武漢大學

Undergraduate Thesis

Tree-Based Sorts: Predicting Future Returns Using Past Returns of A-shares

Name:	Chuyu Xu
Number:	2020301052066
Major:	Financial Engineering
School:	Economics and Management School
Advisor:	Bin Li

5 / 2024

ABSTRACT

Does past returns contain information related to future returns? If so, when faced with multiple past-return-based factors from different formation periods and significant interactions among these factors, how to model them to accurately depict their nonlinear relationship with future returns, thereby maximizing the extraction of hidden information and making more precise predictions? This paper draws on methods from the field of machine learning, using tree-based conditional portfolio sorts to explore the information related to monthly past returns characteristics of A-shares from February 1993 to December 2023 and their correlation with future returns, and then making predictions. To validate the effectiveness of the method, the paper designs a simple long-short investment strategy based on this prediction, conducts simulated investments, obtains returns, and verifies them. Under the condition of method effectiveness, the paper analyzes the mechanisms of variables to identify the most important variables for future returns prediction and their directional effects. The investment strategy based on this prediction achieves approximately a 1.9% risk-adjusted monthly return and a Sharpe ratio of 1.62, outperforming the traditional linear method of Fama-MacBeth regression with superior and stable returns. Moreover, the results of SMB and UMD regression coefficients for risk adjustment using three-factor and four-factor models are consistent with the findings of previous literature, further confirming the effectiveness of the method. Among all the monthly return variables of the past two years, the medium- and long-term past returns are the most important, and it shows a more complicated nonlinear inverse relationship with the future returns due to the important interaction between them. This paper sets a precedent for the effective use of tree-based conditional portfolio sorts framework in the A-share market, exploring the interaction among past returns at a finer scale, enriching the nonlinear research in the field of empirical asset pricing.

Key words: Empirical asset pricing; Reversal effect; Stock market anomalies; Quantitative Investment; Machine Learning

Catalog

1 Introduction

1.1 Research Background

Cross-sectional asset pricing—seeking the reasons for differential returns among different assets—is a continuously explored topic in academia. Its origins can be traced back to the 1960s with the advent of the Capital Asset Pricing Model (CAPM①). The CAPM model posits that the expected returns of assets can be explained by market factors, thereby initiating a wave of subsequent research on factor pricing models. Subsequently, numerous studies found that asset returns are not solely determined by a single market factor, but are also influenced by other factors. Consequently, in the 1970s, the Arbitrage Pricing Theory (APT), which constructs a linear multiple-factor pricing model, was introduced by Ross (1976)[7]. Within the framework of the APT model, scholars have developed various three-factor, four-factor, and even five-factor models. Although consensus on which model is optimal has not been reached, the shared expectation is to maximize the explanation of cross-sectional differences in asset expected returns while adhering to the principle of parsimony. Based on these pricing models, both academia and industry have begun extensive exploration of anomalous factors that cannot be explained by existing pricing models for different purposes.

After decades of development, so far, academia has found over 400 factors. These factors encompass not only financial indicators of listed companies, such as valuation metrics like price-to-earnings ratios, but also volume and price indicators of stock trading, such as variables based on transaction price and volume, as well as alternative data such as text sentiment analysis indicators (such as the content of posts by investors on stock forums) and news indicators (such as themes in Wall Street Journal articles). However, this is a trend that warrants caution. John Cochrane, in his 2011 Presidential Address to the American Finance Association, described this fervent exploration of anomalies as a

①CAPM was independently proposed by Treynor (1961, 1962)[1][2], Sharpe (1964)[3], Lintner (1965a, 1965b)[4][5], and Mossion (1966)[6].

”factor zoo.” He raised key questions that prompted deep reflection in both academia and industry: Which features truly provide independent information about expected returns? Which factors are encompassed by others? How many of these new factors are genuinely significant?[8]

These questions are of paramount importance. For academia, identifying independent and genuinely significant factors helps in assessing whether the market is efficient and in testing whether asset pricing models are biased. This further characterizes and understands the structure of stock markets and asset returns, determines the key risk factors driving asset returns, and develops refined and effective pricing models. For the industry, identifying factors that consistently and reliably provide excess returns from a plethora of anomalies, while avoiding overfitting issues, can bring stable returns to investors, safeguarding the growth of household wealth and investable assets, as well as preserving and increasing asset value.

Traditional methods of testing factors and anomalies, such as portfolio sort methodology and Fama-MacBeth regressions, may become ineffective when faced with a vast and continually growing set of alternative factors. Portfolio sort methodology is a mainstream factor testing method, consisting of three steps: firstly, stocks in a designated pool are sorted based on the value of the factor under examination; secondly, the sorted stocks are divided into N groups (typically, N=10), and the rank correlation between group returns and grouping variables is calculated to assess monotonicity; finally, stocks in the top-ranked first group are bought long, while stocks in the bottom-ranked last group are sold short using equal-weight or market-value-weighted methods, and the hedge portfolio's returns are computed. Academia often evaluates factor or anomaly performance based on monotonicity and hedge portfolio returns. However, when the target factor expands to two or more, this method faces a significant limitation: after multiple groupings, some groups may have too few stocks, leading to unstable results. In response to this issue, Fama-MacBeth regressions estimate at each time period and then calculate the expected values of all estimation results, ultimately presenting the effects of each factor variable clearly in the form of regression equations, thus addressing the problem of multiple explanatory variables. However, this method also has its drawbacks. When studying interactions among multiple explanatory variables, including interaction terms will result in a large number of terms, making the analysis less intuitive and harder to interpret. For example, with 30 individual explanatory variables, including interaction terms will lead to 465 terms, making it difficult to analyze each one individually and potentially leading to insufficient sample size issues in early periods.

How can we flexibly, intuitively, and effectively explore interactions between factors? In recent years, the rapid development of machine learning has provided fresh solutions to the dilemma faced by traditional methodologies. Drawing inspiration from Moritz and Zimmermann (2016)[9], this paper adopts tools from the field of machine learning—decision trees—as the foundational model for research. Decision trees can be applied to conditional portfolio sorting and viewed as an improvement over traditional methods. Compared to rigidly dividing stocks into L equal parts based solely on factor values, conditional portfolio sorting based on tree models allows for flexibility in selecting classification variables and thresholds at different nodes based on historical data and different optimization objectives. This process continues within each node, forming a multi-level tree model. However, tree models have significant risks: firstly, since the selection of classification variables and thresholds for nodes is entirely based on historical data, when the hierarchy becomes more complex, the model is prone to overfitting; secondly, because the classification process is discrete, estimation errors can greatly affect the model's estimation results. To mitigate these risks, we employ ensemble learning, using random forests as an example. Random forests consist of multiple decision trees, each trained on different random samples and random features. The final prediction results are synthesized from the results of all decision trees, greatly reducing the risks of discrete estimation and overfitting inherent in single tree model.

In addition, machine learning methods are often criticized for being less intuitive to interpret. Therefore, we introduce two approaches that facilitate understanding of the mechanisms involved. Firstly, a method for calculating the importance of factor variables can be employed to assess the significant impact of each explanatory variable on the dependent variable, achieving effects similar to t-tests in regression analysis. Secondly, a method for approximating the partial derivatives of factor variables can be utilized to determine the direction of marginal effects for each explanatory variable, further understanding the direction in which factors influence returns.

It's important to note that when the research period spans a long duration, the time variation of the model needs to be considered. Therefore, the use of rolling windows can be employed, wherein the model is trained and tested on out-of-sample data using the same time span in different time intervals. Specifically, this paper models each year during the research period, utilizing data from the past 10 years as the training set, and rolling annually. This approach not only enables the model to capture dynamic trends more effectively, resulting in more robust results, but also allows for examination of whether the most important factor variables change over different time periods.

Forecasting future returns based on past returns has long been a topic of interest in academia, and it is classified into two types of effects based on the contribution direction of the forecast—momentum effect, which refers to the concept that winners continue to win and losers continue to lose, representing a form of inertia, and reversal effect, which refers to the phenomenon that winners turn into losers and losers turn into winners, representing a contrarian effect. This paper uses this set of classic and widespread effects in international, especially developed-country stock markets as the object of study for the model, predicting future returns in the A-share market. Regarding data frequency, the paper selects a moderately sized monthly frequency and divides it into different features based on the length of the time interval before the formation of the strategy. This is done to investigate whether sufficient information can be extracted from past returns to predict future returns.

1.2 Problem Statement and Significance

Based on the aforementioned research background, the problem studied in this paper can be described as follows: Using a tree-based ensemble algorithm, this study models future returns by utilizing past monthly return cross-sectional rankings as explanatory variables. A predictive model is developed, and based on the model's predictions, a simple long-short investment strategy is employed for out-of-sample testing to verify the effectiveness of the method by examining the resulting returns. Additionally, the mechanisms of the variables are analyzed to identify the most important variables for predicting future returns and their directional effects.

The contributions of this paper are primarily divided into the following three points:

First, this paper serves as a precedent for the effective use of a tree-based conditional portfolio sorting framework in the A-share market, providing significant insights for the further application and promotion of this framework in different markets and anomalies. Second, previous literature on momentum and reversal effects in the A-share market has predominantly used the mainstream overlapping method by Jegadeesh and Titman (1993)[10]. This method can only be seen as a single-layer sorting method and does not further explore the impact of interactions between different variables on returns. This paper adopts a two-layer or more tree model, which can effectively capture the nonlinear relationships between anomalous factors, thereby enriching the nonlinear research in the empirical asset pricing field. Furthermore, this paper explores the mechanisms of non-linearity, not only capturing the existence of nonlinear relationships but also further understanding their importance and directional effects. Third, in the overlapping method, the variable J for the formation period is defined as the average return over different past periods, often synthesizing information from multiple past returns. This paper focuses on the individual past monthly returns, extracting more complex information contained within the returns at a smaller scale, and facilitates the study of the interactions between each past month's returns.

1.3 Research Design and Preliminary Results

First, this paper constructs and trains a tree-based portfolio sorting model. Using the predictions on the test set, a simple long-short investment strategy is designed to obtain hedge returns. To verify the method's effectiveness, this paper adjusts these returns for risk, aiming to achieve significantly superior excess returns even after adjustment. Next, to further evaluate the performance, this paper employs both ”kitchen sink” and LASSO regression methods for linear Fama-MacBeth regressions. Using the aforementioned strategy and risk adjustments, the adjusted returns obtained are compared with those of the tree models. Finally, the paper explores the mechanisms of the variables' effects by calculating variable importance and partial derivatives, identifying the most important variables on the results and their directional effects.

Preliminary results indicate that the tree-based portfolio sorting model yields a risk-adjusted monthly hedge return of approximately 1.9%, which is about 0.4% higher than the best-performing Fama-MacBeth regression, and the Sharpe ratio is approximately 0.54 higher than the latter. Additionally, the SMB and UMD regression coefficients obtained from risk adjustment using three-factor and four-factor models are consistent with findings in previous literature, further validating the method's effectiveness. By calculating variable importance, it is found that in A-shares, medium- to long-term past monthly returns have the most significant impact on future returns. Under the influence of complex interactions among these variables, a nonlinear negative correlation with future returns is observed.

The structure of the rest of this paper is arranged as follows: Chapter 2 provides a literature review; Chapter 3 describes the research design methods, presenting the overall framework, data, and specific methods; Chapter 4 analyzes the empirical results, including the performance of strategies based on the model, variable importance, and partial derivatives; and Chapter 5 presents the conclusions.

2 Literature Review

In the 1960s, the Capital Asset Pricing Model (CAPM) was proposed. CAPM was the first to clearly depict the relationship between returns and risk, introducing the unprecedented concept of factors—specifically, the idea that capital returns can be explained by the returns of the market portfolio, which can be considered as the market factor. Subsequent research quickly discovered that capital returns are actually related to other factors as well. Consequently, the Arbitrage Pricing Theory (APT) was developed in the 1970s as a multi-factor linear pricing model. Within its framework, a variety of pricing models emerged. Fama and French (1993)[11] published and introduced the first multi-factor model, which added value and size factors on top of the market factor. Following this, Carhart (1997)[12], Novy-Marx (2013)[13], Fama and French (2015)[14], Hou et al. (2015)[15], Stambaugh-Yuan (2017)[16], and Daniel-Hirshleifer-Sun (2020)[17] constructed various three-factor, four-factor, and even five-factor models from different perspectives, greatly enriching the understanding of using factors for asset pricing.

Accompanying the emergence of various pricing models and based on them, a plethora of market anomalies have been identified. Subrahmanyam (2010)[18], Goyal (2012)[19], and Green et al. (2013)[20] reviewed academic research and identified 330 anomaly factors. Kakushadze (2016)[21] introduced the industry's well-known "101 Alphas," providing formulas for 101 anomalous factors. Unlike traditional approaches that rely on economic and financial theory, these anomalies were constructed through data mining methods. Green et al. (2017)[22] examined 94 anomalies and identified 12 that significantly predict stock returns. Yan and Zheng (2017)[23] used bootstrapping methods to construct over 18,000 factors, including both previously known features and new ones derived from data sampling. John Cochrane commented on this phenomenon: "We used to think that 100% of the variation in expected returns came from CAPM; now we think it' s almost zero, replaced by a plethora of new factors to describe the cross-section"[8]. This paper aims to provide a universal, standardized method to efficiently identify the effectiveness of a vast set of anomalies while considering their interactions.

In recent years, both academia and industry have begun exploring the use of machine learning methods to address the technical challenges posed by the emergence of numerous factors, which have overwhelmed traditional research methods. This paper primarily draws on Moritz and Zimmermann (2016)[9], integrating decision trees from machine learning with portfolio sorting methods, and examining the contribution of past returns to the prediction of future returns. This framework is applied and validated for the first time in the A-share market. Additionally, machine learning methods have been widely applied in other areas of finance. In terms of evaluating factor contributions, Feng et al. (2020)[24] used LASSO regression to measure factor effectiveness, discovering that profitability and investment factors have statistically significant explanatory power compared to the hundreds of factors previously identified. For prediction purposes, Li Bin et al. (2019)[25] constructed stock return prediction models and portfolios based on 96 anomalous factors in the A-share market using 12 different machine learning algorithms. In extracting common factors from a set of factors to explain cross-sectional differences, Kozak et al. (2020)[26] and Kelly et al. (2019)[27] used PCA and IPCA (Instrumented PCA) methods, respectively, to identify the common components among the factors.

The academic community has been highly interested in the predictability of future returns based on past returns, exploring various angles on this issue. Specifically, which frequencies, months, or weeks within the formation period most significantly impact future profitability? Does the impact manifest as a positive momentum effect or a negative reversal effect? How long do these influences affect future holding period returns? Why do these effects occur?

Numerous studies have demonstrated the widespread presence of momentum effects of different duration in the U.S. stock market. Jegadeesh (1990)[28] conducted cross-sectional regressions using all past 12 months of returns against the current month's returns. The results showed that returns from the previous month exhibited a reversal effect, while returns from 12 months prior had a significant momentum effect on the current month's returns. Subsequently, Jegadeesh and Titman (1993, 2001)[10][29] found that going long on the top 10% of stocks with the highest cumulative returns over the past 3 to 12 months while shorting the bottom 10% would yield an average monthly excess return of approximately 1% over the next 3 to 12 months. Additionally, momentum effects have been observed not only in cross-sections but also in time series, as discovered by

Moskowitz et al. (2012)[30]. However, some momentum effects significantly observed in the U.S. stock market do not generalize well to other markets. For example, Novy-Marx (2012)[31] found that in the U.S. stock market, portfolios formed based on returns from 7 to 12 months ago (considered mid-term returns) yield better future momentum effects than those formed based on returns from the past 2 to 6 months (considered short-term returns), a phenomenon he termed the ”echo effect.” Yet, Goyal and Wahal (2015)[32], expanding on Novy-Marx's findings, studied 37 other international markets and did not universally find the same significant mid-term momentum effect as in the U.S. stock market.

In fact, Fama and French (2012)[33] discovered that momentum strategies yield significant and robust returns in all developed country stock markets except Japan, but the overall profitability is lower in emerging East Asian markets, particularly in China, Korea, and other Asian emerging stock markets. According to Chui et al. (2010), this may be related to the level of individualism and cultural differences across countries and regions. As described in these studies, most research on the A-share market using monthly or yearly frequency data has not found significant momentum returns. On the contrary, some studies have identified a significant long-term reversal effect over 2-3 years (Wang Yonghong and Zhao Xuejun (2001)[34]; Liu Bo and Pi Tianlei (2007)[35]; Pan Li and Xu Jianguo (2011)[36]). Lu Zhen and Zou Hengfu (2007)[37] found that there is only a significant mid-term momentum effect (formation period of the past 6 months and holding period of the next 6 months) in the A-share market, while there are many short-term (formation period within the past 3 months) and long-term (formation period of at least the past 12 months) reversal effects. Although momentum effects at monthly and yearly frequencies are not apparent, higher-frequency weekly data seem to show different indications. Gao Qiuming et al. (2014)[38] found that while there is no significant monthly frequency momentum effect in the A-share market, there is a stable weekly frequency momentum return. Lu Zhen and Zou Hengfu (2007)[37] also mentioned the presence of ultra-short-term weekly momentum effects in the A-share market. It is worth noting that most past literature in this field has adopted the overlapping method of Jegadeesh and Titman (1993)[10]. Under this method, the variable representing the formation period J is calculated as the average return of past periods. However, the feature variables in this paper are independent past monthly returns, aiming to precisely analyze which specific periods make the most significant contributions to future returns and whether these contributions are positive momentum or negative reversal effects. Additionally, this paper only examines the future one-month return, focusing on the short-term future scenario when the holding period $K = 1$.

3 Research Design

3.1 Overall Framework

Figure 3.1 illustrates the overall framework for constructing and validating the tree-based conditional portfolio sorting model. In the ”Stock Pool Design” module, all A-share market stocks, excluding ST (Special Treatment) and financial stocks, are selected. The ”Asset Pricing Model” module, enclosed within the dashed box, employs the ensemble learning method based on decision trees—Random Forest—to model 25 past monthly return factors and predict future returns. In the ”Simulation Investment” module, based on the prediction results, long portfolio of each decile group is constructed, along with a long-short portfolio of the highest and lowest deciles. Using past trading data, these constructed portfolio positions are subjected to simulated trading to obtain simulation investment returns. The ”Verification of Strategy Returns” module involves statistical analysis of annual returns, net asset value statistics, and risk-adjusted return tests using the three-factor and four-factor models on the simulation investment returns. Finally, the ”Exploration Mechanism” module calculates the importance of variables and measures approximate partial derivatives. Among these, the ”Asset Pricing Model” module is the focus and innovation of this research.

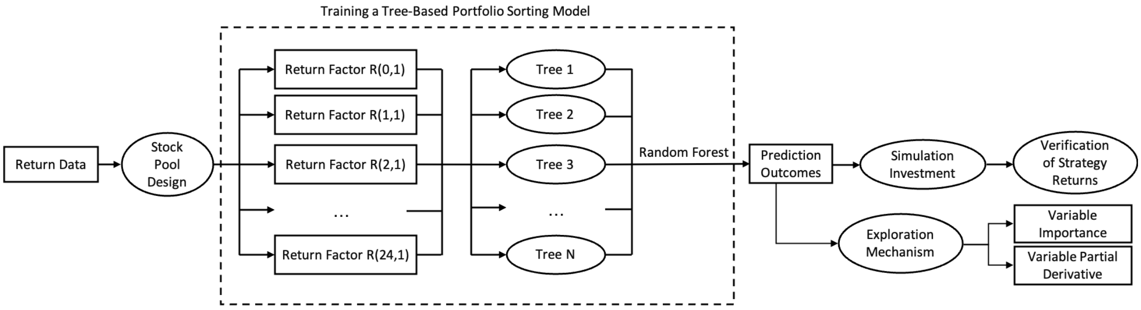


Figure 3.1 The overall framework

Note: The ”Asset Pricing Model” module in the dashed box is the focus and innovation of this paper.

3.2 Data Source, Sampling Method, and Sample Statistics

The foundational data for this study comprises the monthly returns of all A-share market stocks in China, considering reinvested cash dividends. The data spans from January 1991 to December 2023 and is sourced from the stock market series of the CSMAR database. The risk-free rate used as a benchmark originates from the stock market series of the CSMAR database as well. The Fama-French three-factor model, used to examine the returns of long-short portfolios, and the Carhart four-factor model, which includes a momentum factor, are sourced from the factor research series of the CSMAR database.

This article constructs input features based on past returns. Figure 3.2 illustrates the method of feature construction. Suppose an investor wants to form a portfolio at time t_f. Predictive variables based on past returns can be defined using two sub-variables: *gap*, the time interval between portfolio formation and the most recent month included in past return calculations; *length*, the length of the range for past return calculations. We denote the former as g, the latter as l, and the function mapping past returns to cross-sectional decile rankings as $R_{i,t_f}(g, l)$. For example, $R_{i,t_f}(1, 5) = 10$ indicates that company i ranks in the top decile in the cross-section at time t_f based on its past 6-month returns excluding the most recent month. Our factor features focus on all one-month returns in the two years prior to portfolio formation, namely $R_{i,t_f}(g, 1), g = 0, ..., 24$.

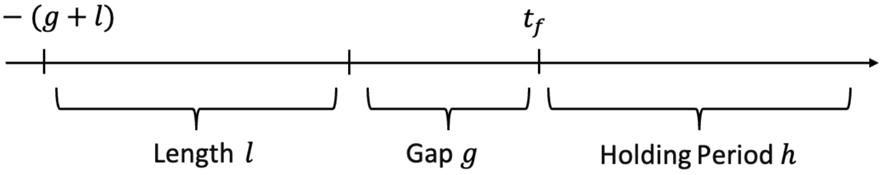


Figure 3.2 Establishment of factor features based on past monthly returns

The output variable in this study is the monthly return following the portfolio formation time point t_f. For example, the input variables used to predict stock returns for January 2023 are the monthly return cross-sectional decile rankings from December 2021 to December 2023. By pairing $R_{i,t_f}(g, 1)$ from period t to $t - 24$ with the monthly return of period $t + 1$, ”feature-label” data pairs can be obtained.

The initial stock pool consists of all stocks in the Chinese A-share market. However, since ST stocks are associated with performance losses and delisting risks, they were excluded from the analysis to ensure the accuracy and reliability of the results. Additionally, financial sector stocks were excluded to maintain consistency, as their metric calculations differ from those of other listed companies. The construction of the database and features led to some missing values in the dataset, addressed as follows: (1) If a stock's return data is missing in month *t* (5282 missing entries in total), typically due to continuous suspension of trading, all data for that stock in month *t* is removed. (2) Since factor features are based on the monthly returns of the past two years, the initial two years of past returns for newly listed stocks and any missing monthly returns result in incomplete factor features. Considering the complexity of relationships between features (the study focuses on cross-sectional differences, and factor generation logic follows a time series), both cross-sectional and time-series imputation methods may introduce significant noise. Therefore, if any stock's factor data is missing in month *t* (199,258 missing entries in total), all data for that stock in month *t* is removed.

After excluding ST and financial stocks and handling the missing values, the valid sample size from January 1991 to December 2023 is 476,703. Data from 1991 and 1992 were entirely excluded, so the data period starts from 1993. Figure 3.3 shows the monthly valid sample size from February 1993 to December 2023. Overall, the number of monthly samples shows an upward trend over the years, increasing from 5 valid samples in February 1993 to 4,262 valid samples in December 2023.

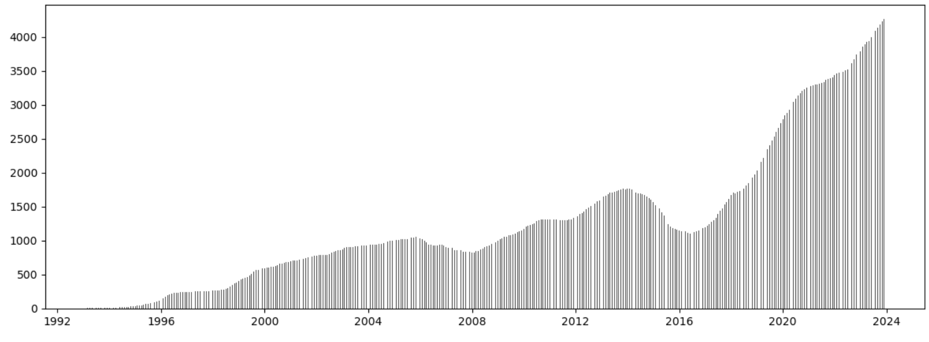


Figure 3.3 Monthly valid sample size of the dataset

3.3 Methodology

The estimation method for cross-sectional asset pricing in this paper combines traditional conditional portfolio sorting with machine learning algorithms. This section begins with an introduction to the conditional portfolio sorting method, followed by an explanation of how machine learning algorithms can further expand and improve upon it. Next, the ensemble learning method used in this study is introduced. This is followed by a brief introduction to the mainstream linear model, the Fama-MacBeth regression, which serves as a benchmark for comparison. Subsequently, the method for exploring the mechanism of factor effects is discussed. Finally, the investment strategy used for out-of-sample testing of model effectiveness and the rolling window method are introduced.

3.3.1 Conditional Portfolio Sorting

In the context of cross-sectional asset pricing, a common model form is shown in equation (3.1): at each time cross-section t, investors model and predict the conditional expectation for the next period $t+1$ using information Θ_{it} about company i. In this study, one month is considered as one period.

$$E_t[r_{i,t+1}|\Theta_{it}] = f_t(\Theta_{it}) \tag{3.1}$$

Where the expectation of $r_{i,t+1}$ is formed at time t, the function $f_t(\cdot)$ maps the information set to expected returns and may vary over time. This function is central to the estimation methods studied in this paper. The information set Θ_{it} can include various variables such as company balance sheet information, past earnings, past return volatility, and past alternative data. Since this paper focuses on the relationship between past returns and future returns, we assume that the information set only includes the cross-sectional rankings of the company's monthly returns for the past two years, i.e., $\Theta_{it} = R_{it}(0, 1), ..., R_{it}(24, 1)$, which are the decile rankings of the most recent 25 monthly returns. We will now focus on the estimation method, specifically the form of the function $f_t(\cdot)$.

The foundation of the estimation method in this paper is conditional portfolio sorting, as illustrated in Figure 3.4. Consider sorting stocks into two portfolios based on the sorting variable $R(g^{(1)}, 1)$ and threshold $\tau^{(1)}$. All stocks with $R(g^{(1)}, 1) \leq \tau^{(1)}$ are grouped into one portfolio, while stocks with $R(g^{(1)}, 1) > \tau^{(1)}$ are grouped into another. For example, if $\tau^{(1)} = 5$ and $g^{(1)} = 0$, this means that we classify all stocks with last month's returns below the cross-sectional median into one portfolio, and those with last month's returns above the cross-sectional median into another portfolio. Now, the expected stock returns for each portfolio are $E[r_{i,t+1}|R(g^{(1)},1) \leq \tau^{(1)}]$ and $E[r_{i,t+1}|R(g^{(1)},1) > \tau^{(1)}]$, respectively. If the expected returns in each portfolio are modeled as constants, the predicted values are simply the average returns of each group for the next month. Next, by sorting the stocks within each portfolio based on another (or the same) feature with corresponding thresholds $\tau^{(2a)}$ and $\tau^{(2b)}$, we obtain four different portfolios, $S1$ to $S4$. For example, the expected returns for stocks in portfolio $S1$ are given by $E[r_{i,t+1}|R(g^{(1)},1) \leq \tau^{(1)}, R(g^{(2a)},1) \leq \tau^{(2a)}]$.

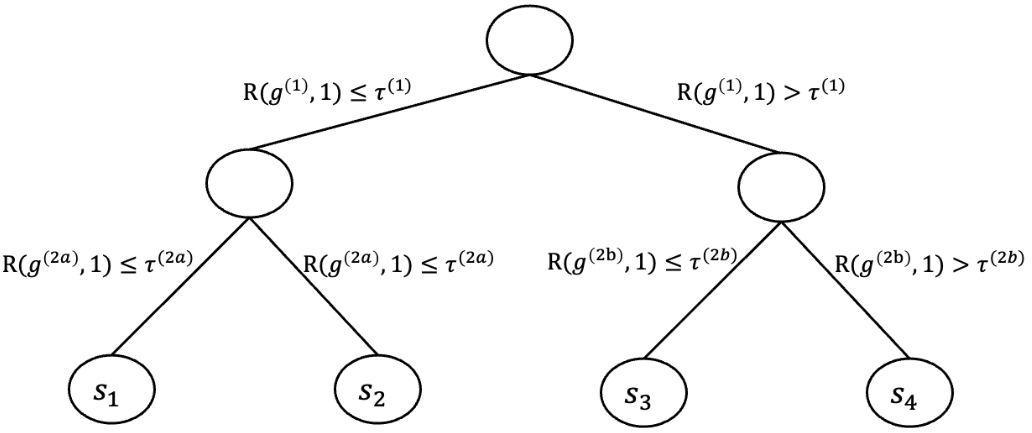

$R(g^{(1)},1) \leq \tau^{(1)}$
$R(g^{(1)},1) > \tau^{(1)}$
$R(g^{(2a)},1) \leq \tau^{(2a)}$
$R(g^{(2a)},1) \leq \tau^{(2a)}$
$R(g^{(2b)},1) \leq \tau^{(2b)}$
$R(g^{(2b)},1) > \tau^{(2b)}$
s_1
s_2
s_3
s_4

Figure 3.4 Illustration of an conditional portfolio sorting

Below, we formally define conditional portfolio sorting using equations. Our goal is to estimate the expected return of company i at period $t+1$ based on information available at period t, as shown in equation (3.1). Consider portfolio S_1 in Figure 3.4, which is defined by the condition that the variable $R(g^{(1)},1)$ of the companies in the portfolio is less than the threshold $\tau^{(1)}$ and the variable $R(g^{(2a)},1)$ is less than the threshold $\tau^{(2a)}$. Other portfolios can similarly be defined based on the relationship between the classification variables of the companies and the corresponding thresholds. For each portfolio S_j, the estimated expected return $\hat{\mu}_l$ is defined as the average return of all companies within that portfolio, as shown in equation (3.2).

$$\hat{\mu}_j = Mean(r_{i,t+1}|Firm\ i \in S_j\ in\ period\ t) \tag{3.2}$$

In other words, similar to linear regression, we are interested in approximating the conditional mean of the dependent variable at given values of the explanatory variables by the average value of the dependent variable for observations with similar values of the explanatory variables. Therefore, conditional portfolio sorting divides more homogeneous company observations into different subsets. Assuming we have divided the more homogeneous companies into different portfolio subsets, the prediction function can be written as equation (3.3):

$$\hat{r}_{i,t+1} = \sum_{J}^{j=1} \hat{\mu}_j \mathbb{1}(Firm\ i \in S_j\ in\ period\ t) \tag{3.3}$$

3.3.2 tree-based conditional portfolio sorting

Building on the foundation of conditional portfolio sorting, this paper's tree-based conditional portfolio sorting method expands and improves upon the traditional approach in three significant ways. First, unlike previous examples that pre-select thresholds and sorting variables, this paper uses a data-driven approach to select the optimal thresholds and sorting variables within each portfolio (where ”optimal” will be defined later). Second, this method applies the process to a deeper level than the two levels usually considered in conditional portfolio sorting. The depth of this process is also determined in advance by the data, resulting in tree-based conditional portfolio sorting. Third, because conditional sorting involves hard thresholds that are sensitive to small data changes, its out-of-sample predictive performance can be less effective. Therefore, this paper employs ensemble learning algorithms, averaging the results of many tree-based conditional portfolio sorting instances to smooth decision boundaries, which significantly improves prediction accuracy. This will be explained in more detail later.

First, we introduce how variables, thresholds, and tree depth are estimated and determined using historical data. Considering the large data volume and the need for computational efficiency, this paper adopts an efficient approximate algorithm for the optimal solution, namely the greedy algorithm. A greedy algorithm refers to the method of making the locally optimal choice at each step with the hope of finding a global optimum. Specifically, in each step, it uses a brute force approach to traverse all possible combinations and then makes the optimal choice for the current state. This process iterates until further progress does not significantly improve the results. More specifically, given a set of variables, the greedy algorithm tries all combinations of variables and thresholds at each step, finding the combination that minimizes the mean squared error calculated according to equation (3.2). The same process is repeated for the divided subsets until the subsets become sufficiently small or further division does not significantly reduce the mean squared error. The final result is a multi-level tree-based portfolio sorting model.

Below, we provide the formal definition using equations. Let $S_1(g, \tau)$ and $S_2(g, \tau)$ represent two portfolios divided by the past return deciles ranking $R(g, 1)$ and the threshold τ. Similar to conditional portfolio sorting, all observations where $R(g, 1) \leq \tau$ are in portfolio S_1, and all observations where $R(g, 1) > \tau$ are in portfolio S_2. At each node, all observations belonging to that node are split into two such portfolios. The greedy algorithm identifies the past return feature $R(g, 1)$ and the threshold τ that minimize:

$$(g^*, \tau^*) = \arg\min_{g,\tau} SC(g, \tau), \tag{3.4}$$

where $SC(g, \tau)$ is the splitting criterion function, a concept borrowed from machine learning literature. There are various splitting criteria for different tasks, such as information gain and Gini impurity. For this regression task, mean squared error (MSE) is chosen as the splitting criterion. The splitting criterion function selects the combination of the predictor variable and threshold that minimizes the variance of the expected returns within the resulting portfolios, as follows:

$$SC(g, \tau) = \min_{\mu_1}\left(\sum_{R_{it}(g,1)\in S_1(g,\tau)} (r_{i,t+1} - \mu_1)^2\right) + \min_{\mu_2}\left(\sum_{R_{it}(g,1)\in S_2(g,\tau)} (r_{i,t+1} - \mu_2)^2\right) \tag{3.5}$$

Here, the result of the internal minimization is the mean of each group, as shown in equation (3.2). Overall, this algorithm breaks down a complex non-linear problem into simpler subsets of linear problems. Each subproblem is solved using brute force, where the splitting criterion function is calculated for each input feature and each threshold. The optimization is repeated for each resulting combination until the sample size at the nodes is too small to be further split or no variable can sufficiently improve the mean squared error in equation (3.5). The result is a multi-level conditional portfolio sorting model.

Figure 3.5 illustrates the tree-based sorting model that improves upon the conditional portfolio sorting shown in Figure 3.4 (the data in the figure is hypothetical, and the entire iteration process is not shown, only the first three node cases are selected). Suppose the first chosen splitting variable is $R(23, 1)$ with a corresponding threshold of 5, meaning all stocks with the lowest 50% monthly returns from 24 months ago are classified into one portfolio, while the remaining stocks are classified into another portfolio. Under this split, $R(23, 1)$ is again selected at the next level in the left branch, and $R(24, 1)$, the monthly return from 25 months ago, is selected in the right branch. The actual iterative sorting goes deeper, but for simplicity, the figure assumes the returns for one month ahead for each of the four subsets have been calculated. According to the hypothetical scenario, the differences are already quite apparent: subset $S1$, which includes stocks in the lower group of the two $R(23, 1)$ groups, shows the highest returns, indicating a long-term reversal. The right branch illustrates a momentum effect—stocks with higher $R(24, 1)$ values have higher subsequent mean returns.

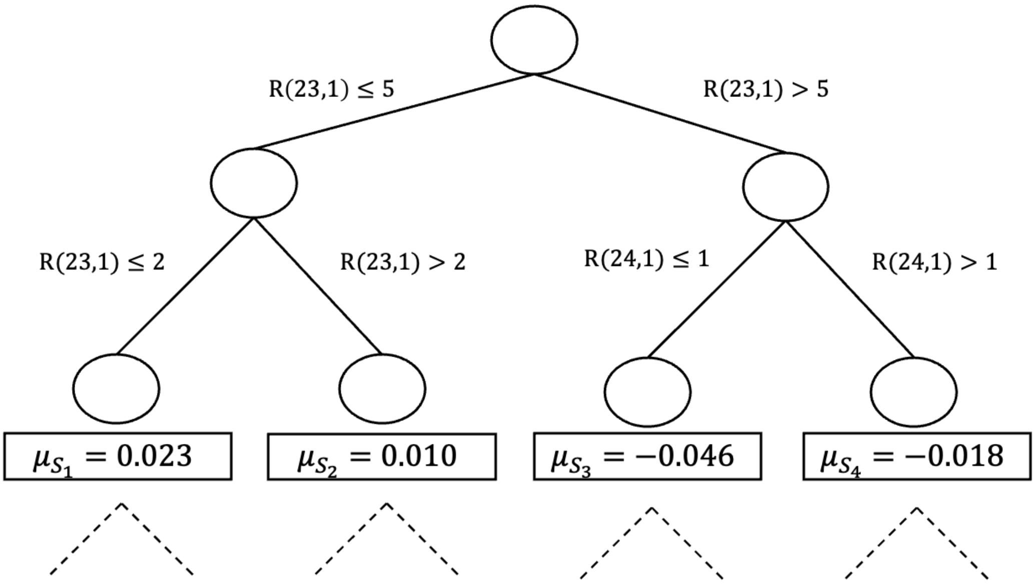


Figure 3.5 Illustration of a tree-based conditional portfolio sorting

3.3.3 Random Forest

Constructing a tree-based conditional portfolio sorting model in the manner described earlier presents several pitfalls. First, as previously mentioned, due to the complexity of optimization, we must use a greedy algorithm to estimate the model. However, since this algorithm uses local optima to approximate global optima, it cannot guarantee that the thresholds and splitting variables selected at each node are globally optimal. Second, the threshold splitting method is discrete, so any errors in the estimation can significantly distort the true value of the expected returns predicted by the estimated model. Third, a large body of literature indicates that while single tree-based conditional portfolio sorting models summarize the estimation data well, they have weak generalization capabilities and do not extend well to new observations. In other words, because each step has many degrees of freedom (variables and thresholds), tree-based conditional portfolio sorting models often overfit the historical sample used for estimation.

These issues are well-known in the machine learning literature, and this paper adopts the general solution proposed by Breiman (2001)[39]. The main idea is to perform multiple estimates of the tree-based conditional portfolio sorting model, each time using only a random subset of variables. This results in multiple sets of models, which are less likely to overfit because they use fewer variables and are relatively less complex. Then, this paper calculates the expected return estimate for each model and averages the estimates of all models to obtain the final prediction. This ensemble method, which uses the mean of multiple decision tree models as the final estimate, is called the Random Forest algorithm.

In addition to Random Forests, there is another common ensemble learning method based on decision trees in the field of machine learning—Gradient Boosting Trees. However, in Random Forests, each decision tree is trained independently, allowing for parallel computation, whereas Gradient Boosting Trees is a serial method where each new tree depends on the previous ones, preventing parallel processing. Moreover, Random Forests reduce model variance and the risk of overfitting by randomly selecting feature subsets and samples for training, and by using the voting mechanism of multiple decision trees for prediction. On the other hand, Gradient Boosting Trees, being a serial method, are more prone to overfitting and require some regularization. Given the large dataset and the sensitivity to overfitting in this study, this paper chooses the Random Forest ensemble algorithm for modeling.

Below is the formal definition of the formula. Let B be the number of trees computed in parallel, and let $\hat{f}_b(\Theta_{it})$ be the expected return of stock i at time t predicted by model b. The final expected return estimate is presented in equation (3.6) (3.6).

$$\hat{r}_{i,t+1} = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(\Theta_{i,t}) \tag{3.6}$$

Furthermore, we employ pre-pruning in pruning techniques to reduce the complexity of decision trees and improve generalization performance, thereby avoiding overfitting. Pre-pruning involves evaluating each node before splitting during the tree construction process. If the split does not improve generalization performance, the split is halted, and the current node is marked as a leaf node. We use the most common pre-pruning parameter, maximum depth (max_depth), to limit the maximum depth of the decision tree and prevent overfitting.

To determine the number of conditional rankings based on the tree B, the number of regression variables used in each ranking n, and the value of the maximum depth parameter max_depth, this paper employs Grid Search and 5-fold Cross-Validation. Specifically, we first use grid search to determine the range of model hyperparameters and construct a grid of hyperparameter combinations. For each hyperparameter combination, the model is evaluated using 5-fold cross-validation. For each fold of the cross-validation, three-quarters of the data are used for training, and the remaining quarter is used for validation. Then, the average performance metric for each hyperparameter combination is calculated. Finally, the hyperparameter combination with the best performance metric is selected as the final model configuration. This method of combining grid search with 5-fold cross-validation allows for a more comprehensive exploration of the hyperparameter space and selection of the best model configuration to enhance model performance and generalization ability.

The parameter pool constructed in this paper is as follows: $B = [100, 150, 200]$; $n = [2, 4, 6, 8]$; $max_depth = [2, 3, 4]$. The final hyperparameter combinations selected for the 22 models (see section 3.3.7 for model generation) are shown in Table **??**:

3.3.4 Fama-MacBeth Regression

To further evaluate the performance of the tree-based conditional portfolio sorting model, this paper uses the Fama-MacBeth regression as a benchmark. Regarding the Fama-MacBeth regression, this paper follows the two methods provided by Moritz and Zimmermann (2016)[9]: The first is the “kitchen sink” Fama-MacBeth estimation, which includes all past return sorting variables in the regression equation, regardless of their individual significance, and uses all of them for prediction. The second method bases entirely on relevant variables, where ”relevant” is defined as variables selected in the LASSO regression.

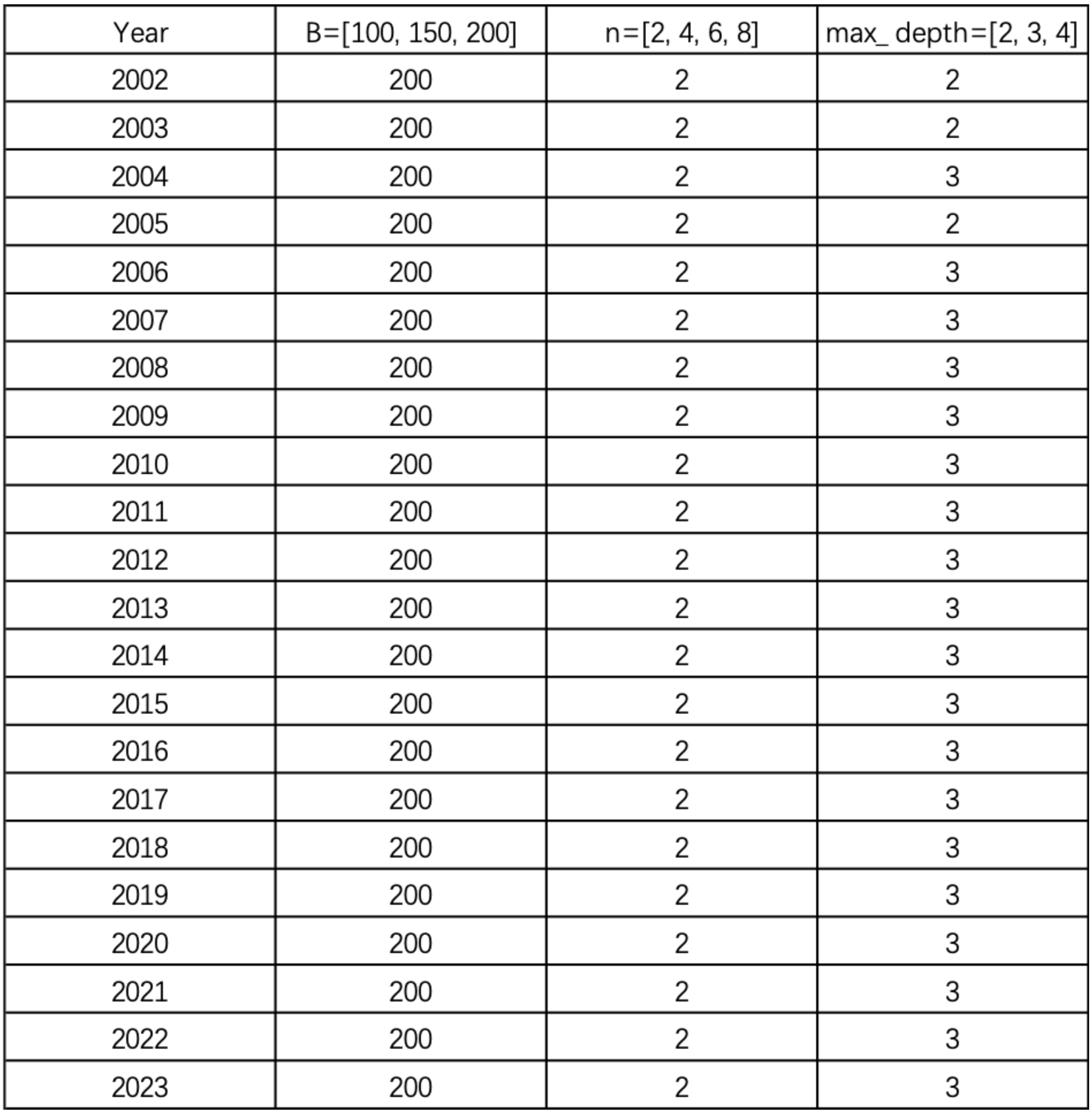
Here, this paper briefly introduces the core method of the Fama-MacBeth regression. The general implementation of the Fama-MacBeth framework is shown in equations (3.7) and (3.8).

First, for each cross-section, the following regression is performed:

$$r_{i,t+1} = \beta^t_{cons} + \sum_{g=0}^{24} \beta^t_g R_{it}(g, 1) + \epsilon_{it} \tag{3.7}$$

including all variables (the kitchen sink method) or using LASSO to select relevant vari-

Table 3.1 Optimal Hyperparameter Combinations for Models



Year	B=[100, 150, 200]	n=[2, 4, 6, 8]	max_ depth=[2, 3, 4]
2002	200	2	2
2003	200	2	2
2004	200	2	3
2005	200	2	2
2006	200	2	3
2007	200	2	3
2008	200	2	3
2009	200	2	3
2010	200	2	3
2011	200	2	3
2012	200	2	3
2013	200	2	3
2014	200	2	3
2015	200	2	3
2016	200	2	3
2017	200	2	3
2018	200	2	3
2019	200	2	3
2020	200	2	3
2021	200	2	3
2022	200	2	3
2023	200	2	3

ables.

Next, the period $t+1$ prediction is based on the moving average of the coefficient estimates up to period $t-1$, and then the coefficients are applied to the corresponding $\boldsymbol{R}_{it}(g,1)$, as follows:

$$\hat{r}i,t+1 = \bar{\beta}cons^{t-1} + \sum_{g=0}^{24} \bar{\beta}g^{t-1}\boldsymbol{Rit}(g,1) \tag{3.8}$$

where $\bar{\beta}g^{t-1} = \frac{1}{m}\sum j = t-1-m^{t-1}\hat{\beta}_g^t$. This paper uses a 60-month moving window. Consistent with the implementation of the model in this paper, the input variables are all one-month returns within the two years prior to portfolio formation. Each period, the model calculates return predictions based on past model estimates and divides the predictions into 10 deciles. Similar to the previous strategy, this paper constructs an equal-weighted hedge portfolio using the Fama-MacBeth predictions, going long on the highest decile and short on the lowest decile of predicted returns.

3.3.5 Variable Importance

Since the relevance of variables in the tree-based portfolio sorting method is determined by their hierarchy and potential interactions with other variables, the simple t-test used in linear models to indicate statistical significance is not applicable here. Instead, we use a measure of relative variable importance, which is similar to the t-statistic in simple regression.

For each predictor variable and each tree-based conditional sorting model, we randomly permute the values of a specific variable, calculate the predicted mean squared error (MSE) in this case, and then compare it to the MSE when all variables are at their original values. This score is averaged over all tree models, and the predictor variables are ranked according to this measure. Variables with higher average scores indicate that randomizing the predictor variable results in a greater increase in MSE, thus suggesting that the predictor variable is more relevant.

3.3.6 Partial Derivatives of Variables

Unlike linear models, our model cannot directly derive the partial derivatives of predictor features with respect to the predicted values. To evaluate the impact direction of specific predictor variables on the predictions, we define a partial derivative measure applicable to tree-based conditional portfolio sorting. Let $R_{it}(g^{-}, 1)$ be the vector of past return decile ranking variables excluding $R_{it}(g, 1)$. The approximation of the partial derivative of the prediction is shown in equation (3.9). More intuitively, we construct the past return rankings as cross-sectional decile rankings, i.e., $R_{it}(g, 1) \in 1, \dots, 10$. For each of these 10 values, we counter-factually set the observed value to $R_{it}(g, 1) = d, \ \forall d = 1, \dots, 10$, and calculate the mean of the predicted values across time, firms, and bootstrap samples:

$$\hat{r}^{g,d}_{i,t+1} = \frac{1}{T}\frac{1}{N}\frac{1}{B}\sum_{i,t,b} \hat{f}_b(R_{it}(g, 1); R_{it}(g^{-}, 1)) \tag{3.9}$$

Repeat this operation for all values of d, and plot the predicted mean results for each past return function $R(g, 1)$ corresponding to each value of d. This method can be easily extended to the case of simultaneously changing two (or more) predictor variables.

3.3.7 Investment Strategy

Next, we examine whether the strategies generated by the tree-based conditional portfolio sorting can provide more accurate return predictions, thus validating the effectiveness of the model. Through model estimation, we predict the returns of each stock every month and then divide the stocks into decile groups based on these predictions. We calculate the average return differences generated in the decile groups each month. Additionally, a simple trading strategy is employed: every month, we long the group with the highest expected returns and short the group with the lowest expected returns, thereby obtaining equally weighted hedge returns.

This prediction needs to be performed out-of-sample. Although actual out-of-sample testing is challenging to achieve, we follow a standard pseudo out-of-sample process, as depicted in Figure 3.6. Considering the temporal variability of the model, the tree-based conditional portfolio sorting is recalculated every year based on the data from the past 10 years, and then used to predict the expected returns for the next 12 months, rolling forward one year for training and prediction. Specifically, considering the limited data volume before 2002, to maintain balance and stability in the model, this paper starts modeling from 2002, and the model for 2002 is trained using data from all previous years. Thus, this paper establishes 22 models spanning from 2002 to 2023. The sliding window method used to divide the training set and out-of-sample test set is illustrated in Figure 3.7. In the months when each model makes predictions, trading is conducted according to the strategy described earlier. This method takes into account the importance of different regression variables and the possibility that their effects may change over time, and addresses the question of whether the tree-based conditional portfolio sorting based on average trees can be used for trading in principle.

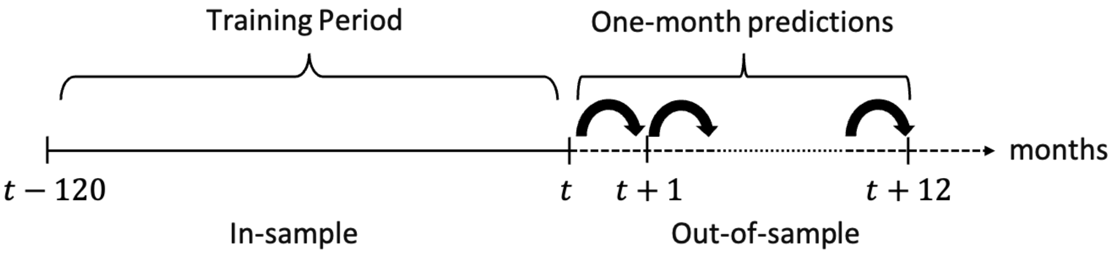


Figure 3.6 Out-of-sample Test

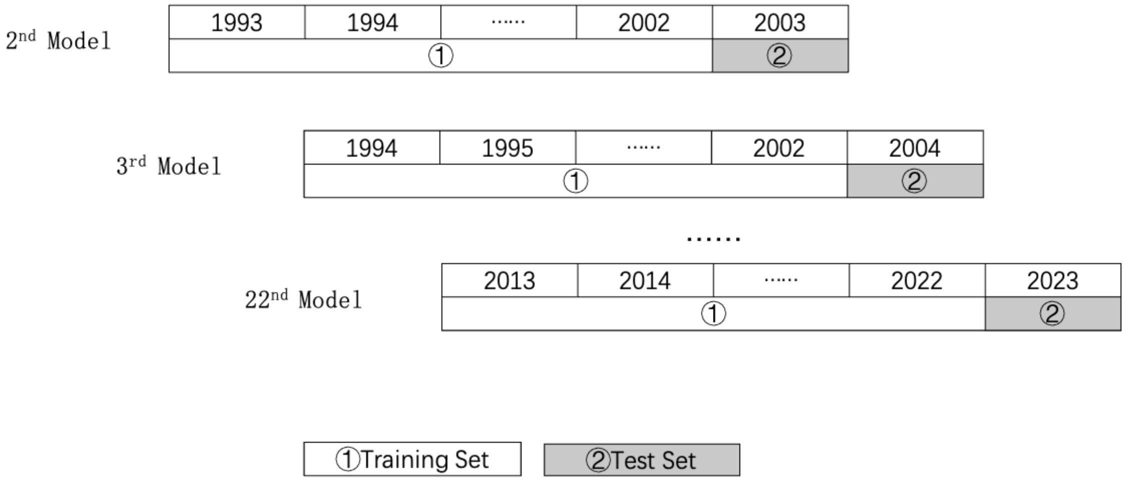


Figure 3.7 Sliding Window Diagram

4 Empirical Results and Analysis

The research methodology of this paper is applied to predict future returns based on past returns. The results of the study can be summarized in three points. First, the strategy returns and Sharpe ratios based on model predictions are significantly higher than those of the market and the linear model Fama-MacBeth regression. This indicates that the tree-based conditional portfolio sorting model performs well and accurately predicts expected future returns. Second, among all 25 past return ranking functions, the functions of mid-to-long-term past returns are more influential on future returns compared to recent returns. Third, the excellent predictive ability can be attributed to the flexible handling of interactions between past returns, thus precisely capturing the nonlinear relationship between past and future returns. The following sections provide evidence and detailed analysis of these results.

4.1 Strategy Performance of Tree-Based Conditional Portfolio Sorting

First, we demonstrate that the strategy described above can generate significant and superior risk-adjusted excess returns. Specifically, as described in Section 3.3.7, except for the year 2002, we trained the model using 10 years of data up to period t, and used the trained model to predict periods $t + 1, ..., t + 12$ (one month per period). This process was repeated every year from 2002 to 2023. We divided the monthly predictions into 10 deciles, from lowest to highest, and constructed an equally weighted long-short portfolio by going long on the stocks in the highest decile of expected returns and short on the stocks in the lowest decile. This ultimately forms a hedged return. Figure 4.1 shows the annual returns of the strategy over the past 22 years. It can be seen that the strategy return is positive every year.

Figure 4.2 shows the returns of investing 1 yuan in the long-term portfolio and the short-term portfolio. It can be seen that the terminal value of the long portfolio is 57.66 yuan, the terminal value of the short portfolio is 0.41 yuan, and the terminal value of the market portfolio is 1.66 yuan. This indicates that the tree-based conditional portfolio sorting is highly effective in both the long and short portfolios, and significantly outperforms

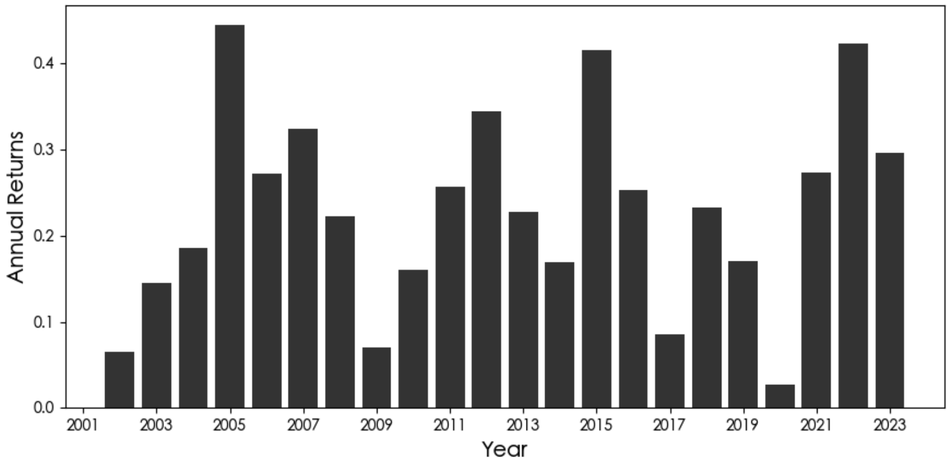


Figure 4.1 Strategy Annual Returns

the market.

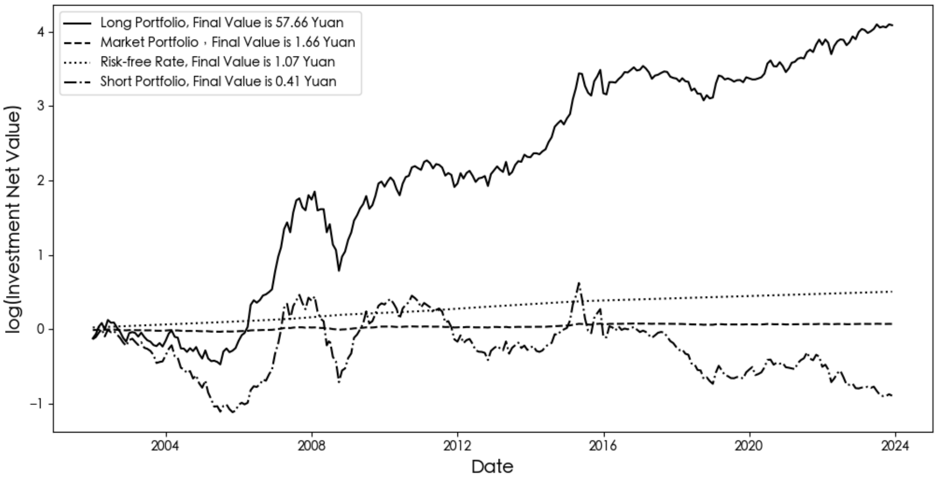


Figure 4.2 Net Value Curve of Investing 1 Yuan in January 2002

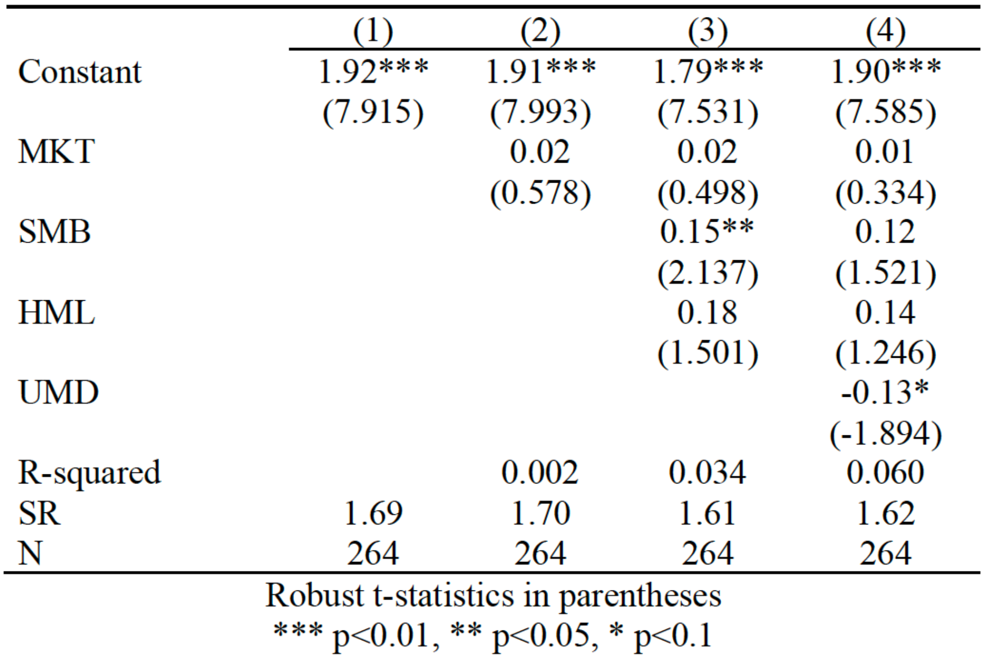
Table 4.1 presents the regression of the long-short strategy returns using the CAPM, three-factor, and four-factor equilibrium models. In column (1), the raw average monthly return is 1.92%. Observing column (2), the positive correlation between strategy returns and market returns is not significant, and the factor exposure is extremely low, which aligns with the hedging nature of the long-short portfolio. Moreover, after regressing strategy returns on market returns, the abnormal average return remains largely unaffected and remains highly significant at 1.91%. In column (3), we find that while the size factor has a relatively low positive exposure, it significantly explains the strategy returns. This supports the findings of Lu Zhen and Zou Hengfu (2007)[37] that smaller stocks are more prone to reversal effects compared to larger stocks (as shown in Section 4.3, the strategy returns mainly come from reversal effects). Additionally, it is noted that the strategy is not dependent on the value factor. Similarly, column (4) shows that the momentum factor has a negative explanatory power despite its low factor exposure. This aligns with previous literature that finds no significant monthly momentum effect in the A-share market, but a strong medium-to-long-term reversal effect. It is also observed that after including the momentum factor, the explanatory power of the size factor disappears.

Regarding R^2, the R^2 of the CAPM model is extremely low. For the three-factor and four-factor models, R^2 gradually increases, and the time variation in strategy returns can be partially explained by the size factor in the three-factor model and by the momentum factor in the four-factor model. However, the intercept remains highly significant, with a monthly value of around 1.85%. Despite the increase in R^2, it ultimately rises to only 0.06, indicating that a substantial portion of the time variation in the strategy cannot be explained by these equilibrium models.

Across all regressions, we observe very high Sharpe ratios, around 1.65. Note that, since the long-short portfolio is essentially a hedged neutral portfolio, the Sharpe ratio in this paper is calculated by dividing the mean of the risk-adjusted returns by the standard deviation, without using the risk-free rate as a benchmark.

Table 4.2 more clearly illustrates the returns of each decile portfolio formed based on model predictions, as well as the factor exposures of each decile portfolio in the four risk models. First, the returns of all decile portfolios are correlated one-to-one with market returns, which is expected since they can all be viewed as long portfolios. Second, compared to the value factor, the size factor exposure is significantly more pronounced, with higher exposure values. For both the value and size factors, the exposure values are higher in higher deciles, showing a clear monotonicity, although the differences between groups are small. Third, for the momentum factor exposure, the returns of the decile groups also show a monotonic relationship. Similarly, in terms of magnitude, these differences are small. Fourth, despite the lack of significant differences in risk factor exposures among these portfolios, there is a strong monotonic relationship between them and their risk-adjusted average returns. In terms of risk exposures, this contrasts sharply with portfolios

Table 4.1 Strategy Factor Loadings: Tree-Based Conditional Portfolio Sorting



	(1)	(2)	(3)	(4)
Constant	1.92***	1.91***	1.79***	1.90***
	(7.915)	(7.993)	(7.531)	(7.585)
MKT		0.02	0.02	0.01
		(0.578)	(0.498)	(0.334)
SMB			0.15**	0.12
			(2.137)	(1.521)
HML			0.18	0.14
			(1.501)	(1.246)
UMD				-0.13*
				(-1.894)
R-squared		0.002	0.034	0.060
SR	1.69	1.70	1.61	1.62
N	264	264	264	264

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

that appear very similar. More importantly, this relationship is not only driven by extreme portfolios (although the monotonic relationship is particularly strong in these portfolios) but is present across all 10 portfolios.

To further test the model's performance, we present the performance of the benchmark Fama-MacBeth regression. As mentioned earlier, this paper employs two methods of Fama-MacBeth regression.

Starting with the kitchen sink model, the first four columns of Table 4.3 show the factor exposures of the strategy to market, size, value, and momentum factors in the time series regression. The strategy's average monthly return is 1.48%, and none of the four factors significantly explain it. The alpha relative to the four-factor model is approximately 1.58% per month, with a Sharpe ratio of about 1.08. When using LASSO regression within the Fama-MacBeth framework, the results significantly deteriorate. The last four columns of Table 4.3 show that the average strategy return is approximately 0.64% per month, and the four-factor alpha is 0.65% per month. The Sharpe ratio is about 0.45, similar to the kitchen sink regression. The noticeable difference between the two arises because, in the case of LASSO regression, many less relevant regression coefficients are shrunk towards zero, even though these regressors might still be important.

Based on the performance of the investment strategy in this paper, the tree-based

Table 4.2 Factor Loadings of Decile Portfolios: Tree-Based Conditional Portfolio Sorting

	Low	2	3	4	5	6	7	8	9	High	High-Low
Average return	0.06	0.70	0.91*	1.14**	1.26**	1.44**	1.51***	1.58***	1.74***	1.74***	1.92***
	(0.105)	(1.276)	(1.670)	(2.031)	(2.291)	(2.586)	(2.708)	(2.831)	(3.114)	(3.114)	(7.915)
					CAPM						
Intercept	-0.53*	0.09	0.31	0.52**	0.66**	0.83***	0.90***	0.98***	1.13***	1.37***	1.91***
	(-1.957)	(0.362)	(1.217)	(2.002)	(2.538)	(3.126)	(3.405)	(3.564)	(4.212)	(4.777)	(7.993)
MKT	1.03***	1.06***	1.06***	1.07***	1.06***	1.06***	1.06***	1.05***	1.06***	1.05***	0.02
	(18.670)	(19.936)	(20.335)	(18.784)	(19.532)	(17.810)	(18.784)	(17.434)	(18.148)	(16.807)	(0.578)
					Three-factor model						
Intercept	-0.92***	-0.30**	-0.11	0.07	0.21**	0.36***	0.43***	0.49***	0.64***	0.64***	1.79***
	(-5.759)	(-2.558)	(-1.062)	(0.680)	(2.037)	(3.335)	(3.711)	(4.059)	(4.935)	(4.935)	(7.531)
MKT	0.96***	0.99***	0.99***	1.00***	0.99***	0.99***	0.99***	0.98***	0.99***	0.99***	0.02
	(29.186)	(37.297)	(42.664)	(34.409)	(44.293)	(36.742)	(40.382)	(40.794)	(38.277)	(38.277)	(0.498)
SMB	0.71***	0.72***	0.75***	0.79***	0.79***	0.83***	0.81***	0.85***	0.83***	0.83***	0.15**
	(14.020)	(17.856)	(19.170)	(17.565)	(21.135)	(20.647)	(18.717)	(18.283)	(17.805)	(17.805)	(2.137)
HML	-0.02	-0.02	0.02	0.05	0.05	0.07	0.08	0.11*	0.15**	0.15**	0.18
	(-0.316)	(-0.318)	(0.384)	(0.722)	(1.014)	(1.288)	(1.568)	(1.931)	(2.262)	(2.262)	(1.501)
					Four-factor model						
Intercept	-0.96***	-0.35***	-0.13	0.08	0.23**	0.38***	0.47***	0.53***	0.68***	0.68***	1.90***
	(-5.843)	(-2.963)	(-1.230)	(0.719)	(2.248)	(3.420)	(4.019)	(4.324)	(5.088)	(5.088)	(7.585)
MKT	0.96***	0.99***	0.99***	1.00***	0.99***	0.99***	0.99***	0.98***	0.99***	0.99***	0.01
	(28.476)	(37.105)	(41.893)	(33.639)	(43.494)	(36.065)	(39.967)	(40.176)	(37.390)	(37.390)	(0.334)
SMB	0.72***	0.73***	0.75***	0.79***	0.78***	0.82***	0.80***	0.84***	0.82***	0.82***	0.12
	(13.965)	(18.240)	(19.423)	(17.296)	(21.242)	(20.340)	(19.006)	(17.815)	(17.017)	(17.017)	(1.521)
HML	-0.01	-0.00	0.03	0.05	0.04	0.07	0.07	0.10*	0.14**	0.14**	0.14
	(-0.132)	(-0.042)	(0.494)	(0.679)	(0.839)	(1.187)	(1.376)	(1.734)	(2.150)	(2.150)	(1.246)
UMD	0.05	0.06**	0.02	-0.01	-0.03	-0.02	-0.05**	-0.05*	-0.05	-0.05	-0.13*
	(1.173)	(2.079)	(0.958)	(-0.203)	(-1.303)	(-0.775)	(-1.977)	(-1.744)	(-1.259)	(-1.259)	(-1.894)

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

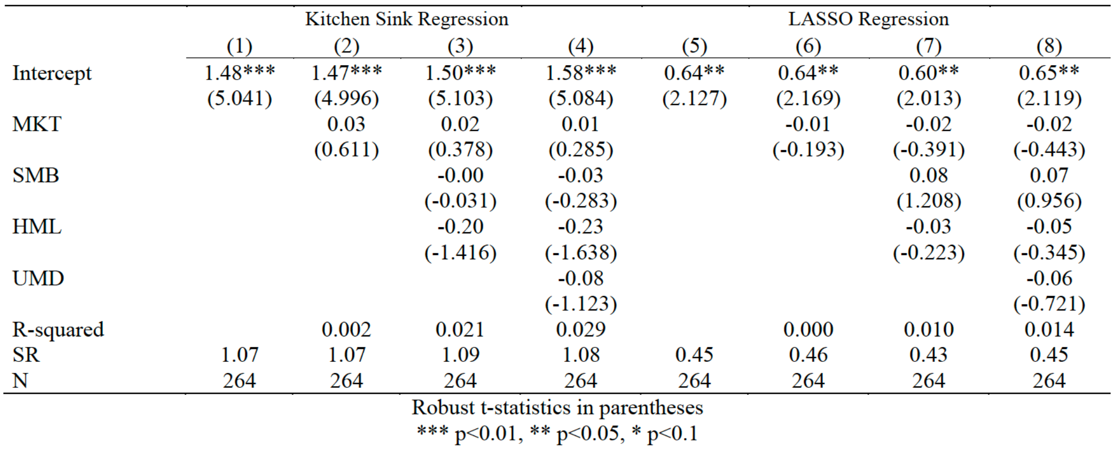
conditional portfolio sorting appears to perform very well, as it generates high and stable out-of-sample excess returns that cannot be explained by the three major equilibrium factor models. This raises the question of what internal mechanisms drive its strong performance. Next, we will analyze the role of the predictive variables to understand this mechanism.

4.2 Variable Importance

Recall that over time, we re-estimate the model every year, resulting in a total of 22 models. When computing the importance of predictive variables each year, we obtain the importance rankings of each variable annually. To summarize, this paper ranks past returns based on the median of these 22 importance rankings. Table 4.4 displays the top ten past return variables along with their upper quartile and lower quartile rankings.

It can be observed that among the top ten important return functions, six are related to medium- and long-term past returns from one year ago. Additionally, some medium-term past returns from 6 to 12 months prior to portfolio formation also appear in the top ten important list. However, all return functions from the most recent six months failed to enter the top ten. This is consistent with previous literature findings that significant performance in the A-share market is only observed in monthly returns from 2-3 years

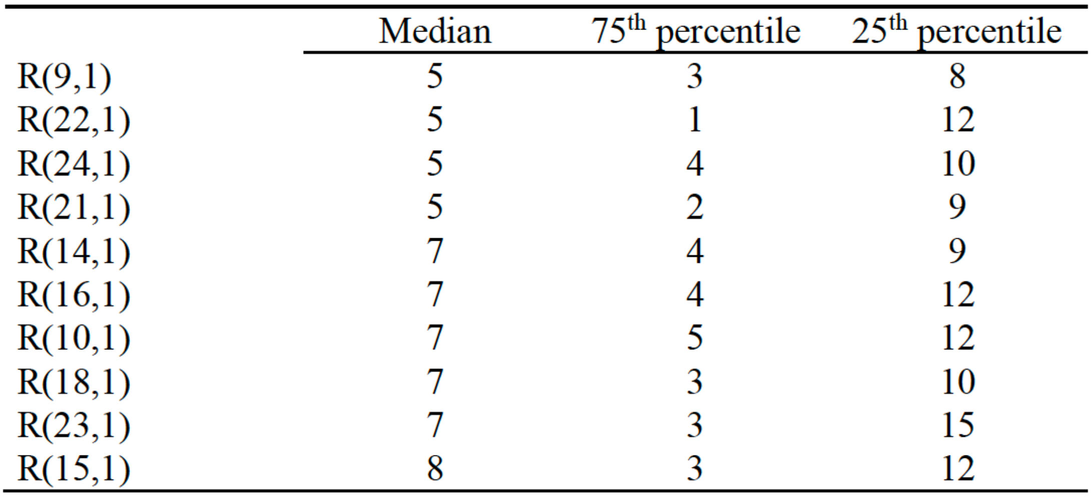
Table 4.3 Strategy Factor Loadings: Fama-MacBeth Regression



	Kitchen Sink Regression				LASSO Regression			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	1.48***	1.47***	1.50***	1.58***	0.64**	0.64**	0.60**	0.65**
	(5.041)	(4.996)	(5.103)	(5.084)	(2.127)	(2.169)	(2.013)	(2.119)
MKT		0.03	0.02	0.01		-0.01	-0.02	-0.02
		(0.611)	(0.378)	(0.285)		(-0.193)	(-0.391)	(-0.443)
SMB			-0.00	-0.03			0.08	0.07
			(-0.031)	(-0.283)			(1.208)	(0.956)
HML			-0.20	-0.23			-0.03	-0.05
			(-1.416)	(-1.638)			(-0.223)	(-0.345)
UMD				-0.08				-0.06
				(-1.123)				(-0.721)
R-squared		0.002	0.021	0.029		0.000	0.010	0.014
SR	1.07	1.07	1.09	1.08	0.45	0.46	0.43	0.45
N	264	264	264	264	264	264	264	264

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.4 Top Ten Past Return Ranking Variables



	Median	75th percentile	25th percentile
R(9,1)	5	3	8
R(22,1)	5	1	12
R(24,1)	5	4	10
R(21,1)	5	2	9
R(14,1)	7	4	9
R(16,1)	7	4	12
R(10,1)	7	5	12
R(18,1)	7	3	10
R(23,1)	7	3	15
R(15,1)	8	3	12

ago.

Furthermore, there are significant changes in rankings over time, as indicated by the quartiles of rankings of each return function in Table 4.4, columns (2) and (3). All of these functions rank in the top half over 50% of the time, with four out of the ten return functions ranking in the top five at least half of the time. However, it is also important to note that each variable experiences periods of lesser relevance for prediction, as shown in column (3).

From this, a conclusion can be drawn that the effectiveness of tree-based conditional portfolio sorting lies in its efficient utilization of important medium- and long-term monthly past returns information. The next section will discuss the direction of these variables' impact.

4.3 Variable Partial Derivatives

Next, this paper analyzes the measurement of approximate partial derivatives introduced in Section 3.3.6. For all observations, the values of the top six ranked monthly return ranking features shown in Section 4.2 are varied from the lowest value (1) to the highest value (10), and predictions are made under this condition. This allows us to track whether variables are monotonically related to predicted returns and approximately estimate the sign of the explanatory variable.

Figure 4.3 presents the results, corresponding to the six most important features considered in tree-based conditional portfolio sorting so far, with each subplot displaying one feature. For each feature, its observations are varied from low to high, and the predicted values for each of the ten values are averaged.

It can be observed that these variables generally exhibit a long-term reversal effect, whereby medium- and long-term past monthly returns are negatively correlated with predicted future returns. In other words, the higher the medium-term past monthly returns, the lower the future returns. For the first three features in row (1), they all exhibit a nonlinear reversal relationship, specifically, extreme groups are more strongly associated with returns, with lower expected returns for low groups and higher expected returns for high groups, but the intermediate groups are less distinct. For the fourth function, i.e., $\boldsymbol{R}(21, 1)$, both high and low values are associated with lower returns, while intermediate values are associated with higher returns. The last two functions generally exhibit a linear monotonic relationship, showing a more pronounced negative correlation with returns. Encouragingly, this is consistent with past literature findings that in the A-share market, there is generally no significant momentum profit at monthly or yearly frequencies, but there is a significant reversal effect at medium to long-term frequencies (Wang and Zhao, 2001[34]; Liu and Pi, 2007[35]; Pan and Xu, 2011[36]).

Figure **??** illustrates contour plots of the top 4 past monthly return ranking functions. In each subplot, darker regions represent higher predicted returns, while lighter regions represent lower predicted returns. Several noteworthy results emerge: Firstly, many interactions of features exhibit a pattern where, once one variable is fixed, some return predictions neither monotonically decrease nor increase within the range of the predicting variable, but rather show a nonlinear relationship with return predictions, as shown in the last

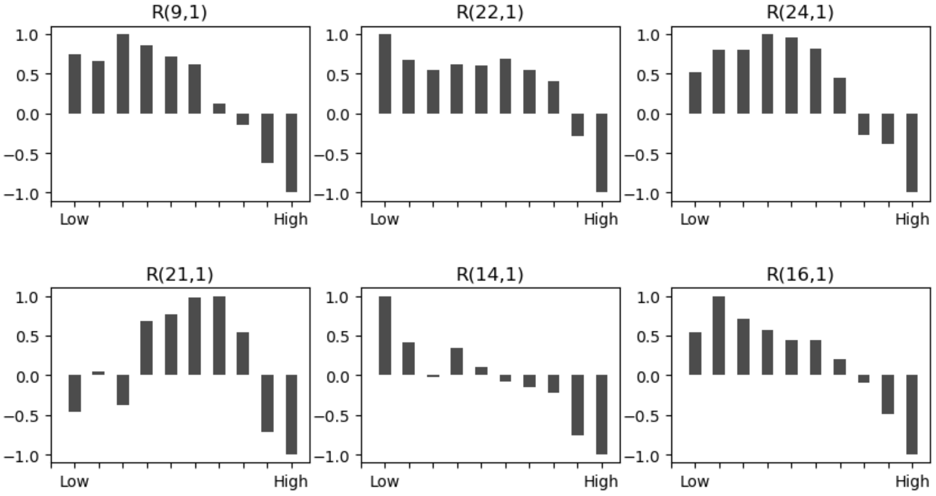


Figure 4.3 Average Partial Derivatives of the Top Six Past Return Ranking Functions

two plots in row (1) and the first two plots in row (2). Specifically, the second plot in row (1) shows the interaction between $R(22,1)$ (monthly return 23 months ago) and $R(24,1)$ (monthly return 25 months ago). When the value of $R(24,1)$ is fixed, the predicted returns are lower at extreme values of $R(22,1)$ but higher at intermediate values. This is in contrast to the relatively monotonic relationship between $R(22,1)$ and predicted returns shown in Figure 4.3, indicating an interaction between $R(22,1)$ and $R(24,1)$. Secondly, for some return features, we find that the predictions of both features are monotonically increasing. For example, the subplot in the top-left corner shows the interaction between $R(21,1)$ (monthly return 22 months ago) and $R(22,1)$ (monthly return 23 months ago). For both variables, higher returns are predicted at lower values and lower returns at higher values, exhibiting reversal effects. Thirdly, some features interact in a nonlinear manner. For instance, the subplot in the bottom-right corner shows the interaction between $R(9,1)$ (monthly return 10 months ago) and $R(24,1)$ (monthly return 25 months ago). In the high range of $R(9,1)$ values, the predicted returns monotonically decrease with $R(24,1)$, while in the low range of $R(9,1)$ values, the predictions exhibit a non-linear relationship with $R(24,1)$, with lower predictions at extreme values and higher predictions at intermediate values.

Currently, there is limited attention in the literature on the nonlinear relationship between past and future returns. However, given that (1) the predictions generated by the tree-based conditional portfolio sorting in this paper yield higher risk-adjusted excess re-

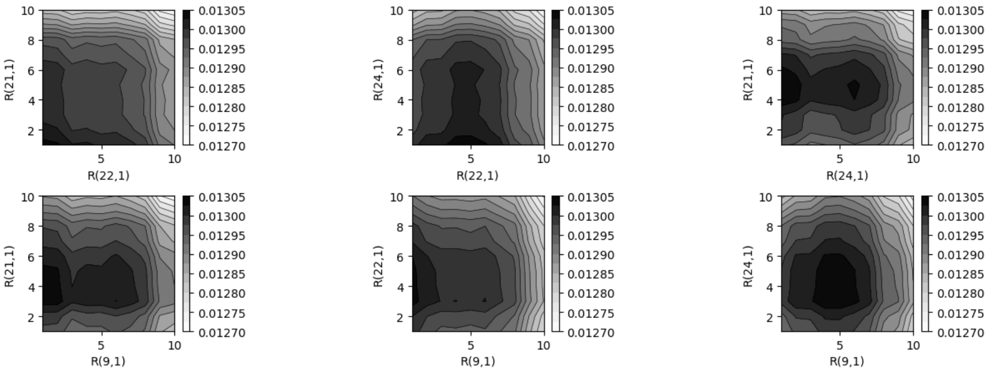


Figure 4.4 Contour plots of the top 4 past monthly return ranking functions

turns than linear models, (2) medium- and long-term past return features exhibit higher performance in the calculation of variable importance in this paper, and (3) the measures of partial derivatives of these variables are not all linearly related to returns, further research into nonlinear relationships is warranted in the future.

5 Conclusion

Since the advent of the Capital Asset Pricing Model (CAPM) in the 1960s and the Arbitrage Pricing Theory (APT) in the 1970s, the academic community has proposed equilibrium pricing models such as the three-factor and four-factor models. However, there is still a lack of consensus on which variables are most relevant to expected stock returns in a comprehensive, important, and stable manner. To date, the literature has identified over 400 anomaly factors that earn returns in a manner not explainable by standard equilibrium models. While many of these factors contain relevant information and some may not hold out of sample, academia has yet to rigorously determine which ones are the most important. Additionally, some factors interact in important ways, making the task of selecting key factors using standard methods challenging. Therefore, this paper draws inspiration from Moritz and Zimmermann (2016)[9] and methods in the field of machine learning, introducing a framework—the tree-based conditional portfolio sorting—that can handle a large number of variables and their potential nonlinearity and interactions, and systematically test all results out of sample. It bridges model evaluation in finance with machine learning in computer science, serving as a bridge between these two fields.

This paper applies this framework to search for information in past monthly returns that may be related to future returns, and designs simple long-short investment portfolios for performance testing. Firstly, the paper obtains positive annual returns for all years, and the net values of long and short positions are both effective. Secondly, the excess monthly returns adjusted by the three-factor and four-factor equilibrium models are significant and close to 2%, with Sharpe ratios significantly higher than simple linear Fama-MacBeth regressions. In other words, using the same variables as linear models in the tree-based conditional portfolio sorting framework, it is found that higher and more stable excess returns can be generated, indicating that the linear framework has not fully utilized all relevant information in the data.

In addition, to address the ”black box” issue of machine learning methods and ensure that nonlinear relationships are not only captured but also precisely understood, this paper adopts methods for computing the importance of predictive variables and measuring approximate partial derivatives. This allows for interpretable information to be extracted from the results of tree-based conditional sorting, even if the results are not presented in the form of equations. Firstly, the paper finds that among the monthly return functions in the two years prior to portfolio formation, the medium to long-term return functions are most important for accurate predictions. This is consistent with past literature indicating that monthly data in the A-share market is effective only on a 2-3 year scale. Secondly, through the calculation of partial derivative measures, it is found that important medium to long-term past returns exhibit a reversal effect on future returns, which is also consistent with previous literature. Additionally, it innovatively discovers that some of these past returns exhibit significant interaction effects, leading to a more complex nonlinear relationship between past and future returns. The discovery of such nonlinear relationships provides insights and suggestions for future research.

Finally, the tree-based conditional portfolio sorting model can be easily extended to incorporate new anomaly factors. If an anomaly factor is highly correlated with predictions, it should appear as one of the most important variables. Therefore, important variables can be added to the existing set of variables. Predictions made on this expanded set of variables effectively control for correlations with other variables and take into account potential interactions and nonlinear relationships. This paper, serving as a precedent for the effective use of the tree-based conditional portfolio sorting framework in the A-share market, holds positive implications for the continued expansion and acceleration of cross-sectional asset pricing research using this method in different markets and with different anomalies.

Reference

[1] TREYNOR J L. Market Value, Time, and Risk[EB/OL]. 1961. https://papers.ssrn.com/abstract=2600356.

[2] TREYNOR J L. Jack Treynor's 'Toward a Theory of Market Value of Risky Assets' [EB/OL]. 1962. https://papers.ssrn.com/abstract=628187.

[3] SHARPE W F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*[J]. The Journal of Finance, 1964, 19(03): 425-442.

[4] LINTNER J. Security Prices, Risk, and Maximal Gains From Diversification[J]. The Journal of Finance, 1965, 20(04): 587-615.

[5] LINTNER J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets[J]. The Review of Economics and Statistics, 1965, 47(01): 13.

[6] MOSSIN J. Equilibrium in a Capital Asset Market[J]. Econometrica, 1966, 34(04): 768-783.

[7] ROSS S A. The arbitrage theory of capital asset pricing[J]. Journal of Economic Theory, 1976, 13(03): 341-360.

[8] COCHRANE J H. Presidential Address: Discount Rates[J]. The Journal of Finance, 2011, 66(04): 1047-1108.

[9] MORITZ B, ZIMMERMANN T. Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns[EB/OL]. 2016. https://papers.ssrn.com/abstract=2740751.

[10] JEGADEESH N, TITMAN S. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency[J]. The Journal of Finance, 1993, 48(01): 65-91.

[11] FAMA E F, FRENCH K R. Common risk factors in the returns on stocks and bonds [J]. Journal of Financial Economics, 1993, 33(01): 3-56.

[12] CARHART M M. On Persistence in Mutual Fund Performance[J]. The Journal of Finance, 1997, 52(01): 57-82.

[13] NOVY-MARX R. The other side of value: The gross profitability premium[J]. Jour-

nal of Financial Economics, 2013, 108(01): 1-28.

[14] FAMA E F, FRENCH K R. A five-factor asset pricing model[J]. Journal of Financial Economics, 2015, 116(01): 1-22.

[15] HOU K, XUE C, ZHANG L. Digesting Anomalies: An Investment Approach[J]. The Review of Financial Studies, 2015, 28(03): 650-705.

[16] STAMBAUGH R F, YUAN Y. Mispricing Factors[J]. The Review of Financial Studies, 2017, 30(04): 1270-1315.

[17] DANIEL K, HIRSHLEIFER D, SUN L. Short- and Long-Horizon Behavioral Factors[J]. The Review of Financial Studies, 2020, 33(04): 1673-1736.

[18] SUBRAHMANYAM A. The Cross-Section of Expected Stock Returns: What Have We Learnt from the Past Twenty-Five Years of Research?[J]. European Financial Management, 2010, 16(01): 27-42.

[19] GOYAL A. Empirical cross-sectional asset pricing: a survey[J]. Financial Markets and Portfolio Management, 2012, 26(01): 3-38.

[20] GREEN J, HAND J, ZHANG X. The Supraview of Return Predictive Signals[J]. Review of Accounting Studies, 2013, 18(03): 692-730.

[21] KAKUSHADZE Z. 101 Formulaic Alphas[J]. Wilmott Magazine, 2016, 84(07): 72-80.

[22] GREEN J, HAND J, ZHANG X. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns[J]. Review of Financial Studies, 2017, 30(12): 4389-4436.

[23] YAN X S, ZHENG L. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach[J]. The Review of Financial Studies, 2017, 30(04): 1382-1423.

[24] FENG G, GIGLIO S, XIU D. Taming the Factor Zoo: A Test of New Factors[J]. The Journal of Finance, 2020, 75(03): 1327-1370.

[25] 李斌. 机器学习驱动的基本面量化投资研究[J]. 中国工业经济, 2019, 377(08): 61-79.

[26] KOZAK S, NAGEL S, SANTOSH S. Shrinking the cross-section[J]. Journal of Financial Economics, 2020, 135(02): 271-292.

[27] KELLY B T, PRUITT S, SU Y. Characteristics are covariances: A unified model of

risk and return[J]. Journal of Financial Economics, 2019, 134(03): 501-524.

[28] JEGADEESH N. Evidence of Predictable Behavior of Security Returns[J]. The Journal of Finance, 1990, 45(03): 881-898.

[29] JEGADEESH N, TITMAN S. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations[J]. The Journal of Finance, 2001, 56(02): 699-720.

[30] MOSKOWITZ T J, OOI Y H, PEDERSEN L H. Time series momentum[J]. Journal of Financial Economics, 2012, 104(02): 228-250.

[31] NOVY-MARX R. Is momentum really momentum?[J]. Journal of Financial Economics, 2012, 103(03): 429-453.

[32] GOYAL A, WAHAL S. Is Momentum an Echo?[J]. The Journal of Financial and Quantitative Analysis, 2015, 50(06): 1237-1267.

[33] FAMA E F, FRENCH K R. Size, value, and momentum in international stock returns [J]. Journal of Financial Economics, 2012, 105(03): 457-472.

[34] 王永宏, 赵学军. 中国股市“惯性策略”和“反转策略”的实证分析[J]. 经济研究, 2001, 36(06): 56-61.

[35] 刘博, 皮天雷. 惯性策略和反转策略：来自中国沪深 A 股市场的新证据[J]. 金融研究, 2007, 326(08A): 154-166.

[36] 潘莉, 徐建国. A 股个股回报率的惯性与反转[J]. 金融研究, 2011, 367(01): 149-166.

[37] 鲁臻, 邹恒甫. 中国股市的惯性与反转效应研究[J]. 经济研究, 2007, 42(09): 145-155.

[38] 高秋明, 胡聪慧, 燕翔. 中国 A 股市场动量效应的特征和形成机理研究[J]. 财经研究, 2014, 40(02): 97-107.

[39] BREIMAN L. Random Forests[J]. Machine Learning, 2001, 45(01): 5-32.

Acknowledgements

For this research, I would like to express my gratitude to many individuals.

First and foremost, I would like to thank my advisor, Professor Bin Li, a pioneer in the field of quantitative investment research, for guiding me in selecting a topic that I am interested in and that holds research value. Next, I want to express my heartfelt appreciation to my parents, who have provided me with almost all material support and more importantly the unconditional moral support. They are my eternal harbor. Furthermore, I would like to thank my friends for their emotional warmth, encouragement, and vitality. They have consistently infused me with fresh insights and perspectives. Lastly, I want to thank myself. I have witnessed my own progress through each step of the research, overcoming every obstacle and gradually growing both professionally and mentally. I have become a wiser person and have learned to accept everything that comes my way, naturally exploring new interests along the journey. How wonderful it is!

I extend my heartfelt thanks to all mentioned above, and I wish us all health, happiness, and success.