-0.41**	-0.45**	-0.32	-0.42*	-0.42*
	(-1.70)	(-2.18)	(-2.13)	(-1.49)	(-1.85)	(-1.69)
High – Low	0.29**	0.47**	0.46**	0.41**	0.49**	0.45**
	(2.12)	(2.32)	(2.55)	(2.00)	(2.15)	(2.10)
Abs(SUE)	**Excess Return**	**CAPM**	**FF6**	**SY**	**HXZ**	**DHS**
Low	-0.61***	-0.97***	-0.97***	-0.76***	-0.94***	-0.90***
	(-3.83)	(-4.70)	(-4.59)	(-3.42)	(-4.04)	(-3.92)
Medium	-0.47***	-0.62***	-0.73***	-0.64**	-0.70**	-0.80***
	(-3.47)	(-3.68)	(-3.91)	(-2.45)	(-2.56)	(-2.88)
High	-0.24*	-0.41*	-0.37*	-0.31	-0.40	-0.37
	(-1.65)	(-1.91)	(-1.91)	(-1.35)	(-1.49)	(-1.61)
High – Low	0.37**	0.55***	0.60***	0.44**	0.54***	0.53**
	(2.40)	(3.07)	(2.95)	(2.28)	(2.80)	(2.54)

Table 11: Limits-to-arbitrage and MFD

This table splits the stock sample into three tercile subsamples based on proxies of arbitrage costs; (*i*) idiosyncratic volatility (IVOL), (*ii*) illiquidity (ILLIQ), and (*iii*) market capitalization (SIZE). Idiosyncratic volatility is constructed following Ang et al. [2006]. The illiquidity measure is calculated following Amihud [2002]. SIZE is the log value of market capitalization at the end of the previous month. I conduct the dependent double sorting. I first sort stocks into tercile portfolios every month based on an arbitrage cost proxy – IVOL, ILLIQ, or SIZE. Then, I divide each arbitrage cost tercile into deciles based on the MFD. Table reports the excess returns and alphas of long-short decile portfolios of stocks sorted by MFD within IVOL, ILLIQ, and SIZE terciles. Portfolio sorts are conducted by every calendar month, and the time-series standard errors are corrected for heteroskedasticity and autocorrelation. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

IVOL	Excess Return	CAPM	FF6	SY	HXZ	DHS
Low	-0.22	-0.30	-0.35	-0.24	-0.29	-0.30
	(-1.41)	(-1.44)	(-1.46)	(-1.09)	(-1.25)	(-1.27)
Medium	-0.47***	-0.62***	-0.83***	-0.65***	-0.66***	-0.67***
	(-3.52)	(-3.94)	(-3.99)	(-3.03)	(-2.98)	(-2.58)
High	-0.63***	-1.01***	-1.00***	-0.78***	-0.99***	-0.98***
	(-4.29)	(-4.74)	(-4.88)	(-3.74)	(-4.17)	(-3.97)
High – Low	-0.41***	-0.71***	-0.66***	-0.54***	-0.70***	-0.69***
	(-3.16)	(-3.64)	(-3.76)	(-2.92)	(-3.21)	(-2.97)
ILLIQ	**Excess Return**	**CAPM**	**FF6**	**SY**	**HXZ**	**DHS**
Low	-0.33*	-0.55*	-0.54**	-0.40	-0.52*	-0.47*
	(-1.65)	(-1.95)	(-2.10)	(-1.44)	(-1.75)	(-1.75)
Medium	-0.36***	-0.65***	-0.59***	-0.46***	-0.57***	-0.67***
	(-2.97)	(-3.34)	(-3.92)	(-2.86)	(-2.69)	(-3.13)
High	-0.53***	-0.83***	-0.85***	-0.61***	-0.85***	-0.78***
	(-3.50)	(-3.82)	(-4.06)	(-2.95)	(-3.21)	(-3.36)
High – Low	-0.20**	-0.28**	-0.31**	-0.21*	-0.32	-0.31*
	(-2.04)	(-2.06)	(-2.16)	(-1.66)	(-1.61)	(-1.78)
SIZE	**Excess Return**	**CAPM**	**FF6**	**SY**	**HXZ**	**DHS**
Low	-0.62***	-0.97***	-0.99***	-0.77***	-1.00***	-0.91***
	(-4.13)	(-4.64)	(-4.74)	(-3.54)	(-4.02)	(-3.86)
Medium	-0.41***	-0.59***	-0.81***	-0.63**	-0.81***	-0.63***
	(-3.47)	(-2.84)	(-3.36)	(-2.41)	(-2.97)	(-2.87)
High	-0.23	-0.32*	-0.38*	-0.29	-0.34	-0.36
	(-1.47)	(-1.81)	(-1.73)	(-1.28)	(-1.54)	(-1.48)
High – Low	0.39***	0.65***	0.61***	0.48**	0.66***	0.55**
	(2.80)	(2.97)	(3.15)	(2.37)	(2.60)	(2.50)

Table 12: Quarterly Earnings Prediction

This table reports the results from Fama and MacBeth [1973] regressions of individual firm's one-quarter- to four-quarter-ahead SUE on the past MFD and control variables. All independent variables are calculated using the last non-missing observations prior to each quarter. I classify each firm's industry peers based on the Fama-French 48-industry classifications. I winsorize all variables at the 1% and 99% levels and standardize all independent variables to have a zero mean and standard deviation of one. The control variables are from Table 6, but their estimated coefficients are not shown here. Cross-sectional regressions are run every calendar quarter, and the time-series standard errors are corrected for heteroskedasticity and autocorrelation. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Independent Variables	SUE_{q+1}	SUE_{q+2}	SUE_{q+3}	SUE_{q+4}
MFD	-0.27***	-0.21**	-0.14	-0.04
	(-3.05)	(-2.10)	(-1.54)	(-0.23)
SUE_q	0.69***	0.55***	0.48***	0.64***
	(5.23)	(3.56)	(3.51)	(5.14)
Controls	Yes	Yes	Yes	Yes
Industry FEs	Yes	Yes	Yes	Yes
N	555716	539273	522894	507466
Adj. R2	0.24	0.19	0.16	0.12

Table 13: Predicting Future Earnings Announcement Returns

This table reports the long-short value-weighted portfolio return and alpha spreads on MFD-sorted portfolios during the periods with and without earnings announcement. I divide the whole sample period into months with earnings announcement and months without earnings announcement. Following Engelberg et al. [2018], I obtain earnings announcement dates from the Compustat quarterly database. I define the months with earnings announcement when there is earnings announcement in that month and define the months without earnings announcement when there is no earnings announcement in that month. Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Earnings announcement	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	1.01***	0.27	0.28	0.01	0.25	0.21
	(4.12)	(1.33)	(1.25)	(0.55)	(1.31)	(1.34)
P10	0.16	-0.73***	-0.75***	-0.76***	-0.74***	-0.74***
	(0.03)	(-3.80)	(-2.71)	(-3.45)	(-3.76)	(-4.06)
L/S	-0.86***	-1.00***	-1.03***	-0.77***	-0.99***	-0.95***
	(-3.68)	(-4.61)	(-3.60)	(-3.56)	(-4.55)	(-4.83)
Non-earnings announcement	**Excess Return**	**CAPM**	**FF6**	**SY**	**HXZ**	**DHS**
P1	0.63**	0.11	0.11	0.01	0.11	0.09
	(2.57)	(0.52)	(0.51)	(0.22)	(0.53)	(0.53)
P10	0.31	-0.30	-0.29	-0.30	-0.29	-0.30
	(0.05)	(-1.45)	(-1.14)	(-1.37)	(-1.58)	(-1.59)
L/S	-0.32**	-0.41**	-0.40*	-0.30*	-0.40**	-0.39**
	(-2.01)	(-2.07)	(-1.73)	(-1.66)	(-2.21)	(-2.22)

Table 14: Testing Risk-based Explanations

This table presents the results from testing potential risk-based explanations. Table reports the average portfolio risk attributes for each decile portfolio sorted by MFD, and the differences for the risk attributes between deciles 10 and 1 and the associated Newey and West [1987] adjusted t-statistics. P1 is the portfolio of stocks with the lowest MFD and P10 is the portfolio of stocks with the highest MFD. The CAPM implied measures of total volatility (TVOL), idiosyncratic volatility (IVOL), and market (MKT) Beta are estimated for each month using the past 60-month individual stock returns. The individual stock exposures (Betas) to the change in VIX, consumption growth rate, and each factor are estimated for each month using the past 60-month observations. The sample period is from July 1976 to December 2019.

Risk	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1	t-stat
CAPM												
TVOL	0.09	0.09	0.10	0.10	0.10	0.11	0.11	0.11	0.12	0.12	0.03	8.32
IVOL	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01	4.24
MKT Beta	0.97	0.93	1.01	1.01	0.98	0.97	0.99	0.97	0.96	1.00	0.03	2.11
ICAPM												
VIX Beta	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03	-0.04	-0.04	-0.04	-0.05	-0.03	-1.28
CCAPM												
Consumption Growth Beta	0.40	0.42	0.47	0.48	0.50	0.51	0.52	0.52	0.53	0.54	0.14	2.02
Factor Exposures												
SMB Beta	0.87	0.88	0.81	0.89	0.93	0.96	0.98	0.96	0.94	0.91	0.04	2.75
HML Beta	-0.22	-0.22	-0.07	-0.06	-0.10	-0.03	-0.08	-0.01	-0.02	0.04	0.26	14.62
RMW Beta	-0.26	-0.29	-0.15	-0.07	-0.12	-0.15	-0.16	-0.20	-0.22	-0.20	0.06	2.48
CMA Beta	0.02	0.18	0.08	0.13	0.07	0.09	0.16	0.05	0.04	0.09	0.07	1.64
MOM Beta	-0.14	-0.12	-0.14	-0.14	-0.17	-0.14	-0.13	-0.15	-0.14	-0.13	0.01	0.13
LIQ Beta	0.01	0.02	0.03	0.03	0.03	0.04	0.04	0.05	0.06	0.07	0.06	3.14
MGMT Beta	-0.26	-0.23	-0.14	-0.11	-0.21	-0.15	-0.15	-0.18	-0.20	-0.18	0.08	2.86
PERF Beta	-0.13	-0.15	-0.16	-0.16	-0.18	-0.19	-0.18	-0.21	-0.23	-0.25	-0.12	-2.42
IA Beta	-0.17	-0.05	-0.05	0.00	-0.08	0.01	-0.02	-0.03	-0.01	0.06	0.23	9.25
ROE Beta	-0.36	-0.38	-0.35	-0.33	-0.34	-0.34	-0.43	-0.44	-0.41	-0.42	-0.06	-1.54
FIN Beta	-0.41	-0.38	-0.26	-0.26	-0.31	-0.29	-0.31	-0.31	-0.31	-0.29	0.12	7.19
PEAD Beta	-0.09	-0.13	-0.14	-0.20	-0.17	-0.17	-0.21	-0.21	-0.24	-0.18	-0.09	-1.36

Machine Forecast Disagreement

Online appendix

Table OA1 reports out-of-sample R-squared (R^2_{OS}, in percentage) of predicting stock returns with 12 machine learning models and 310 stock characteristics.

Table OA2 compares the monthly out-of-sample prediction among 12 machine learning models using Diebold-Mariano tests.

Table OA3 reports the excess returns of the long-short stock portfolios constructed based on the expected return forecasts obtained from machine learning models.

Table OA4 reports the six-factor alphas of the long-short stock portfolios constructed based on the expected return forecasts obtained from machine learning models.

Table OA5 reports results from the univariate portfolios of stocks sorted by machine forecast error disagreement (MFED).

Table OA6 reports results from the univariate portfolios of stocks sorted by machine MAX-MIN difference (MAX-MIN).

Section OB presents 310 firm characteristics used to forecast stock returns.

Section OC provides a description of the machine learning models.

Table OA1: Out-of-sample R-squared performance using the 310 stock characteristics

This table reports the out-of-sample R-squared (R^2_{OS}, in percentage) from predicting one-month-ahead stock returns using 12 machine learning models and 310 stock characteristics. The models include (*i*) dimension reduction models: principal components analysis (PCA), Scaled PCA (SPCA), and Partial Least squares (PLS); (*ii*) penalized linear regressions: LASSO, Ridge, and Elastic Net (E-Net); (*iii*) regression trees: random forests (RF), Gradient Boosted Regression Tree (GBRT), and Extreme Tree (ET); and (*iv*) neural networks: Feed Forward Neural Network (FNN), Recurrent Neural Network (RNN), and Long Short-Term Memory Neural Network (LSTM). The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} \left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{i,t+1}}$$

where p-values associated with R^2_{OS} are reported using one-sided test. The full sample covers the period from July 1966 to December 2019. I use ten years (first six years as the training sample $\mathcal{T}_1$ and subsequent four years as the validation sample $\mathcal{T}_2$) as the rolling window for estimating the parameters and tuning the hyperparameters of machine learning forecasting models. The out-of-sample "test" subsample (from July 1976 to December 2019, $\mathcal{T}_3$) is used to evaluate a model's forecasting performance. All of the R^2_{OS} associated with machine learning models are statistically significant with p-values less than 1%.

	PCA	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
R^2_{OS}	0.37	0.48	0.46	0.13	0.08	0.13	0.51	0.54	0.56	0.57	0.58	0.59

Table OA2: Comparison of monthly out-of-sample prediction using Diebold-Mariano tests

This table reports pairwise Diebold-Mariano test statistics comparing the out-of-sample stock-level prediction performance (R^2_{OS}) among the models used in Table OA1. Positive numbers indicate the column model outperforms the row model. Numbers in bold denote statistical significance at the 5% level or better.

	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
PCA	0.45	0.40	-1.29	-1.32	-1.37	1.36	1.39	1.35	**1.89**	**1.82**	**1.87**
SPCA		-0.13	-1.37	-1.37	-1.41	1.30	1.33	1.29	1.72	1.71	1.72
PLS			-1.34	-1.36	-1.39	1.35	1.37	1.30	1.75	1.77	**1.78**
LASSO				-0.42	0.19	**2.24**	**2.10**	**2.13**	**3.16**	**3.11**	**3.03**
Ridge					0.43	**2.36**	**2.20**	**2.19**	**3.29**	**3.18**	**3.17**
E-Net						**2.22**	**2.08**	**2.11**	**3.14**	**3.09**	**3.01**
RF							0.21	0.25	1.36	1.41	1.41
GBRT								0.19	1.34	1.30	1.34
ET									1.32	1.26	1.29
FNN										0.12	0.14
RNN											0.12

Table OA3: Excess returns of machine learning portfolios sorted by expected returns

This table reports the monthly excess returns of equal-weighted and value-weighted decile portfolios sorted by the out-of-sample machine learning expected return forecasts using the 310 stock characteristics (i.e., $\hat{r}_{i,t+1}$ where (i,t)$\in \mathcal{T}_3$, the test subsample)). At the end of each month, I calculate one-month-ahead out-of-sample stock return predictions for each method. In Panel A (Panel B), I sort stocks into deciles based on each model's forecasts and construct the equal-weighted (value-weighted) portfolio based on the out-of-sample forecasts. "Low" corresponds to the portfolio with the lowest expected return (decile 1), "High" corresponds to the portfolio with the highest expected return (decile 10), and "High–Low" corresponds to the long short portfolio that buys the highest expected return stocks (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively.

Panel A: Excess returns of equal-weighted machine learning portfolios

Equal-weight	PCA	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
Low	-0.39	-0.68	-0.53	0.05	0.03	-0.13	-0.61	-0.82	-0.87	-0.97	-1.06	-1.00
2	0.44	0.16	0.32	0.57	0.48	0.41	0.24	0.22	0.23	0.16	0.18	0.18
3	0.59	0.41	0.55	0.62	0.62	0.52	0.58	0.49	0.55	0.45	0.54	0.45
4	0.76	0.57	0.63	0.79	0.65	0.65	0.61	0.61	0.58	0.68	0.72	0.55
5	0.64	0.70	0.72	0.75	0.56	0.78	0.65	0.77	0.71	0.78	0.73	0.71
6	0.84	0.81	0.76	0.86	0.76	0.78	0.77	0.81	0.77	0.80	0.83	0.79
7	0.67	1.02	0.75	0.92	0.53	0.84	0.77	0.86	0.85	0.82	0.94	0.83
8	0.83	1.06	0.87	0.93	0.75	1.02	0.95	0.85	0.99	0.89	0.96	1.07
9	1.13	1.32	1.22	1.20	0.99	1.18	1.27	1.13	1.14	1.24	1.11	1.26
High	1.75	1.90	1.90	1.58	1.39	1.47	2.21	2.24	2.38	2.47	2.42	2.49
High – Low	2.14***	2.58***	2.43***	1.52***	1.35***	1.60***	2.81***	3.06***	3.26***	3.43***	3.48***	3.49***
t-stat	(4.14)	(5.00)	(4.69)	(2.95)	(2.62)	(3.09)	(5.44)	(5.91)	(6.30)	(6.64)	(6.72)	(7.10)

Panel B: Excess returns of value-weighted machine learning portfolios

Value-weight	PCA	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
Low	-0.21	-0.40	-0.28	0.03	0.02	-0.07	-0.35	-0.45	-0.47	-0.54	-0.62	-0.53
2	0.25	0.09	0.18	0.33	0.24	0.21	0.13	0.11	0.13	0.10	0.09	0.09
3	0.34	0.22	0.30	0.35	0.32	0.30	0.30	0.25	0.28	0.25	0.31	0.26
4	0.38	0.34	0.37	0.47	0.33	0.38	0.31	0.36	0.32	0.35	0.42	0.28
5	0.33	0.42	0.38	0.41	0.33	0.40	0.38	0.40	0.40	0.41	0.37	0.39
6	0.44	0.47	0.44	0.46	0.39	0.43	0.39	0.42	0.46	0.44	0.47	0.47
7	0.37	0.56	0.39	0.55	0.30	0.46	0.43	0.49	0.44	0.49	0.48	0.47
8	0.46	0.60	0.45	0.50	0.42	0.54	0.54	0.44	0.60	0.51	0.49	0.55
9	0.60	0.76	0.69	0.69	0.53	0.65	0.69	0.64	0.66	0.72	0.62	0.75
High	0.92	1.02	0.99	0.89	0.71	0.78	1.25	1.14	1.37	1.32	1.43	1.36
High – Low	1.14**	1.42***	1.27**	0.86	0.69	0.85	1.60***	1.59***	1.83***	1.86***	2.05***	1.89***
t-stat	(2.27)	(2.71)	(2.46)	(1.47)	(1.33)	(1.40)	(2.67)	(2.68)	(2.91)	(3.08)	(3.39)	(3.70)

Table OA4: Six-factor alphas of machine learning portfolios sorted by expected returns

This table reports the monthly Fama and French [2018] six-factor alphas of equal-weighted and value-weighted decile portfolios of stocks sorted by the out-of-sample machine learning expected return forecasts using the 310 stock characteristics (i.e., $\hat{r}_{i,t+1}$ where (i,t)$\in \mathcal{T}_3$, the test subsample). At the end of each month, I calculate one-month-ahead out-of-sample stock return predictions for each method. In Panel A (Panel B), I sort stocks into deciles based on each model's forecasts and construct the equal-weighted (value-weighted) portfolio based on the out-of-sample forecasts. "Low" corresponds to the portfolio with the lowest expected return (decile 1), "High" corresponds to the portfolio with the highest expected return (decile 10), and "High–Low" corresponds to the long short portfolio that buys the highest expected return stocks (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively.

Panel A: FF6 alphas of equal-weighted machine learning portfolios

Equal-weight	PCA	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
Low	-0.48	-0.82	-0.66	0.07	0.04	-0.16	-0.77	-1.02	-1.13	-1.22	-1.31	-1.20
2	0.40	0.17	0.29	0.59	0.53	0.44	0.25	0.24	0.24	0.16	0.17	0.18
3	0.63	0.40	0.52	0.67	0.57	0.52	0.55	0.53	0.56	0.45	0.54	0.47
4	0.75	0.55	0.57	0.86	0.62	0.71	0.61	0.63	0.60	0.66	0.79	0.57
5	0.62	0.67	0.66	0.77	0.56	0.79	0.71	0.77	0.74	0.76	0.76	0.75
6	0.90	0.81	0.72	0.86	0.81	0.72	0.69	0.81	0.84	0.78	0.82	0.74
7	0.71	1.08	0.74	0.90	0.54	0.82	0.70	0.80	0.76	0.91	0.86	0.83
8	0.78	1.09	0.85	1.00	0.78	1.02	0.99	0.78	1.09	0.86	0.99	1.09
9	1.23	1.36	1.11	1.18	1.08	1.29	1.24	1.21	1.18	1.31	1.08	1.24
High	1.37	1.48	1.44	1.20	0.98	1.17	1.73	1.69	1.90	1.78	1.72	1.92
High – Low	1.85***	2.30***	2.10***	1.13***	0.94***	1.33***	2.50***	2.71***	3.03**	2.99***	3.03***	3.12***
t-stat	(5.04)	(6.57)	(5.66)	(3.62)	(3.28)	(4.09)	(6.80)	(7.85)	(7.80)	(8.19)	(8.43)	(9.18)

Panel B: FF6 alphas of value-weighted machine learning portfolios

Value-weight	PCA	SPCA	PLS	LASSO	Ridge	E-Net	RF	GBRT	ET	FNN	RNN	LSTM
Low	-0.15	-0.38	-0.13	0.01	0.01	-0.02	-0.38	-0.49	-0.46	-0.49	-0.56	-0.56
2	0.15	0.07	0.06	0.07	0.08	0.06	0.10	0.09	0.09	0.08	0.09	0.08
3	0.24	0.17	0.09	0.10	0.07	0.09	0.25	0.19	0.27	0.20	0.23	0.18
4	0.28	0.23	0.14	0.11	0.08	0.13	0.31	0.30	0.23	0.30	0.34	0.24
5	0.21	0.29	0.11	0.12	0.06	0.13	0.32	0.32	0.29	0.35	0.32	0.33
6	0.28	0.37	0.14	0.15	0.13	0.16	0.35	0.38	0.38	0.31	0.32	0.38
7	0.23	0.47	0.13	0.16	0.08	0.11	0.42	0.41	0.41	0.32	0.42	0.32
8	0.29	0.49	0.17	0.15	0.14	0.15	0.43	0.39	0.45	0.37	0.48	0.54
9	0.38	0.60	0.19	0.13	0.13	0.23	0.56	0.50	0.58	0.56	0.43	0.47
High	0.51	0.58	0.18	0.11	0.16	0.21	0.75	0.71	0.77	0.88	0.84	0.87
High – Low	0.66***	0.96***	0.31	0.10	0.15	0.23	1.13***	1.20***	1.24***	1.37***	1.40***	1.42***
t-stat	(2.74)	(3.24)	(0.97)	(0.75)	(0.65)	(0.62)	(3.92)	(4.06)	(4.45)	(4.90)	(5.04)	(4.70)

Table OA5: Univariate portfolios of stocks sorted by machine forecast error disagreement

This table reports the average monthly excess returns and alphas on the value-weighted and equal-weighted portfolios of stocks sorted by the machine forecast error disagreement (MFED). For each month t from July 1976 to December 2019, individual stocks are sorted into decile portfolios based on MFED at month $t-1$, and are held for the next one month. P1 is the portfolio of stocks with the lowest MFED and P10 is the portfolio of stocks with the highest MFED. L/S is a zero-cost portfolio that buys stocks in decile 10 (highest MFED) and sells stocks in decile 1 (lowest MFED). All returns and alphas are expressed in percentage. Excess return is the raw return of the portfolio over the risk-free rate. Alpha is the intercept from a time-series regression of monthly excess returns on the factors of alternative models: the CAPM, Fama and French [2018] six-factor model (FF6), Stambaugh and Yuan [2017] mispricing-factor model (SY), Hou et al. [2015] q4-factor model (HXZ), and Daniel et al. [2020] behavioral factor model (DHS). Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Panel A: Value-weighted MFED-sorted decile portfolios

Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.56**	0.15	0.15	0.01	0.14	0.11
	(2.21)	(0.72)	(0.65)	(0.30)	(0.69)	(0.74)
P2	0.46*	0.14	0.13	-0.02	0.12	0.10
	(1.79)	(0.43)	(0.40)	(-0.06)	(0.67)	(0.39)
P3	0.44*	0.04	0.03	-0.12	-0.03	-0.08
	(1.71)	(0.64)	(0.18)	(-0.05)	(-0.28)	(-0.25)
P4	0.44*	0.13	0.07	-0.06	0.09	-0.02
	(1.75)	(0.33)	(0.40)	(-0.06)	(0.32)	(-0.29)
P5	0.34	-0.05	-0.09	-0.13	-0.14	-0.27
	(1.62)	(-0.82)	(-0.26)	(-0.65)	(-0.16)	(-0.78)
P6	0.34	-0.08	-0.10	-0.15	-0.28	-0.30
	(1.06)	(-0.89)	(-0.47)	(-0.78)	(-0.39)	(-1.48)
P7	0.28	-0.34	-0.29	-0.32	-0.38	-0.38*
	(0.47)	(-1.52)	(-1.35)	(-1.16)	(-1.58)	(-1.68)
P8	0.28	-0.23	-0.25	-0.20	-0.37	-0.29
	(0.70)	(-1.37)	(-0.61)	(-0.77)	(-1.58)	(-1.54)
P9	0.25	-0.39*	-0.35	-0.35	-0.37*	-0.39**
	(0.25)	(-1.67)	(-1.43)	(-1.57)	(-1.75)	(-2.16)
P10	0.21	-0.40**	-0.39	-0.42*	-0.41**	-0.40**
	(0.04)	(-1.98)	(-1.54)	(-1.94)	(-2.07)	(-2.21)
L/S	-0.35**	-0.55***	-0.54***	-0.43*	-0.55**	-0.52**
	(-2.28)	(-2.64)	(-2.69)	(-1.91)	(-2.24)	(-2.25)

Panel B: Equal-weighted MFED-sorted decile portfolios						
Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.67**	0.21	0.25	-0.01	0.16	0.13
	(2.52)	(0.77)	(0.28)	(-0.44)	(0.34)	(0.77)
P2	0.59**	0.09	0.13	-0.25	0.06	-0.03
	(2.16)	(0.29)	(0.03)	(-0.64)	(0.33)	(-1.07)
P3	0.46	0.07	0.09	-0.29	0.03	-0.17
	(1.64)	(0.25)	(-0.31)	(-0.98)	(-0.04)	(-1.34)
P4	0.44	0.04	-0.01	-0.28	-0.08	-0.23
	(1.43)	(0.08)	(-0.63)	(-1.13)	(-0.62)	(-1.41)
P5	0.43	-0.08	-0.05	-0.35	-0.08	-0.30
	(1.37)	(-0.03)	(-0.98)	(-1.53)	(-1.46)	(-1.46)
P6	0.39	-0.17	-0.11	-0.38**	-0.11*	-0.30
	(1.26)	(-0.23)	(-1.05)	(-2.11)	(-1.66)	(-1.53)
P7	0.31	-0.19	-0.24	-0.42**	-0.29*	-0.32*
	(1.22)	(-1.10)	(-1.10)	(-2.12)	(-1.88)	(-1.72)
P8	0.30	-0.21	-0.29	-0.42**	-0.44**	-0.36**
	(1.21)	(-1.18)	(-1.53)	(-2.17)	(-2.18)	(-2.05)
P9	0.24	-0.33	-0.34**	-0.45**	-0.47**	-0.43***
	(1.17)	(-1.61)	(-2.02)	(-2.40)	(-2.27)	(-3.07)
P10	0.19	-0.47**	-0.56***	-0.53***	-0.55**	-0.49***
	(0.97)	(-2.09)	(-2.80)	(-2.66)	(-2.38)	(-3.10)
L/S	-0.48***	-0.68***	-0.81***	-0.52***	-0.71***	-0.62***
	(-3.14)	(-3.17)	(-3.35)	(-2.81)	(-3.48)	(-3.22)

Table OA6: Univariate portfolios of stocks sorted by machine MAX-MIN difference

This table reports the average monthly excess returns and alphas on the value-weighted and equal-weighted portfolios of stocks sorted by the machine MAX-MIN difference (MAX-MIN). For each month t from July 1976 to December 2019, individual stocks are sorted into decile portfolios based on MAX-MIN at month $t-1$, and are held for the next one month. P1 is the portfolio of stocks with the lowest MAX-MIN and P10 is the portfolio of stocks with the highest MAX-MIN. L/S is a zero-cost portfolio that buys stocks in decile 10 (highest MAX-MIN) and sells stocks in decile 1 (lowest MAX-MIN). All returns and alphas are expressed in percentage. Excess return is the raw return of the portfolio over the risk-free rate. Alpha is the intercept from a time-series regression of monthly excess returns on the factors of alternative models: the CAPM, Fama and French [2018] six-factor model (FF6), Stambaugh and Yuan [2017] mispricing-factor model (SY), Hou et al. [2015] q4-factor model (HXZ), and Daniel et al. [2020] behavioral factor model (DHS). Newey and West [1987] adjusted t-statistics are given in parentheses. Coefficients marked with *, **, and *** are significant at the 10%, 5%, and 1% level, respectively. The sample period is from July 1976 to December 2019.

Panel A: Value-weighted MAX-MIN-sorted decile portfolios

Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.50*	0.13	0.14	0.01	0.12	0.10
	(1.83)	(0.66)	(0.58)	(0.26)	(0.61)	(0.64)
P2	0.48*	0.14	0.13	-0.02	0.13	0.10
	(1.72)	(0.44)	(0.41)	-(0.06)	(0.65)	(0.37)
P3	0.44	0.04	0.02	-0.12	-0.03	-0.08
	(1.63)	(0.65)	(0.18)	(-0.05)	(-0.27)	(-0.26)
P4	0.46*	0.13	0.07	-0.06	0.09	-0.02
	(1.79)	(0.32)	(0.39)	(-0.06)	(0.31)	(-0.30)
P5	0.33	-0.05	-0.09	-0.13	-0.15	-0.28
	(1.55)	(-0.85)	(-0.27)	(-0.61)	(-0.16)	(-0.78)
P6	0.34	-0.09	-0.10	-0.15	-0.28	-0.29
	(1.11)	(-0.93)	(-0.48)	(-0.77)	(-0.40)	(-1.47)
P7	0.28	-0.33	-0.29	-0.32	-0.38	-0.38
	(0.45)	(-1.46)	(-1.36)	(-1.21)	(-1.61)	(-1.64)
P8	0.28	-0.22	-0.25	-0.20	-0.38	-0.30
	(0.67)	(-1.35)	(-0.61)	(-0.80)	(-1.62)	(-1.52)
P9	0.25	-0.37*	-0.35	-0.36	-0.36*	-0.37**
	(0.25)	(-1.65)	(-1.48)	(-1.63)	(-1.79)	(-2.06)
P10	0.19	-0.36*	-0.34	-0.37	-0.37*	-0.34*
	(0.03)	(-1.77)	(-1.33)	(-1.61)	(-1.84)	(-1.84)
L/S	-0.31**	-0.49**	-0.48**	-0.38*	-0.49**	-0.44**
	(-2.09)	(-2.31)	(-2.50)	(-1.78)	(-2.07)	(-2.09)

Panel B: Equal-weighted MAX-MIN-sorted decile portfolios						
Rank	Excess Return	CAPM	FF6	SY	HXZ	DHS
P1	0.59**	0.18	0.22	-0.01	0.15	0.12
	(2.22)	(0.68)	(0.24)	-(0.38)	(0.29)	(0.67)
P2	0.56**	0.09	0.14	-0.25	0.06	-0.04
	(2.23)	(0.30)	(0.03)	(-0.62)	(0.33)	(-1.03)
P3	0.48	0.07	0.09	-0.29	0.03	-0.16
	(1.61)	(0.25)	(-0.31)	(-0.98)	(-0.04)	(-1.34)
P4	0.42	0.04	-0.01	-0.29	-0.09	-0.23
	(1.39)	(0.08)	(-0.60)	(-1.10)	(-0.60)	(-1.43)
P5	0.43	-0.08	-0.05	-0.34	-0.08	-0.30
	(1.36)	(-0.03)	(-1.00)	(-1.53)	(-1.52)	(-1.43)
P6	0.39	-0.18	-0.11	-0.39**	-0.11*	-0.29
	(1.26)	(-0.22)	(-1.01)	(-2.04)	(-1.71)	(-1.49)
P7	0.31	-0.19	-0.24	-0.41**	-0.28*	-0.31*
	(1.19)	(-1.10)	(-1.14)	(-2.08)	(-1.95)	(-1.75)
P8	0.31	-0.22	-0.29	-0.40**	-0.44**	-0.37*
	(1.19)	(-1.21)	(-1.54)	(-2.10)	(-2.20)	(-1.95)
P9	0.25	-0.33*	-0.34**	-0.46**	-0.48**	-0.44***
	(1.17)	(-1.67)	(-2.03)	(-2.37)	(-2.25)	(-3.18)
P10	0.17	-0.41*	-0.48**	-0.48**	-0.47**	-0.41***
	(0.88)	(-1.94)	(-2.46)	(-2.34)	(-2.18)	(-2.76)
L/S	-0.42***	-0.59***	-0.70***	-0.47***	-0.62***	-0.53***
	(-2.93)	(-2.79)	(-3.02)	(-2.66)	(-3.06)	(-2.92)

Section OB. Firm characteristics

Table : **Firm characteristics**

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
Panel A: Momentum (30)			
ABR	Earnings announcement abnormal returns	R36	Long-term reversal, 36 months
Alli	Strategic alliance peer momentum	R6	6-month momentum
Board	Common board peer momentum	R6	Long-term reversal, 60 months
Cust	customer momentum	RE	Long-term reversal, 60 months
DEF	Change in forecasted earnings per share	RESID11	Residual momentum, prior 11-month returns
DR6	Change in 6-month momentum	RESID6	Residual momentum, prior 6-month returns
EAR	Earnings announcement returns	RINT	Intermediate momentum
ES1	left-tail momentum, 1% expected shortfall	RS	Revenue surprise
ES5	left-tail momentum, 5% expected shortfall	Share	Shared analyst peer momentum
Geo	Geographic peer momentum	Stand	Standalone-conglomerate momentum
IM	Industry momentum	SUE	Standardized unexpected earnings
NEI	Number of consecutive earnings increases	Tech	Technological peer momentum
R1	1-month momentum reversal	TES	Tax expense surprise
R11	12-month momentum	VAR1	left-tail momentum, 1% VaR
R18	Momentum reversal	VAR5	left-tail momentum, 5% VaR
Panel B: Value vs. growth (44)			
BM	Book to market equity	IRC	Intangible return, cash flow to price
BMIA	Industry-adjusted book to market equity	IRE	Intangible return, earnings to price
BMJ	Book to June-end market equity	IRS	Intangible return, sales to price
BMQ	Quarterly book to market equity	LTG	Analyst long-term growth forecasts
CD	Cash flow to debt	NDP	Net debt to price
DM	Debt to market equity	NDPQ	Quarterly net debt to price
DMQ	Quarterly debt to market equity	NOP	Net payout yield
DP	Dividend yield	NOPQ	Quarterly net payout yield
DPQ	Quarterly dividend yield	OCP	Operating cash flow to price
DUR	Equity duration	OCPQ	Quarterly operating cash flow to price
EBP	Enterprise book to price	OP	Payout yield
EBPQ	Quarterly enterprise book to price	OPQ	Quarterly payout yield
EFP	Analyst earnings forecast to price	Q	Tobinâ€™s q
EM	Enterprise multiple	SG	Annual sales growth
EMQ	Quarterly enterprise multiple	SGQ	Quarterly sales growth
EP	Earnings to price	SGR	5-year sales growth rank
EPQ	Quarterly earnings to price	SP	Sales to price

Table : **Firm characteristics** (continued)

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
FCB	Free cash flow to book equity	SPQ	Quarterly sales to price
G5Y	Forecasted growth in 5-year earnings per share	VHP	Intrinsic value to market equity
IRB	Intangible return, book to market		
Panel C: Investment (35)			
ACI	Abnormal corporate investment	DSO	Change in shares outstanding
CDI	Composite debt issuance	DSTI	Change in short-term investments
CEI	Composite equity issuance	DWC	Change in net noncash working capital
CI	Corporate investment	IA	Investment to assets
DA	Asset growth	IAQ	Quarterly investment to assets
DBE	Change in common equity	IG	Investment growth
DCOA	Change in current operating assets	IG2	2-year investment growth
DCOL	Change in current operating liabilities	IG3	3-year investment growth
DDEP	Percent change in depreciation	IGIA	Industry-adjusted percent change in investment
DEP	Depreciation to PP&E	IPO	New equity issues
DFIN	Change in financial assets	IVC	Inventory changes
DFNL	Change in financial liabilities	IVG	Inventory growth
DLNO	Change in long-term net operating assets	NDF	Net debt finance
DLTI	Change in long-term investments	NEF	Net equity finance
DNCA	Change in noncurrent operating assets	NOA	Net operating assets
DNCL	Change in noncurrent operating liabilities	NSI	Net stock issues
DNCO	Change in net noncurrent operating assets	NXF	Net external finance
DNOA	Change in net operating assets		
Panel D: Profitability (52)			
ATO	Asset turnover	DPMIA	Industry-adjusted change in profit margin
ATOQ	Quarterly asset turnover	DROAQ	4-quarter change in return on assets
BL	Book leverage	DROEQ	4-quarter change in return on equity
BLQ	Quarterly book leverage	EPS	Earnings per share
CLA	Cash-based operating profits to lagged book	F	Piotroski fundamental score
CLAQ	Quarterly cash-based operating profits to lagged book assets	FP	Failure profitability
COP	Cash-based operating profits to book assets	FQ	Quarterly Piotroski fundamental score
CR	Credit rating	G	Mohanram growth score
CTO	Capital turnover	GLA	Gross profits to lagged assets
CTOQ	Quarterly capital turnover	GLAQ	Quarterly gross profits to lagged assets
DATO	Change in asset turnover	GPA	Gross profits to assets

Table : **Firm characteristics** (continued)

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
DATOIA	Industry-adjusted change in asset turnover	IPM	Pre-tax income to sales
DPM	Change in profit margin	O	Ohlson O-score
OLA	Operating profits to lagged book assets	RNAQ	Quarterly return on net operating assets
OLAQ	Quarterly operating profits to lagged book assets	ROA	Return on assets
OLE	Operating profits to lagged book equity	ROAQ	Quarterly return on assets
OLEQ	Quarterly operating profits to lagged book equity	ROC	Cash productivity
OPA	Operating profits to book assets	ROE	Return on equity
OPE	Operating profits to book equity	ROEQ	Quarterly return on equity
OQ	Quarterly Ohlson O-score	ROIC	Return on invested capital
PCM	Sales minus costs of goods sold to sales	SAT	Sales to total assets
PM	Profit margin	SATIA	Industry-adjusted sales to total assets
PMIA	Industry-adjusted profit margin	TBI	Taxable income to book income
PMQ	Quarterly profit margin	TBIQ	Quarterly taxable income to book income
PROF	Gross profitability to book equity	Z	Altman Z-score
RNA	Return on net operating assets	ZQ	Quarterly Altman Z-score
Panel E: Intangibles (92)			
ACQ	Accrual quality	DIS	Dispersion in analyst earnings forecasts
ADM	Advertising expense to market equity	DIVI	Dividend initiation
AGE	Firm age	DIVO	Dividend omission
ALA	Liquidity of book assets	DLD	Growth in long-term debt
ALAQ	Quarterly liquidity of book assets	DLG	Dispersion in analyst long-term growth forecasts
ALM	Liquidity of market assets	DLS	Disparity between long- and short-term earnings growth forecasts
ALMQ	Quarterly liquidity of market assets	DQUICK	Percent change in quick ratio
ANA	Analyst coverage	DSA	Percent change in sales minus percent change in accounts receivable
AOA	Absolute value of operating accruals	DSI	Percent change in sales minus percent change in inventories
AOP	Analyst optimism	DSIV	Percent change in sales to inventories
BCA	Brand capital to book assets	DSS	Percent change in sales minus percent change in SG&A
CAL	Current ratio	ECS	Earnings conservatism
CDIND	Convertible debt indicator	EPER	Earnings persistence
CTA	Cash to assets	EPRD	Earnings predictability
DAC	Discretionary accruals	ESM	Earnings smoothness
DANA	Change in analyst coverage	ETL	Earnings timeliness
DCAL	Percent change in current ratio	ETR	Effective tax rate

Table : **Firm characteristics** (continued)

(1) Abbrev.	(2) Characteristic	(3) Abbrev.	(4) Characteristic
DGS	Percent change in gross margin minus percent change in sales	EVR	Value relevance of earnings
FRA	Pension plan funding to book assets	RA610	Years 6-10 lagged returns, annual
FRM	Pension plan funding to market equity	RCA	R&D capital to book asset
GAD	Growth in advertising expense	RDIND	R&D increase
GIND	Corporate governance	RDM	R&D expense to market equity
HA	Industry concentration in total assets	RDMQ	Quarterly R&D expense to market equity
HE	Industry concentration in book equity	RDS	R&D expense to sales
HN	Hiring rate	RDSQ	Quarterly R&D expense to sales
HNIA	Industry-adjusted hiring rate	RER	Real estate ratio
HS	Industry concentration in sales	RN1	Year 1 lagged return, nonannual
KZ	Kaplan-Zingales index of financing constraints	RN1115	Years 11-15 lagged returns, nonannual
KZQ	Quarterly Kaplan-Zingales index of financing constraints	RN1620	Years 16-20 lagged returns, nonannual
LBP	Leverage component of book to price	RN25	Years 2-5 lagged returns, nonannual
LFE	Labor force efficiency	RN610	Years 6-10 lagged returns, nonannual
OA	Operating accruals	SA	SA index of financing constraints
OB	Order backlog	SC	Sales to cash
OCA	Organizational capital to book assets	SDD	Secured debt to total debt
OCAIA	Industry-adjusted organizational capital to book assets	SDIND	Secured debt indicator
OL	Operating leverage	SIN	Sin stocks
OLQ	Quarterly operating leverage	SIV	Sales to inventories
PAFE	Predicted analyst forecast error	SR	Sales to receivables
PDA	Percent discretionary accruals	TA	Total accruals
POA	Percent operating accruals	TAN	Tangibility of assets
PTA	Percent total accruals	TANQ	Quarterly tangibility of assets
QUCIK	Quick ratio	VCF	Cash flow volatility
RA1	Year 1 lagged return, annual	VOA	Accrual volatility
RA1115	Years 11-15 lagged returns, annual	VROA	Earnings volatility
RA1620	Years 16-20 lagged returns, annual	WW	Whited-Wu index of financing constraints
RA25	Years 2-5 lagged returns, annual	WWQ	Quarterly Whited-Wu index of financing constraints
Panel F: Trading frictions (57)			
AMI	Absolute return to volume	BETADAILY	CAPM beta using daily returns
AT	Total assets	BETADOWN	Downside beta
ATQ	Quarterly total assets	BETAEW	CAPM beta using daily returns and equal-weighted market excess return

Table : **Firm characteristics** (continued)

(1)	(2)	(3)	(4)
Abbrev.	Characteristic	Abbrev.	Characteristic
BETAC	CAPM beta	BETAEWSQ	CAPM beta squared
BETAD	Dimson beta	BETAFF	Fama-French 3-factor beta
BETAFP	Frazzini-Pedersen beta	IVCA	Idiosyncratic volatility from the CAPM
BETAHS	Hong-Sraer beta	IVEW	Idiosyncratic volatility using equal-weighted market excess return
BETALCC	Acharya-Pedersen liquidity beta, illiquidity-illiquidity	IVFF	Idiosyncratic volatility from the Fama-French 3-factor model
BETALCR	Acharya-Pedersen liquidity beta, illiquidity-return	IVQ	Idiosyncratic volatility from the q-factor model
BETALEV	Financial intermediary leverage beta	LM1	Turnover-adjusted number of zero daily trading volume
BETALRC	Acharya-Pedersen liquidity beta, return-illiquidity	LM12	Prior 12-month turnover-adjusted number of zero daily trading volume
BETALSY	Liu-Stambaugh-Yuan beta	LM7	Prior 6-month turnover-adjusted number of zero daily trading volume
BETANET	Acharya-Pedersen net liquidity beta	MDR	Maximum daily return
BETAPS	PÃ¡stor-Stambaugh liquidity beta	ME	Market equity
BETARET	Acharya-Pedersen liquidity beta, return-return	MEIA	Industry-adjusted market equity
CS1	Coskewness, 1 month	PIN	Probability of information-based trading
CS60	Coskewness, 60 months	PPS	Price per share
CVD	Coefficient of variation of dollar trading volume	SBA	Bid-ask spread
CVT	Coefficient of variation of share turnover	SHL	High-low bid-ask spread
D1	Price delay based on R2	SUV	Standardized unexplained volume
D2	Price delay based on slopes	SV	Systematic volatility risk
D3	Price delay based on adjusted slopes	TAIL	Tail risk
DTO	Detrended turnover minus market turnover	TS	Total skewness
DTV	Dollar trading volume	TUR	Share turnover
HIGH52	52-week high price	TV	Total volatility
ISC	Idiosyncratic skewness from the CAPM	VDTV	Volatility of dollar trading volume
ISFF	Idiosyncratic skewness from the Fama-French 3-factor model	VEA	Abnormal earnings announcement volume
ISQ	Idiosyncratic skewness from the q-factor model	VT	Volume trend
		VTUR	Turnover volatility

Section OC. Details of the Machine Learning Models

In this section, I briefly describe the basic principles and strengths of machine learning models considered in my paper. More details of models can be found from Hastie et al. [2009] and Goodfellow et al. [2016].

A. Dimension reduction models

The principal component analysis (PCA) is to maximize the common variation across all the characteristics and its first K principal components represent the strongest variables that explain the variations of the P characteristics. The partial least square (PLS) approach extracts some stock characteristics from all stock characteristics according to its covariance with future stock returns and chooses a linear combination of the stock characteristics that is optimal for forecasting. While the PCA model maximally represents the total variations of predictors, it ignores the forecasting target and therefore, is an unsupervised learning technique for dimension reduction. In contrast, the scaled PCA (SPCA) is designed to use the target information to guide dimension reduction. A characteristic with strong forecasting power receives a larger weight, whereas a characteristic with weak forecasting power receives a smaller weight.

B. Penalized linear regressions

The penalized linear model is a generalization of the OLS linear regression model. When there is a large number of predictors, the OLS tends to have good in-sample performance (small bias in the terms of machine learning) and bad out-of-sample performance (large variation in the terms of machine learning). Furthermore, the OLS can generate significant loadings on a large number of independent variables, making the interpretation of the model difficult. One class of models, the shrinkage models, generalize the OLS by imposing a penalty on the number and size of non-zero coefficients in the estimation, effectively limiting the model to focus on a subset of the independent variables and achieving dimension reduction.

The Elastic-Net model, introduced by Zou and Hastie [2005], is a shrinkage model in which the penalty function is a linear combination of L1 and L2 norms of the coefficients. The Elastic-Net model is also a generalization of the well-known LASSO and Ridge regression models. In general, the LASSO model tends to select a few strong predictors while setting the coefficients of other predictors to essentially zero, but can make random choices among several strong and correlated variables. The Ridge model usually includes more predictors and shrink the coefficients of correlated variables together. The Elastic-Net model strikes a balance between these characteristics, allowing both a selection of strong features and the averaging of correlated features.

C. Regression trees

Regression trees can approximate any a priori unknown function while keeping the interpretation from a recursive binary tree. However, with more than two inputs, the interpretation is less obvious as trees. Finding the optimal partition by using a least squares procedure is generally infeasible, however. I thus follow Friedman et al. [2001] and implement a gradient boosting procedure. Gradient boosting in a tree context boils down to combining several weak trees of shallow depth.

Boosting is a technique for reducing the variance of the model estimates and increasing precision. However, trees are "grown" in an adaptive way to reduce the bias, and thus are not identically distributed. An alternative procedure would be to build a set of de-correlated trees which are estimated separately and then averaged out. Such modeling framework is known in the machine learning literature as "Random Forests" (see Bushee [2001]). It is a substantial modification of bagging (or bootstrap aggregation) whereby the outcome of independently drawn processes is averaged to reduce the variance estimates. Bagging implies that the regression trees are identically distributed – that is the variance of the average estimates, as the number of simulated trees increases, depends on the variance of each tree times the correlation among the trees. Random forests aim to minimize the variance of the average estimate by minimizing the correlation among the simulated regression trees.

I also consider an extended version of the random forest procedure which is called "Extremely Randomized Trees" (Geurts et al. [2006]). While similar to ordinary random forests, in that they still represent an ensemble of individual trees, extreme trees have two main distinguishing features: first, each tree is trained using the whole training sample (rather than a bootstrap sample); and second, the top-down splitting in the tree learner is randomized. That means that instead of computing the optimal cut-point locally for each input variable under consideration, a random cut-point is selected. In other words, with extreme trees the split of the trees is stochastic; with random forests the split is instead deterministic.

Tree-based methods such as Gradient Boosted Regression Trees or Random Forests are essentially modifications of a universal underlying algorithm utilized for the estimation of regression trees, commonly, that is the Classification and Regression Tree (CART) algorithm (Breiman et al.

[1984]). Random Forests consist of trees populated following an algorithm like CART, but randomly select a sub-set of predictors from the original data. In this manner, the individual trees in the forest are de-correlated and overall predictive performance relative to a single tree is increased. The hyperparameters to be determined by cross-validation include first and foremost the number of trees in the forest, the depth of the individual trees and the size of the randomly selected sub-set of predictors. Generally, larger forests tend to produce better forecasts in terms of predictive accuracy. GBRTs are based on the idea of combining the forecasts of several weak learners. The GBRT comprises of trees of shallow depth that produce weak predictions stand-alone, however, tend to deliver powerful forecasts when aggregated adequately.

D. Neural networks

The neural networks models, initial motivated by the neuron structures in the brains of humans and animals, blossomed after breakthroughs in algorithms and computing power (LeCun et al. [2015]. Neural networks models, also called deep learning models, have become some of the most powerful models and achieved near- or super-human capabilities in a wide variety of applications, such as natural language processing, speech recognition, computer vision, game playing, and autonomous driving.

There are many different architectures of neural networks, such as the simplest Feed-forward Neural Networks for straightforward classification tasks, the Convolutional Neural Networks for image and pattern recognition, and Recurrent Neural Networks (RNN) that can process sequential data such as speech and text. Long-Short Term Memory (LSTM) Neural Networks are a special type of RNN that is the key to the many successes of RNN, including speech recognition, language modeling, and translation.

In a neural network, there are nodes (neurons) that are connected to each other. There are three types of nodes: input nodes that are used to receive data; output nodes that produce desired outcomes or predictions; and intermediate nodes that process the data from input nodes and convert them to outputs. The connections of the nodes determine the structure of the neural network and its features. RNNs are neural networks with loops, or nodes that are connected to themselves.

LSTM networks are introduced by Hochreiter and Schmidhuber [1997] to solve the problem that standard RNNs have trouble retaining "memory" of the much earlier parts of sequential input data, when processing the later parts of the data. Since sequential data may have long-term dependencies, i.e., parts far away in the sequence may be related, it is important to have "long-term memory" to handle them. LSTM networks have a sequence of nodes that are specifically designed to retain long-term information and update it continuously with new information in a flexible way. As a result, LSTM can capture both short-term and long-term relations in sequential or time-series data very well, suggesting its potential applications in financial economics given the abundance of time-series financial data.