Interpreting Cross-Section Returns of Machine Learning Models: Firm Characteristics and Moderation Effect through LIME *

Zequn Li† Xiaoxia Lou‡ Ying Wu§ Steve Yang¶

May 15, 2024

Abstract

Our study introduces a novel framework to interpret machine learning asset pricing models through the Local Interpretable Model-agnostic Explanations (LIME) method. This methodology illuminates how the inclusion of LIME local coefficients, representing the interaction among characteristics within ML models, modifies the relationship between a firm characteristic and stock returns. The empirical results underscore the significance of incorporating moderation effects into portfolio analysis. Our results present that certain firm characteristics exhibit varying long-short portfolio performance across LIME groups, suggesting their predictive power is specific to certain asset segments. These findings deepen our understanding of the complexities in cross-sectional stock returns, uncovering the detailed dynamics between firm characteristics and their return effects, and distinguishing our research from existing studies.

*We express our deepest gratitude to Professor Majeed Simaan and Professor Hui Wang for their invaluable feedback and comments. Special thanks to Shihao Gu, Bryan Kelly, and Dacheng Xiu for their generosity in sharing their data and code.

†School of Business, Stevens Institute of Technology, Email: zli61@stevens.edu

‡Alfred Lerner College of Business and Economics, University of Delaware, Email: lous@udel.edu

§School of Business, Stevens Institute of Technology, Email: ywu4@stevens.edu

¶School of Business, Stevens Institute of Technology, Email: syang14@stevens.edu

Keywords: Asset pricing, Machine learning, Factor model, Interpretability
JEL Codes: G12, G17, C12, C23, C55, C58

1 Introduction

In the field of empirical asset pricing, an important component of understanding why different assets have different returns is exploring how interactions among firm characteristics contribute to the cross-section of expected stock returns. This challenge becomes more complex within the “factor zoo” (Cochrane, 2011), where a myriad of characteristics and their potential interactions complicate the analysis. With the rise of machine learning methods in asset pricing, researchers have started to focus on the interpretation of the complex non-linearity of such models. Specifically, this paper focused on the moderation effect, which examines how the relationship between one firm characteristics and stock returns can be changed when other characteristics are considered.

Machine learning has emerged as a promising method in empirical asset pricing literature due to its proficiency in managing high-dimensional data. Additionally, it accommodates nonlinearity and interactions among variables and delivers superior predictive performance (Gu, Kelly, & Xiu, 2020). This capability makes it particularly adept at addressing the challenges presented by the “factor zoo” (Jensen, Kelly, & Pedersen, 2023; Freyberger, Neuhierl, & Weber, 2020; Feng, Giglio, & Xiu, 2020). However, the application of machine learning in asset pricing introduced the “black-box” problem, where the inner mechanism of many models remains opaque. Recognizing the critical need for transparency, researchers in asset pricing have also begun to prioritize the interpretability of the machine learning applications (Cong, Tang, Wang, & Zhang, 2021; Cheng, Dong, & Lapata, 2016; Demirbaga & Xu, 2023). In our research, we employed Local Model-agnostic Explanations(LIME) to enhance the interpretability of machine learning models used in empirical asset pricing. The implementation of LIME involves using a linear model to locally approximate the predictions of the "black-box" machine learning models for each instance. The coefficients of these linear models serve as a measure of sensitivity, indicating how changes in the features affect the prediction. This process allows for a detailed and economically meaningful interpretation of how each firm characteristics, through interactions among others, contribute to stock return predictions.

Our approach to addressing this challenge involves employing machine learning and Local Interpretable Model-agnostic Explanations (LIME) method to interpret the contribution of firm characteristics to cross-section asset returns. Machine learning is promising in empirical asset pricing for its ability to handle complex, high-dimensional data and uncover non-linear relationships and interactions among variables. LIME enhances this by improving the transparency of machine learning models, offering insights into how individual firm characteristics predict stock returns at the local level. Through LIME, we interpret the moderation effects among firm characteristics, providing a clearer understanding of their roles in explaining the cross-section of expected stock returns. This methodology allows us to delve deeper into the "factor zoo", shedding light on the nuanced dynamics that explain the cross-section of stock returns.

Additionally, we explore the concept of moderation effects within the realm of empirical asset pricing. Moderation effects occur when the influence of one variable on another is modified by a third variable, known as the moderator (Baron & Kenny, 1986). This concept is critical for understanding how one firm characteristic's impact on stock returns may change when accounting for other characteristics. Empirical asset pricing literature demonstrates that variables such as industry (Asness, Porter, & Stevens, 2000) and firm size (Chan & Chen, 1991) can act as moderators but often in a simplified structure, where interactions between size or industry and a limited number of firm characteristics were analyzed in a one-to-one linear fashion. In contrast, our study adopts a novel approach by exploring how a single firm characteristic can interact simultaneously and nonlinearly with multiple others within a machine learning model, offering a more complex and comprehensive analysis of interactions. This method stands in contrast to the more common additive models, such as those assumed by Freyberger et al. (2020), which limit interactions to firm size and overlook the complexity of broader interactions. Furthermore, our analysis employs LIME to explore the interactions among characteristics and their impact on stock returns. It diverges from the work of Green, Hand, and Zhang (2017), which focuses on the independent predictive power of individual characteristics without examining their interactions. Specifically, we use the LIME local coefficients derived from machine learning models as moderators, offering a nuanced perspective on model interpretation that distinguishes our work from the existing literature.

Our empirical investigation starts with a comparative analysis between machine learning models and Fama-Macbeth regression. We observe enhanced predictive performance from machine learning, attributed to its capacity for handling nonlinearity and interactions, setting the stage for a deeper exploration of moderation effects. Subsequently, we apply LIME to our machine learning models as a foundational step, streamlining our approach for the nuanced analysis ahead.

Building on the groundwork, our main contribution rests on the introduction of a LIME-adjusted moderation regression framework to understand the moderation effects between firm characteristics and stock returns. This framework involves performing a cross-section moderation regression of stock returns on individual firm characteristics, with LIME's local coefficients acting as moderators. This methodology allows us to dissect both the direct and moderation effects of each firm characteristics on stock returns systematically.

Second, to ensure the integrity and reliability of our findings, we employ rigorous multiple testing and p-value adjustment techniques. This methodological safeguard is essential for discerning genuinely significant predictors of stock returns from the myriad of tested firm characteristics, thus mitigating the risk of false discoveries inherent in high-dimensional data analysis. This process underscores our commitment to statistical rigor and robustness in uncovering meaningful insights into asset pricing.

Finally, we conduct a bivariate dependent sort portfolio analysis, which further demonstrates the practical implications of the moderation effects identified in our study. Specifically, for each firm characteristic, we initially sort stocks based on their LIME local coefficients, and then further sort stocks within each of these LIME coefficient groups based on the firm characteristic. This two-step sorting process is meticulously performed for each firm characteristic, allowing us to construct portfolios that reflect the moderated influences on stock returns.

Our research aims to understand the complex interactions and moderation effects of individual firm characteristics, considering a specific set of characteristics simultaneously included in the machine learning model. Our exploration of firm characteristics similar to Green et al. (2017), simultaneously includes a set of characteristics to assess cross-sectional predictive power. However, our approach diverges by incorporating an analysis of their interactions, an aspect not addressed in their study. Our study explores the complex dynamics among firm characteristics. It diverges from Feng et al. (2020); Freyberger et al. (2020), which examines the incremental contribution of new factors within a high-dimensional set of existing factors.

Our primary focus of utilizing machine learning models is to unpack the complex interactions and moderation effects among firm characteristics, rather than pursuing the "best" predictive model. Building on the premise established by Gu et al. (2020), which highlighted machine learning's potential in empirical asset pricing through comparative analysis, our study seeks to deepen the interpretation of these models' outputs. Unlike other studies, such as those referenced by Moritz and Zimmermann (2016); Feng, He, and Polson (2018); L. Chen, Pelger, and Zhu (2023); Cong, Feng, He, and He (2022), our objective is not to propose new or superior machine learning applications within empirical asset pricing. Instead, we aim to provide a nuanced understanding of the underlying dynamics revealed by machine learning, setting our work apart by focusing on interpretation based on established research foundations.

Our framework is rooted in a cross-sectional examination of stock returns, employing a methodological framework inspired by the Fama-Macbeth style regression for comparative purposes. This approach not only anchors our methodology but also provides a context for highlighting how our interpretations, based on machine learning outputs, diverge from existing literature. While Cong et al. (2022) introduces a novel, interpretable machine learning method tailored for empirical asset pricing, our interpretation strategy is model-agnostic, applicable across various machine learning models that demonstrate strong predictive performance. This flexibility allows us to explore moderation effects inherent in the models, a feature not confined to a single method. Furthermore, although studies like Gu et al. (2020); Demirbaga and Xu (2023) offer model-agnostic interpretations focusing on identifying the most important firm characteristics in predictions, our analysis uniquely concentrates on the direct and moderation effects within the cross-section of stock returns. We do not prioritize the importance of one characteristic over another; instead, we deem all characteristics significant to the extent that they exhibit statistically significant direct and moderate effects. This nuanced interpretation sets our work apart, emphasizing a comprehensive understanding of the roles these characteristics play in asset pricing models.

Additionally, our framework evaluates both the collective and individual impacts of firm characteristics, acknowledging that moderation effects depend on the specific combination of characteristics in the model. This individualized testing approach necessitates careful consideration of multiple testing and data-snooping challenges. In line with Harvey, Liu, and Zhu (2016), we adopt a multiple-testing framework to address these issues diligently, ensuring the integrity of our findings. Similar to Green et al. (2017), we apply p-value adjustments to control the false discovery rate, thereby reinforcing the statistical robustness of our results.

Anchored in the Fama-MacBeth regression framework and enhanced by Local Interpretable Model-agnostic Explanations (LIME), our study gains further depth by incorporating moderation effects as interpreted from machine learning models. Our study reveals that an expanded set of firm characteristics becomes significant after considering these moderation effects, indicating that machine learning models with enhanced predictive performance tend to identify a greater number of characteristics as significant. Building on this foundation, our analysis uncovers that certain firm characteristics, initially not significant in traditional analyses, exhibit substantial significance within the context of moderation effects. Characteristics such as dollar trading volume, 12-month momentum, size, return volatility, maximum daily return, and 6-month momentum, emerge as notably significant when analyzed through the lens of machine learning models considering moderation effects. This shift highlights the added value of incorporating moderation effects in uncovering the nuanced roles these characteristics play in predicting cross-sectional stock returns.

Our bivariate dependent sort portfolio analysis highlights how moderation effects intricately influence firm characteristics' predictive power. For example, size exhibits divergent performances across LIME groups: a univariate sort yields a modest return of -0.726 (t-statistic of -2.394), but dependent sorting within specific groups reveals an average return of 3.791 with a t-statistic of 7, indicating that its predictive power is concentrated in particular asset segments. Similarly, 6-month momentum, initially showing weak predictive strength with an average return of 0.36 (t-statistic of 1.4), demonstrates significant variances through dependent sorting: the lowest LIME coefficient group sees a -1.03 return (t-statistic of -3.13), while the highest group achieves a 1.36 return (t-statistic of 4.39). Such findings indicate that univariate sorts may obscure substantial, opposing effects within different segments. This nuanced analysis, extendable to more firm characteristics, underscores moderation effects' vital role in accurately evaluating firm characteristics' predictive capabilities.

Our paper offers a methodological contribution by developing an innovative framework that combines interpretable machine learning to systematically analyze the moderation effects of firm characteristics on predicting cross-sectional stock returns. This new approach makes the application of machine learning in this field more transparent, revealing complex interactions and non-linear relationships that were previously overlooked. Our empirical findings offer novel insights, especially on how moderation effects can inform future asset pricing research. These insights add new dimensions to our understanding of how firm characteristics explain cross-sectional stock returns. Furthermore, we demonstrate the practical value of our research through a bivariate dependent sort portfolio analysis. This method illustrates how our findings can enhance investment strategies and portfolio management.

2 Firm Characteristics and Cross-Sectional Expected Returns

Consider an economy with N_t stocks at each discrete time $t = 1, \cdots, T$. For any given firm i at time t, we denote its next-period excess return from time t to time $t+1$ by $r_{i,t+1}$. The general factor model for an excess return is of the form

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t}' f_{t+1} + \epsilon_{i,t+1} \tag{1}$$

where $\beta_{i,t}$ is a J-dimensional vector representing the risk exposures at time t, f_{t+1} is a J-dimensional vector of common factor returns, and $\epsilon_{i,t+1}$ represents the scalar idiosyncratic risk for firm i at time $t+1$. The idiosyncratic risk is assumed to have a mean of zero for all firms at any time and is uncorrelated with the common factor returns, i.e., $\mathbb{E}_t[\epsilon_{i,t+1}] = \mathbb{E}_t[\epsilon_{i,t+1} f_{t+1}] = 0$, and the expected value of the common factor returns is denoted as $\mathbb{E}_t[f_{t+1}] = \lambda_t$.

Taking the conditional expectation of both sides of Equation (1)and imposing the non-arbitrage restriction (which states that $\alpha_{i,t} = 0$ to prevent arbitrage opportunities), the model can be simplified to:

$$\mathbb{E}_t[r_{i,t+1}] = \beta_{i,t}' \lambda_t = \sum_{j=1}^{J} \beta_{i,t}^{(j)} \lambda_t^{(j)}, \tag{2}$$

where $\mathbb{E}_t[r_{i,t+1}]$ represents the expected excess return of firm i from time t to $t+1$.

Informed by the factor model described in Equation (2), we understand that the variations in expected returns across assets come from the differences in risk exposures. This framework forms the basis of our investigation into firm characteristics. Previous studies have established a strong connection between observable firm characteristics and the variation in expected stock returns. Our research aims to extend these findings by thoroughly examining how these firm characteristics, serving as proxies for risk exposures, help explain the observed differences in stock returns, particularly through their interactions.

Let us denote the firm-level characteristics observed at time t for firm i as $c_{i,t}$, where $c_{i,t} = (c_{i,t}^{(1)}, \cdots, c_{i,t}^{(K)})'$ represents a vector of K characteristics for each firm. These characteristics are assumed to be orthogonal to the idiosyncratic shock $\epsilon_{i,t+1}$, implying that the expected value of $\epsilon_{i,t+1}$ given $c_{i,t}$ is zero, i.e., $E[\epsilon_{i,t+1}|c_{i,t}] = 0$. It is important to note that the number of observable firm characteristics (K) does not necessarily equate to the number of factors (J) in our model.

2.1 Current Methods

We first introduce the conventional methodology employed to examine the cross-sectional relationship between future stock returns $r_{i,t+1}$ and individual firm characteristics $c_{i,t}^{(k)}$. This lays the groundwork for our proposed framework, wherein we evaluate each firm characteristics, while also considering their interactions with other characteristics.

2.1.1 Univariate Portfolio Analysis

Portfolio analysis is widely used to explore the relationship between firm characteristics and future stock returns cross-sectionally. This method involves creating portfolios based on varying levels of a firm characteristic and analyzing the returns of these portfolios over time. Specifically, we focus on the performance difference between the portfolios at the extreme ends of the characteristic spectrum for each time period t. The goal is to determine if this difference, observed over time, significantly deviates from zero. A significant deviation from zero suggests a consistent relationship between the chosen firm characteristic and future stock returns across time periods.

Portfolio sorting, while straightforward and intuitive, is not without significant drawbacks, such as the curse of dimensionality. This issue arises as the quantity of portfolios increases exponentially with the addition of more characteristics. To illustrate, consider the scenario where there are 10 distinct characteristics, each divided into 10 quantiles. In this case, the total number of possible portfolios escalates to 10^{10}.

2.1.2 Fama-MacBeth Regression

Cochrane (2011) highlight that portfolio sorting and nonparametric cross-sectional regressions share fundamental similarities. Freyberger et al. (2020) advocate for the utilization of cross-sectional rankings of firm characteristic values. Specifically, for each characteristic, they introduce the notation $\tilde{c}_{i,t}^{(k)}$ to represent the rank transformation of $c_{i,t}^{(k)}$, normalizing the distribution of characteristics across firms within the unit interval, such that $\tilde{c}_{i,t}^{(k)} \in [-1, 1]$. This mathematical framework underpins our analysis, aligning with the established equivalence of the methodologies discussed.

Parallel to the portfolio analysis, nonparametric Fama-MacBeth regression analysis is an alternative statistical method designed to examine the cross-sectional relation between firm characteristics and future returns. Fama-MacBeth regression analysis is implemented using a two-step procedure. The first step is to run periodic cross-sectional regression of the future return $r_{i,t+1}$ on transformed firm characteristics $\tilde{c}_{i,t}^{(k)}$ for each period t.

$$r_{i,t+1} = \delta_{0,t} + \sum_{k=1}^{K} \delta_{k,t} \tilde{c}_{i,t}^{(k)} + \epsilon_{i,t} \tag{3}$$

The second step is to compute the time-series average of the cross-sectional regression coefficients $\delta_{k,t}$ for each firm characteristic k. To examine whether a firm characteristic $c_{i,t}^{(k)}$ can predict the cross-sectional stock return, we test whether the average coefficient $\bar{\delta}_k = \frac{1}{T} \sum_{t=1}^{T} \delta_{k,t}$ is statistically different from zero. The economic magnitude of the cross-sectional relation between firm characteristic $c_{i,t}^{(k)}$ and expected future return is reflected by the average of the coefficient $\bar{\delta}_k$. This average represents the expected change in future returns resulting from a one-unit change in $\tilde{c}_{i,t}^{(k)}$.

This prevailing framework typically assumes an additive model. A critical limitation of any additive model is articulated in the following restriction:

$$\frac{\partial r_{i,t+1}}{\partial c^{(k)} \partial c^{(k')}} = 0 \tag{4}$$

for all $k \neq k'$. This constraint signifies that additive models do not accommodate interdependencies between characteristics unless explicitly specified.

2.2 Machine Learning in Empirical Asset Pricing

In empirical asset pricing, machine learning offers a markedly different approach. Rather than assume a restricted functional form between the future return and the firm characteristics, machine learning directly targets the estimation of the conditional expected return function $g(\cdot)$ as a black box, devoid of any presupposed functional form. This is encapsulated in the equation:

$$\mathbb{E}_t[r_{i,t+1}] = g(\tilde{c}_{i,t}) \tag{5}$$

Here we impose the same assumptions as Gu et al. (2020):

a. The function $g(\cdot)$ depends neither on i nor t. The model maintains the same form over time and across different stocks.

b. The function $g(\cdot)$ depends on $\tilde{c}$ only through $\tilde{c}_{i,t}$. The prediction does not use information from history before time t or from individual stocks other than i.

The machine learning model, by approximating the conditional mean function without assuming any specific functional form, generalizes the non-parametric Fama-MacBeth regression. Its flexibility surpasses that of the portfolio sorts method, as it does not confine returns within a portfolio. Additionally, it transcends the limitations of the regression approach by not adhering to an additive structure between characteristics and excess return, and by allowing for cross-dependencies between characteristics. However, this flexibility comes at the cost of economic interpretability, notably in terms of explicating the nature of risk exposures β and risk premium λ.

Gu et al. (2020) conducted an extensive analysis of various machine learning methods within the context of empirical asset pricing. Their findings indicate the promise of machine learning in terms of predictive performance. Nonetheless, the challenge of interpretability remains a significant hurdle in the application of these models. Our research contributes to this discourse by introducing a novel approach to interpreting the cross-sectional relationship between firm characteristics and expected returns.

2.3 Interpreting Machine Learning Models with LIME

LIME represents a groundbreaking algorithm that elucidates the predictions of any machine learning model by approximating it locally with an interpretable, typically linear, model (Ribeiro, Singh, & Guestrin, 2016b). Formally, we define an explanation of a prediction from a machine learning model g for stock i at time t as an interpretable linear model $g_{i,t}(z)$. z is a

K-dimensional vector representing the same set of observable firm characteristics as $\tilde{c}_{i,t}$. However, z is not observed from the real world but simulated from the historical distribution. The proximity measure $\pi_{\tilde{c}_{i,t}}(z)$ quantifies the closeness of an instance z to $\tilde{c}_{i,t}$, so as to define the locality $\pi_{\tilde{c}_{i,t}}$ around $\tilde{c}_{i,t}$.

$$\mathbb{E}[g(z)] = g_{i,t}(z) \quad z \in \pi_{\tilde{c}_{i,t}} \tag{6}$$

$$g_{i,t}(z) = a_{i,t} + \sum_{k=1}^{K} b_{i,t}^{(k)} z^{(k)} \tag{7}$$

We employ the loss function $\mathcal{L}(f, g, \pi)$ be a measure of how effectively $g_{i,t}(\cdot)$ approximates $g(\cdot)$ within the defined locality π_x. The explanation $g_{i,t}(\cdot)$ is obtained by minimizing $\mathcal{L}$. So $\mathcal{L}$ is a loss function, and here we use l_2 loss.

For each observation $\tilde{c}_{i,t}$, we aim to estimate $K+1$ unknown parameters. This involves generating random samples z_m, $m = 1, \cdots, M$ in the vicinity of $\tilde{c}_{i,t}$. Each sample z_m yields a machine learning prediction $y_m = f(z_m)$, serving as a target for the explanatory model $g_{i,t}$. Using the generated dataset $Z_{i,t} = (z_1, z_2, \cdots, z_M)'$ around $\tilde{c}_{i,t}$ and associated labels $Y_{i,t} = (y_1, y_2, \cdots, y_M)'$, where $y_m = g(z_m)$, an OLS regression is conducted to obtain the local intercept $a_{i,t}$, and local coefficients $b_{i,t}$. The regression equation is represented as:

$$y_m = a_{i,t} + \sum_{k=1}^{K} b_{i,t}^{(k)} z^{(k)} + e_m \quad z_m \in \pi_{c_{i,t}} \tag{8}$$

It's important to note that while z_m represents a set of firm characteristics, an actual firm possessing this exact set may not exist. The value y_m denotes the prediction from the machine learning mode g differing from the true excess return r_m, which remains unobservable to us.

In the context of a complex, black-box machine learning mode g, the functional form and parameters remain elusive and non-interpretable. However, with the application of LIME, the local linear surrogate model and its parameters become accessible for interpretation. A machine learning model g approximates the conditional mean function of the expected return, while the local surrogate model $g_{i,t}$ approximates g in the vicinity of observation $c_{i,t}$. The coefficients $b_{i,t}$ in local model $g_{i,t}$ are not time-varying variables but constant parameters linked to the specific observation $c_{i,t}$

$$\mathbb{E}_t[r_{i,t+1}] = g(c_{i,t}) = g_{i,t}(c_{i,t}) + e_{i,t} \tag{9}$$

While the local surrogate model $g_{i,t}(\cdot)$ provides a means to approximate expected excess returns, it potentially introduces a larger error term compared to the original machine learning model $g(\cdot)$. Thus, while we utilize the machine learning model for predicting expected returns, the local coefficients $b_{i,t}^{(k)}$ offer valuable insights for analyzing firm characteristics.

2.4 Moderation Effect

Moderation analysis, similar to including an interaction term in regression, explores when or under what conditions a relationship between two variables exists or changes in strength or direction. In this context, the moderator, $\mathbf{W}$, interacts with the relationship between $\mathbf{X}$ and $\mathbf{Y}$ to identify when $\mathbf{X}$ influences $\mathbf{Y}$. In our asset pricing framework, $\mathbf{Y}$ represents future stock returns ($r_{i,t+1}$), and $\mathbf{X}$ denotes a specific firm characteristic ($c^{(k)}$). The LIME local coefficient for the same firm characteristic, $b_{i,t}^{(k)}$, acts as a moderator by summarizing the feature interactions within the machine learning model. Rather than directly regressing $r_{i,t+1}$ on $c_{i,t}^{(k)}$ to examine the relationship between stock returns and firm characteristics, we also incorporate the LIME local coefficient $b_{i,t}^{(k)}$ as a moderator.

The baseline regression model is univariate Fama-MacBeth regression in Equation (3)

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \varepsilon \tag{10}$$

where $\delta_{k,t}$ quantifies the direct impact of firm characteristic k on the one-month-ahead stock return. The change in the estimate of excess return for a one-unit change in firm characteristic k is $\delta_{k,t}$. If $\delta_{k,t}$ is statistically significantly different from 0, we conclude that the firm characteristic k significantly predicts future expected returns.

To account for the moderation effect suggested by the machine learning model, we conduct a LIME-adjusted regression:

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \gamma_{k,t} b_{i,t}^{(k)} c_{i,t}^{(k)} + \xi_{k,t} b_{i,t}^{(k)} + \varepsilon \tag{11}$$

where the adjusted estimate of excess return for a one-unit change in firm characteristic k becomes $\delta_{k,t} + \gamma_{k,t} b_{i,t}^{(k)}$. The influence of firm characteristic k on future excess returns depends on other firm characteristics, as captured by the LIME local coefficient $b_{i,t}^{(k)}$.

In this case, we examine two coefficients: $\delta_{k,t}$ for the direct impact, and $\gamma_{k,t}$ for the moderation effect.

3 Empirical Results

Our empirical analysis aims to identify how many and which among a set of firm characteristics significantly influence one-month-ahead returns, particularly when we challenge the traditional assumptions of independence and linearity. To achieve this, we utilize machine learning models, which allow nonlinearity and interaction to include all firm characteristics in the model simultaneously, and interpret the black-box model with the LIME local coefficient. Following this, we perform a cross-section LIME-adjusted moderation regression, as detailed in Equation (13).

3.1 Dataset

Our sample period for the empirical study spans 58 years, from January 1964 to December 2021. We strategically divide the full dataset into three distinct segments: a 13-year training sample (1964-1976), a 12-year validation sample (1977-1988), and a 33-year out-of-sample testing sample (1989-2021).

For our machine learning models, we adopt an annual refitting strategy. Each year, we expand the training sample by one year and roll the validation sample forward by the same duration, maintaining a consistent validation period of 12 years.

Monthly equity return data is sourced from CRSP, and we use the one-month Treasury bill rates from the Kenneth French Data Library as the risk-free rate for calculating excess returns.

The firm-level characteristics used in our analysis match those referenced by Gu et al. (2020) and Green et al. (2017)[1]. For these characteristics, we conduct a monthly cross-sectional ranking, transforming these rankings into quantiles and normalizing them to fit within the $[-1, 1]$ range. To manage missing values, we substitute them with the cross-sectional median for each stock within the same month, ensuring our dataset's consistency and integrity. It's important to highlight that early historical records do not include some firm characteristics of all stocks. When we address this by imputing a median value of 0 for missing entries before performing cross-sectional linear regression, we encounter a challenge: the firm characteristics become uniformly valued (i.e., missing = 0). To enable a meaningful comparison between our machine learning models and linear models with the same set of characteristics, we selectively reduced our dataset from 94 to 72 characteristics. This selection process, which deviates from the approaches of Gu et al. (2020) and Green et al. (2017), maximizes our sample size. The 72 chosen characteristics are detailed in Table C.1.

3.2 Machine Learning Models

In our empirical analysis, we utilized two examples of machine learning models to interpret the implied moderation effects. The first, a Neural Network model with three hidden layers (NN3), excels at identifying complex nonlinearities and interactions, making it highly effective for analyzing the intricate

[1] The firm-level characteristics data are available from Dacheng Xiu's website.

dynamics of stock returns. The second, a Random Forest model (RF), combines multiple decision trees to enhance prediction accuracy. This model is celebrated for its robust approach to nonlinearity and ability to integrate varied predictions, offering a comprehensive view of the outcomes.

In the study conducted by Gu et al. (2020), both the NN3 and RF models are identified as top-performing techniques. Our research focuses on interpreting these advanced black-box models. It's important to note that when we mention the NN3 and RF models, we are referring to the specific versions we have trained, rather than the general methodologies. Our conclusions are drawn from the results of our specific implementations, not from the general techniques themselves. Although our NN3 and RF models were trained using methods very similar to those employed by Gu et al. (2020), they may differ significantly from Gu et al. (2020)'s models and consequently yield distinct predictions.

We compare the predictive performance of traditional linear models with machine learning approaches, specifically Neural Networks (NN3) and Random Forest (RF), against a naive benchmark. Consistent with existing literature, our findings confirm the efficacy of machine learning models in financial prediction. Both NN3 and RF models demonstrated superior predictive ability over linear models, with NN3 exhibiting the highest overall predictive performance based on out-of-sample R-squared values. Detailed results of the comparative analysis in terms of predictive performance are included in the Appendix A. The enhanced performance of machine learning models underscores their potential for capturing complex, non-linear relationships in stock returns, motivating our further exploration through LIME for interoperability.

3.3 Multiple Testing and p-value Adjustment

When evaluating if firm characteristics are consistently related to expected returns, we consider the potential for inferential biases due to data snooping. Our analysis risks incorrectly rejecting the null hypothesis for certain characteristics by chance, given our aim to examine the significance of individual firm characteristics. This issue of multiple testing is well-documented in research literature (Harvey et al., 2016; Green et al., 2017). Similar to Green et al. (2017), our concern lies in the possibility of mistakenly identifying a characteristic as significant due to conducting multiple analyses with the same dataset. To mitigate this bias from multiple testing, we adjust p-values for the false discovery rate using Benjamini-Hochberg-Yekutieli's method.

The Benjamini-Hochberg-Yekutieli adjustment is a sequential technique to control the false discovery rate (FDR). It involves comparing each p-value, starting from the smallest, against a threshold determined by its rank and the total number of tests, adjusting for dependencies among test statistics. Specifically, a null hypothesis is rejected if its p-value is less than $\frac{i}{K}\alpha_d$, where i is the p-value's rank. This step-up procedure sorts p-values in ascending order and sequentially tests them against adjusted thresholds. This approach, less stringent than methods controlling the family-wise error rate (FWER), seeks a balance between minimizing Type I and Type II errors.

We adopt a 5% significance level for our analysis, aiming to rigorously identify truly significant firm characteristics while controlling for the risk of false discoveries.

3.4 Lime-adjusted Regression Analysis

We begin by establishing a conventional baseline for comparison with our LIME-adjusted Moderation results, which integrates Machine Learning (ML) for interpretation. This baseline is crucial, serving as a benchmark to gauge the added insights our ML approach contributes. It involves performing univariate Fama-MacBeth regressions as in Equation (12) for each of the 72 firm characteristics, conducting a monthly cross-sectional regression during the out-of-sample period and averaging the coefficients $\delta_{k,t}$ over time, $\bar{\delta}_k = \sum_{\mathcal{T}_3} \delta_{k,t}$. $\mathcal{T}_3$ is the set of testing samples, where the data never enter into model estimation or tuning.

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \varepsilon \tag{12}$$

The analysis presented in Table 1 reveals that, after adjusting for multiple testing through p-values, 16 firm characteristics emerge as statistically significant in the Fama-MacBeth regression analysis (left column of the table). The table also shows the average direct impacts ($\bar{\delta}_k$) and the corresponding t-statistics, analyzed through NN3 and RF models considering the moderation effects (middle and right columns, respectively), for the same set of firm characteristics.

$$r_{i,t+1} = a + \delta_{k,t} c_{i,t}^{(k)} + \gamma_{k,t} b_{i,t}^{(k)} c_{i,t}^{(k)} + \xi_{k,t} b_{i,t}^{(k)} + \varepsilon \tag{13}$$

Upon incorporating moderation effects as interpreted by the NN3 model, 11 out of the 16 characteristics remain significant in their direct effects. However, Industry-adjusted change in profit margin(**chempia**), growth in common shareholder equity(**egr**), employee growth rate(**hire**), industry momentum(**indmom**), and sales to price(**sp**) loses their significance after adjusting for the false discovery rate and considering moderation effects. Similarly, when evaluating direct effects through the RF model, 14 characteristics retain their significance. In this context, industry-adjusted cash flow to price ratio(**cfp_ia**) and 1-month momentum(**mom1m**) do not maintain their statistical significance after adjustments for the false discovery rate and accounting for moderation effects in the RF model.

Table 1: Firm Characteristics Significant in Direct Effect

Fama-MacBeth			NN3			RF		
	mean	t-stats		mean	t-stats		mean	t-stats
agr	0.656	**5.012**	agr	0.358	**3.425**	agr	0.732	**4.834**
cfp_ia	0.319	**3.107**	cfp_ia	0.359	**3.958**	cfp_ia	0.33	3.194
chcsho	-0.389	**-3.792**	chcsho	-0.413	**-3.419**	chcsho	-0.401	**-3.96**
chempia	-0.289	**-3.591**	chempia	-0.194	-2.446	chempia	-0.285	**-3.59**
chinv	-0.343	**-4.441**	chinv	-0.278	**-4.037**	chinv	-0.372	**-4.477**
egr	-0.454	**-3.736**	egr	-0.283	-2.583	egr	-0.459	**-3.615**
grcapx	-0.339	**-4.341**	grcapx	-0.356	**-4.502**	grcapx	-0.344	**-4.369**
grltnoa	-0.418	**-4.279**	grltnoa	-0.337	**-3.854**	grltnoa	-0.435	**-4.152**
hire	-0.397	**-4.564**	hire	-0.251	-3.000	hire	-0.406	**-4.729**
indmom	0.408	**3.273**	indmom	0.452	2.905	indmom	0.648	**3.796**
invest	-0.411	**-4.024**	invest	-0.364	**-4.215**	invest	-0.459	**-3.841**
lgr	-0.359	**-5.573**	lgr	-0.289	**-4.433**	lgr	-0.38	**-5.568**
mom1m	-0.531	**-4.522**	mom1m	0.667	**3.447**	mom1m	-0.843	-2.084
pctacc	-0.243	**-3.294**	pctacc	-0.222	**-3.206**	pctacc	-0.251	**-3.363**
sgr	-0.39	**-5.401**	sgr	-0.332	**-4.707**	sgr	-0.431	**-5.116**
sp	0.517	**3.193**	sp	0.394	2.391	sp	0.594	**3.233**

This table reports the average univariate regression coefficients ($\bar{\delta}_k$) for firm characteristics that are statistically significant with the baseline model at the 5% level, after Benjamini-Hochberg-Yekutieli (BHY) p-value adjustment for multiple testing. The columns under Fama-MacBeth, NN3, and RF display the average regression coefficients $\bar{\delta}_k$ and their respective t-statistics, averaged over the out-of-sample period. These coefficients quantify the estimated direct linear impact of each firm characteristic on stock returns. T-statistics are provided to assess the statistical significance of the coefficients.

Table 2 showcases additional firm characteristics deemed significant by the NN3 and RF models for their direct effects on future stock returns, taking into account moderation effects. Both models conclusively found that size(**mvel1**) significantly explains one-month-ahead stock returns. Beyond these, the NN3 model highlighted 3 other firm characteristics, dollar trading volume(**dolvol**), 12-month momentum(**mom12m**), and return volatility(**retvol**) with significant direct impacts.

Table 2: Firm Characteristics Significant in Direct Effect

	NN3			RF	
	mean	t-stats		mean	t-stats
dolvol	0.996	**4.733**	mvel1	-0.967	**-4.027**
mom12m	0.686	**4.025**			
mvel1	0.774	**4.426**			
retvol	0.952	**3.619**			

This table presents additional firm characteristics that have been identified as significant at the 5% significance level, after adjusting for multiple testing using the Benjamini-Hochberg-Yekutieli (BHY) procedure, within NN3 and RF LIME-adjusted moderation regressions. The table enumerates the mean coefficients, which represent the direct effects of these characteristics on stock returns, alongside their associated t-statistics. The values are averaged over the out-of-sample period. The reported t-statistics aid in determining the reliability of the coefficients' estimates, with all listed characteristics surpassing the significance threshold set forth by the BHY adjustment method.

Table 3 illustrates the firm characteristics identified by the NN3 model (Neural Network 3) as having significant moderation effects. According to the NN3 model, 13 firm characteristics exhibit significant moderation effects, with 4 of these, dollar trading volume(**dolvol**), 1-month momentum(**mom1m**), size(**mvel1**), and return volatility(**retvol**) also showing significance in both direct and moderation impacts.

Table 3: Firm Characteristics Significant in Moderation Effect(NN3)

	NN3 $\bar{\delta}_k$		NN3 $\bar{\gamma}_k$	
	mean	t-stats	mean	t-stats
betasq	0.075	0.313	-1.538	**-4.069**
chmom	0.239	2.28	0.533	**4.302**
dolvol	0.996	**4.733**	2.456	**4.221**
ep	0.328	1.253	-1.8	**-3.401**
ill	-0.124	-0.784	-1.152	**-3.246**
maxret	0.737	3.132	1.023	**3.592**
mom12m	0.686	4.025	1.176	**3.335**
mom1m	0.667	**3.447**	1	**4.195**
mom6m	-0.287	-1.688	2.272	**5.878**
mvel1	0.774	**4.426**	1.779	**4.877**
retvol	0.952	**3.619**	2.019	**4.175**
std_dolvol	0.032	0.298	-1.688	**-3.569**
turn	0.575	2.878	0.827	**4.307**

This table presents the NN3 LIME-adjusted regression coefficients, focusing on the moderation effects ($\bar{\gamma}_k$) of firm characteristics that are statistically significant at the 5% level after BHY p-value adjustments for multiple comparisons. We report both the direct linear effects ($\bar{\delta}_k$) and the significant moderation effects ($\bar{\gamma}_k$), along with their corresponding mean coefficients and t-statistics. The bold figures indicate statistically significant values. It is important to note that the table exclusively reports firm characteristics significant in the moderation effect ($\bar{\gamma}_k$), providing a focused view of how these characteristics interact with other variables to influence stock returns within the RF model framework.

Similarly, Table 4 details the findings from the RF model, highlighting 5 firm characteristics with significant moderation effects. They are industry momentum(**indmom**), maximum daily return(**maxret**), 6-month momentum(**mom6m**), size(**mvel1**), and return volatility(**retvol**). Out of which, industry momentum(**indmom**) and size(**mvel1**) exhibit significance in both direct and moderation effects.

Table 4: Firm Characteristics Significant in Moderation Effect(RF)

	RF $\bar{\delta}_k$		RF $\bar{\gamma}_k$	
	mean	t-stats	mean	t-stats
indmom	0.648	**3.796**	-5.817	**-3.384**
maxret	0.873	2.824	5.924	**4.3**
mom6m	-0.057	-0.365	9.137	**3.787**
mvel1	-0.967	**-4.027**	9.507	**4.325**
retvol	0.839	2.704	13.671	**4.964**

This table presents the NN3 LIME-adjusted regression coefficients, focusing on the moderation effects ($\bar{\gamma}_k$) of firm characteristics that are statistically significant at the 5% level after BHY p-value adjustments for multiple comparisons. We report the direct linear effects ($\bar{\delta}_k$) and the moderation effects ($\bar{\gamma}_k$) of firm characteristics. The analysis reports the mean coefficients and their associated t-statistics, which are derived from firm characteristics deemed significant at a 5% level after BHY p-value adjustments for multiple comparisons. Bold figures represent statistically significant coefficients.

Notably, maximum daily return(**maxret**), 6-month momentum(**mom6m**), size(**mvel1**), and return volatility(**retvol**), are recognized for their significant moderation effects by both the NN3 and RF models.

Despite the differences in direct and moderation effect outcomes, which stem from the inherent differences between the NN3 and RF models, a consistent pattern emerges. Certain firm characteristics are robustly identified as significant across both models. This observation underscores a foundational consistency. These characteristics likely represent fundamental aspects of stock returns that are robust across different analytical approaches. However, the divergence in the number of characteristics identified and the specifics of direct and moderation effects reflects the models' unique capabilities and biases. NN3's broader identification suggests a sensitivity to complex, non-linear relationships, while RF's more conservative identification hints at its emphasis on broad, generalizable patterns.

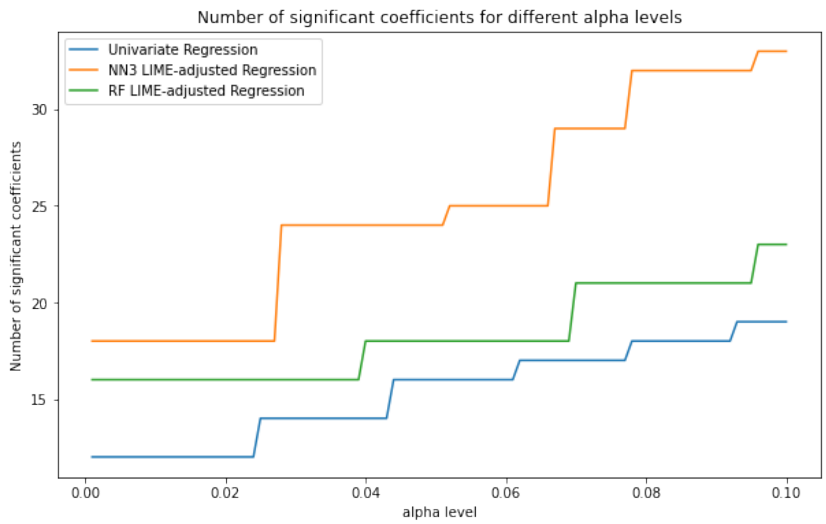

Figure 1: Significant Level Cut Off

This figure depicts the number of firm characteristics identified as statistically significant, either through direct or moderation effects, at varying false discovery rate (FDR) significance levels ranging from 0 to 10%. The analysis compares the baseline univariate regression against more complex models, including NN3 LIME-adjusted moderation regression and RF LIME-adjusted moderation regression. The y-axis quantifies the count of significant characteristics, while the x-axis represents the incrementally adjusted significance thresholds.

Figure 1 illustrates how the number of coefficients recognized as significant varies across different levels of statistical significance. Near the 5% significance level, the trend lines stabilize and become relatively flat. Within a significance level range of $(0, 10\%)$, the NN3 LIME-adjusted moderation regression consistently identifies more significant firm characteristics than both the RF LIME-adjusted regression and the baseline univariate regression model. This finding underscores that machine learning LIME-adjusted regression analysis is capable of uncovering nonlinear and interactive effects that predict future cross-sectional returns—effects that conventional univariate regression analyses may miss. Importantly, this insight remains robust across different significance levels, indicating that the effectiveness of machine learning models in identifying these effects is not significantly influenced by the choice of significance threshold.

3.5 Bivariate Dependent Sort Portfolio Analysis

Portfolio analysis is another common method used to assess the predictability of future returns based on firm characteristics and to understand the cross-sectional relationships among these characteristics. It's established that sorting portfolios by characteristics equates to nonparametric regression (Freyberger et al., 2020). In this section, we aim to incorporate the LIME local coefficient derived from machine learning models(NN3 and RF) into our portfolio analysis to demonstrate the moderation effects.

Parallel to regression analysis, we set a baseline model. Throughout our out-of-sample period, we sort stocks into equal-weighted 5 portfolios based on one firm characteristics. We focus on the performance of zero-cost long-short portfolios, which involve going long on the top quintile and short on the bottom quintile. Significant differences from zero in these long-short portfolio returns suggest the predictive power of the characteristic.

To effectively integrate LIME local coefficients as moderators in our portfolio analysis, we employ bivariate dependent-sort portfolio analysis. This approach allows us to examine the relationship between firm characteristics, LIME local coefficients, and expected returns. Specifically, our analysis focuses on those firm characteristics deemed significant within a LIME-adjusted moderation regression framework, listed in Table 3 and Table 4 for NN3 model and RF model respectively.

For each period analyzed, we started by dividing stocks into 5 distinct groups based on their LIME local coefficient, $b_{i,t}^{(k)}$, and then further sorting them into 5 subgroups by firm characteristics,$c_{i,t}^{(k)}$ within each LIME group. This sorting results in the formation of 5×5 dependent sort equal-weighted portfolios.

We then proceed to calculate the average excess return for each of these 5×5 portfolios over the out-of-sample period, including their respective t-statistics. For each LIME group, we also analyze the average return and t-statistics for the zero-cost long-short portfolios, which are based on going long the highest and short the lowest firm characteristics subgroups.

Table 5 presents the result from the bivariate dependent sort portfolio analysis, focusing on firm characteristics identified as significant in a moderation effect with NN3 model. Each column shows the average return of the long-short portfolio within each LIME group and the corresponding t-statistics. Each row represents a level of LIME local coefficients, with the final row showing the results of the baseline long-short portfolio result without this control. This detailed analysis demonstrates how LIME coefficients adjust the cross-sectional relationship between the firm characteristics and expected returns.

Table 5: Bivariate Dependent Sort Portfolio Analysis(NN3)

	maxret	retvol	betasq	ep	turn	mom6m	mom12m
LIME1	-0.325	-1.011	0.235	0.38	-1.691	-1.027	-0.711
	(-1.013)	(-2.804)**	(0.524)	(0.926)	(-5.427)***	(-3.133)*	(-1.493)
LIME2	0.372	0.041	0.187	0.144	-0.237	0.255	1.29
	(1.445)	(0.133)	(0.516)	(0.485)	(-0.785)	(1.143)	(4.388)***
LIME3	0.512	0.635	0.186	0.164	0.292	0.276	1.231
	(2.046)*	(2.114)*	(0.503)	(0.618)	(0.922)	(1.343)	(5.07)***
LIME4	0.893	0.876	0.093	-0.091	0.325	0.59	1.293
	(3.609)**	(3.191)**	(0.233)	(-0.328)	(1.308)	(2.377)*	(5.225)***
LIME5	1.412	2.823	-0.415	-1.194	0.393	1.36	1.287
	(3.656)**	(5.651)***	(-1.193)	(-2.554)*	(2.259)*	(4.397)***	(3.379)***
baseline	0.020	0.166	0.214	-0.236	-0.281	0.36	0.515
	(0.052)	(0.402)	(0.554)	(-0.667)	(-0.962)	(1.398)	(1.639)

	ill	std_dolvol	chmom	mvel1	dolvol	mom1m
LIME1	2.096	1.095	-0.773	-3.791	-2.776	-3.249
	(4.365)***	(3.948)***	(-3.261)**	(-7.039)***	(-7.646)***	(-5.539)***
LIME2	0.545	0.155	-0.054	-0.17	-0.234	-0.904
	(2.089)*	(0.923)	(-0.266)	(-0.747)	(-0.907)	(-4.161)***
LIME3	0.312	0.05	-0.136	0.072	0.332	-0.384
	(1.335)	(0.357)	(-0.789)	(0.406)	(1.418)	(-2.875)**
LIME4	0.127	0.055	-0.091	0.15	0.113	-0.027
	(0.63)	(0.338)	(-0.503)	(0.996)	(0.572)	(-0.193)
LIME5	0.396	-0.225	0.208	0.096	0.343	0.356
	(1.94)	(-1.254)	(1.093)	(0.754)	(2.351)*	(1.525)
baseline	0.427	0.338	-0.471	-0.726	-0.7	-0.956
	(1.862)	(1.951)	(-2.305)*	(-2.394)*	(-3.191)**	(-4.55)***

This table presents simple average returns and t-statistics for long-short portfolios, generated through bivariate dependent sorting. Stocks are first grouped into five categories based on their LIME coefficients generated based on the NN3 model, then further sorted within these groups by specific firm characteristics. The table specifically reports performance metrics for long-short portfolios within each LIME grouping. This table incorporates 13 firm characteristics identified by the NN3 model as significant moderators, the table is structured with each column dedicated to one of these characteristics. Rows correspond to different LIME groups, offering a clear view on how varying levels of LIME coefficients interact with firm characteristics to influence stock returns. A baseline row at the bottom provides comparison results.
Significance codes: *, **, and ***Significant at the 5%, 1%, and 0.1% levels, respectively.

There are 13 firm characteristics identified as significantly moderated by the LIME local coefficient, as derived from the NN3 model. Among these, only 4 characteristics show significant predictability in the baseline univariate sort long-short portfolios with a 5% confidence level without p-value adjustment for multiple testing. When adjusting for the LIME local coefficient from the NN3 model, the link between firm characteristics and expected returns becomes more pronounced. The baseline model of change of 6-month momentum(**chmom**) indicates a negative relationship between size and expected return, with a long-short portfolio having an average monthly return of -0.471% and a t-statistic of -2.305. However, under LIME adjustment, the strongest negative relationship at the lowest LIME local coefficient group(LIME1), showing an average return of -0.773% with a t-statistic of -3.261. For other LIME groups, variations are not significant under a 5% significant level. Dollar trading volume(**dolvol**) in the baseline model yields an average excess return of -0.7% with a t-statistics of -3.494. The strongest negative relationship at the lowest LIME local coefficient group(LIME1), showing an average return of -2.776% with a t-statistic of -7.646. Remarkably, at the highest LIME coefficient group (LIME5), this relationship reversed to an average return of 0.343% with a t-statistics of 2.351. Similar patterns are observed with 1-month momentum(**mom1m**). The baseline model presents a significant negative relationship with long-short portfolios yielding an average monthly return of -0.956% and a t-statistics of -4.55. With LIME adjustment, the negative relationship intensified in the lowest LIME group(LIME1) resulting in an average return of -3.249% and a t-statistics of -5.539. Conversely, in the highest LIME group(LIME5), the relationship flips. Size(**mvel1**) also follows this pattern. The baseline model indicates a negative relationship between size and expected return, with a long-short portfolio having an average monthly return of -0.726% and a t-statistics of -2.394. This difference primarily originates from the higher return of the lowest size level, while returns from groups 2 to 5 remain similar. However, under LIME adjustment, the lowest LIME group(LIME1) shows a significant decrease in portfolio returns as size increases, with long-short portfolios averaging -3.791 monthly return and a t-statistics of -7.039. For other LIME groups, variations are not significant.

The baseline model reveals distinct patterns for illiquidity (**ill**), 12-month momentum (**mom12m**), 6-month momentum (**mom6m**), and volatility of dollar trading volume (**std_dolvol**) with a 10% confidence level without p-value adjustment for multiple testing. However, when we apply multiple testing corrections across 72 firm characteristics for sorted long-short portfolios, these returns no longer appear significant after adjusting p-values. Yet, by adjusting for LIME local coefficients, some LIME groups display a strengthened relationship, evidenced by higher t-statistics for the long-short portfolios. Specifically, illiquidity (**ill**) in the LIME1 group exhibits a notable positive relationship, with a 2.096% return and a 4.365 t-statistic, surpassing the baseline's 0.427% return and 1.862 t-statistic. For 12-month momentum, LIME groups 2 to 5 show enhanced positive relationships, each outperforming the baseline in terms of returns and t-statistics. However, the LIME1 group's relationship with 12-month momentum is negatively aligned but insignificantly so. The baseline model's positive association with 6-month momentum, indicating a 0.36% return and a 1.398 t-statistic, is significantly strengthened in the LIME5 group, which shows a 1.36% return and a 4.397 t-statistic, while a negative relationship is observed in the LIME1 group, with a -1.027% return and a -3.133 t-statistic. Lastly, for the volatility of dollar trading volume (**std_dolvol**), the baseline model's positive relationship, yielding a 0.338% return and a 1.951 t-statistic, is markedly stronger in the LIME1 group, showing a 1.095% return and a 3.948 t-statistic after adjusting for LIME coefficients.

In the baseline model, earnings to price (**ep**), maximum daily return (**maxret**), return volatility (**retvol**), and share turnover (**turn**) displayed no significant patterns. Adjusting for LIME local coefficients, however, reveals significant patterns in certain LIME groups. For Earnings to Price (**ep**), groups LIME1 to LIME3 exhibited a negative relationship with expected returns, which reverses in groups LIME4 and LIME5, notably within LIME5, where the long-short portfolio returned -1.194% with a t-statistic of -2.554. With maximum daily return (**maxret**), LIME1 displayed a negative pattern, whereas LIME4 and LIME5 showed a positive, significant relationship, with portfolio returns of 0.893% and 1.412%, and t-statistics of 3.609 and 3.656, respectively. For return volatility (**retvol**), LIME1 presented a notable negative pattern, with the long-short portfolio returning -1.011% and a t-statistic of -2.804, while LIME5 indicated a strong positive relationship, with a return of 2.823% and a t-statistic of 5.651. Lastly, share turnover (**turn**) in LIME1 exhibited a significant negative relationship, with returns of -1.691% and a t-statistic of -5.427, whereas LIME5 demonstrated a significant positive relationship, with a return of 0.393% and a t-statistic of 2.259.

In the baseline analysis, beta squared (**betasq**) does not exhibit significant cross-sectional patterns. However, the scenario shifts dramatically upon adjusting for LIME local coefficients, which illuminate significant patterns across diverse LIME groups. Specifically, while the baseline analysis for beta squared(**betasq**) presents a slightly positive long-short portfolio return, the LIME-adjusted findings depict a more complex scenario. LIME1 through LIME4 groups show varying degrees of positive H-L differences, suggesting a mild positive moderation effect, particularly in LIME3 with a 0.186 H-L difference and a 0.503 t-statistic. However, in a significant shift, the LIME5 group exhibits a negative long-short portfolio return of -0.415 with a -1.193 t-statistic, indicating a pronounced negative impact on returns as Beta Squared increases, a stark contrast to the other groups.

Table 6: Bivariate Dependent Sort Portfolio Analysis(RF)

	indmom	maxret	mom6m	mvel1	retvol
LIME1	0.867	-0.563	-0.380	-3.723	-0.780
	(3.227)**	(-2.450)*	(-1.256)	(-7.319)***	(-2.782)**
LIME2	0.576	-0.01	0.409	-0.800	-0.187
	(2.339)*	(-0.027)	(1.481)	(-4.192)***	(-0.493)
LIME3	0.347	0.799	0.325	-0.022	0.628
	(1.353)	(2.060)*	(1.185)	(-0.214)	(1.396)
LIME4	0.265	0.583	0.599	0.011	1.160
	(1.269)	(2.452)*	(2.304)*	(0.072)	(3.462)***
LIME5	0.086	0.523	0.623	0.100	1.062
	(0.513)	(2.620)**	(2.218)*	(0.806)	(3.889)***
baseline	0.705	0.020	0.360	-0.726	0.166
	(2.900)**	(0.052)	(1.398)	(-2.394)*	(0.402)

This table presents simple average returns and t-statistics for long-short portfolios, generated through bivariate dependent sorting. Stocks are first grouped into five categories based on their LIME coefficients generated by RF model, then further sorted within these groups by specific firm characteristics. The table specifically reports performance metrics for long-short portfolios within each LIME grouping.
This table incorporates 5 firm characteristics identified by the RF model as significant moderators, the table is structured with each column dedicated to one of these characteristics. Rows correspond to different LIME groups, offering a clear view of how varying levels of LIME coefficients interact with firm characteristics to influence stock returns. A baseline row at the bottom provides comparison results.
Significance codes: *, **, and ***Significant at the 5%, 1%, and 0.1% levels, respectively.

Table 6 present findings from a bivariate dependent sort portfolio analysis, focusing on firm characteristics with significant moderation effects when using the RF model. This analysis parallels the NN3 model's findings, revealing similar patterns across four specific firm characteristics. In terms of the industry momentum(**indmom**), the baseline model reveals significant patterns. However, after p-value adjustment for multiple testing 72 firm characteristics for sorted long-short portfolios, these returns no longer appear significant. Yet, by adjusting for LIME local coefficients, the LIME1 group demonstrates the most pronounced positive relationship, showing a long-short portfolio return of 0.867% and a t-statistic of -3.227, compared to the baseline's 0.705% return and 2.9 t-statistic.

Notably, both models identify these characteristics as significantly moderated by LIME coefficients. For maximum daily return(**maxret**), in contrast to the baseline model's lack of significance, the LIME1 group reveals a notable negative relationship, while the LIME4 and LIME5 groups exhibit positive relationships. Specifically, the LIME1 group's long-short portfolio return difference is -0.563% with a t-statistic of -2.45, showcasing a significant departure from the baseline. Conversely, LIME4 and LIME5 groups show positive long-short portfolio returns of 0.583% and 0.523%, with t-statistics of 2.452 and 2.62, respectively, indicating robust positive moderation effects. For 6-month momentum(**mom6m**), the LIME1 group displays a negative association, whereas a stronger positive relationship emerges in the LIME5 group. The long-short portfolio return for LIME5 is 0.623% with a t-statistic of 2.218, underscoring the significant positive effect in higher LIME groups. In terms of size(**mvel1**), a pronounced negative relationship is observed in the LIME1 and LIME2 groups, with the LIME1 group showing a striking long-short portfolio return of -3.723% and a t-statistic of -7.319, highlighting a strong negative moderation effect. Lastly, for return volatility(**retvol**), LIME1 exhibits a significant negative pattern, whereas LIME4 and LIME5 groups demonstrate significant positive relationships. The LIME5 group, in particular, shows a long-short portfolio return of 1.062% with a t-statistic of 3.889, indicating a significant positive moderation effect.

4 Conclusion

In this paper, we explore the concept of moderation effects—specifically, how the relationship between a firm characteristic and cross-sectional stock returns changes when considering many other characteristics. We introduce a framework that combines machine learning models with Local Interpretable Model-agnostic Explanations (LIME) to dissect these effects. Our methodology begins by training a machine learning model to predict stock returns using a comprehensive set of firm characteristics. We then apply LIME to this model to systematically uncover the moderation effects. The culmination of our approach is a LIME-adjusted moderation regression, which allows for a detailed quantitative analysis of how the interactions between firm characteristics influence stock returns. This process provides a clearer picture of the complex dynamics of empirical asset pricing.

We apply the proposed methodologies to a comprehensive set of firm characteristics identified in the literature as predictors of cross-sectional stock returns, utilizing two distinct machine learning models: the Neural Network (NN3) and the Random Forest (RF) model. Our empirical investigation reveals several notable findings. First, we find that several firm characteristics demonstrate statistical significance in predicting cross-section stock returns with moderation effects. However, it's important to note that some of these characteristics lose their significance when moderation effects are not taken into account. Moreover, our bivariate dependent sort portfolio analysis reveals how moderation effects influence the predictive power of these firm characteristics on cross-sectional stock returns. Additionally, we discover that machine learning models with enhanced predictive capabilities are inclined to incorporate additional characteristics and their interactions. Finally, our findings highlight distinct differences from those that would emerge from employing the traditional Fama-MacBeth regression method, underscoring the unique insights provided by our approach.

Our study advances the empirical asset pricing literature by introducing a novel analytical framework that combines the predictive power of machine learning models with the interpretability of Local Interpretable Model-agnostic Explanations (LIME). This new framework not only sheds light on the complex moderation effects among firm characteristics on stock returns but also advances the transparency and applicability of “black-box” machine learning models within empirical asset pricing.

Despite these contributions, our study acknowledges several limitations, particularly our analysis is constrained by the scope of firm characteristics and machine learning models employed. The inclusion of additional predictors or alternative modeling techniques may yield different results.

Future research should endeavor to expand our framework, offering deeper insights into asset pricing models. Questions such as how the recognition of nonlinearity and interactions in moderation effects can refine our estimates of common risk factors, and risk exposures, or even augment the portfolio optimization process, are ripe for exploration. By addressing these limitations and exploring these avenues, subsequent studies can build on our work to further demystify the “factor zoo”, ultimately contributing to a more comprehensive and nuanced understanding of empirical asset pricing.

References

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics*, *17*(2), 223–249.

Asness, C. S., Porter, R. B., & Stevens, R. L. (2000). Predicting stock returns using industry-relative firm characteristics. *Available at SSRN 213872*.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K.-R. (2010). How to explain individual classification decisions. *The Journal of Machine Learning Research*, *11*, 1803–1831.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, *9*(1), 3–18.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Chan, K. C., & Chen, N.-F. (1991). Structural and return characteristics of small and large firms. *The journal of finance*, *46*(4), 1467–1484.

Chen, A. Y. (2021). The limits of p-hacking: Some thought experiments. *The Journal of Finance*, *76*(5), 2447–2480.

Chen, A. Y., & Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, *11*(2), 207–264.

Chen, L., Pelger, M., & Zhu, J. (2023). Deep learning in asset pricing. *Management Science*.

Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. *arXiv preprint arXiv:1601.06733*.

Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. *arXiv preprint arXiv:1506.07503*.

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Cong, L. W., Feng, G., He, J., & He, X. (2022). Asset pricing with panel tree under global split criteria. *NBER Working Paper*(w30805).

Cong, L. W., Tang, K., Wang, J., & Zhang, Y. (2021). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Available at SSRN 3554486*.

Connor, G., Hagmann, M., & Linton, O. (2012). Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica*, *80*(2), 713–754.

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *the Journal of Finance*, *52*(1), 1–33.

Demirbaga, U., & Xu, Y. (2023). Empirical asset pricing using explainable artificial intelligence. *Available at SSRN*.

Dimopoulos, Y., Bourret, P., & Lek, S. (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters*, *2*, 1–4.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (1996). The capm is wanted, dead or alive. *The Journal of Finance*, *51*(5), 1947–1958.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, *18*(3), 25–46.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fan, J., Liao, Y., & Wang, W. (2016). Projected principal component analysis in factor models. *Annals of statistics*, *44*(1), 219.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, *75*(3), 1327–1370.

Feng, G., He, J., & Polson, N. G. (2018). Deep learning for predicting asset returns. *arXiv preprint arXiv:1804.09314*.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.*, *20*(177), 1–81.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Giglio, S., Kelly, B., & Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics*, *14*, 337–368.

Giglio, S., Liao, Y., & Xiu, D. (2021). Thousands of alpha tests. *The Review of Financial Studies*, *34*(7), 3456–3496.

Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, *129*(7), 1947–1990.

Giglio, S., Xiu, D., & Zhang, D. (2021). *Test assets and weak factors* (Tech. Rep.). National Bureau of Economic Research.

Goyal, A. (2012). Empirical cross-sectional asset pricing: a survey. *Financial Markets and Portfolio Management*, *26*(1), 3–38.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, *30*(12), 4389–4436.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Gu, S., Kelly, B. T., & Xiu, D. (2019). Autoencoder asset pricing models.

Hao, Y., Dong, L., Wei, F., & Xu, K. (2020). Self-attention attribution: Interpreting information interactions inside transformer. *arXiv preprint arXiv:2004.11207*.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, *78*(5), 2465–2518.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

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

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Messmer, M. (2017). Deep learning and the cross-section of expected returns. *Available at SSRN 3081555*.

Moritz, B., & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN 2740751*.

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political economy*, *111*(3), 642–685.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016a). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016b). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Robnik-Šikonja, M., & Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering*, *20*(5), 589–600.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis*, *9*(2), 263–274.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16.

Ross, S. (1976). The arbitrage pricing theory. *Journal of Economic Theory*, *13*(3), 341–360.

Shalit, H. (2020). The shapley value of regression portfolios. *Journal of Asset Management*, *21*(6), 506–512.

Shapley, L. S., et al. (1953). A value for n-person games.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning* (pp. 3145–3153).

Suermondt, H. J. (1992). *Explanation in bayesian belief networks* (Unpublished doctoral dissertation). Stanford University.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328).

Tarashev, N., Tsatsaronis, K., & Borio, C. (2016). Risk attribution using the shapley value: Methodology and policy applications. *Review of Finance*, *20*(3), 1189–1213.

Thrun, S. (1995). Extracting rules from artificial neural networks with distributed representations. *Advances in neural information processing systems*, 505–512.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning* (pp. 2048–2057).

A The Predictive Performance of Machine Learning Models

We assess the predictive performance of machine learning models to confirm that they can outperform linear models. This implies that machine learning models, which allow for nonlinearity and feature interactions, can yield superior results compared to linear models, and the interpretations derived from the estimated machine learning models can offer valuable new insights.

A.1 Out-of-Sample R-square

We use out-of-sample R-squared from (Gu et al., 2020) to assess the predictive performance on a panel level of each model.

$$R_{oos}^2 = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,t+1})^2} \tag{14}$$

where $\mathcal{T}_3$ is the set of testing samples, where the data never enter into model estimation or tuning. In our case, the out-of-sample predictions of monthly returns are calculated from January 1989 to December 2021. Different from the traditional R^2 measure, the denominator is the sum of squared excess returns without demeaning. We compare the model with the naive forecast of zero.

Table A.1: Out-of-Sample Predictive Performance

	Naive	Linear	NN3	RF
Out-of-Sample R^2	0.00%	-1.02%	0.49%	0.20%
Out-of-Sample MSE	344.63	348.14	342.96	343.94

This table summarizes the out-of-sample R_{oos}^2 and Mean Squared Error(MSE) for four prediction models: Naive, Linear, Neural Network with three hidden layer(NN3), and Random Forest(RF). The R_{oos}^2 is calculated as specified in the Equation (14), which spans from January 1989 to December 2021. The MSE values represent the average squared differences between the observed and predicted monthly returns for each model.

Table A.1 presents the out-of-sample R^2 values and Mean Squared Errors(MSE) for four different predictive models: Naive, Linear, Neural Network with three hidden layers(NN3), and Random Forest(RF). The R^2_{oos} for the Naive model serves as a baseline at 0.00%, indicating no predictive capability. The Linear model shows a slight underperformance relative to the naive forecast with negative R^2_{oos}. Both NN3 and RF models show positive out-of-sample R^2_{oos} values, suggesting that they have predictive power with NN3 being the most predictive. Correspondingly, the out-of-sample Mean Squared Error(MSE) is listed for each model, offering a measure of the average magnitude of the errors in the predictions, where the NN3 model demonstrates the lowest error, implying a more accurate predictive performance relative to the other models.

A.2 Predictive Accuracy

To assess the predictive accuracy across stocks, following Lewellen et al. (2015), we regress the actual realized return r_i^{t+1} on the predicted expected returns $\hat{r}_i^{t+1}$

$$r_{i,t+1} = \gamma_0 + \gamma_1 \hat{r}_{i,t+1} + \epsilon_{i,t} \tag{15}$$

With an accurate prediction, this regression should have a slope $\gamma_1 = 1$, and an intercept $\gamma_0 = 0$.

Table A.2: Predictive Accuracy

	Linear	NN3	RF
Intercept	0.666	0.325	0.467
Coefficient	0.195	0.676	0.514
R^2	0.001	0.004	0.000

This table reports the regression results of actual realized returns on predicted returns for three prediction models: Linear, Neural Network with three hidden layers(NN3), and Random Forest(RF). The table presents the estimated intercept(γ_0) and the slope(γ_1) from the regression specified in Equation (15), alongside the R^2 statistic for each regression.

Table A.2 illustrates the predictive accuracy of the expected returns across individual stocks for three different models: Linear, Neural Network with three hidden layers (NN3), and Random Forest(RF). Following the methodology outlined by Lewellen et al. (2015), we report the intercept γ_0 and the slope γ_1, as well as the R^2. The OLS model yields a significantly lower slope coefficient of 0.195 and an intercept of 0.666, indicating a considerable deviation from the ideal. The NN3 model performs better with a coefficient of 0.676 and an intercept of 0.325, while the RF model demonstrates a coefficient of 0.514 and an intercept of 0.467. The R^2 values for each regression are low, suggesting limited explanatory power of the predictive returns for the actual realized returns. The NN3 model has the highest R^2, albeit still very low, indicating a marginally better fit compared to the OLS and RF models.

A.3 Prediction-Sorted Portfolio

We sort all firms into 10 portfolios based on their model-predicted returns and compute the holding period equal-weighted returns for each portfolio. Additionally, we construct a zero-investment portfolio that involves buying stocks with the highest predicted returns and selling those with the lowest. We report the average equal-weighted portfolio returns and the Sharpe Ratio.

Table A.3: Prediction-Sorted Portfolio

	Linear				NN3				RF			
	Mean	SD	t-stats	SR	Mean	SD	t-stats	SR	Mean	SD	t-stats	SR
Low(L)	-0.34	0.38	-0.90	-0.15	-0.99	0.47	-2.09	-0.41	-0.23	0.42	-0.55	-0.10
2	0.31	0.31	1.01	0.16	0.17	0.33	0.53	0.09	0.31	0.31	1.00	0.18
3	0.58	0.28	2.09	0.34	0.44	0.29	1.53	0.30	0.37	0.29	1.25	0.22
4	0.64	0.26	2.49	0.42	0.62	0.24	2.55	0.48	0.58	0.26	2.21	0.39
5	0.76	0.24	3.17	0.53	0.75	0.23	3.21	0.61	0.70	0.25	2.76	0.50
6	0.87	0.24	3.60	0.62	0.99	0.22	4.44	0.78	0.87	0.24	3.60	0.64
7	1.02	0.26	3.84	0.71	1.02	0.23	4.46	0.78	1.00	0.25	3.98	0.72
8	1.23	0.29	4.26	0.81	1.28	0.26	5.02	0.90	1.13	0.26	4.41	0.81
9	1.45	0.32	4.49	0.87	1.51	0.29	5.16	0.96	1.34	0.27	4.92	0.88
High(H)	2.21	0.45	4.91	1.01	2.89	0.51	5.72	1.13	2.66	0.48	5.54	1.17
H-L	2.55	0.41	6.27	1.05	3.88	0.49	7.93	2.69	2.89	0.43	6.80	2.02

This table presents the performance metrics for portfolios sorted based on model-predicted returns, spanning from the lowest (Low) to the highest (High) deciles. We evaluate the portfolios using three predictive models: Linear, Neural Network (NN3), and Random Forest (RF). The metrics reported include the mean monthly out-of-sample return (Mean), standard deviation of returns (SD), t-statistics (t-stats), and the Sharpe Ratio (SR). The portfolios are constructed on an equal-weighted basis, reflecting the average performance across all firms within each decile. Additionally, the table features the High minus Low (H-L) portfolio, which represents a zero-investment strategy of buying stocks with the highest predicted returns and selling those with the lowest. This strategy's performance is indicative of the predictive models' ability to capture return patterns across the spectrum of portfolio deciles. The analysis aims to provide insights into the efficacy of each predictive model in terms of risk-adjusted returns and statistical significance.

Table A.3 showcases the results for prediction-sorted portfolios using three predictive models: Linear, Neural Network with three layers (NN3), and Random Forest (RF). Each model demonstrates a consistent increase in mean returns from the lowest to the highest quintiles, indicating their effectiveness in distinguishing stocks with varying levels of predicted returns. Notably, the zero-investment strategy—long positions in the highest quintile stocks and short positions in the lowest—yields significant mean returns across all models, with NN3 leading in performance. The high T-statistics for the H-L portfolios signify robust statistical significance in this strategy's performance, and the elevated Sharpe Ratios, particularly for the NN3 model, suggest superior risk-adjusted returns.

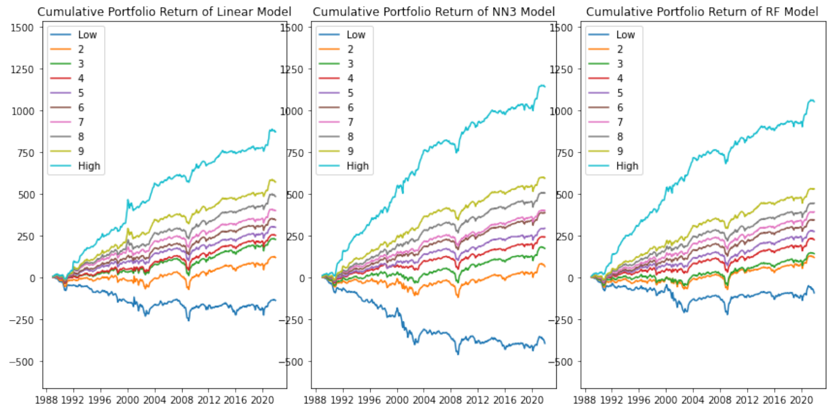

Figure A.1: Prediction-Sorted Portfolio

The figure illustrates the cumulative returns of ten equal-weighted portfolios, which are sorted based on their predicted returns using three predictive models: Linear, Neural Network (NN3), and Random Forest (RF). Each line represents the growth of one dollar invested in the respective decile portfolio, beginning from the start of the out-of-sample period and tracking the accumulation of returns over time. The portfolios are rebalanced monthly, reflecting the updated predictions. The 'High' portfolio comprises stocks with the highest expected returns, while the 'Low' portfolio contains stocks with the lowest expected returns for the forthcoming month.

Figure A.1 displays the cumulative returns of five different portfolios sorted by their predicted returns, which is in line with the previous table results on prediction-sorted portfolios. The portfolios range from "Low" (predicted to have the lowest returns) to "High" (predicted to have the highest returns). Similar to the results indicated in the table, there is a clear upward gradient in cumulative returns from the "Low" to the "High" portfolios across all three graphs. This suggests that the models are effectively ranking stocks based on their predicted returns, which translates into actual realized returns over time. The cumulative returns graph visually supports the quantitative findings in the table. It shows that the predictive models can sort stocks into portfolios that realize different levels of returns corresponding to their predictions, with the NN3 model, in particular, showing the most promising results.

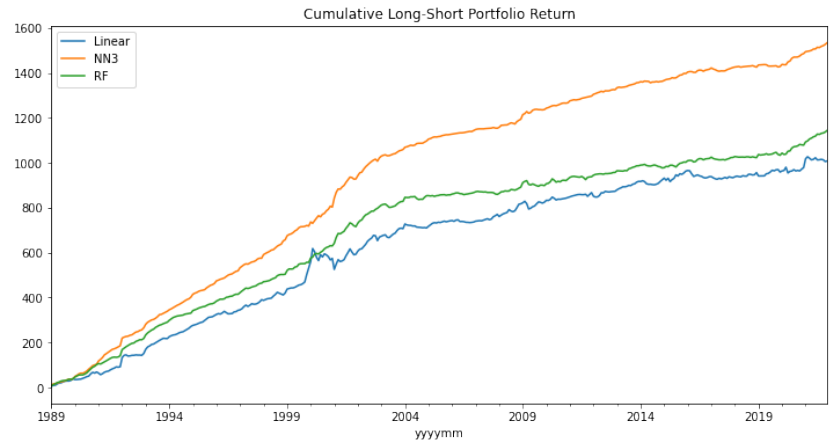

Figure A.2: Prediction-Sorted Portfolio

This figure shows the cumulative returns of the High minus Low (H-L) investment strategy, which involves going long on stocks predicted to have the highest returns and shorting stocks predicted to have the lowest returns for the next month, based on three different predictive models: Linear, Neural Network (NN3), and Random Forest (RF). The returns start from a neutral point and are accumulated by holding and rebalancing the portfolio monthly according to each model's predictions. The graph tracks these cumulative returns over the out-of-sample period, showcasing the efficacy of each model in generating excess returns through this zero-investment strategy. The diverging paths of the models reflect their varying predictive accuracies and risk profiles in an out-of-sample context.

Figure A.2 displays the cumulative returns of the zero-investment strategy, which involves buying stocks in the "High" portfolio and selling those in the "Low" portfolio. The cumulative returns are consistently positive across all three models, with the NN3 model showing the highest returns. This suggests that the zero-investment strategy can generate significant returns, especially when the predictions are derived from the NN3 model.

B Bivariate Dependent Sort Portfolio Analysis

Table B.1: Bivariate Dependent Sort Portfolio Analysis(NN3)

Panel A:Beta Squared(**betasq**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.663	0.925	0.956	0.993	0.898	0.235
	(2.396)	(2.77)	(2.636)	(2.45)	(1.648)	(0.524)
LIME2	0.577	0.755	0.829	0.729	0.763	0.187
	(3.21)	(3.278)	(3.071)	(2.451)	(1.784)	(0.516)
LIME3	0.718	0.822	0.848	0.895	0.904	0.186
	(4.207)	(4.039)	(3.658)	(3.275)	(2.157)	(0.503)
LIME4	0.875	1.013	0.913	0.916	0.969	0.093
	(4.778)	(4.824)	(3.716)	(3.049)	(2.102)	(0.233)
LIME5	1.54	1.37	1.202	1.166	1.125	-0.415
	(5.437)	(4.441)	(3.636)	(2.733)	(2.145)	(-1.193)
baseline	0.875	0.977	0.95	0.94	0.932	0.214
	(4.656)	(4.009)	(3.304)	(2.723)	(1.911)	(0.554)

Panel B: Change in 6-month Momentum (**chmom**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	2.172	1.862	1.663	1.579	1.399	-0.773
	(4.345)	(4.063)	(3.999)	(3.365)	(2.582)	(-3.261)
LIME2	0.794	0.777	0.736	0.663	0.739	-0.054
	(2.473)	(2.944)	(2.793)	(2.193)	(1.824)	(-0.266)
LIME3	0.794	0.834	0.709	0.568	0.659	-0.136
	(3.041)	(3.847)	(3.266)	(2.377)	(2.227)	(-0.789)
LIME4	0.798	0.8	0.667	0.529	0.707	-0.091
	(3.116)	(4.25)	(3.264)	(2.348)	(2.293)	(-0.503)
LIME5	0.757	0.584	0.636	0.628	0.965	0.208
	(2.37)	(2.22)	(2.634)	(2.276)	(2.459)	(1.093)
baseline	1.063	0.972	0.882	0.793	0.894	-0.471
	(3.184)	(3.595)	(3.339)	(2.719)	(2.347)	(-2.305)

Panel C: Dollar Trading Volume(**dolvol**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	3.178	1.452	0.925	0.538	0.402	-2.776
	(5.665)	(2.803)	(1.764)	(1.077)	(0.857)	(-7.646)
LIME2	0.791	0.676	0.667	0.55	0.557	-0.234
	(2.194)	(1.838)	(1.814)	(1.699)	(1.471)	(-0.907)
LIME3	0.624	0.883	0.97	0.949	0.956	0.332
	(2.15)	(3.03)	(3.45)	(3.4)	(3.351)	(1.418)
LIME4	0.816	0.771	0.939	0.975	0.928	0.113
	(3.257)	(3.36)	(4.076)	(4.725)	(4.122)	(0.572)
LIME5	0.477	0.568	0.63	0.749	0.82	0.343
	(2.174)	(3.059)	(3.408)	(4.073)	(4.756)	(2.351)
baseline	1.177	0.87	0.826	0.752	0.733	-0.7
	(4.074)	(2.92)	(2.611)	(2.427)	(2.33)	(-3.191)

Panel D: Earnings to Price(**ep**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.297	0.114	0.407	0.516	0.677	0.38
	(0.51)	(0.254)	(0.994)	(1.775)	(2.172)	(0.926)
LIME2	0.741	0.663	0.787	0.791	0.885	0.144
	(1.688)	(2.265)	(3.452)	(3.406)	(3.356)	(0.485)
LIME3	0.781	0.779	0.901	0.897	0.945	0.164
	(1.932)	(2.955)	(4.367)	(3.993)	(3.357)	(0.618)
LIME4	1.102	0.885	0.968	0.976	1.01	-0.091
	(2.425)	(2.98)	(3.981)	(4.122)	(3.483)	(-0.328)
LIME5	2.708	2.039	1.745	1.42	1.514	-1.194
	(3.763)	(4.58)	(4.761)	(4.93)	(4.405)	(-2.554)
baseline	1.126	0.896	0.961	0.92	1.006	-0.236
	(2.298)	(2.704)	(3.619)	(3.693)	(3.425)	(-0.667)

Panel E: Illiquidity(**ill**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	-0.032	-0.131	0.194	0.443	2.064	2.096
	(-0.076)	(-0.298)	(0.401)	(0.868)	(3.487)	(4.365)
LIME2	0.635	0.801	0.76	0.787	1.18	0.545
	(1.909)	(2.507)	(2.222)	(2.088)	(2.818)	(2.089)
LIME3	0.918	0.932	0.965	1.038	1.23	0.312
	(3.477)	(3.798)	(3.621)	(3.602)	(3.529)	(1.335)
LIME4	0.902	0.975	1.005	1.022	1.029	0.127
	(4.397)	(4.574)	(4.663)	(4.358)	(3.489)	(0.63)
LIME5	0.891	0.964	0.924	1.008	1.287	0.396
	(4.709)	(5.499)	(5.118)	(5.121)	(5.083)	(1.94)
baseline	0.663	0.708	0.769	0.86	1.358	0.427
	(2.388)	(2.424)	(2.505)	(2.677)	(4.006)	(1.862)

Panel F: Maximum Daily Return(**maxret**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.35	0.174	0.105	0.127	0.025	-0.325
	(1.04)	(0.43)	(0.243)	(0.247)	(0.043)	(-1.013)
LIME2	0.814	0.778	0.868	0.993	1.186	0.372
	(3.27)	(2.82)	(2.864)	(2.86)	(2.752)	(1.445)
LIME3	0.897	0.924	0.966	1.034	1.409	0.512
	(4.673)	(4.524)	(4.097)	(3.689)	(3.78)	(2.046)
LIME4	0.688	0.896	0.949	1.023	1.581	0.893
	(3.993)	(4.869)	(4.465)	(4.113)	(4.541)	(3.609)
LIME5	0.631	0.797	1.079	1.489	2.208	1.412
	(3.776)	(3.9)	(4.44)	(4.744)	(4.647)	(3.656)
baseline	0.676	0.714	0.794	0.933	1.263	0.02
	(3.079)	(2.715)	(2.754)	(2.818)	(2.99)	(0.052)

Panel G: 12-month momentum(**mom12m**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	2.466	1.264	1.226	1.343	1.754	-0.711
	(3.419)	(2.439)	(2.806)	(3.476)	(4.141)	(-1.493)
LIME2	-0.01	0.633	0.718	0.899	1.28	1.29
	(-0.023)	(2.282)	(3.103)	(4.37)	(4.72)	(4.388)
LIME3	0.075	0.606	0.698	0.979	1.307	1.231
	(0.218)	(2.508)	(3.677)	(5.288)	(5.396)	(5.07)
LIME4	0.121	0.658	0.882	1.068	1.414	1.293
	(0.349)	(2.789)	(4.436)	(5.113)	(5.049)	(5.225)
LIME5	-0.002	0.555	0.803	1.005	1.285	1.287
	(-0.006)	(2.061)	(3.283)	(3.497)	(3.454)	(3.379)
baseline	0.53	0.743	0.865	1.059	1.408	0.515
	(1.222)	(2.476)	(3.46)	(4.391)	(4.431)	(1.639)

Panel H: 1-month momentum(**mom1m**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	3.665	1.43	1.006	0.916	0.416	-3.249
	(5.542)	(3.184)	(2.48)	(2.047)	(0.785)	(-5.539)
LIME2	0.846	0.804	0.692	0.644	-0.058	-0.904
	(2.522)	(3.1)	(2.669)	(2.235)	(-0.14)	(-4.161)
LIME3	0.782	0.805	0.726	0.697	0.397	-0.384
	(3.059)	(3.666)	(3.252)	(3.255)	(1.334)	(-2.875)
LIME4	0.793	0.775	0.722	0.777	0.765	-0.027
	(3.043)	(3.639)	(3.737)	(3.824)	(2.751)	(-0.193)
LIME5	0.724	0.719	0.77	0.822	1.079	0.356
	(2.173)	(2.715)	(3.047)	(2.908)	(3.018)	(1.525)
baseline	1.362	0.907	0.783	0.771	0.52	-0.956
	(3.948)	(3.324)	(2.982)	(2.73)	(1.419)	(-4.55)

Panel I: 6-month momentum(**mom6m**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	2.403	1.305	1.031	1.13	1.377	-1.027
	(4.488)	(3.51)	(3.319)	(3.486)	(3.668)	(-3.133)
LIME2	0.988	0.824	0.813	0.922	1.243	0.255
	(2.804)	(3.768)	(4.185)	(4.378)	(4.229)	(1.143)
LIME3	0.934	0.77	0.836	0.962	1.21	0.276
	(2.862)	(3.547)	(4.065)	(4.361)	(4.041)	(1.343)
LIME4	0.644	0.682	0.74	0.859	1.234	0.59
	(1.729)	(2.651)	(3.278)	(3.479)	(3.833)	(2.377)
LIME5	-0.491	-0.113	0.357	0.59	0.869	1.36
	(-0.969)	(-0.277)	(1.1)	(1.921)	(2.265)	(4.397)
baseline	0.896	0.694	0.755	0.893	1.186	0.36
	(2.125)	(2.448)	(3.085)	(3.494)	(3.546)	(1.398)

Panel J: Size(**mvel1**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	4.166	1.068	0.61	0.495	0.375	-3.791
	(5.823)	(2.089)	(1.335)	(1.088)	(0.853)	(-7.039)
LIME2	0.732	0.577	0.557	0.603	0.562	-0.17
	(2.098)	(1.681)	(1.551)	(1.73)	(1.577)	(-0.747)
LIME3	0.713	0.786	0.902	0.883	0.786	0.072
	(2.884)	(2.795)	(3.318)	(3.369)	(2.919)	(0.406)
LIME4	0.74	0.817	0.897	0.871	0.89	0.15
	(3.329)	(3.519)	(3.713)	(3.705)	(4.011)	(0.996)
LIME5	0.717	0.689	0.771	0.808	0.813	0.096
	(3.608)	(3.685)	(3.836)	(3.996)	(3.954)	(0.754)
baseline	1.414	0.787	0.747	0.732	0.685	-0.726
	(4.7)	(2.614)	(2.389)	(2.363)	(2.272)	(-2.394)

Panel K: Return Volatility(**retvol**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.448	0.255	0.007	-0.031	-0.563	-1.011
	(1.49)	(0.652)	(0.015)	(-0.06)	(-0.961)	(-2.804)
LIME2	0.815	0.812	0.755	0.891	0.856	0.041
	(3.541)	(2.863)	(2.478)	(2.408)	(1.85)	(0.133)
LIME3	0.787	0.83	0.929	1.022	1.422	0.635
	(4.105)	(3.926)	(3.946)	(3.671)	(3.452)	(2.114)
LIME4	0.544	0.81	0.928	1.106	1.421	0.876
	(3.257)	(4.124)	(4.531)	(4.533)	(3.91)	(3.191)
LIME5	0.625	0.808	1.176	1.692	3.447	2.823
	(3.269)	(3.681)	(4.627)	(5.374)	(6.149)	(5.651)
baseline	0.644	0.703	0.759	0.936	1.316	0.166
	(3.191)	(2.775)	(2.655)	(2.758)	(2.924)	(0.402)

Panel L: Volatility of Dollar Trading Volume(**std_dolvol**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	1.184	1.412	1.45	1.837	2.279	1.095
	(3.052)	(3.468)	(3.286)	(3.835)	(4.724)	(3.948)
LIME2	0.919	0.897	0.907	0.958	1.074	0.155
	(4.017)	(3.633)	(3.523)	(3.323)	(3.561)	(0.923)
LIME3	0.832	0.789	0.817	0.857	0.882	0.05
	(3.966)	(3.45)	(3.321)	(3.603)	(3.312)	(0.357)
LIME4	0.663	0.662	0.7	0.675	0.718	0.055
	(2.759)	(2.762)	(2.674)	(2.717)	(2.592)	(0.338)
LIME5	0.341	0.365	0.27	0.193	0.115	-0.225
	(0.979)	(1.037)	(0.717)	(0.533)	(0.303)	(-1.254)
baseline	0.788	0.825	0.829	0.904	1.014	0.338
	(2.849)	(2.742)	(2.667)	(2.895)	(3.295)	(1.951)

Panel M: Share Turnover(**turn**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	1.708	1.576	1.486	1.207	0.017	-1.691
	(3.509)	(3.301)	(2.742)	(2.306)	(0.031)	(-5.427)
LIME2	0.587	0.725	0.715	0.717	0.35	-0.237
	(1.785)	(2.116)	(2.073)	(1.898)	(0.807)	(-0.785)
LIME3	0.613	0.88	1.022	1.008	0.906	0.292
	(2.044)	(3.36)	(3.877)	(3.634)	(2.463)	(0.922)
LIME4	0.692	0.887	0.967	1.028	1.017	0.325
	(2.928)	(4.138)	(4.38)	(4.902)	(3.649)	(1.308)
LIME5	0.529	0.627	0.799	0.884	0.922	0.393
	(2.593)	(3.871)	(4.691)	(4.771)	(4.115)	(2.259)
baseline	0.826	0.939	0.998	0.969	0.642	-0.281
	(3.071)	(3.414)	(3.346)	(3.046)	(1.729)	(-0.962)

Table B.2: Bivariate Dependent Sort Portfolio Analysis(RF)

Panel A: Industry Momentum(**indmom**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.531	0.822	1.12	1.277	1.353	0.867
	(1.669)	(2.654)	(3.45)	(4.128)	(3.189)	(3.227)
LIME2	0.542	0.805	1.011	1.118	1.035	0.576
	(1.888)	(2.866)	(3.682)	(3.815)	(3.308)	(2.339)
LIME3	0.69	0.757	0.873	1.141	1.06	0.347
	(2.426)	(2.561)	(3.035)	(3.977)	(3.113)	(1.353)
LIME4	0.577	0.908	0.929	1.077	0.833	0.265
	(2.091)	(3.308)	(2.835)	(3.646)	(2.2)	(1.269)
LIME5	0.556	0.819	0.713	0.801	0.642	0.086
	(1.812)	(2.516)	(2.247)	(2.237)	(1.18)	(0.513)
baseline	0.579	0.822	0.931	1.092	1.011	0.705
	(1.837)	(2.776)	(3.05)	(3.382)	(2.781)	(2.9)

Panel B: Maximum Daily Return(**maxret**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.703	0.636	0.557	0.481	0.14	-0.563
	(2.377)	(1.776)	(1.392)	(1.153)	(0.297)	(-2.45)
LIME2	0.827	0.874	0.979	1.106	0.818	-0.01
	(3.273)	(2.969)	(2.748)	(2.641)	(1.545)	(-0.027)
LIME3	0.811	0.986	0.955	1.03	1.609	0.799
	(4.222)	(4.214)	(3.667)	(3.358)	(3.147)	(2.06)
LIME4	0.699	0.881	0.978	0.97	1.282	0.583
	(3.801)	(4.379)	(4.421)	(3.753)	(3.755)	(2.452)
LIME5	0.607	0.784	0.968	1.01	1.142	0.523
	(3.401)	(4.196)	(4.737)	(4.561)	(3.98)	(2.62)
baseline	0.729	0.832	0.887	0.92	0.997	0.02
	(3.372)	(3.163)	(3.001)	(2.776)	(2.404)	(0.052)

Panel C: 6-month momentum(**mom6m**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	1.67	1.151	1.077	1.085	1.29	-0.38
	(3.317)	(3.776)	(4.143)	(3.837)	(3.589)	(-1.256)
LIME2	0.889	0.882	0.929	0.942	1.298	0.409
	(2.124)	(3.112)	(4.008)	(3.987)	(3.539)	(1.481)
LIME3	0.725	0.768	0.842	0.874	1.05	0.325
	(1.7)	(2.826)	(3.724)	(3.932)	(3.334)	(1.185)
LIME4	0.453	0.629	0.709	0.848	1.052	0.599
	(1.108)	(2.257)	(3.101)	(3.937)	(3.785)	(2.304)
LIME5	0.239	0.585	0.632	0.639	0.862	0.623
	(0.576)	(2.032)	(2.574)	(2.866)	(3.273)	(2.218)
baseline	0.795	0.803	0.838	0.878	1.11	0.36
	(1.735)	(2.74)	(3.472)	(3.611)	(3.464)	(1.398)

Panel D: Size(**mvel1**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	3.911	1.198	0.646	0.51	0.188	-3.723
	(5.795)	(2.597)	(1.679)	(1.331)	(0.51)	(-7.319)
LIME2	1.289	0.839	0.741	0.685	0.488	-0.8
	(4.109)	(2.71)	(2.388)	(2.143)	(1.486)	(-4.192)
LIME3	0.808	0.662	0.695	0.764	0.786	-0.022
	(3.1)	(2.379)	(2.642)	(2.858)	(3.122)	(-0.214)
LIME4	0.733	0.771	0.768	0.75	0.744	0.011
	(2.858)	(2.998)	(3.151)	(3.21)	(3.244)	(0.072)
LIME5	0.701	0.804	0.738	0.807	0.8	0.1
	(2.781)	(3.214)	(3.121)	(3.499)	(3.579)	(0.806)
baseline	1.414	0.787	0.747	0.732	0.685	-0.726
	(4.7)	(2.614)	(2.389)	(2.363)	(2.272)	(-2.394)

Panel E: Return Volatility(**retvol**)

	Low(L)	2	3	4	High(H)	H-L
LIME1	0.764	0.704	0.597	0.325	-0.016	-0.78
	(2.797)	(2.142)	(1.526)	(0.743)	(-0.033)	(-2.782)
LIME2	0.793	0.921	0.887	0.903	0.606	-0.187
	(3.866)	(3.237)	(2.676)	(2.209)	(1.165)	(-0.493)
LIME3	0.73	0.905	0.981	1.084	1.359	0.628
	(3.799)	(4.168)	(3.704)	(3.036)	(2.442)	(1.396)
LIME4	0.613	0.877	0.937	1.074	1.773	1.16
	(3.676)	(4.334)	(4.226)	(4.193)	(4.247)	(3.462)
LIME5	0.546	0.769	0.983	1.066	1.608	1.062
	(3.297)	(3.959)	(4.799)	(4.38)	(4.674)	(3.889)
baseline	0.689	0.835	0.877	0.89	1.066	0.166
	(3.561)	(3.361)	(3.046)	(2.601)	(2.31)	(0.402)

C Firm-Level Characteristics

Table C.1: Details of Firm-Level Characteristics

No.	Acronym	Firm characteristic	Paper's author(s)	Year, Journal	Data Source	Frequency
1	absacc	Absolute accruals	Bandyopadhyay, Huang & Wirjanto	2010, WP	Compustat	Annual
2	acc	Working capital accruals	Sloan	1996, TAR	Compustat	Annual
3	age	# years since first Compustat coverage	Jiang, Lee & Zhang	2005, RAS	Compustat	Annual
4	agr	Asset growth	Cooper, Gulen & Schill	2008, JF	Compustat	Annual
5	beta	Beta	Fama & MacBeth	1973, JPE	CRSP	Monthly
6	betasq	Beta squared	Fama & MacBeth	1973, JPE	CRSP	Monthly
7	bm	Book-to-market	Rosenberg, Reid & Lanstein	1985, JPM	Compustat+CRSP	Annual
8	bm_ia	Industry-adjusted book to market	Asness, Porter & Stevens	2000, WP	Compustat+CRSP	Annual
9	cashdebt	Cash flow to debt	Ou & Penman	1989, JAE	Compustat	Annual
10	cashpr	Cash productivity	Chandrashekar & Rao	2009, WP	Compustat	Annual
11	cfp	Cash flow to price ratio	Desai, Rajgopal & Venkatachalam	2004, TAR	Compustat	Annual
12	cfp_ia	Industry-adjusted cash flow to price ratio	Asness, Porter & Stevens	2000, WP	Compustat	Annual
13	chatoia	Industry-adjusted change in asset turnover	Soliman	2008, TAR	Compustat	Annual
14	chcsho	Change in shares outstanding	Pontiff & Woodgate	2008, JF	Compustat	Annual
15	chempia	Industry-adjusted change in employees	Asness, Porter & Stevens	1994, WP	Compustat	Annual
16	chinv	Change in inventory	Thomas & Zhang	2002, RAS	Compustat	Annual
17	chmom	Change in 6-month momentum	Gettleman & Marks	2006, WP	CRSP	Monthly
18	chpmia	Industry-adjusted change in profit margin	Soliman	2008, TAR	Compustat	Annual
19	currat	Current ratio	Ou & Penman	1989, JAE	Compustat	Annual
20	depr	Depreciation / PP&E	Holthausen & Larcker	1992, JAE	Compustat	Annual
21	dolvol	Dollar trading volume	Chordia, Subrahmanyam & Anshuman	2001, JFE	CRSP	Monthly
22	dy	Dividend to price	Litzenberger & Ramaswamy	1982, JF	Compustat	Annual
23	egr	Growth in common shareholder equity	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
24	ep	Earnings to price	Basu	1977, JF	Compustat	Annual

Table C.1: Details of Firm-Level Characteristics(Continued)

No.	Acronym	Firm characteristic	Paper's author(s)	Year, Journal	Data Source	Frequency
25	gma	Gross profitability	Novy-Marx	2013, JFE	Compustat	Annual
26	grcapx	Growth in capital expenditures	Anderson & Garcia-Feijoo	2006, JF	Compustat	Annual
27	grltnoa	Growth in long term net operating assets	Fairfield, Whisenant & Yohn	2003, TAR	Compustat	Annual
28	herf	Industry sales concentration	Hou & Robinson	2006, JF	Compustat	Annual
29	hire	Employee growth rate	Bazdresch, Belo & Lin	2014, JPE	Compustat	Annual
30	idiovol	Idiosyncratic return volatility	Ali, Hwang & Trombley	2003, JFE	CRSP	Monthly
31	ill	Illiquidity	Amihud	2002, JFM	CRSP	Monthly
32	indmom	Industry momentum	Moskowitz & Grinblatt	1999, JF	CRSP	Monthly
33	invest	Capital expenditures and inventory	Chen & Zhang	2010, JF	Compustat	Annual
34	lev	Leverage	Bhandari	1988, JF	Compustat	Annual
35	lgr	Growth in long-term debt	Richardson, Sloan, Soliman & Tuna	2005, JAE	Compustat	Annual
36	maxret	Maximum daily return	Bali, Cakici & Whitelaw	2011, JFE	CRSP	Monthly
37	mom12m	12-month momentum	Jegadeesh	1990, JF	CRSP	Monthly
38	mom1m	1-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
39	mom36m	36-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
40	mom6m	6-month momentum	Jegadeesh & Titman	1993, JF	CRSP	Monthly
41	mvel1	Size	Banz	1981, JFE	CRSP	Monthly
42	mve_ia	Industry-adjusted size	Asness, Porter & Stevens	2000, WP	Compustat	Annual
43	operprof	Operating profitability	Fama & French	2015, JFE	Compustat	Annual
44	orgcap	Organizational capital	Eisfeldt & Papanikolaou	2013, JF	Compustat	Annual
45	pchcapx_ia	Industry adjusted % change in capital expenditures	Abarbanell & Bushee	1998, TAR	Compustat	Annual
46	pchcurrat	% change in current ratio	Ou & Penman	1989, JAE	Compustat	Annual
47	pchdepr	% change in depreciation	Holthausen & Larcker	1992,JAE	Compustat	Annual
48	pchgm_pchsale	% change in sales - % change in A/R	Abarbanell & Bushee	1998, TAR	Compustat	Annual

Table C.1: Details of Firm-Level Characteristics(Continued)

No.	Acronym	Firm characteristic	Paper's author(s)	Year, Journal	Data Source	Frequency
49	pchquick	% change in quick ratio	Ou & Penman	1989, JAE	Compustat	Annual
50	pchsale_pchinvt	% change in sales - % change in inventory	Abarbanell & Bushee	1998, TAR	Compustat	Annual
51	pchsale_pchrect	% change in sales - % change in A/R	Abarbanell & Bushee	1998, TAR	Compustat	Annual
52	pchsale_pchxsga	% change in sales - % change in SG&A	Abarbanell & Bushee	1998, TAR	Compustat	Annual
53	pchsaleinv	% change sales-to-inventory	Ou & Penman	1989, JAE	Compustat	Annual
54	pctacc	Percent accruals	Hafzalla, Lundholm & Van Winkle	2011, TAR	Compustat	Annual
55	pricedelay	Price delay	Hou & Moskowitz	2005, RFS	CRSP	Monthly
56	ps	Financial statements score	Piotroski	2000, JAR	Compustat	Annual
57	quick	Quick ratio	Ou & Penman	1989, JAE	Compustat	Annual
58	rd_mve	R&D to market capitalization	Guo, Lev & Shi	2006, JBFA	Compustat	Annual
59	rd_sale	R&D to sales	Guo, Lev & Shi	2006, JBFA	Compustat	Annual
60	retvol	Return volatility	Ang, Hodrick, Xing & Zhang	2006, JF	CRSP	Monthly
61	roic	Return on invested capital	Brown & Rowe	2007, WP	Compustat	Annual
62	salecash	Sales to cash	Ou & Penman	1989, JAE	Compustat	Annual
63	saleinv	Sales to inventory	Ou & Penman	1989, JAE	Compustat	Annual
64	salerec	Sales to receivables	Ou & Penman	1989, JAE	Compustat	Annual
65	sgr	Sales growth	Lakonishok, Shleifer & Vishny	1994, JF	Compustat	Annual
66	sp	Sales to price	Barbee, Mukherji, & Raines	1996, FAJ	Compustat	Annual
67	std_dolvol	Volatility of liquidity (dollar trading volume)	Chordia, Subrahmanyam & Anshuman	2001, JFE	CRSP	Monthly
68	std_turn	Volatility of liquidity (share turnover)	Chordia, Subrahmanyam, &Anshuman	2001, JFE	Compustat	Annual
69	tang	Debt capacity/firm tangibility	Almeida & Campello	2007, RFS	Compustat	Annual
70	tb	Tax income to book income	Lev & Nissim	2004, TAR	Compustat	Annual
71	turn	Share turnover	Datar, Naik & Radclif	1998, JFM	CRSP	Monthly
72	zerotrade	Zero trading days	Liu	2006, JFE	CRSP	Monthly