Mosaics of Predictability

Lin William Cong Guanhao Feng Jingyu He Yuanzhi Wang*

First draft: Feb 2024; this draft: August 2024

Abstract

Existing studies on asset return predictability focus on aggregate performance. We examine the oft-overlooked grouped heterogeneity in return predictability across different assets and macroeconomic regimes. A novel tree-based asset clustering methodology is introduced to partition the panel of asset-return observations according to return predictability, using high-dimensional asset characteristics and aggregate time series predictors. When implemented on U.S. equities over the past five decades, we find that some characteristics-managed (dollar trading volumes, unexpected earnings, earnings-to-price, and cashflow-to-price) and/or macro-based (dividend yield and default yield) clusters are more predictable, resulting in a heterogeneous predictive model with outperformance. Finally, less predictable clusters generally exhibit lower risk-adjusted investment performance, revealing an important empirical link between return predictability and trading profitability.

Key Words: Heterogeneity, High-dimensional Characteristics, Regime Switches, Return Predictability, Tree-based Clustering.

*We are grateful to Chunrong Ai, Doron Avramov, Daniele Bianchi, Jian Chen (Discussant), Qihui Chen, Victor DeMiguel, Fuwei Jiang, Hao Ma, Dacheng Xiu, Lei Yang (Discussant), Yi Zhang, and the seminar and conference participants at CUHK Shenzhen, 2024 EFMA Annual Meeting, 2024 FMA Asia/Pacific Conference, HKUST Guangzhou, Queen Mary University of London, 4th Big Data and Econometric Conference at Xiamen University, 16th Annual SoFiE Meeting, 8th PKU-NUS Annual International Conference on Quantitative Finance and Economics, Summer Institute of Finance 2024, 2024 Academic Conference on Digital Economy Development and Governance, SIDF 2024, 17th ARMC 2024. Cong (E-mail: will.cong@cornell.edu) is at Cornell University SC Johnson College of Business (Johnson) and NBER; Feng (E-mail: gavin.feng@cityu.edu.hk), He (E-mail: jingyuhe@cityu.edu.hk), and Wang (E-mail: yuanzwang5-c@my.cityu.edu.hk) are at the City University of Hong Kong.

1 Introduction

Forecasting asset returns is a long-standing topic in asset pricing, with various methods developed to study the complex return dynamics. Existing studies on individual stock return predictability (e.g., Fama and French, 2008; Lewellen, 2015) focus on whether individual stock returns are predictable *on average*. Using the entire stock universe, researchers evaluate the significance of predictor coefficients (through Fama-Macbeth or panel regression models) or implement a forecast-implied strategy for risk-adjusted performance. The answer is affirmative, as reconfirmed by recent literature on empirical asset pricing via advanced machine learning methods (e.g., Gu et al., 2020).

However, the heterogeneous nature of asset return predictability, which is about the return forecasting difficulty or signal-to-noise ratio, is less explored. Certain assets exhibit higher return predictability than others, and the predictability of individual asset returns varies with market conditions. For example, Avramov et al. (2023) find that return predictability concentrates in micro-caps, distressed stocks, or high volatility periods. Green et al. (2017) also find a sharp decline (i.e., less significant characteristics) of individual stock return predictability since 2003. To this end, we treat predictability as an unobservable property of individual stock returns. We investigate a novel and general research question: What types of assets exhibit higher return predictability, and under which macroeconomic regimes?

"Return predictability" is more nuanced than "return levels." Notably, the literature has no unanimous conclusion on whether high predictability implies high return levels. Instead, return predictability is closely related to the signal-to-noise (S2N) ratio or the regression R^2. Our research, therefore, concentrates on R^2 to identify stock types or macroeconomic regimes that exhibit higher predictability rather than focusing on expected return levels. According to Rapach et al. (2010) and Kelly et al. (2024), there is no direct relationship between the out-of-sample predictive R^2 and investment gains (i.e., levels of returns). However, we illustrate that highly predictable stocks consis-

tently outperform the less predictable ones in risk-adjusted performance.

For the long-documented stock market return predictability linked to business cycle predictors (e.g., Keim and Stambaugh, 1986; Fama and French, 1989), researchers have investigated the time-varying (heterogeneous in time series) return predictability and found stronger predictability during economic recessions (e.g., Henkel et al., 2011; Dangl and Halling, 2012). A recent study by Farmer, Schmidt, and Timmermann (2023) also finds "pockets of predictability," where return predictability is intermittent, with short periods of significant predictability alternating with long periods of no predictability. The time-varying coefficient model is commonly used for these empirical investigations to examine the predictability of heterogeneous market returns over time. The predictor coefficient for market returns can change over time in different macroeconomic regimes and may exhibit sparsity over long periods.

Our research explores the heterogeneous predictability of individual asset returns, significantly broadening the scope of the current literature on time-varying predictability and market returns. This leads to the concept of "mosaics" of predictability in the panel data of individual stocks, which builds from "pockets" of predictability of market returns (Farmer et al., 2023). The return predictability varies over time periods and thousands of stocks, revealing cluster-wise heterogeneous patterns that look like "mosaics" on the panel. As demonstrated by Cong, Feng, He, and Li (2023), these heterogeneous panel data analyses can be treated as a clustering problem, identifying group observations with similar patterns. Beyond clustering, heterogeneous models can be more effective than homogeneous models.

We introduce a novel clustering approach based on a single decision tree with a customized objective to tackle this new empirical challenge. Specifically, our approach identifies optimal clustering structures that capture variations of stock return predictability – stock returns with similar levels of predictability are put in the same cluster. Furthermore, the decision tree structure organizes cluster patterns in mosaics on the panel, described by firm characteristics and aggregate predictors. It preserves the economic interpretations and identifies critical variables related to the heteroge-

neous level of predictability. By utilizing the asymmetric interactions between aggregate predictors and firm characteristics, we uncover heterogeneity in stock returns, fit cluster-wise heterogeneous predictive models, and enhance the forecasting power.

Methodological Innovations. We present a "divide-and-conquer" goal-oriented clustering approach that sequentially separates asset-return observations based on their predictability. Unlike traditional clustering algorithms such as K-means, which minimize within-cluster variation of observable characteristics, our method aims to distinguish clearly between highly and less predictable stock returns, even though the predictability is not directly observable from data. This partitioning is achieved by maximizing the difference of S2N ratios, R^2, between clusters through cluster-wise heterogeneous return forecast models. Our study is related to the broader investigation of asset heterogeneity alongside Cong et al. (2023) and Cong et al. (2023) but differs in its economic objective for the clustering. The former focuses on maximizing the collective Sharpe ratio to span the efficient frontier, and the latter aims to maximize the marginal likelihood of the heterogeneous factor models. Our paper focuses on differentiating stock return predictability.

Moreover, the return prediction literature (e.g., Gu et al., 2020) typically fits a homogeneous global model that applies to all individual stocks and time periods. However, Feng and He (2022) and Evgeniou et al. (2023) argue that such homogeneous modeling largely ignores the heterogeneity of asset returns. Our framework addresses this issue by fitting a cluster-wise heterogeneous predictive model based on the clustering structure. The clusters are defined by firm characteristics and/or aggregate predictors. Notably, clustering and fitting local predictive models are seamlessly integrated within a unified framework rather than decomposed as two separate steps. Furthermore, this clustering framework does not depend on the specific choice of the workhorse prediction model employed. It can be integrated with various machine learning models: regression, Lasso, PCA, random forest, and deep learning.

Empirical Highlights. We empirically analyze a large panel of individual U.S. stock returns from 1973 to 2022, utilizing 58 monthly firm-level characteristics from eight

major categories alongside eight monthly aggregate predictors.

We present clustering results on the cross-section, utilizing all characteristics, yielding about 15 clusters (distinct leaf nodes) within a decision tree. Notably, the S2N ratios R^2 vary dramatically from 0.78% (Figure A.8, stocks with non-high SUE, non-low DOLVOL, non-high SVAR, and non-high RDM) to 12.08% (Figure A.2, stocks with high SUE, high EP, low DOLVOL, and high BM_IA). Similar to small-growth or high-value stocks in Fama and French (1992), characteristic ranges define our clusters, and we know the stock types for high or low return predictability. Characteristics such as dollar trading volumes, unexpected earnings, earnings-to-price, and cashflow-to-price play a crucial role in distinguishing the predictability of stock returns. These findings of SUE align with the post-earning-announcement drift of Ball and Brown (1968) and Bernard and Thomas (1989). We validate the predictability clustering by finding consistent out-of-sample performance, where highly and less predictable clusters remain consistent in the out-of-sample data.

Second, we extend our clustering analysis with macroeconomic regime changes and structural breaks. We observe the mosaics of stock return predictability in both time series and cross-sectional dimensions. Notably, stock return predictability drops dramatically when the dividend yield, an important business cycle predictor (Campbell and Shiller, 1988), is not high. Conversely, stock returns are highly predictable when the dividend yield is high and the default yield is low, conditions typical of recession periods. Under three macroeconomic regimes, each decision tree selects different firm characteristics, but value characteristics consistently emerge as important variables. Structural breaks by calendar months exhibit similar phenomena.

Furthermore, we rank clusters according to their predictability, demonstrating that the highly predictable clusters consistently outperform the less predictable ones in risk-adjusted performance. This finding complements Rapach et al. (2010) and Kelly et al. (2024), indicating that clustered return predictability is associated with investment gains. Additionally, excluding these highly predictable clusters will significantly decrease the overall predictability of the remaining stock universe and lower

the risk premiums of multiple long-short risk factors. These findings support the idea that a small subset of predictable stock returns significantly influences the overall predictability, confirming the existence of mosaics of predictability.

Finally, we create forecast-implied long-short portfolios to assess cluster-wise heterogeneous predictive modeling enhancement. We find consistent trends in various financial metrics, including average returns, Sharpe ratio, market alphas, and maximum draw-downs. For example, in the highly predictable cluster identified in the cross-sectional clustering, the monthly average return of value-weighted portfolios reaches 3.49%. However, the performance of the less predictable clusters is significantly worse than that of others, with an average return of only 0.53%. Additional time series clustering shows a more stable and clearly declining performance trend. These findings indicate that cluster-wise heterogeneous predictive modeling significantly outperforms traditional global homogeneous modeling.

Literature. Our paper contributes to the extensive empirical literature on return predictability. For aggregate markets of equities and bonds, early studies (e.g., Keim and Stambaugh, 1986; Fama and French, 1989) identify useful market-wide predictors (e.g., term spread and default spread) when studying time series return predictability over business cycles. For individual stock returns, a zoo of characteristics, anomalies, or long-short factors (e.g., market equity values, book-to-market ratios, and prior cumulative returns) have been documented (e.g., Fama and French, 1992, 1993; Jegadeesh and Titman, 1993) while investigating cross-sectional return predictability.[1] However, many of these empirical findings appear unstable in out-of-sample or post-publication evaluations (Pesaran and Timmermann, 1995; Welch and Goyal, 2008; Harvey et al., 2016; McLean and Pontiff, 2016). Our study complements this by investigating the heterogeneity in return predictability, which may help explain the inconsistencies in asset return predictions.[2]

[1] Green et al. (2013) find that characteristics discovered in the 2000s have similar return properties to those discovered in earlier decades, indicating that investment technologies have improved over time to offset a shrinking pool of high-performing but not yet discovered characteristics.

[2] More specifically on predictability testing, studies such as Stambaugh (1999) and Lewellen (2004) focus on statistical properties of time series predictability of the market-wide indices. For the cross-

Our paper joins the emerging studies on heterogeneous return predictability. Studies on time-varying return predictability demonstrate that market returns are more predictable during economic recessions when analyzed using a regime-switching VAR (Henkel et al., 2011) and a time-varying coefficient model (Dangl and Halling, 2012). Farmer et al. (2023) document calendar-time pockets of differential predictability. In the cross-section, Avramov (2002) finds that small-cap value stocks are more predictable than large-cap growth stocks, and Green et al. (2017) show that hedge returns from exploiting characteristics-based predictability have been insignificant outside of micro-caps since 2003. Avramov et al. (2023) further discover that predictability is concentrated in micro-caps, distressed stocks, or during periods with high volatility. Our model more systematically analyzes the heterogeneity in return predictability, expanding the traditional time-varying coefficient model by incorporating high-dimensional characteristics and macroeconomic variables.

This paper, along with Cong et al. (2023) and Cong et al. (2023), is among the first to develop economically guided clustering (i.e., panel trees), which belongs to the emerging AI literature on goal-oriented search – a data-driven approach to optimizing an economic goal in a large and flexible modeling space (e.g., Cong et al., 2020). The "divide-and-conquer" approach of our panel trees mimics how humans solve complex problems by completing constituent tasks. Cong et al. (2023) and Bryzgalova et al. (2023) focus on portfolio estimation for endogenously and exogenously generated leaf-basis portfolios. Cong et al. (2023) and Feng et al. (2024) extend the panel-tree framework to fit heterogeneous models by maximizing the marginal likelihood, while ours focuses on separating observations for heterogeneous signal-to-noise ratios. Closely related by also analyzing endogenous grouped heterogeneity in financial markets, Ahn et al. (2009) use unsupervised clustering based on return correlations, and Patton and Weller (2022) generalize K-means to group assets based on within-

sectional return predictability of individual assets, a standard test for anomaly predictability entails risk premium estimate from the cross-sectional regression (Fama and MacBeth, 1973; Fama and French, 2008; Lewellen, 2015). Similar to the recent literature of machine learning models using panel regression for individual stock returns and risk-adjusted performance of the long-short portfolios (e.g., Gu et al., 2020; Freyberger et al., 2020), our clustering objective focuses on the fitness of panel return forecasts.

group slopes and averages, and find risk-price heterogeneity pervasive and essential. More recently, Evgeniou et al. (2023) applies unsupervised K-means to cluster firms based on characteristics and estimates post-cluster heterogeneous predictive models. Our clustering differs in being endogenously guided by economic objectives (largest predictability dispersion) and further incorporating time series splits (e.g., clustering for predictability "mosaics" in panels of asset returns).

The remainder of the paper is organized as follows: Section 2 illustrates the clustering model with an economically guided split criterion, while Section 3 describes the data and model evaluation. Section 4 presents the empirical findings from U.S. equities, and Section Section 5 reports the predictability-implied investment performance. Finally, Section 6 concludes, with the appendices containing finer details of the algorithms and data.

2 Methodology

2.1 Measurement of Return Predictability for Clustering

As discussed earlier, predictability is generally associated with the signal-to-noise ratio, or R^2. However, it is unobservable and lacks a consensus definition in the literature. Furthermore, no conclusive evidence indicates that assets with higher predictability yield higher returns. We choose in-sample R^2 as a measure of return predictability to train the clustering models for several reasons. First, we revisit the calculations of R^2 to justify its use as a measure of return predictability. In the literature (e.g., Fama and French, 2008; Lewellen, 2015; Gu et al., 2020), a predictive model is typically represented as:

$$r_{i,t+1} = \mathrm{E}_t[r_{i,t+1}] + \epsilon_{i,t+1}, \tag{1}$$

with assumption $\mathrm{E}[\epsilon_{i,t+1}] = 0$ such that the prediction of expected return is unbiased. The information regarding the heterogeneous predictable difficulty can be represented by the signal-to-noise ratio in Eq. (2):

$$R^2_{i,t} = 1 - \mathrm{Var}(\epsilon_{i,t})/\mathrm{Var}(r_{i,t}) := 1 - \sigma^2_{\epsilon,i,t}/\sigma^2_{i,t}, \tag{2}$$

where $\sigma^2_{\epsilon,i,t}$ and $\sigma^2_{i,t}$ are the variance of $\epsilon_{i,t}$ and $r_{i,t}$ respectively. We add subscript i and t for this metric to highlight its cross-sectional and time series variability. Conceptually, when $R^2_{i,t}$ is high for asset i at a specific period t, it is relatively easier for a predictive model to capture the conditional expectation. If the noise is large for asset i, even knowing the true $\mathrm{E}_t[r_{i,t+1}]$ can still yield small R^2, let alone learning the conditional expectation from super noisy data. Therefore, with certain model regularization, the in-sample $R^2_{i,t}$ is a reasonable measure of the signal-to-noise (S2N) ratios, reflecting the return predictability for various assets across different periods. Even without model regularization, the in-sample R^2 is barely over 5% during the last fifty years.

Second, estimating $R^2_{i,t}$ for each asset i and time t is challenging due to the lack of data. Typically, the literature addresses this issue by fitting a pooled model and calculating a single R^2 for all assets across all periods, ignoring the panel structure (Gu et al., 2020). However, this metric ignores the heterogeneity of predictability. Our approach balances these two extremes: we cluster asset returns into a few subsets and calculate R^2 for each cluster to measure the cluster-wise predictability. This clustering method is specifically designed to maximize the differences in predictability among clusters. Details of the clustering approach will be discussed in Section 2.3.

Third, readers may question why we do not use out-of-sample (OOS) $R^2_{i,t}$ to guide our clustering approach. The primary reason is that focusing on the OOS metric could prevent our clustering algorithm from detecting time heterogeneity patterns during training periods. Additionally, notice that the expected OOS mean squared error has the well-known bias-variance decomposition as:

$$\mathrm{E}_t[(r_{i,t+1} - \mathrm{E}_t[r_{i,t+1}])^2] = \Big(\mathrm{Bias}\{\widehat{\mathrm{E}}_t[r_{i,t+1}]\}\Big)^2 + \sigma^2_{\epsilon,i,t+1} + \mathrm{Var}\{\widehat{\mathrm{E}}_t[r_{i,t+1}]\}. \tag{3}$$

When predicting stock returns, $\sigma^2_{\epsilon,i,t+1}$ dominates this measure due to the meager S2N ratio. Robust prediction benchmarks, such as zero for individual stocks or the historical average for the market index, may not reduce predictive bias but typically lead to almost zero predictive variance. Despite these challenges, as reported in the literature, the OOS $R^2_{i,t}$ is generally less than 1%, suggesting its limited effectiveness in

environments characterized by low S2N ratio.

Last, we use the in-sample metric solely to identify clustering patterns in stock returns. We still evaluate the out-of-sample performance once the clustering pattern is detected and heterogeneous predictive models are fitted.

2.2 Heterogeneous Predictive Modeling

Next, we introduce our methodology. We denote the data as $\mathcal{D} = \{(r_{i,t+1}, \mathbf{z}_{i,t}, \mathbf{x}_t) \mid i = 1, \ldots, N \text{ and } t = 1, \ldots, T_i\}$, where $r_{i,t+1}$ is the excess return of stock i at time period $t+1$. Common predictors used in stock return prediction literature include $\mathbf{z}_{i,t}$, a C-dimensional vector of firm characteristics, and $\mathbf{x}_t$, a M-dimensional vector of aggregate predictors.

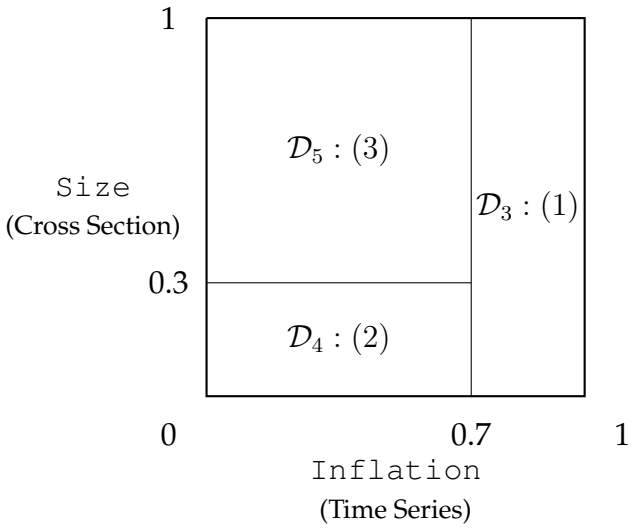
Ideally, the best way to model heterogeneous and time-varying expected excess return $\mathrm{E}_t[r_{i,t+1}]$ is to express it as $g_{i,t}(\mathbf{z}_{i,t}, \mathbf{x}_t)$ that varies for different assets across different periods. However, due to the limited observations for individual stock return data, estimating each $g_{i,t}(\cdot)$ separately is hard. Instead, many studies (e.g., Gu et al., 2020) estimate a homogeneous predictive function $g_t(\cdot)$ and update the time-varying coefficients through a rolling-window scheme. However, Feng and He (2022) and Evgeniou et al. (2023) argue that such homogeneous modeling of $g_t(\cdot)$ largely ignores the predictive power heterogeneity for different assets. More importantly, this homogeneous predictive modeling implicitly assumes homogeneous return predictability (same R^2 for every asset i), conflicting with the empirical findings (Hou et al., 2020; Avramov et al., 2023). Moreover, while rolling-window estimation is robust, it naively overlooks potential macroeconomic-driven regime shifts in the stock market.

We propose a heterogeneous predictive modeling approach that lies in between an individual model and a pooled model, where we fit a cluster-wise heterogeneous predictive model. Unlike the two-step clustering and estimation of Evgeniou et al. (2023), we adopt the panel tree framework proposed by Cong et al. (2023) and customize it for clustering according to asset return predictability. The clustering approach aims to obtain multiple small rectangles (leaf nodes or clusters) by partitioning the entire stock return panel. The entire panel is partitioned through the cross-

sectional and time series dimensions according to values of firm characteristics $\mathbf{z}_{i,t}$ and aggregate predictors $\mathbf{x}_t$. See below Figure 1 for an illustration. There are three non-overlapped clusters for the panel of stock return observations[3] : (1) high inflation, (2) non-high inflation and small-cap, (3) non-high inflation and non-small cap.

Figure 1: **Clustering illustration via partitions**

This figure partitions the panel of stock returns into three rectangular $\mathcal{D}_3$, $\mathcal{D}_4$, and $\mathcal{D}_5$. The first partition is inflation at 0.7, and the second partition is size at 0.3 when inflation is not high.



Rather than fitting individual predictive models, $g_{i,t}(\cdot)$, for every asset i in period t or a pooled model, we only estimate a small number of heterogeneous models for each cluster. The cluster-wise predictive model is:

$$E_t[r_{i,t+1}] = g_j(\mathbf{z}_{i,t}, \mathbf{x}_t), \tag{4}$$

where stock return observations in the j-th cluster follow the same predictive model $g_j(\cdot)$. Our approach clusters observations and estimates local predictive models simultaneously, putting stock return observations with similar return predictability in the same cluster. This is in contrast to the two-step approach of Evgeniou et al. (2023) that clusters on firm IDs $\{r_{i,t+1}\}_{t=1}$. Remarkably, the clustering approach allows the user's

[3]Note that a company can change clusters if the associated characteristic value changes over time. For instance, if a company transitions from a small firm to a large one in market equity values, its cluster membership may change based on the partition outcome.

arbitrary choice of workhorse predictive model $g_j(\cdot)$. For simplicity, we illustrate our approach using Ridge regression, which is suggested to be robust under weak signal scenarios by Shen and Xiu (2024). Consequently, the in-sample R_j^2 is calculated with stock returns in the same j-th cluster. Next, we illustrate our tree-based clustering approach step-by-step.

2.3 Clustering: First Split

Standard clustering algorithms, such as K-means, optimize within-cluster observation distances based on characteristic space. These algorithms group returns with similar characteristic values together. However, our clustering problem aims to partition the data sample according to the unobservable predictability, distinguishing between highly and less predictable observations. This implementation involves optimizing the separation of the data sample into two groups to maximize the difference in their signal-to-noise (S2N) ratios or R^2 values.[4]

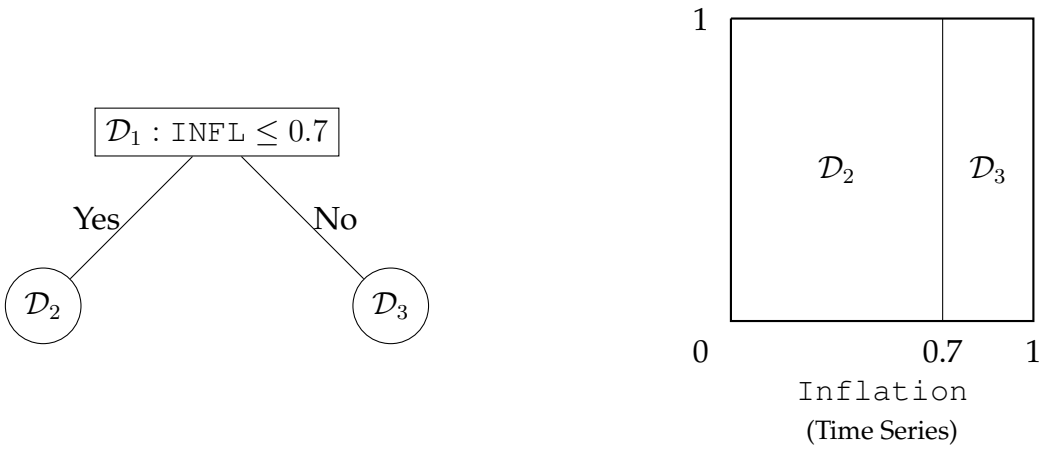
Relying on $\mathbf{z}_{i,t}$ and $\mathbf{x}_t$, we employ an iterative approach to partition the panel of stock return observations into clusters sequentially, adding one cluster at a time and visualizing the results with a decision tree. The initial step involves dividing the data to identify the optimal split predictor and cutpoint value that effectively separates the data into two clusters, fitting predictive models and calculating the R^2 for every split candidate, and finally, searching for the optimal one that maximizes the R^2 difference.

Figure 2 illustrates one *candidate* for the first split through a decision tree, which splits the panel of stock return observations in the root node $\mathcal{D}_1$ into two clusters. They are also referred to as *leaf nodes* in the machine learning terminologies, representing any node without subsequent node under it. $\mathcal{D}_2$ and $\mathcal{D}_3$ according to a split rule "$\text{var}_p(\texttt{Inflation}) \leq c_k(0.7)$". Here, the variable "var" is a characteristic or aggregate predictor. To gauge the quality of the split candidate, we define a goal-oriented split criterion, evaluating whether the candidate can successfully distinguish returns with high predictability from those with lower predictability.

[4]There might be other clustering objectives for achieving this goal of separating partitioned subsamples but maximizing the R^2 difference is interpretable and feasible.

Figure 2: **Illustration for the first split**

To calculate the split criterion and search for the optimal split predictor and cutpoint value, we consider one of the first split candidates, INFL≤ 0.7. The left figure shows a decision tree that splits the sample, and the right figure shows the corresponding partition plot that only partitions over time.



The first candidate divides the entire return samples $\mathcal{D}_1$ into two clusters $\mathcal{D}_2$ and $\mathcal{D}_3$. We fit two cluster-wise predictive models to these clusters separately, denoted as $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$. Generally, for the j-th leaf node, we fit a cluster-wise predictive model $\widehat{g}_j(\cdot)$, and the return predictions are denoted as $\widehat{r}_{i,t+1} = \widehat{g}_j(\mathbf{z}_{i,t}, \mathbf{x}_t)$. Fama and French (2008) criticize that small-caps with high return variance largely dominate the panel regression, and Hou et al. (2020) show that many anomalies are replicable due to the dominance of micro-caps (about 60% of all firms) in the cross-sectional regression. Therefore, we consider using the volatility-weighted Ridge regression for the split criterion calculation.

$$
\begin{aligned}
& g_j(\cdot) = \beta_0 + \beta^\intercal \mathbf{s}_{i,t} + \epsilon_{i,t}, \\
& \text{and } \widehat{\beta}_j = \arg\min_{\beta_0,\beta} \left\{ \frac{1}{N_{\text{leaf}_j}} \sum_{\text{leaf}_j} w_{i,t} \left(r_{i,t+1} - \beta_0 - \beta^\intercal \mathbf{s}_{i,t} \right)^2 + \lambda ||\beta||_2^2 \right\},
\end{aligned} \tag{5}
$$

where $\mathbf{s}_{i,t} = \{\mathbf{z}_{i,t}, \mathbf{x}_t\}$ and $w_{i,t} = 1/\sigma_{i,t}^2$ is the inverse of idiosyncratic return variance. The volatility, $\sigma_{i,t}^2$, is estimated on a rolling-window basis, which helps to incorporate both the time series and cross-sectional variation for observation weights within the leaf cluster. The hyper-parameter λ is determined by cross-validation. Therefore, $\widehat{r}_{i,t+1} = \widehat{\beta}_{j,0} + \widehat{\beta}_j^\intercal \mathbf{s}_{i,t}$ is the heterogeneous return forecast for calculating the correspond-

ing S2N ratios, R^2. Within the j-th leaf node:

$$R_j^2 = 1 - \frac{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t+1} - \widehat{r}_{i,t+1})^2}{\sum_{\{i,t\}\in \text{leaf}_j} r_{i,t+1}^2}. \tag{6}$$

Since our goal is to separate returns with high predictability from those with low predictability, it is natural to use the absolute value of the R^2 difference between the two clusters as the split criterion,

$$S_{\{\text{leaf}_2,\text{leaf}_3\}}(\text{var}_p, c_k) = |R_2^2 - R_3^2|. \tag{7}$$

Intuitively, this criterion evaluates whether the candidate can differentiate the R^2 between two leaf nodes, no matter which one is higher. It focuses on detecting a subset of stock returns that are more predictable than others.

Different pairs of split candidates $\{\text{var}_p, c_k\}$ result in various partitions of the data, leading to non-overlapping sub-samples in leaf nodes $\mathcal{D}_2$ and $\mathcal{D}_3$ and different cluster-wise predictive models $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$. These ultimately lead to different values of the split criterion in Eq. (7). A successful split candidate will make the criterion as large as possible. Thus, for P candidate variables and K potential cutpoint values for each partition, we consider a total of $P \times K$ possible segment combinations for the first split. Furthermore, the split variable can be either firm characteristics or aggregated variables for detecting heterogeneity in either cross-section and time series.

After calculating the criterion for all candidates, we pick the optimal one as the first split, and the root node is divided into two child nodes (clusters). Additional separations may be required to capture the heterogeneity of predictability further.

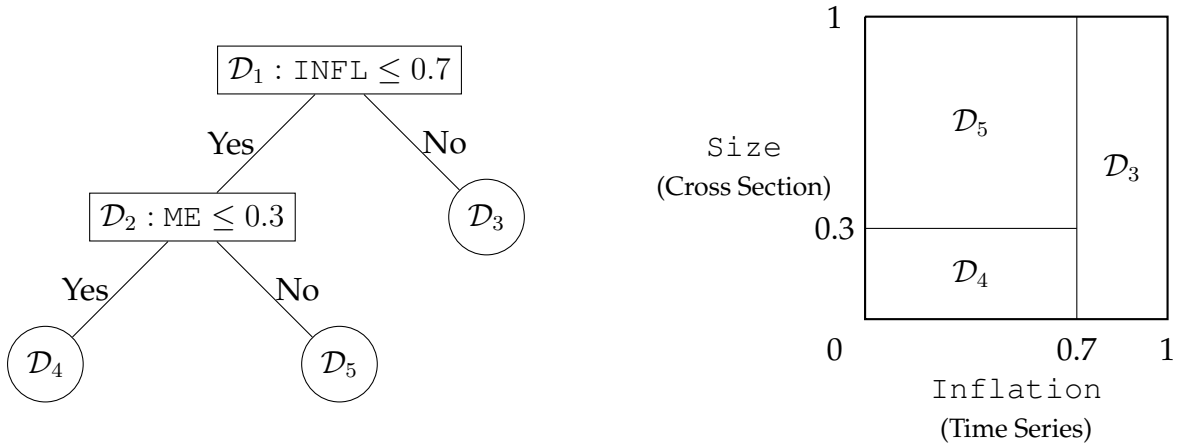
2.4 Clustering: Subsequent Splits and Stop

Our clustering method uses a sequential iterative partitioning approach to create non-overlapping clusters one at a time. Suppose the first split is determined as inflation at value 0.7; this forms two leaf nodes, and the subsequent division can occur at either of these nodes to further separate the data sub-samples. Figure 3 illustrates

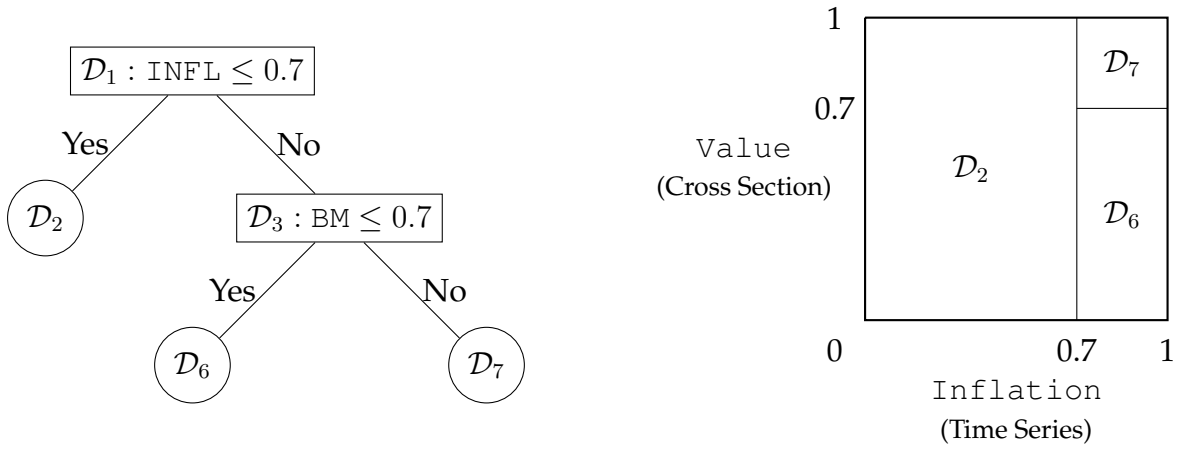
two potential splitting candidates for the second split, which can happen in the left (non-high inflation) or right (high inflation) leaf with different characteristics, demonstrating the asymmetric interaction of split predictors.

Figure 3: **Illustration for the second split**

This figure illustrates two example candidates for the second split, which can happen on the left or right child node, demonstrating the asymmetric interaction of split predictors.



(a) If splitting node $\mathcal{D}_2$ at ME



(b) If splitting node $\mathcal{D}_3$ at BM

The second split can occur within $\mathcal{D}_2$, partitioning it into $\mathcal{D}_4$ and $\mathcal{D}_5$, or within $\mathcal{D}_3$, partitioning it into $\mathcal{D}_6$ and $\mathcal{D}_7$. Each split still processes $P \times K$ candidates, totaling $2 \times P \times K$ candidates. When evaluating candidate splits within $\mathcal{D}_2$, we fit cluster-wise predictive models $\widehat{g}_4(\cdot)$ and $\widehat{g}_5(\cdot)$ according to the resulting leaf nodes $\mathcal{D}_4$ and $\mathcal{D}_5$, respectively, and calculate the split criterion. The procedure is the same when justifying candidate splits for leaf node $\mathcal{D}_3$. Among all the split candidates for $\mathcal{D}_2$

and $\mathcal{D}_3$, we choose the one that maximizes the split criterion as the second split. This procedure takes a global approach, evaluating the benefits of splitting either leaf node and selecting the one that can further differentiate return predictability.

All subsequent splits are determined in the same manner. Each time, we examine all existing leaf nodes, search for all split possible candidates, and choose the one with the maximum value as the best partitioning of that leaf node. Without prior knowledge of the "correct" clustering pattern, this self-supervised clustering approach partitions stock return observations into multiple clusters, maximizing the split criterion, predictability heterogeneity between clusters, and then fitting post-cluster heterogeneous predictive models.

Stopping Criteria. Criteria for ending tree growth help regularize in-sample model training, prevent over-fitting, and maintain the interpretability of decision tree plots. The clustering process stops when it meets certain predetermined conditions. We set a minimum sample size for each leaf cluster, while split candidates that cannot satisfy this minimum leaf size for resulting leaf nodes are eliminated. It is crucial to ensure that the heterogeneous predictive model within each cluster can be fit with enough data observations. Moreover, we also limit the maximum depth and the number of terminal leaves for the tree structure's growth. Finally, we stop splitting a node if all split candidates fail to improve the predictability – the R^2 values of both child nodes are smaller than that of the parent node.

Heterogeneous Predictions. As mentioned in previous sections, during the iterative steps splitting the panel of returns, we fit Ridge regression for return predictions and growing the trees. Shen and Xiu (2024) point out that Ridge regression enjoys robust predictions under weak signals compared to Lasso and is computationally efficient.

Once the tree growth stops, the entire panel of stock return observations is partitioned into multiple non-overlapping clusters based on firm characteristics and macro predictors. We refit cluster-wise heterogeneous predictive models using observations within each cluster. Notably, while we illustrate clustering results using the decision tree structure, this clustering framework is independent of the choice of the prediction

model. Therefore, our clustering approach provides a general framework for heterogeneous predictions with various machine learning models: regression, Lasso, PCA, random forest, etc.

3 Data and Model Design

3.1 Data

We apply our approach to the U.S. individual stock returns to learn their heterogeneous predictability.[5] The monthly sample spans from 1973 to 2022, where the first 30 years are used for model estimations and the recent 20 years for out-of-sample tests. The average and median number of stocks in the training sample are 5,592 and 5,626, respectively, while in the test sample, these numbers are 4,055 and 3,850.[6]

Characteristics. As listed in Table A.4, our dataset includes 58 firm-level characteristics categorized into eight major groups: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. We standardize each characteristic cross-sectionally and uniformly in the range of [0, 1] for every month. We use two cut-points $\{0.3, 0.7\}$ as split value candidates for each characteristic to mimic the "top-middle-bottom" sorting approach. We use all these characteristics for forecasts to construct tree-based clustering and cluster-wise heterogeneous predictive models.

Aggregate Predictors. We examine eight aggregate predictors to construct and choose macroeconomic regimes with time-varying return predictability. As detailed in Table A.3, these variables comprise the 3-month treasury bill rate, inflation, term spread, default yield, and market aggregate characteristics (dividend yield, volatility, net equity issues, and liquidity). We standardize these aggregate predictors into the $[0, 1]$ range based on their empirical percentile values in a rolling 10-year window.[7] This standardization enables us to compare the performance of each predictor on the same variation

[5] We apply the standard filters (see, e.g., Fama and French, 1992) to (1) only include stocks listed on NYSE, AMEX, or NASDAQ for more than one year, (2) use observations of firms with CRSP share codes of 10 and 11, and (3) exclude stocks with negative book equity or negative lagged market equity.

[6] Note that our algorithm allows the panel data to be unbalanced.

[7] For example, if the value of inflation exceeds 0.7, it indicates that the current inflation level is higher than 70% of observations in the past ten years.

scale without introducing look-ahead biases. Similar to the firm-level characteristics, we use two cut-points $\{0.3, 0.7\}$ as split value candidates for each aggregate predictor, employing a "top-middle-bottom" sorting approach.

Model Design. The baseline analysis for the cross-sectional split uses the first 30 years for model estimations and the recent 20 years for out-of-sample tests. We update the tree-based clustering and cluster-specific predictive models every five years using a 30-year window as in-sample data for retraining the decision tree structure. This process is repeated four times over a 20-year out-of-sample data span. We implement a full-sample analysis for the extended analysis with time series splits because of the long and overlapping business cycles. Finally, we employ cross-validation to select the optimal hyperparameters for the post-cluster predictive model training.

3.2 Performance Evaluation

We can evaluate two types of outcomes in tree-based clustering: (i) in-sample and out-of-sample R^2 for heterogeneous return predictability and (ii) forecast-implied investment strategies for heterogeneous predictive modeling. Our first goal is to examine any heterogeneous return predictability and determine if the results obtained from the in-sample data remain consistent when applied to out-of-sample data. The second outcome is to build the link between return predictability and economic gains, which is determined by the heterogeneous predictive modeling. Accordingly, we introduce two types of measurements.

In addition to the in-sample R^2 in Eq. (6), we follow Gu et al. (2020) and use R^2_{OOS} with using zero forecast as the benchmark:

$$R^2_{OOS,j} = 1 - \frac{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t} - \widehat{r}_{i,t})^2}{\sum_{\{i,t\}\in \text{leaf}_j} r^2_{i,t}}. \tag{8}$$

We are introducing the forecast-implied portfolio as an addition to long-short strategies, such as equal-weighted and value-weighted portfolios. This new approach considers all stocks, rather than just the tails of sorted portfolios, and thinks of the return forecast values instead of focusing solely on directions or deciles. Within each

leaf cluster j, we define the forecast-implied portfolio as:

$$R_{j,t} = \sum_{\{i,t\} \in \text{leaf}_j} \widehat{w}_{i,t} r_{i,t}, \tag{9}$$

where the following (10) are the weights for stock i at period t in the j-th portfolio.

$$\begin{aligned} \text{Equal-weighted: } \widehat{w}_{i,t} &= \begin{cases} 1/N^{\text{Pos}}_{t,\text{leaf}_j}, & \text{if } \widehat{r}_{i,t} \geq 0 \\ -1/N^{\text{Neg}}_{t,\text{leaf}_j}, & \text{if } \widehat{r}_{i,t} < 0 \end{cases} \\ \text{Forecast-weighted: } \widehat{w}_{i,t} &= \frac{\widehat{r}_{i,t}}{\sum_{k \in \text{leaf}_j} |\widehat{r}_{k,t}|} \end{aligned} \tag{10}$$

4 Heterogeneous Return Predictability

4.1 Cross-Sectional Heterogeneity

The baseline model investigates heterogeneous return predictability across the cross-section: what types of stock returns are predictable. Rather than restricting a stock to only one cluster for all periods (e.g., Evgeniou et al., 2023), our approach determines cluster assignments based on characteristics. It allows stocks to change cluster membership over time. This cross-sectional decision tree approach also generalizes the security sorting in empirical asset pricing (Cong et al., 2023).

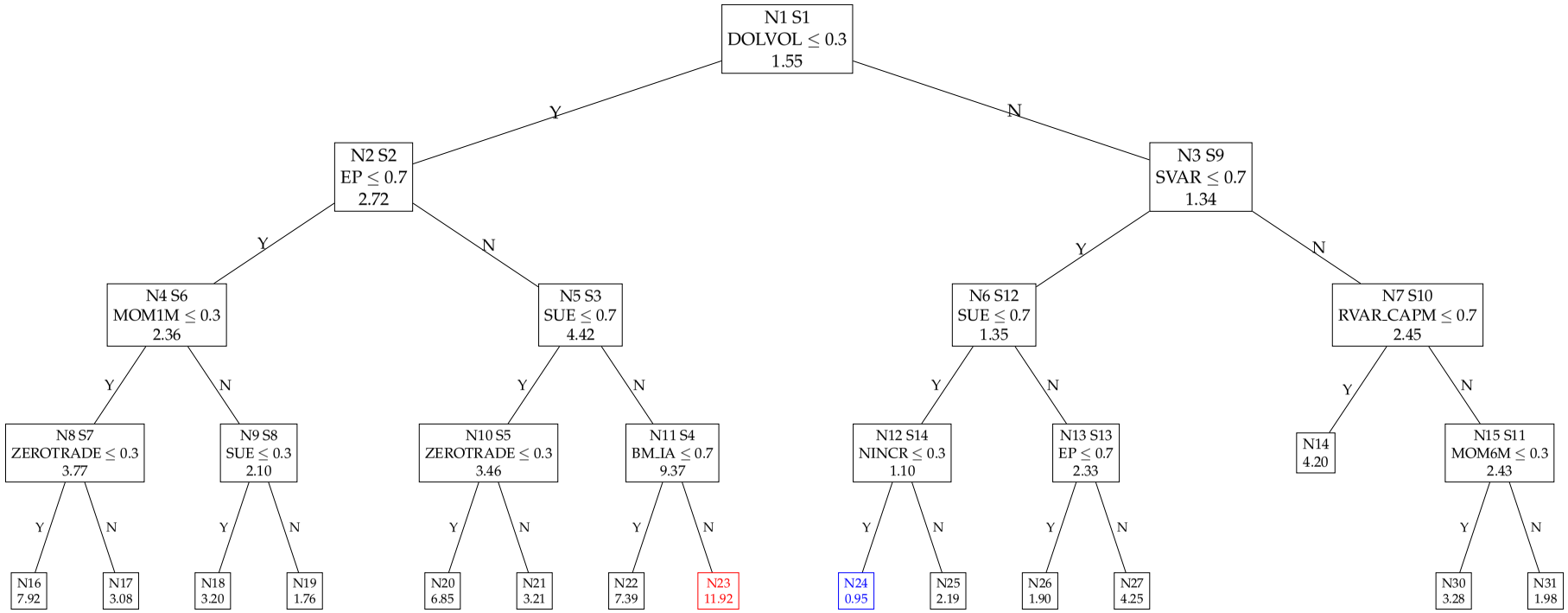
Tree-based Clustering. Figure 4 showcases the clustering results, derived from training the model on data spanning the first 30-year period (1973 - 2002), and is visually presented by a decision tree structure.[8] This tree stops growing with 15 terminal leaves, meeting our predetermined stopping criteria.[9] Specifically, each leaf node in Figure 4 displays two or three rows of information: (1) The first line signifies the

[8] For robustness check and out-of-sample evaluations, the rolling-window updated tree clusters for other periods (1978 - 2007, 1983 - 2012, and 1988 - 2017) can be found in Appendix II..1. One can reveal a similar selection set of characteristics, including dollar trading volumes, unexpected earnings, earnings-to-price, and cashflow-to-price.

[9] A moderately deep decision tree with large leaves (sample size) is robust and easy to understand. Accordingly, we set the maximum tree depth to 5 (at most 16 leaves) and specify the minimum leaf size of 10,000 stock return observations. Under these settings, the algorithm automatically stops after 14 splits.

Figure 4: **Tree-Based Cluster (Cross Section, 1973 - 2002)**

This figure shows the cross-sectional split tree-based clustering structure using monthly data from 1973 to 2002. The tree splits the panel of individual stock returns based on monthly cross-sectional standardization of firm characteristic ranks in [0,1]. The terminal leaves correspond to clusters identified by the interaction of firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. All nodes are labeled with cluster-wise model R^2.



intermediate node's leaf number and the split order. For example, the top root leaf represents the initial partition, and its first row of records corresponds to N1 and S1. (2) Our algorithm determines the optimal split rule, reported in the middle line. Observations satisfying the conditions are directed to the left child node, while those not meeting the criteria are directed to the right-hand side. The terminal leaves, which do not undergo additional splitting, are represented by their index number alone in the first line. Lastly, (3) the S2N ratio, R^2, of the cluster-wise return predictability is displayed for every leaf node in the complete tree structure.

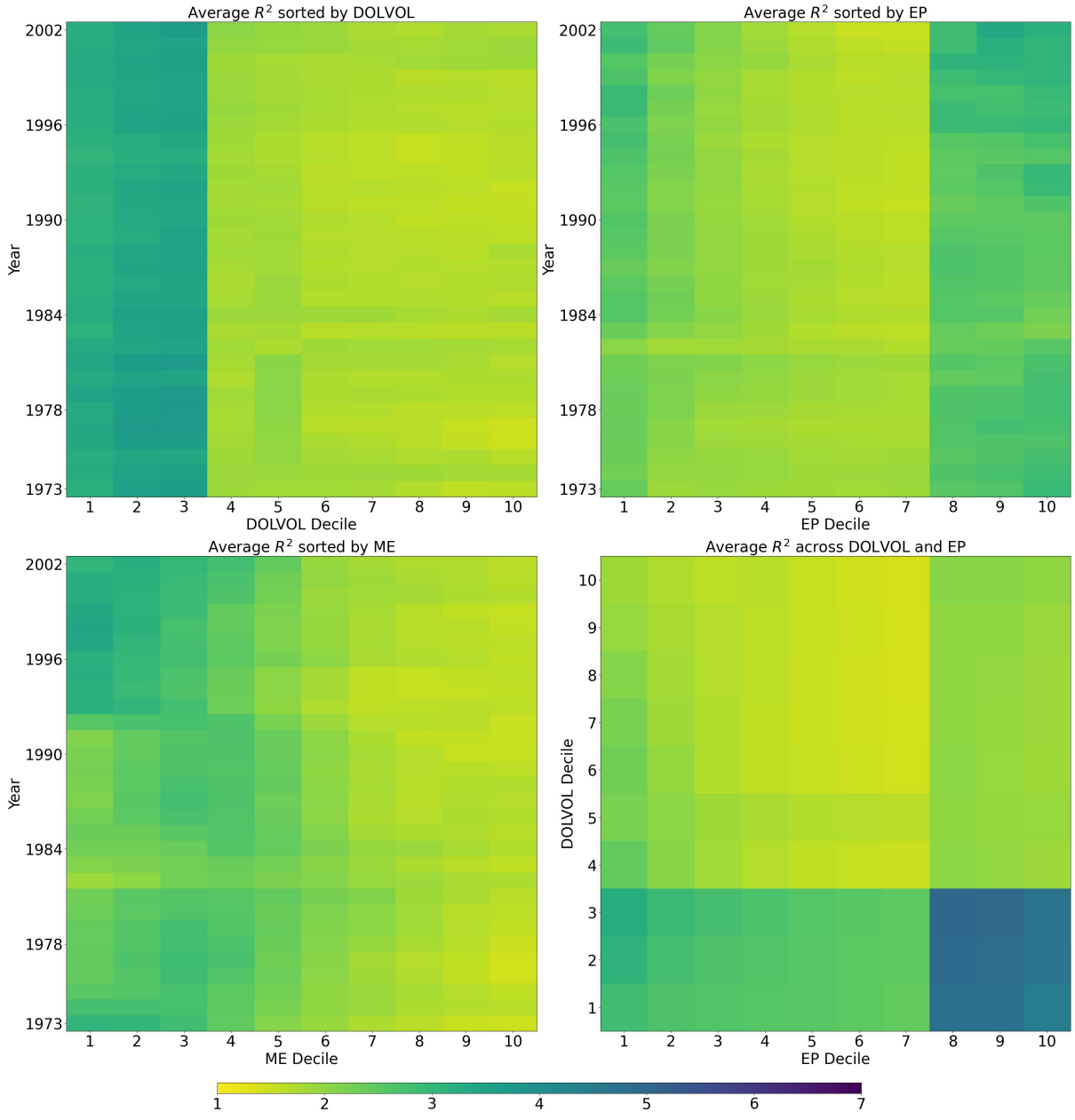
Before any separation, the aggregate return predictive ability (R^2) using a homogeneous model is 1.55%, which aligns with the overall predictability reported in the existing literature. After the first split, this metric improves significantly for stock returns with low dollar trading volume ($\texttt{DOLVOL} \leq 0.3$, reaching 2.72% at N2), while it declines slightly in the complement set (reducing to 1.34% at N3). The R^2 difference (2.72-1.34)% is the maximal value discovered through the complete search by our algorithm. The low-volume stocks may present low liquidity, being small-caps or distressed, related to the findings of Avramov et al. (2023). Figure 5 illustrates the average R^2 values for each year and each decile cluster sorted by different characteristics.[10] It is evident that there is a distinct pattern of decreasing predictability in returns for stocks with higher dollar trading volumes, and this trend has remained consistent over time. Additionally, Figure 5 provides only the marginal information derived from the decision tree depicted in Figure 4.

The second partition selects the high earnings-to-price stocks ($\texttt{EP} > 0.7$), leading to a R^2 value of 4.42% at N5, within the low dollar trading volumes cluster. Value stocks characterized by low volumes demonstrate a higher level of return predictability, while non-value stocks with small trading volumes witness a lower R^2 of 2.36%. The maximum difference in R^2 (4.42-2.36)% is determined by the comprehensive search for the left leaf N2 (stocks with low dollar trading volumes, $\texttt{DOLVOL} \leq 0.3$) and the right leaf N3 (stocks without low dollar trading volumes, $\texttt{DOLVOL} > 0.3$). We can

[10]For robustness check, the rolling-window updated characteristics-sorted return predictability heat maps on other periods (1978 - 2007, 1983 - 2012, and 1988 - 2017) are shown in Appendix II..1.

Figure 5: **Mosaics of Predictability by Predictors (Cross Section, 1973 - 2002)**

We present four heat maps to summarize the average return predictability, R^2 values (% in the color bar), for the panel of individual stock returns corresponding to the tree-based clustering results from Figure 4. The first three illustrate the average R^2 values for groups categorized by various years and deciles based on different characteristics (dollar trading volumes, earnings-to-price, and market equity value). The last one displays the average R^2 values for the 10×10 groups by bivariate-sorted deciles for the top two characteristics.



also observe that the EP-sorted decile cluster exhibits an ascending trend for higher R^2 in Figure 5. The sub-figure located in the bottom right, sorted by DOLVOL and EP, highlights the interaction pattern for stocks with low DOLVOL and high EP, exhibiting the highest return predictability on the bottom right region (the darkest shading).

Upon reviewing the rolling-window updated results in Appendix II..1, it is evident that stocks with low dollar trading volumes and high value (such as earnings-to-price or cashflow-to-price) demonstrate the highest level of return predictability. The decision tree further identifies the less predictable clusters with an R^2 value of 0.95% by $\{\texttt{DOLVOL} > 0.3\}$ and $\{\texttt{SVAR} \leq 0.7\}$: stocks without low trading volumes but non-high volatility.

Figure 6: **Mosaics of Predictability (Cross Section, 1973 - 2002)**

This heat map summarizes the predictability, R^2 values (% in the color bar), for the panel of individual stock returns by the tree-based clustering in Figure 4. The vertical axis represents months, and colors from light to dark indicate ascending levels of return predictability of each cluster within each month. Horizontally, the length of each color bin corresponds to the proportion of observations for each cluster.

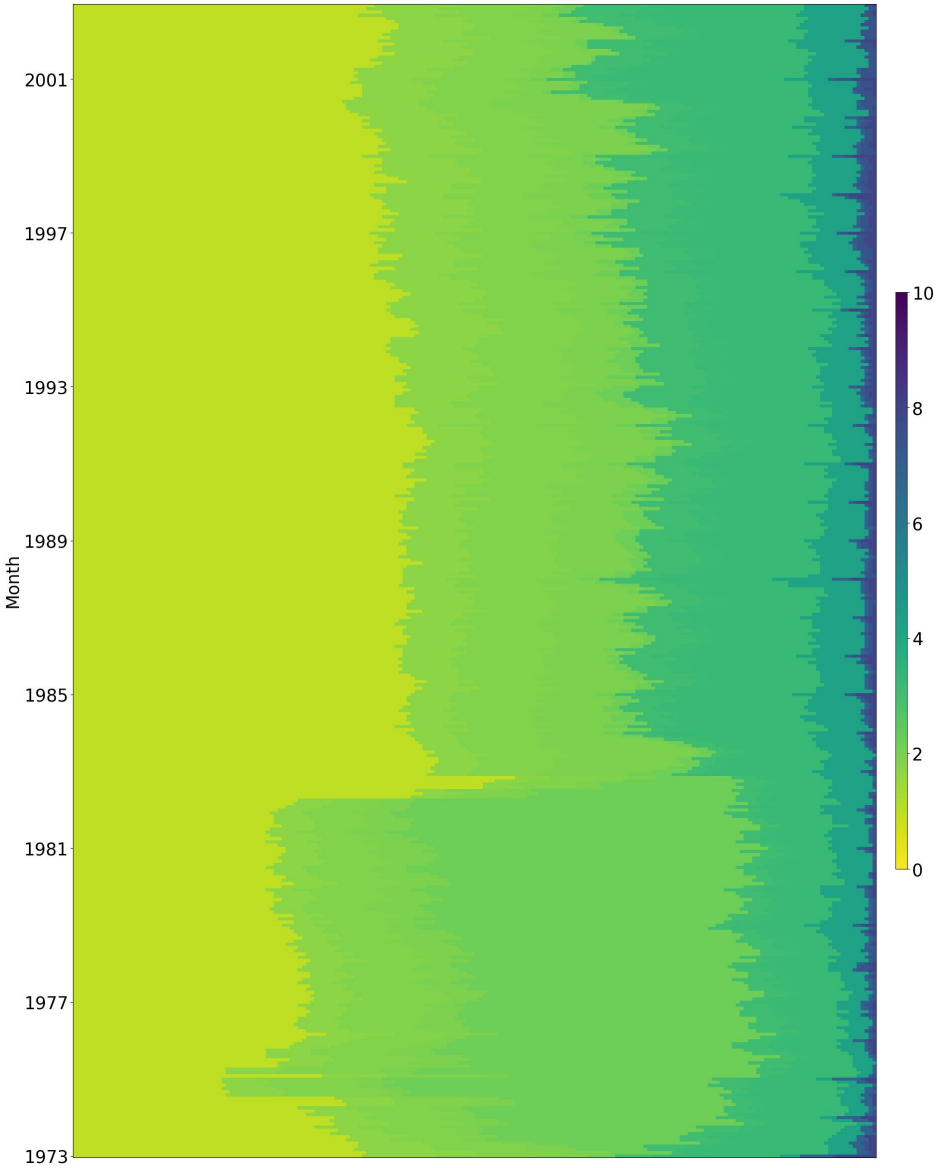


Figure 6 summarizes the return predictability from the tree-based clustering, resembling a mosaic painting. The sample period, encompassing all 360 months, is aligned with the tree structure in Figure 4. There are approximately 4,000 to 6,000 stock return observations in each month. We arrange clusters horizontally based on their respective R^2 values with an ascending order from light (left) to dark (right). The length of each color bin represents the proportion of observations for each cluster within each month. As the R^2 values increase, the proportions of observations tend to decrease, indicating that most stock returns have low S2N ratios. Conversely, only a tiny percentage of observations exhibit high predictability, reflected by R^2 values around 10%. This vertical mountain-like cascading appearance of the consistent mosaic structure can be characterized by its variations in cross-section, time series, and colors, offering insights into the heterogeneous predictability of stock returns. Figure 4 summarizes these facts by displaying an organized decision tree that partitions the cross-section of stock returns based on firm characteristics.

Return Difference Between Clusters. We now explore the cluster-wise predictive performance for four trees grown with different periods in Table 1. In addition to the selected splitting criterion R^2 from Eq. (6), we report the number of observations and the forecast-implied portfolio investment performance by Eq. (9). The table indicates that each tree structure has a specific count of terminal nodes (15, 15, 15, and 16 leaves for those four trees, respectively), allowing for further growth without triggering the termination criteria.

First, when interpreting this table horizontally in the descending order of in-sample R^2, it becomes evident that there is substantial heterogeneity among the clusters. The varying values of R^2 indicate that each cluster exhibits a different level of predictability, highlighting the diversity within the dataset. Across all the tree models generated by each sample, the clusters with the highest R^2 values consistently exceed 9.5% (the second tree, ranging from 1978 to 2007, leaf N29 contains the highest R^2 value of 12.08%), while the lowest values drop nearly below 1% (the fourth tree, ranging from 1988 to 2017, leaf N20 contains the lowest R^2 value of 0.78%). It is also ob-

vious that the number of observations almost remains increasing when predictability decreases, which reflects less predictable facts for most of those sub-samples. Besides, even though there may be a few exceptions among the leaves in the middle rankings, it is notable that the investment performances generally remain relatively stable, following a descending trend similar to the predictability.[11] As an example, in Figure 4, the leaf with the largest R^2 value (N23, labeled in red with $R^2 = 11.92\%$) possesses relatively higher monthly average return and annualized Sharpe ratio, especially for the value-weighted portfolio (VW, 3.68% and 2.14). Oppositely, the cluster with the lowest predictability (N24, labeled in blue with $R^2 = 0.95\%$) exhibits nearly the smallest values (0.68% and 0.93 for VW). Regardless of distances or orders, it is evident that observations belonging to clusters with higher R^2 values are more likely to be easily predicted.

Second, if we illustrate the results vertically, except for the different number of leaves mentioned above, the leaf node indexes and the interactive paths vary among those four tree structures. Even though the leaf nodes may be labeled with the same index numbers, most are determined by different interactions of firm characteristics and possess distinct summary statistics. Therefore, this table transforms the tree architectures by the predictability orders on horizontal and vertical dimensions, offering an alternative perspective for the mosaics of stock returns.

Out-of-Sample Evaluations. Motivated by the previous cluster-wise predictability rankings analysis, we have temporarily identified potential predictable clusters of specific stocks. Hence, examining the combinations of stock returns at different levels of predictive capacity offers an additional perspective for discerning the mosaics corresponding to our tree-based clustering algorithm. Gu et al. (2020) have documented overall return predictability by advanced machine learning techniques. In contrast, simpler models are sufficient for our purpose of evaluating clustering. We mainly focus on assessing heterogeneity within clusters and drawing meaningful insights rather

[11] Figure A.1 in the Appendix provides a clearer visualization of the positive relationships between cluster-wise predictability and forecast-implied portfolio investment performance. The points in different shapes reflect the corresponding positions of predictability (R^2 on the horizontal axis) with average returns (left vertical axis) and Sharpe ratios (right vertical axis), respectively.

Table 1: **Cluster-Wise Performance (Cross Section)**

This table presents the performance of the cross-sectional tree-based clustering model for four sub-samples. The in-sample return predictability, R^2 (in %), are calculated by Eq. (6). "# obs" represents the number of stock returns for each cluster. "Avg" and "SR" denote the monthly average return (in %) and annualized Sharpe ratio for both cluster-wise equal-weighted and value-weighted forecast-implied portfolios, respectively. Each panel of values is arranged in the descending order of R^2 from left to right.

1973-2002	N23	N16	N22	N20	N27	N14	N30	N21	N18	N17	N25	N31	N26	N19	N24
# obs	11,627	14,695	16,341	12,116	79,752	14,402	115,528	97,131	49,275	87,763	179,412	126,050	178,691	133,911	625,549
R^2	11.92	7.92	7.39	6.85	4.25	4.20	3.28	3.21	3.20	3.08	2.19	1.98	1.90	1.76	0.95
Avg_{EW}	4.43	4.18	3.23	2.50	1.97	2.49	3.24	2.11	2.46	3.20	1.26	2.15	1.51	2.02	0.99
SR_{EW}	2.36	1.43	1.91	1.44	1.41	1.21	1.76	2.71	1.50	2.13	1.14	1.38	1.36	2.17	1.40
Avg_{VW}	3.68	3.50	2.57	2.12	1.20	2.31	2.76	1.77	1.98	2.18	0.68	1.68	0.97	1.57	0.68
SR_{VW}	2.14	1.43	1.87	1.31	0.93	0.92	1.46	2.52	1.57	2.03	0.58	0.94	0.98	1.88	0.93

1978-2007	N29	N28	N30	N27	N20	N31	N9	N25	N23	N21	N17	N26	N22	N24	N16
# obs	12,544	18,622	21,639	12,434	29,460	74,262	12,206	45,833	22,340	176,835	122,191	21,053	170,450	234,971	827,758
R^2	12.08	7.52	7.40	5.78	3.77	3.35	3.24	2.77	2.69	2.62	2.42	2.34	1.70	1.45	1.03
Avg_{EW}	4.18	3.11	2.18	3.05	3.33	1.83	2.01	2.63	2.35	2.98	1.14	1.92	1.94	1.65	1.00
SR_{EW}	2.57	2.19	2.02	1.97	1.77	1.29	0.95	1.70	1.87	2.09	1.24	1.86	1.32	1.37	1.71
Avg_{VW}	3.36	2.58	1.41	1.56	2.81	1.09	1.91	1.29	1.69	2.48	0.75	0.51	1.67	0.93	0.61
SR_{VW}	2.38	2.18	1.36	1.07	1.28	0.87	0.78	0.85	0.84	1.53	0.68	0.46	0.87	0.84	0.92

1983-2012	N23	N16	N21	N22	N28	N29	N13	N20	N30	N19	N25	N31	N17	N18	N24
# obs	15,346	15,281	12,290	14,954	10,849	10,615	10,550	72,082	153,579	273,409	282,920	125,163	25,364	66,266	714,301
R^2	9.98	6.92	6.56	6.06	4.06	2.98	2.88	2.68	2.39	2.05	1.97	1.62	1.26	1.24	0.80
Avg_{EW}	4.33	2.92	2.38	3.14	2.35	1.85	2.23	2.58	2.84	2.11	1.50	2.15	1.60	1.51	0.92
SR_{EW}	2.47	1.05	2.04	2.27	0.97	0.79	1.24	1.67	1.35	3.42	1.27	1.42	1.28	1.59	1.19
Avg_{VW}	3.68	2.40	2.10	2.71	2.08	2.10	1.70	1.88	2.68	1.64	0.86	1.73	1.39	1.29	0.65
SR_{VW}	2.34	1.08	1.98	2.00	0.77	0.81	0.73	1.60	1.17	3.23	0.90	1.02	1.07	1.76	0.80

1988-2017	N29	N26	N16	N28	N30	N22	N27	N25	N19	N21	N31	N17	N23	N18	N24	N20
# obs	15,993	21,853	14,170	19,257	38,550	15,295	31,490	51,754	108,337	48,271	34,881	22,903	197,952	237,702	241,664	637,011
R^2	10.64	7.41	6.95	6.28	4.32	3.31	2.97	2.95	2.31	1.74	1.57	1.56	1.53	1.38	1.23	0.78
Avg_{EW}	4.13	1.99	2.98	2.69	2.93	1.67	1.58	2.73	1.62	1.84	1.27	1.54	2.01	1.87	1.48	0.92
SR_{EW}	2.79	2.07	1.09	2.18	1.81	0.76	1.23	1.49	2.83	1.11	0.88	1.04	1.09	2.60	1.19	1.09
Avg_{VW}	2.96	1.52	2.54	1.94	2.17	1.88	1.34	1.58	1.36	1.18	0.89	1.46	1.76	1.42	0.96	0.68
SR_{VW}	1.99	1.34	1.06	1.46	1.50	0.73	1.13	0.99	2.48	0.69	0.77	0.95	0.88	2.41	0.96	0.79

than model comparisons. In line with the five-year rolling update clustering setup, we adopt the same horizon with a 2-fold cross-validation tuning strategy for hyper-parameter optimizations to achieve the most accurate predictions of stock returns.[12]

According to the tree structures and the initial performance in Table 1, we categorize all the leaves into two (highly predictable leaves against all other clusters) or three (high, medium, and low) groups corresponding to the rankings of return predictability.[13] Next, we aggregate the observations across multiple clusters within each part as a new sub-sample. For example, in Figure 4, we identify N23, N16, N22, N20, and N27 as the highly predictable clusters, while N24 represents the less predictable sub-

[12] We divide the in-sample data into two equally continuous periods, training the model on one period while validating it on the other. We identify the best parameter by comparing the average MSE and retraining the model using all the in-sample data. Finally, we input the coefficients to predict the next five years of out-of-sample values.

[13] In all subsequent tables and figures, the sub-sample labeled as "Others" refers to the combination of "Medium" and "Low," i.e., "Others" = "Medium" + "Low."

sample. All nine remaining leaves are aggregated to form the sample with medium predictability.[14] Since we lack prior knowledge regarding which stock returns with specific characteristics are predictable in advance, we could not explore the evaluations, such as classification performance. Therefore, we assess the R^2 in Eq. (8) to further justify the mosaics of stock return predictability on the cross-section.

Table 2: **Out-of-Sample Evaluations (Cross Section)**

This table reports the return predictability, R^2 (in %), based on different predictive methods. We present in-sample (1973 - 2002) and out-of-sample (2003 - 2022, updated every five years) results. We provide six samples: Global (no clustering), Overall (aggregation clustering results), High, Others (Medium + Low), Medium, and Low, determined by the predictive rankings within the tree clusters.

	Panel A: In-Sample (1973 - 2002)			Panel B: Out-of-Sample (2003 - 2022)		
Sample	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Global	1.56	0.57	1.00	0.49	0.35	0.47
Overall	2.48	1.71	1.60	0.28	0.55	0.62
High	6.92	5.96	5.66	1.71	2.07	2.26
Others	2.16	1.40	1.31	0.20	0.46	0.53
Medium	2.56	1.77	1.60	0.22	0.53	0.60
Low	0.95	0.29	0.43	0.13	0.26	0.34

Table 2 provides the R^2 statistics based on three model predictions. Before performing any clustering procedures, our data presents similar predictable patterns corresponding to previous research using the entire sample (the first row, "Global"), particularly noticeable signal similarities over the recent 20 years. On the contrary, the "Overall" sample represents the consolidation of cluster-wise predictions, and almost all the values slightly improve, which reflects the helpfulness of segmenting forecasts. Subsequently, regardless of the techniques employed, the highly predictable portion outperforms any other samples significantly (out-of-sample R^2 values are all at least over 1.7%). If we remove the observations from the relatively predictable leaves from the entire dataset, the predictability will be reduced (the drop from "Overall" to "Others"). The clusters with the lowest predictability consistently exhibit the worst per-

[14] We determine the level of predictability of each leaf by balancing the R^2 values and the proportion of observations. We prioritize identifying highly predictable observations by looking for relatively significant gaps in the R^2 values, while most stock returns cannot be predicted well.

formance in the last line. Moreover, even though the magnitudes of improvements or declines may vary across different methods, it does not affect that they exhibit consistent trends. These phenomena can be further confirmed within the in-sample analysis. All the results allow us to detect the cross-sectional mosaics of return predictability among stocks by different firm characteristics.

4.2 Time Series Regime Change

From our initial analysis of Table 1, we have observed evidence suggesting that return predictability can be time-varying (e.g., Henkel et al., 2011; Dangl and Halling, 2012). Building upon this finding, we extend our analysis by introducing additional partitions along the time dimensions. To ensure interoperability, we first employ macroeconomic variables to divide in time horizons. They can separate the entire period into multiple *discontinuous* regimes, encouraging us to assess the predictability of stock returns across different regimes without the need for rolling window analysis. By restricting the utilization of macroeconomic variables for the first two splits[15] and continuing to employ firm characteristics for cross-sectional partitions, we create a comprehensive approach that facilitates the implementation of a timing and stock selection framework. This approach allows us to concurrently identify the types of stocks that exhibit improved predictability during specific time horizons.

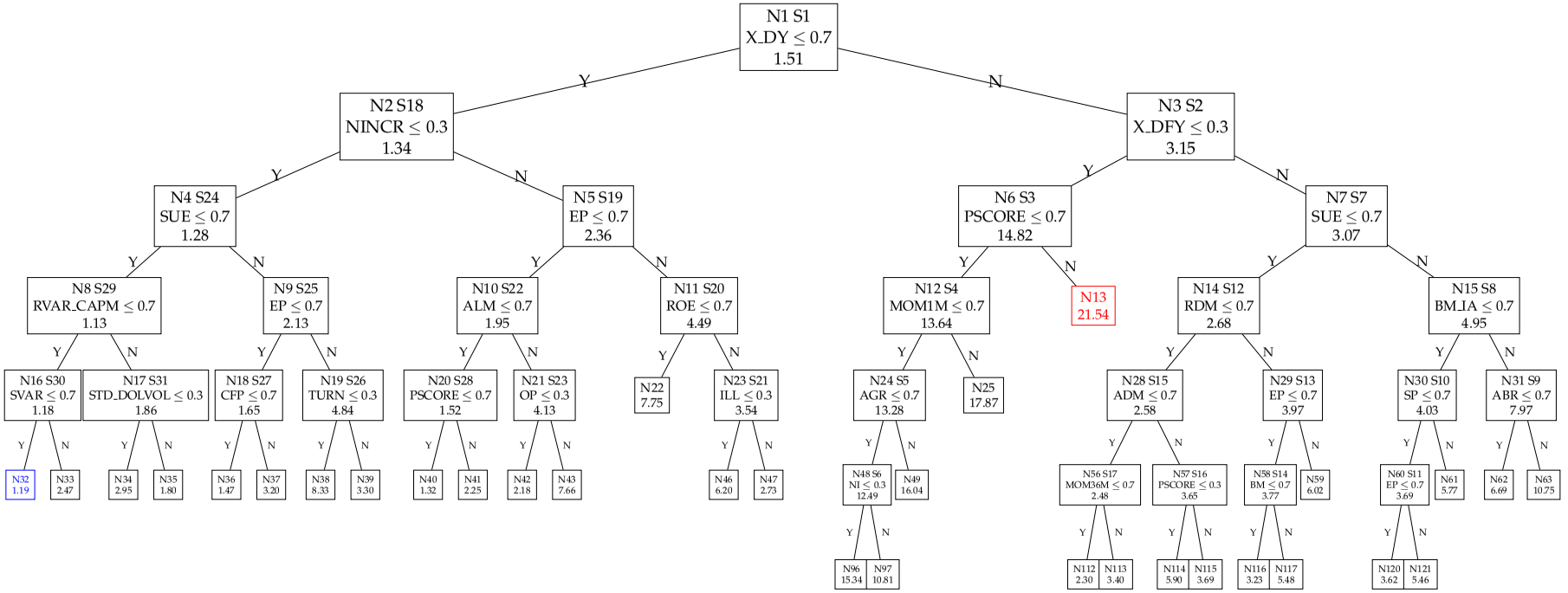
Clustering Presentation. Because we utilize macroeconomic variables to detect time-varying predictability instead of the previous rolling window strategy, there will be only one tree-based clustering structure for the time series split by the entire 50-year sample. Figure 7 consolidates each cluster's quantitative measurements and visualizes the decision tree structure. This tree eventually produces 32 leaves[16] within different predictive capacities. Similar to the cross-section, each leaf node in the tree provides

[15]We have attempted to incorporate all macroeconomic variables and firm characteristics into our tree-based clustering algorithm. Regardless of the setups for parameters of trees, they all choose macroeconomic variables for the first two layers' partitions, indicating a preference for initially separating based on time series. To ensure that the frequency of regime changes is not excessive, we limit the first two divisions to occur along the time dimension and retain others for subsequent cross-sectional partitions.

[16]We limit the maximum tree depth under each regime to 5 (48 leaves at most) and the minimum leaf size to 10,000 stock return observations.

Figure 7: **Tree-Based Cluster (+ Time Series)**

This figure shows the time series + cross-sectional tree-based clustering structure using monthly data from 1973 to 2022. The tree splits the panel of individual stock returns based on aggregate predictors by a 10-year rolling percentile standardized to [0,1] for the first two partitions. Then, it further splits based on monthly cross-sectional standardization of firm characteristic ranks within [0,1]. The terminal leaves represent distinct clusters identified by the interaction of aggregate predictors and firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has a unique ID denoted as $N\#$, while the order of the splits is indicated by $S\#$. All nodes are labeled with cluster-wise model R^2 values.



specific information: (1) The first row indicates the node index (including or excluding the division number); (2) The final line displays the cluster-wise model R^2 from Eq. (6); (3) The middle part of the dividing leaf represents the variable used for partitions and the selected threshold.

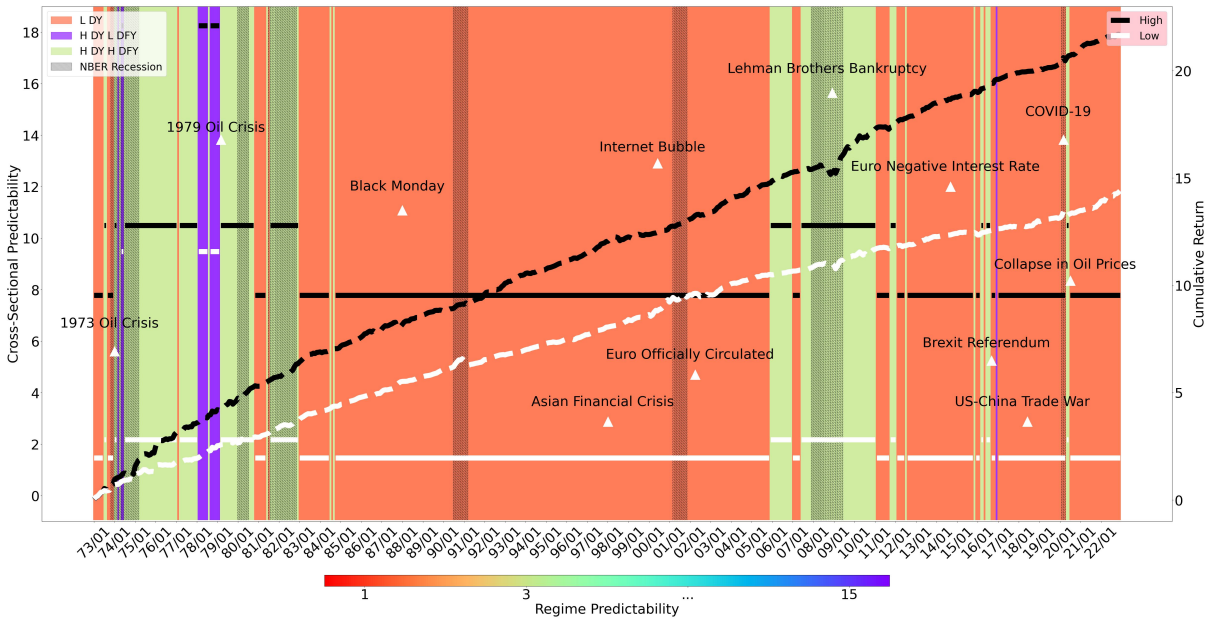
Specifically, the overall in-sample R^2 of the entire sample is 1.51% before any segmentation. After the first two splits, this number improves significantly in the higher dividend yield and lower default yield period ($\texttt{X_DY} > 0.7$ and $\texttt{X_DFY} \leq 0.3$, reaching 14.82% at N6). Meanwhile, it slightly decreases in the regime with non-high dividend yield, reducing to 1.34% at N2. Furthermore, further partitions based on cross-sectional dimensions can amplify the heterogeneity of stock return predictability. For example, the leaf node N13 (labeled in red) has the highest R^2 (21.54%), while the worst cluster N32 (labeled in blue) shows the lowest value (1.19%). This span is more extensive than only cross-sectional clustering results in section 4.1, which reflects that keeping partitions on the cross-section after the time series can enlarge the distance to achieve a more significant mosaic effect.

Figure 8 below presents the same information differently. The color switches reflect regime changes based on various combinations of macroeconomic variables, with the color bar conveying messages of heterogeneity. The span is much larger than that of the cross-sectional case (around 1% in orange and almost 15% in purple). Under each regime, the further cross-sectional divisions by firm characteristics widen the gaps across different sub-samples. Highly predictable clusters consistently display higher predictability and profitability than those with the lowest predictive ability. They always sandwich the time series heterogeneity (before further cross-sectional clustering) between them. Besides, regime switches may be triggered by several global events, including the Oil Crisis (1973, 1979), the Brexit Referendum (2016), and the Collapse in Oil Prices (2020). Other events are more likely to occur during periods of relatively lower predictability. For example, the regime associated with a dividend yield opposite to the upper 30% level (shown in orange) encompasses events such as Black Monday (1987), the Asian Financial Crisis (1997 - 1998), the Internet Bubble (2000),

the Euro Negative Interest Rate (2014), and so on. These figures and statistics confirm that our algorithm can consistently differentiate between predictable observations and poorly predicted stock returns across different dimensions, including cross-sectional and time series. It is essential to consider these factors together and their potential impact on the global economic environment.

Figure 8: **Time Series Regimes (Macro Variables)**

This figure shows macroeconomic variables' detected time series regime switches (see Figure 7) from 1973 to 2022. Three different colors represent different regimes: non-high dividend yield (X_DY ≤ 0.7 in orange, 426 months), high dividend yield with low default yield (X_DY > 0.7 & X_DFY ≤ 0.3 in purple, 16 months), and high dividend yield with non-low default yield (X_DY > 0.7 & X_DFY > 0.3 in green, 158 months), reflecting heterogeneous predictability based on their relative positions in the bottom color bar. The shaded areas represent NBER recession periods, while the labeled texts show global events that affect the world economy. Two vertical axes represent the highly (black) and less (white) predictable clusters' heterogeneous predictability (solid line) and profitability (dashed line) after further cross-sectional partitions.



Compared to the cross-section, the mosaics of stock return predictability on time series in Figure 9 appear messier. The vertical axis still represents the months across 50 years, while not all clusters are visible along the horizontal dimension due to time series partitions. By sorting the cluster-wise model R^2 values for predictability in ascending order and then aggregating monthly again, we can observe that the mosaics are layered. There are totally three hierarchical types and most of the clusters show low predictability in light green. Conversely, specific periods, around 1979 and 2017,

Figure 9: **Mosaics of Predictability (+ Time Series)**

This heat map summarizes the predictability, R^2 values (% in the color bar), for the panel of individual stock returns based on the tree-based clustering in Figure 7. The vertical axis represents months, and colors from light to dark indicate ascending levels of return predictability of each cluster within each month. The colors shown to differ over time correspond to the macroeconomic regimes detected. Horizontally, the length of each color bin denotes the proportion of observations for each cluster.

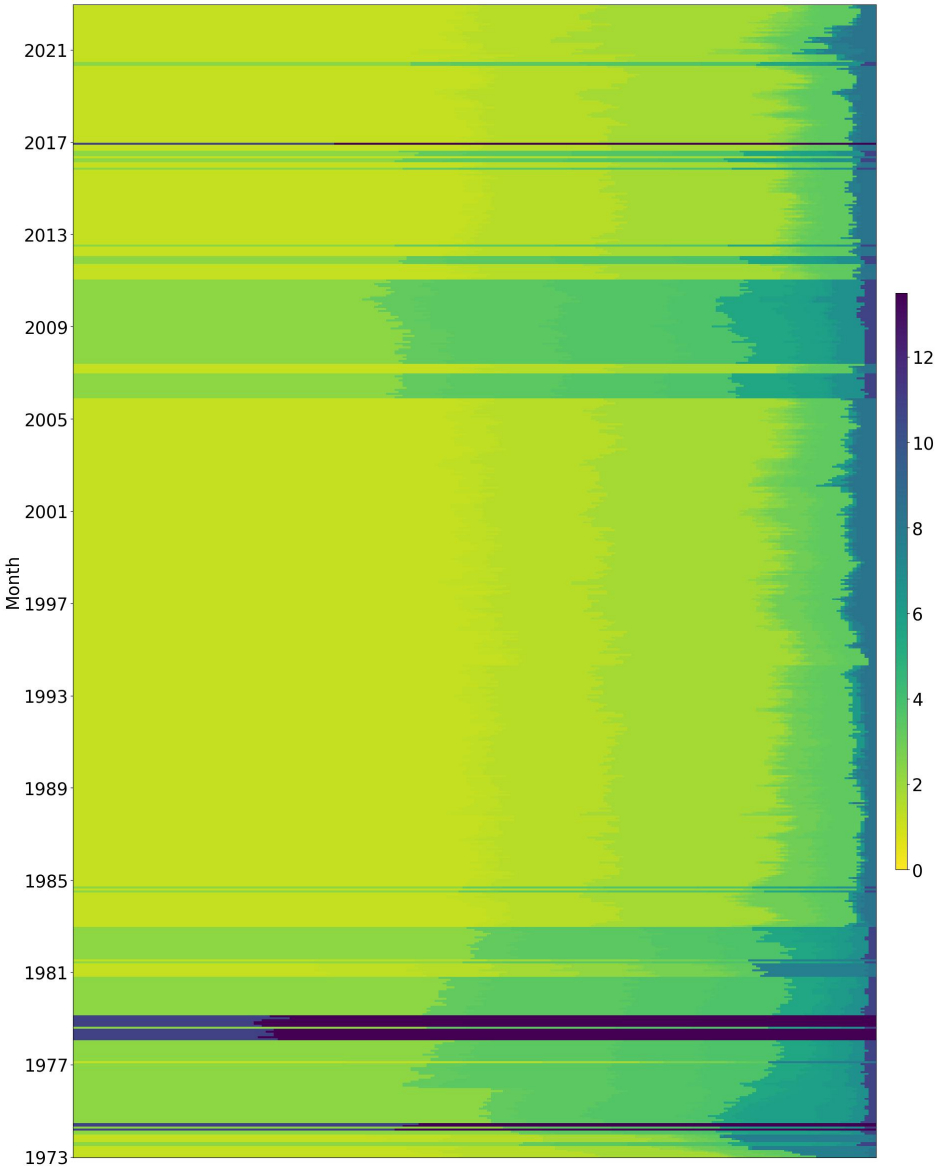
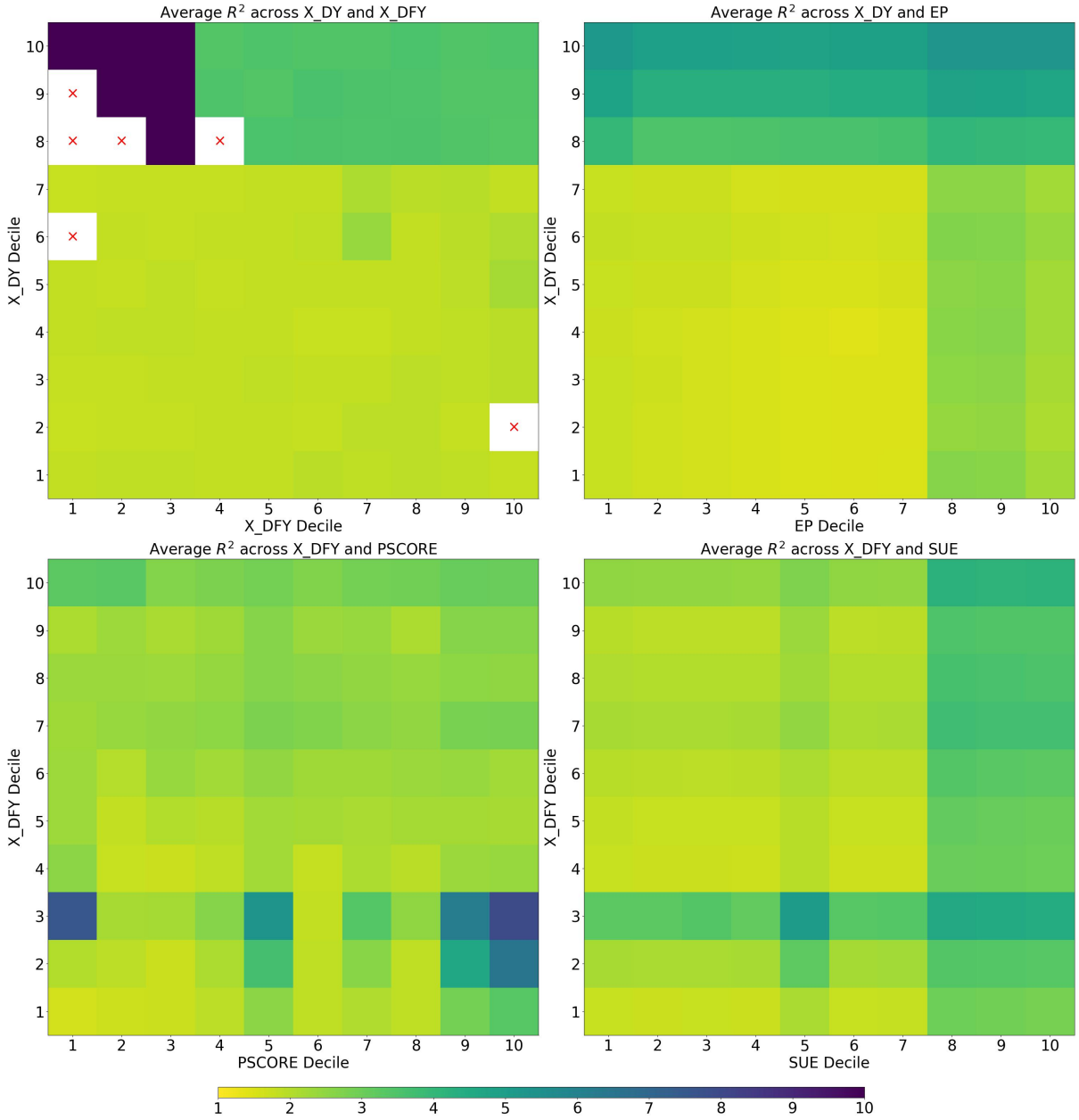


exhibit darker colors, indicating higher predictive accuracy. The colors frequently shift among these clusters, highlighting variations in different regimes. Furthermore, we incorporate macroeconomic predictors and generate four heat maps by bivariate-sorted deciles (see Figure 10). It's easy to identify the highly predictable periods from the top-left sub-figure and cross-section with high earnings-to-price, performance score, and

earnings surprise. Before determining their precise interpretations, we could observe the mosaic patterns when adding in time series considerations.

Figure 10: **Mosaics of Predictability by Predictors (+ Time Series)**

We present four heat maps to summarize the average return predictability, R^2 values (% in the color bar), for the panel of individual stock returns corresponding to the tree-based clustering results from Figure 7. Each sub-figure displays the average R^2 values for the 10×10 groups by bivariate-sorted deciles for different pairs of predictors. The empty grids with "$\times$" represent no observations falling in those specific intervals.



Difference Among Clusters. Similar to the cross-section, we also detect single-leaf statistics of the predictive abilities. By including information from time series parti-

tions, Table 3 also converts the tree structure shown in Figure 7 into the predictability rankings for each regime to identify heterogeneity.

Table 3: **Cluster-Wise Performance (+ Time Series)**

This table shows the performance of the tree-based cluster model based on time series and cross-sectional splits in Figure 7. The in-sample return predictability, R^2 (in %), are calculated by Eq. (6). "# obs" represents the stock returns count for each cluster. "Avg" and "SR" denote the monthly average return (in %) and annualized Sharpe ratio for both cluster-wise equal-weighted and value-weighted forecast-implied portfolios, respectively. Each regime of values is arranged in the descending order of R^2 from left to right.

Regime I	N38	N22	N43	N46	N39	N37	N34	N47	N33	N41	N42	N35	N36	N40	N32
# obs	44,913	17,380	13,344	15,584	77,804	33,025	47,992	19,904	14,486	21,982	11,929	397,816	273,572	64,672	938,700
R^2	8.33	7.75	7.66	6.20	3.30	3.20	2.95	2.73	2.47	2.25	2.18	1.80	1.47	1.32	1.19
Avg_{EW}	2.69	1.60	1.74	1.84	1.92	2.28	2.48	1.88	2.01	1.71	1.89	2.27	1.73	1.49	1.06
SR_{EW}	2.57	1.33	1.05	1.65	1.40	1.88	1.05	1.27	0.93	1.30	1.00	1.47	1.45	1.29	1.29
Avg_{VW}	1.47	1.04	1.20	1.44	1.16	0.80	2.05	1.46	1.89	1.20	2.19	1.87	0.95	0.82	0.83
SR_{VW}	1.30	0.67	0.58	1.15	0.94	0.69	0.76	1.06	0.74	0.84	0.87	1.06	0.90	0.65	1.04

Regime II	N13	N25	N49	N96	N97
# obs	11,785	13,568	10,119	10,659	13,481
R^2	21.54	17.87	16.04	15.34	10.81
Avg_{EW}	4.46	3.93	3.99	4.01	3.65
SR_{EW}	2.71	2.45	2.21	2.65	2.36
Avg_{VW}	3.23	2.66	2.92	2.71	2.41
SR_{VW}	2.77	2.45	2.16	2.14	2.44

Regime III	N63	N62	N59	N114	N61	N117	N121	N115	N120	N113	N116	N112
# obs	10,928	18,667	12,040	12,910	17,589	14,648	19,622	48,112	74,177	95,937	31,856	271,598
R^2	10.75	6.69	6.02	5.90	5.77	5.48	5.46	3.69	3.62	3.40	3.23	2.30
Avg_{EW}	4.58	3.74	2.77	2.81	3.27	3.76	2.48	2.70	2.43	1.89	2.79	1.98
SR_{EW}	2.05	1.72	1.50	1.34	1.71	1.97	1.62	1.58	1.77	1.58	1.60	1.61
Avg_{VW}	2.26	2.05	2.48	2.08	2.11	3.12	1.76	1.76	1.91	1.54	2.21	1.15
SR_{VW}	1.34	1.18	1.30	0.92	1.59	1.56	1.36	1.32	1.70	1.30	1.27	1.03

In this table, the gaps between two-side extremes are more comprehensive than the cross-sectional situation in Table 1, where the highest value is over 20% (N13 with 21.54%) and the lowest one is around 1% (N32 with 1.19%). Vertically, each regime has its wideness of predictability, which differs enormously. Besides, arranging the R^2 statistics in descending order within each group makes the mosaics of stock return predictability across the clusters more obvious. While the forecast-implied portfolio average returns do not consistently decrease across different categories as in the previous analysis, there is still a slight positive relationship between average returns and R^2 values.

Aggregate Evaluations. We base our time series segmentation on the entire 50-year sample and no longer separate in-sample and out-of-sample statistics. The first two

splits by aggregate predictors separate the whole period into three regimes, allowing us to detect predictability within each one. By balancing the proportion of observations with R^2 values, we aggregate the clusters into several sub-samples and report the predictability in Table 4.

Table 4: **Evaluating Return Predictability (+ Time Series)**

This table reports the return predictability, R^2s (in %), based on different predictive methods under various regimes. We present the full-sample results based on the tree-based cluster model incorporating both time series and cross-sectional splits. Under each regime, we present five samples: Global (no cross-sectional clustering, with Ridge results in brackets), Overall (aggregation clustering results), High, Medium, and Low, determined by the predictive rankings within the tree clusters.

	Sample A: All Stocks			Sample B: Large-Cap		
	Regime I (1.34): $\mathbb{1}\{\texttt{X_DY} \leq 0.7\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	1.84	1.11	1.79	1.55	1.08	1.53
High	7.79	6.63	7.78	5.20	5.00	5.41
Medium	2.86	2.07	2.84	2.38	1.81	2.37
Low	1.52	0.81	1.47	1.26	0.80	1.24
	Regime II (14.82): $\mathbb{1}\{\texttt{X_DY} > 0.7\}\mathbb{1}\{\texttt{X_DFY} \leq 0.3\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	16.52	12.20	14.23	21.26	16.43	18.48
High	21.97	15.56	18.25	25.73	19.24	21.55
Medium	16.89	12.98	14.73	19.26	15.61	17.11
Low	10.97	7.10	9.48	19.39	14.10	17.19
	Regime III (3.07): $\mathbb{1}\{\texttt{X_DY} > 0.7\}\mathbb{1}\{\texttt{X_DFY} > 0.3\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	3.63	2.78	3.47	3.60	2.81	3.41
High	10.76	10.12	10.49	6.94	6.43	6.72
Medium	4.31	3.18	4.12	4.21	3.18	4.01
Low	2.31	1.83	2.17	2.43	2.05	2.26

When comparing performance without cross-sectional clustering, most aggregation clustering results can improve the accuracy of predictions. Filtering out large-cap stocks also reveals specific predictability trends in different market conditions. Across these three regimes, each sub-sample in the second period ($\texttt{X_DY} > 0.7$ and $\texttt{X_DFY} \leq 0.3$) performs extremely better than the same group of others. Regardless of the techniques and data portions used, declining trends remain stable in the last three rows within each period. The highly predictable sub-sample consistently achieves the largest R^2 values, while the lowest performs worse than all other samples, includ-

ing the value without further cross-sectional clustering ("Global"). Despite differing magnitudes, the improvement and decline patterns across different methods exhibit similarities. These phenomena help us further understand the mosaics of full-sample stock return predictability or heterogeneity among stocks in different characteristics and market conditions.

4.3 Calendar Months Structural Break

In addition to using macroeconomic information for regime changes, we incorporate contemporary methodologies for time series segmentation. This approach involves directly examining structural breaks based on calendar months (e.g., Smith and Timmermann, 2021). It can be conducted by continuous months under each regime and without imposing restrictions on variables, which still aims to assess whether dividing continuous periods can capture the heterogeneity of stock return predictability.

Figure 11: **Time Series Regimes (Calendar Months)**

This figure shows the detected time series regime switches by calendar months from 1973 to 2022. There are a total of eight continuous colors that represent different regimes, reflecting heterogeneous predictability based on their relative positions in the bottom color bar. The shaded areas represent NBER recession periods, while the labeled texts show global events that affect the world economy. Two vertical axes represent the highly (black) and less (white) predictable clusters' heterogeneous predictability (solid line) and profitability (dashed line) after further cross-sectional partitions.

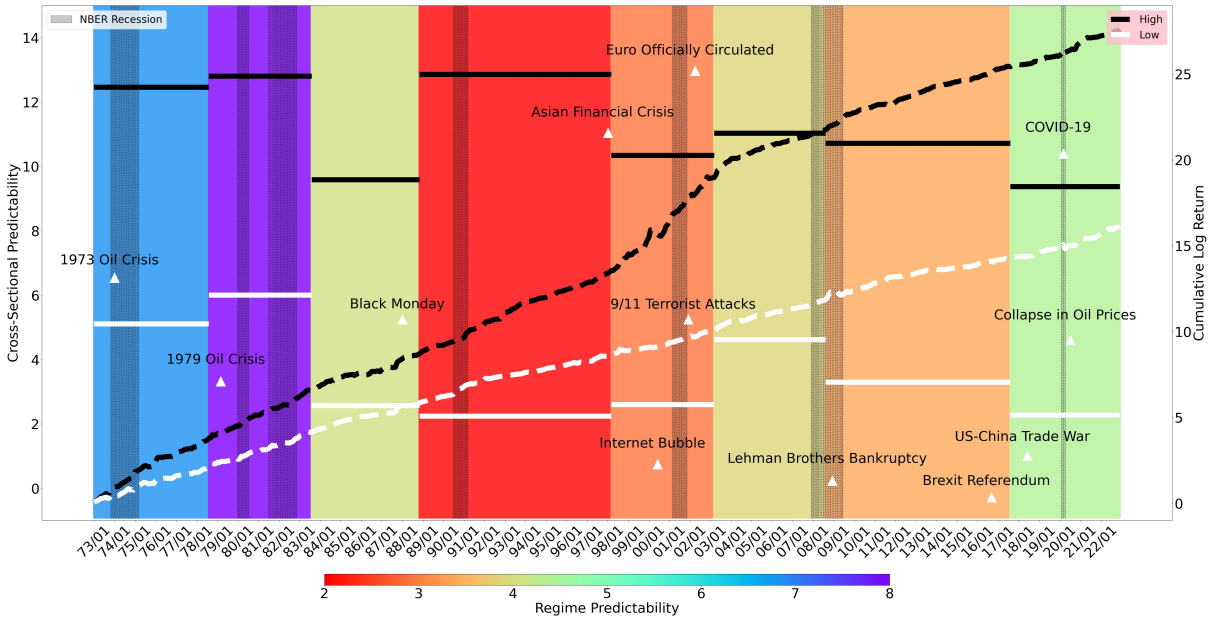


Figure 11 shows that the predictability of stock returns varies across different time

horizons, even though the span is not as much as using aggregate predictors. For example, the predictability is relatively highest from September 1978 to August 1983 (7.99%, labeled in purple), whereas it is only 2.01% from December 1988 to March 1998 (labeled in red). Some of the occurrences, such as the Black Monday (1987), the Asian Financial Crisis (1997 - 1998), the Euro Officially Circulated (2002), and the Brexit Referendum (2016), may have triggered regime changes to other predictability levels. Conversely, structural breaks would also contribute to other incidents (the Oil Crisis in 1979, the Lehman Brothers Bankruptcy in 2008, the US-China Trade War in 2018, and so on) Additionally, the heterogeneity gap will be more pronounced if we continue to segment in the cross-section under each time series period. Different regimes' predictive abilities and investment performance are still in the middle compared with relatively high (black) and low (white) predictable sub-samples. The trends remain similar across various periods to some extent. These findings again confirm our assumptions regarding the heterogeneous predictability of stock returns, highlighting their mosaic nature.

4.4 Heterogeneous Predictability Interpretations

In summarizing the analysis and reviewing the tree structures across both cross-sectional and time series dimensions, we have identified significant disparities in predictability results across different sample intervals.

Upon examining the time series dimension, we have observed that the interactions between characteristics and macroeconomic variables or calendar months vary significantly across three different regimes. Unsurprisingly, the most and least predictable clusters emerge from regimes characterized by high and low predictability (Regime II and I as defined in Table 3 and 4), respectively. Stock returns with higher predictability tend to occur in periods of higher dividend yield ($\texttt{X_DY} > 0.7$). Delving deeper into this subset, the regime with lower default yield ($\texttt{X_DFY} \leq 0.3$) demonstrates greater predictability than the opposite. These two periods outperform those when the dividend yield is non-high (Regime I). Each regime overlaps with the division into calendar months for several periods.

Furthermore, incorporating the cross-sectional considerations, highly predictable market conditions prioritize stocks with higher performance scores (PSCORE $>$ 0.7) or greater momentum (MOM1M $>$ 0.7). Conversely, observations lacking high earnings-to-price ratios (EP $\leq$ 0.7) and earnings surprises (SUE $\leq$ 0.7) are typically categorized into medium or low predictability sub-samples.

In addition, by exclusively focusing on the cross-sectional divisions and disregarding the time series factors, the highly predictable leaf consistently converges on the similar splitting candidate: stocks with lower dollar trading volumes (DOLVOL ≤ 0.3) and larger unexpected earnings (SUE > 0.7). For the other components in those four clustering structures, the first two trees select observations with high earnings-to-price ratios (EP > 0.7), while the latter two prefer stocks with high cash flows (CFP > 0.7), all of which represent high-value stocks. Oppositely, several characteristics commonly appear in less predictable clusters across different sample periods, such as non-high unexpected earnings (SUE ≤ 0.7), lower volatility (SVAR ≤ 0.7), or being in the top 30% of dollar trading volumes (DOLVOL > 0.3). Some of these findings are consistent with the results obtained from time-based splitting.

All these results confirm that stock return predictability can be viewed as a mosaic. The interactive paths of firm characteristics and/or aggregate predictors are constructive in interpreting clustering outputs.

5 Economic Gains on Cluster-wise Models

The previous discussions have mainly concentrated on predicting stock returns. This section will broaden our perspective on their relationship: whether highly predictable stock returns can lead to higher economic gains. It is crucial to distinguish between high stock return predictability and high actual returns. The former evaluates the accuracy of the prediction, while the latter assesses the profitability that can be gained from investing in specific stocks.

Following the same aggregation strategy as Table 2, we categorize stocks into

deciles based on Ridge predictions[17] and construct equal-weighted, value-weighted long-short portfolios, and forecast-implied portfolios based on all observations. The investment performance is detailed in Table 5.

Table 5: **Forecast-Implied Investment Performance (Cross Section)**

This table reports the baseline investment performance for value-weighted, equal-weighted, and forecast-implied portfolios. We present in-sample (1973 - 2002) and out-of-sample (2003 - 2022, updated every five years) results. We provide six samples: Global (no clustering), Overall (aggregation clustering results), High, Others (Medium + Low), Medium, and Low, determined by the predictive rankings within the tree clusters. Four columns in each panel display monthly average return (Avg, in %), annualized Sharpe ratio (SR), market alpha (in %), and maximum draw-down (MDD, in %), respectively.

	In-Sample (1973-2002)				Out-of-Sample (2003-2022)			
	Panel A: Value-weighted Long-short Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	3.61	2.63	3.64***	17.08	0.92	0.78	1.14***	38.25
Overall	4.88	2.93	5.01***	39.22	2.20	1.38	2.33***	31.05
High	4.90	1.88	4.79***	41.84	3.49	1.28	3.01***	57.26
Others	4.63	3.21	4.70***	22.46	1.79	1.10	1.96***	35.75
Medium	5.30	3.55	5.31***	7.36	2.68	1.36	2.78***	40.33
Low	2.27	1.27	2.45***	57.12	0.53	0.41	0.60**	30.28
	Panel B: Equal-weighted Long-short Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	5.52	4.04	5.56***	44.92	2.70	2.29	2.79***	23.36
Overall	6.40	5.74	6.43***	22.80	3.77	2.99	3.91***	20.31
High	6.04	2.26	5.90***	35.56	4.30	1.59	3.98***	57.54
Others	6.04	5.73	6.03***	16.84	3.47	2.68	3.58***	24.44
Medium	7.12	6.58	7.10***	5.90	4.36	2.91	4.55***	19.86
Low	2.86	1.99	2.95***	59.35	1.41	1.45	1.41***	23.84
	Panel C: Forecast-implied Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	2.28	2.21	2.16***	17.84	1.44	1.13	0.75***	44.84
Overall	3.09	2.77	2.95***	12.39	1.83	1.45	1.22***	45.09
High	3.63	1.66	3.18***	28.56	3.00	1.60	2.07***	39.46
Others	2.83	2.52	2.76***	23.83	1.63	1.35	1.12***	45.50
Medium	3.20	2.64	3.15***	25.60	1.94	1.40	1.61***	45.86
Low	1.18	1.33	1.02***	15.84	0.86	0.62	0.06	44.69

The first row of all the panels demonstrates the predictive facts for our entire data set ("Global"). For example, Panel A shows an out-of-sample Sharpe ratio of 0.78

[17]This is to differentiate our paper from other machine learning literature and maintain consistency with the same model used for tree-based clustering.

and a market alpha of 1.14%. These numbers slightly improve when we aggregate clustering results ("Overall" with 1.38 and 2.33%). Portfolios constructed from high predictability stock returns consistently yield superior profitability. For instance, the average return and market alpha display multiplicative improvements compared to the overall sample, reaching 3.49% and 3.01% in the out-of-sample. Although highly predictable stock returns can generate more economic gains, they also come with more risks. Lower Sharpe ratios and higher maximum draw-downs both reflect the evidence of larger volatility for highly predictable clusters. On the other hand, clusters with the lowest predictability consistently perform the worst in terms of profitability, with more than half the average return (0.53%), Sharpe ratio (0.41), and market alphas (0.6% with a relatively lower significance level) compared to the others. This trend is consistent across equal-weighted long-short portfolios. However, if we apply a long-short strategy based on all of the observations (Panel C), the forecast-implied portfolios consistently show that highly predictable stocks can bring us more profits than all other sub-samples.

Nonetheless, all evaluations can demonstrate that those highly predictable observations dominate the economic profitability of the entire sample. Long-short portfolios are more sensitive to those extreme situations. The mosaics of stock return predictability significantly influence investment gains. Removing highly predictable observations from the sample may substantially reduce the profits that investors can achieve.

Using a similar segmentation mechanism as in the cross-sectional analysis, we gather sub-sample information according to time series divisions and present the economic gains in Table 6.[18] Surprisingly, the relationship between return predictability and profitability improves when considering time series information in addition to cross-sectional partitions. This inclusion makes the trend more stable and precise. The clusters with higher predictability consistently show the highest profits and the lowest volatility compared to others, especially for large-cap sub-samples. For example,

[18]Because the periods are no longer continuous, we use the minimum value of monthly portfolio return as the maximum draw-down rather than the way calculated in the cross-section. A negative value means there are no losses under that specific period.

Table 6: **Forecast-Implied Investment Performance (+ Time Series)**

This table reports the investment performance for value-weighted, equal-weighted, and forecast-implied portfolios from 1973 to 2022. The data is split based on time series and cross-sectional dimensions and covers two blocks (All stocks and Large-Cap). We present six samples: Global (no clustering), Overall (aggregation clustering results), High, Others (Medium + Low), Medium, and Low, determined by the predictive rankings within the tree clusters. Four columns in each panel display monthly average return (Avg, in %), annualized Sharpe ratio (SR), market alpha (in %), and monthly maximum draw-down (MDD, in %), respectively.

	Sample A: All Stocks				Sample B: Large-Cap			
	Panel A: Value-weighted Long-short Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	2.45	1.93	2.52***	21.49	1.71	1.56	1.75***	14.24
Overall	3.80	2.29	3.91***	20.77	1.99	1.67	2.02***	14.62
High	4.28	4.88	4.29***	1.95	3.18	5.80	3.14***	0.21
Others	3.83	2.27	3.92***	20.77	2.02	1.69	2.05***	14.62
Medium	4.48	2.29	4.59***	20.18	2.72	1.47	2.87***	30.29
Low	3.72	2.13	3.95***	20.53	1.47	1.33	1.53***	10.30
	Panel B: Equal-weighted Long-short Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	4.91	3.96	4.99***	19.44	2.25	1.92	2.35***	26.94
Overall	5.73	4.70	5.77***	12.74	2.59	2.04	2.69***	28.06
High	6.48	10.48	6.47***	-1.99	3.72	10.03	3.71***	-1.87
Others	5.71	4.65	5.74***	12.74	2.59	2.03	2.68***	28.06
Medium	6.30	3.88	6.46***	16.70	3.19	1.72	3.40***	35.55
Low	5.39	4.64	5.50***	8.63	2.11	1.81	2.25***	24.33
	Panel C: Forecast-implied Portfolio							
	Avg	SR	Alpha	MDD	Avg	SR	Alpha	MDD
Global	2.30	1.83	2.28***	22.73	1.48	1.23	1.36***	18.58
Overall	2.87	2.15	2.83***	20.50	1.80	1.51	1.66***	16.15
High	4.68	3.00	4.58***	2.17	3.26	2.43	3.16***	4.35
Others	2.86	2.15	2.82***	20.50	1.79	1.51	1.66***	16.15
Medium	3.26	2.37	3.17***	22.64	2.21	1.58	2.12***	24.15
Low	2.45	2.05	2.59***	20.05	1.37	1.27	1.25***	14.31

in Panel A, the Sharpe ratio of the highly predictable sub-sample reaches 4.88 with an average return of 4.28%. In contrast, the lowest predictability clusters achieve almost only half of these values (2.13 for "SR" and 3.72% for "Avg").[19] There are also similar situations for market alphas and maximum draw-downs. The large-cap samples

[19]It is possible that "Others" and "Medium" possess higher Sharpe ratio compared with highly predictable sub-sample due to more observations, resulting in less volatility. The average returns have already indicated our estimated trend.

present relatively more stable trends among all three kinds of investment strategies.

Therefore, when we integrate the evidence from cross-sectional and time series analysis, stocks with greater return predictability can lead to higher economic gains. By utilizing tree-based clustering techniques, these strategies can inform investors about specific time periods and stocks that are likely to generate higher profits.

In general, the predictability of stock returns demonstrates a certain level of timeliness in both the cross-sectional and time series dimensions. It's important to continuously update clustering results in a rolling manner for the cross-section without regime changes by macroeconomic predictors or calendar months. The consistent features or determinations within the same periods underscore the universality of predictability. We can develop more robust and rational investment strategies by considering the mosaics of stock return predictability.

6 Conclusion

What types of stocks exhibit higher return predictability, and under which time regimes? Answering this important question leads to discovering "mosaics of predictability" and adds to our understanding of asset return predictability in multiple ways. For example, we provide a systematic framework for studying the heterogeneity of individual asset return predictability, complementing the focus of extant literature on average return predictability. Second, we generalize studies on heterogeneous predictability, such as "pockets of predictability," that show the time-varying return predictability of market returns to accommodate an unbalanced panel of individual asset returns influenced by different macroeconomic regimes. In both innovations, we utilize a sparse and interpretable AI methodology to accommodate high-dimensional characteristics and time series predictors and their interactions for the first time.

More specifically, we construct a tree-based clustering algorithm to distinguish between highly predictable and less predictable ones. The conceptual framework speaks to what observation units are more predictable and when. In an empirical application to U.S. equities, we find supportive evidence that some characteristics-

managed (dollar trading volumes, unexpected earnings, earnings-to-price, cashflow-to-price) and/or macro-based (dividend yield and default yield) clusters are more predictable than others. We then explore the return heterogeneity between these clusters and exploit the predictability for cluster-wise predictive models. We find that heterogeneous predictability models outperform models under homogeneous predictability. Finally, we show that highly predictable clusters have better investment performance than less predictable clusters, linking return predictability to investment gains, which is of great practical relevance.

References

Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009). Basis assets. *Review of Financial Studies 22*(12), 5133–5174.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D., S. Cheng, and L. Metzker (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science 69*(5), 2587–2619.

Ball, R. and P. Brown (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research 6*(2), 159–178.

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

Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. *Journal of Finance, Forthcoming*.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*(3), 195–228.

Cong, L. W., G. Feng, J. He, and X. He (2023). Growing the Efficient Frontier on Panel Trees. Technical report, City University of Hong Kong.

Cong, L. W., G. Feng, J. He, and J. Li (2023). Uncommon Factors and Asset Heterogeneity in the Cross Section and Time Series. Technical report, National Bureau of Economic Research.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2020). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Available at SSRN 3554486*.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Evgeniou, T., A. Gueciouer, and R. Prieto (2023). Uncovering sparsity and heterogeneity in firm-level return predictability using machine learning. *Journal of Financial and Quantitative Analysis 58*(8), 3384–3419.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

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

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

Fama, E. F. and K. R. French (2008). Dissecting anomalies. *Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Farmer, L. E., L. Schmidt, and A. Timmermann (2023). Pockets of predictability. *Journal of Finance 78*(3), 1279–1341.

Feng, G. and J. He (2022). Factor investing: A bayesian hierarchical approach. *Journal of Econometrics 230*(1), 183–200.

Feng, G., J. He, J. Li, L. Sarno, and Q. Zhang (2024). Currency Return Dynamics: What Is the Role of US Macroeconomic Regimes? Technical report, City University of Hong Kong.

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

Green, J., J. R. Hand, and X. F. Zhang (2013). The supraview of return predictive signals. *Review of Accounting Studies 18*, 692–730.

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

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

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

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560–580.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

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

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics 17*(2), 357–390.

Kelly, B., S. Malamud, and K. Zhou (2024). The virtue of complexity in return prediction. *Journal of Finance 79*(1), 459–503.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*(2), 209–235.

Lewellen, J. (2015). The Cross-section of Expected Stock Returns. *Critical Finance Review 4*(1), 1–44.

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

Patton, A. J. and B. M. Weller (2022). Risk price variation: The missing half of empirical asset pricing. *Review of Financial Studies 35*(11), 5127–5184.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance 50*(4), 1201–1228.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Shen, Z. and D. Xiu (2024). Can Machines Learn Weak Signals? *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2024-29).

Smith, S. C. and A. Timmermann (2021). Break risk. *Review of Financial Studies 34*(4), 2045–2100.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

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

Appendix

I. Panel Regression Tree Algorithm

Section 2.3 and 2.4 present the step-by-step tree growing examples, while this section illustrates the complete growing algorithm in pseudo-codes.

Algorithm Panel Regression Tree

procedure PANEL REGRESSION TREE
Input: Asset returns $r_{i,t}$, firm characteristics $z_{i,t-1}$, aggregate predictors x_{t-1}, and tree parameters.
Output: A tree architecture with many split rules.
  for i from 1 to num_iter **do** ▷ Loop over number of iterations
    if current depth $\geq d_{\text{max}}$ **then**
      return.
    else
      Search the tree, find all potential leaf nodes $\mathcal{N}$
      for each leaf node N in $\mathcal{N}$ **do** ▷ Loop over all current leaf nodes
        for each split candidate $\tilde{c}_{p,k,N}$ in $\mathcal{C}_N$ **do**
          Partition data temporally in N according to $\tilde{c}_{p,k,N}$.
          if Left or right child node cannot satisfy minimal leaf size **then**
            continue.
          else
            Obtain cluster-wise return predictions as in (1).
            Calculate the cluster-based R_j^2 by (6).
          end if
        end for
      end for
      Find the best leaf node and split rule that maximizes split criteria for this iteration

$$\tilde{c}_i = \max_{N \in \mathcal{N}, \tilde{c}_{p,k,N} \in \mathcal{C}_N} |R^2_{\text{left}} - R^2_{\text{right}}|$$

      Compare globally for this iteration's split candidates among all leaf nodes.
      Split the node selected at the i-th split rule of the tree $\tilde{c}_i$.
    end if
  end for
  return
end procedure

Note: p, k, N in $\tilde{c}_{p,k,N}$ represent the p-th variable with the k-th value used for the N-th leaf node (Figure 2).

II. Additional Empirical Results

II..1 Cross Section

Figure A.1: **Cluster-Wise Relationships**

This figure shows the cluster-wise relationships between predictability and profitability corresponding to Table 1. Four sub-figures represent each sample period for cross-sectional partitions. The horizontal axis shows the cluster-wised predictability (R^2), while two vertical axes represent the forecast-implied portfolio investment performance by average returns (red) and Sharpe ratios (blue), respectively.

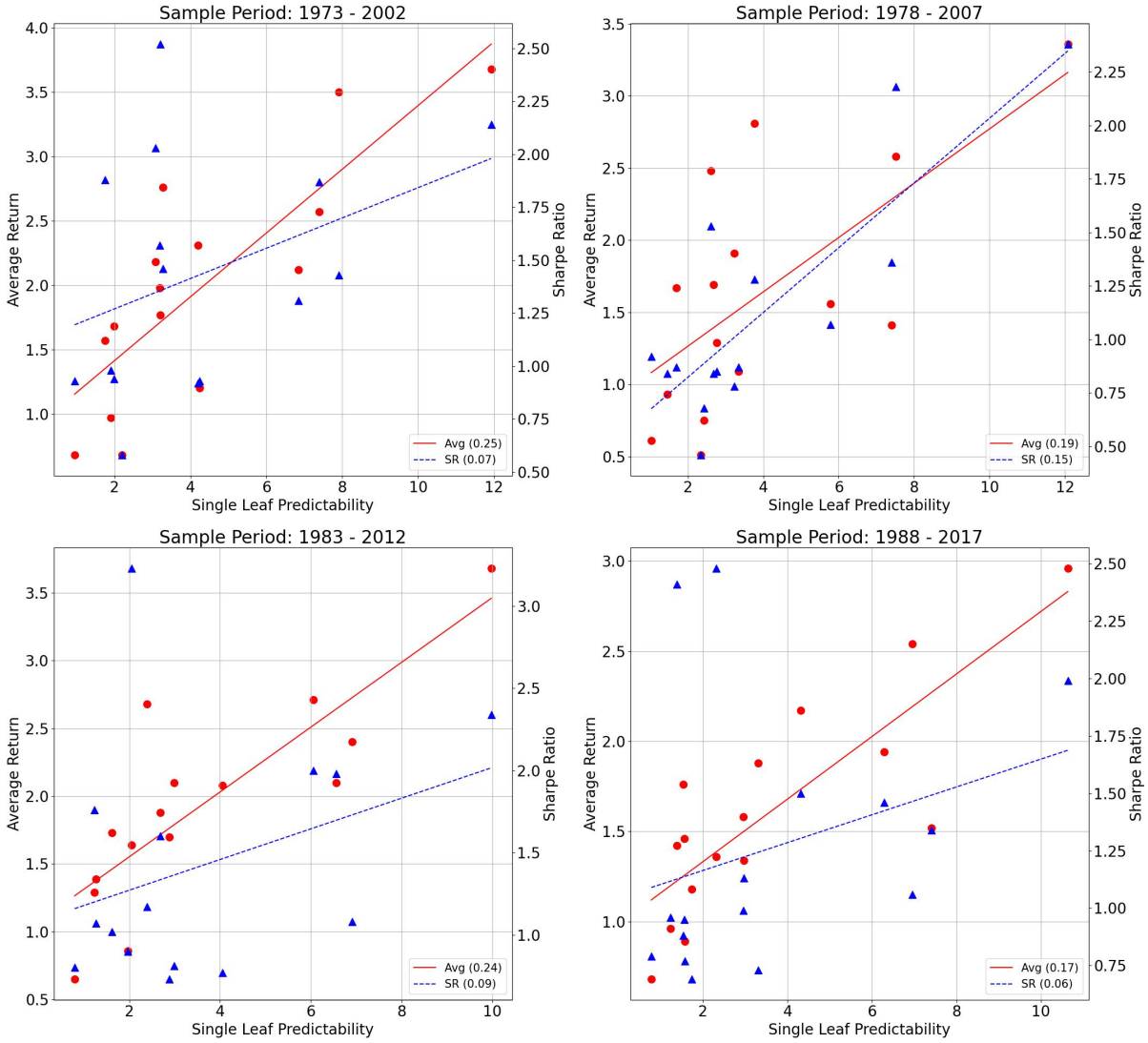


Figure A.2: **Tree-Based Cluster (Cross Section, 1978 - 2007)**

This figure shows the cross-sectional split tree-based clustering structure using monthly data from 1978 to 2007. The tree splits the panel of individual stock returns based on monthly cross-sectional standardization of firm characteristic ranks in [0,1]. The terminal leaves correspond to clusters identified by the interaction of firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. All nodes are labeled with cluster-wise model R^2.

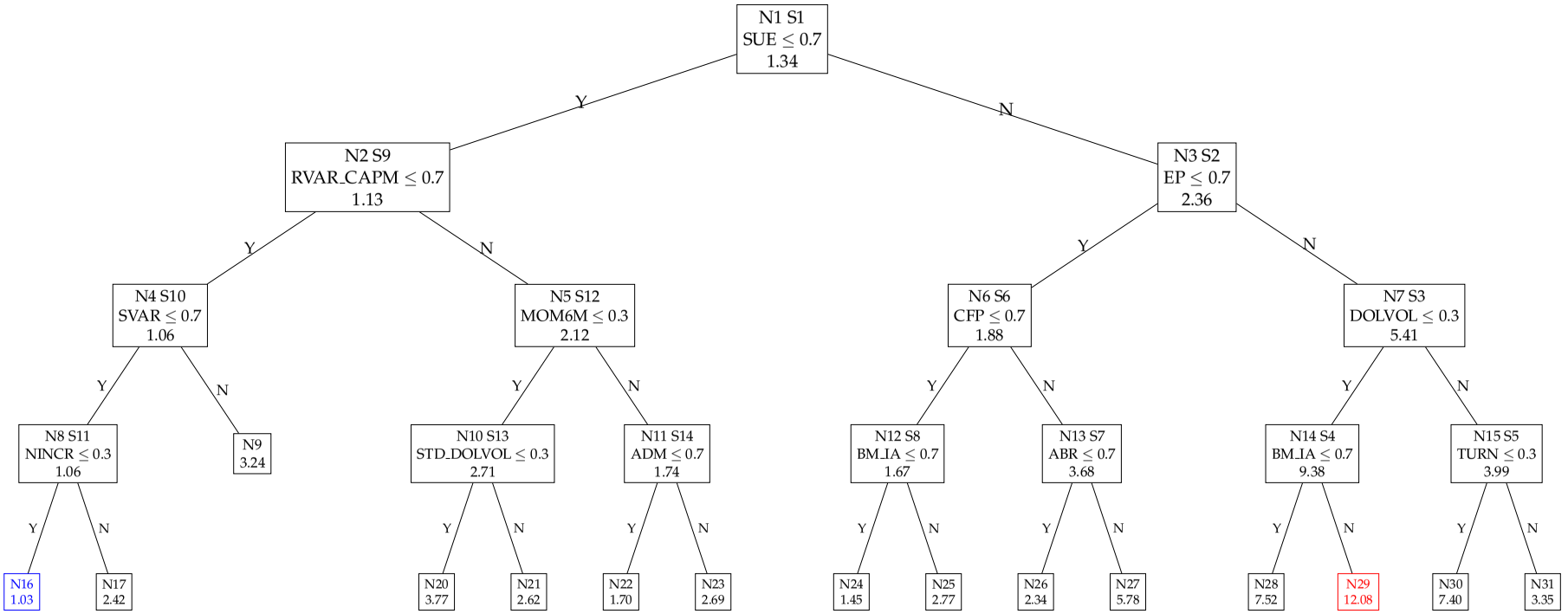


Figure A.3: **Mosaics of Predictability by Predictors (Cross Section, 1978 - 2007)**

We present four heat maps to summarize the average return predictability, R^2 values (% in the color bar), for the panel of individual stock returns corresponding to the tree-based clustering results from Figure A.2. The first three illustrate the average R^2 values for groups categorized by various years and deciles based on different characteristics (earning surprise, earnings-to-price, and market equity value). The last one displays the average R^2 values for the 10×10 groups by bivariate-sorted deciles for the top two characteristics.

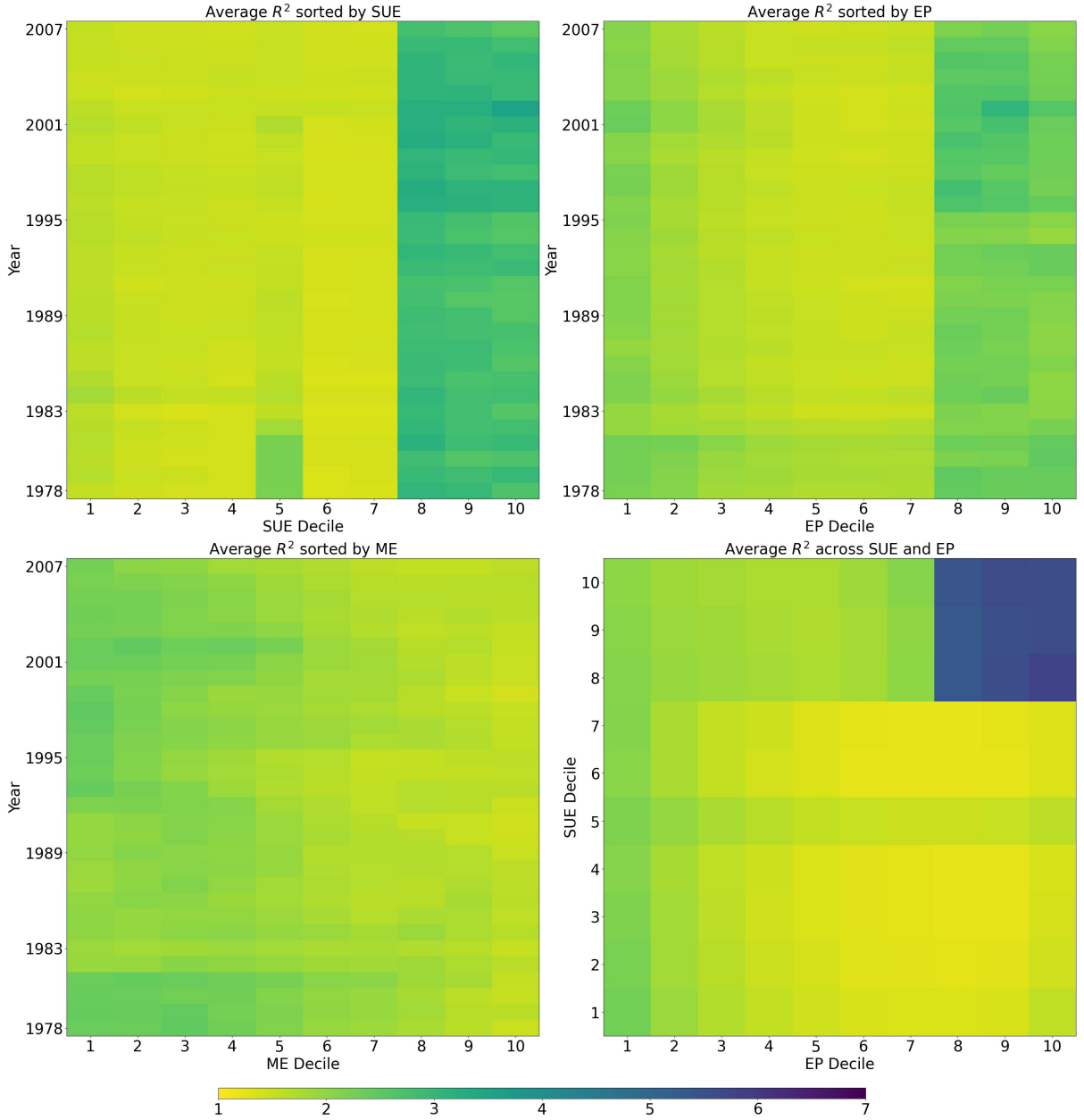


Figure A.4: **Mosaics of Predictability (Cross Section, 1978 - 2007)**

This heat map summarizes the predictability, R^2 values (% in the color bar), for the panel of individual stock returns by the tree-based clustering in Figure A.2. The vertical axis represents months, and colors from light to dark indicate ascending levels of return predictability of each cluster within each month. Horizontally, the length of each color bin corresponds to the proportion of observations for each cluster.

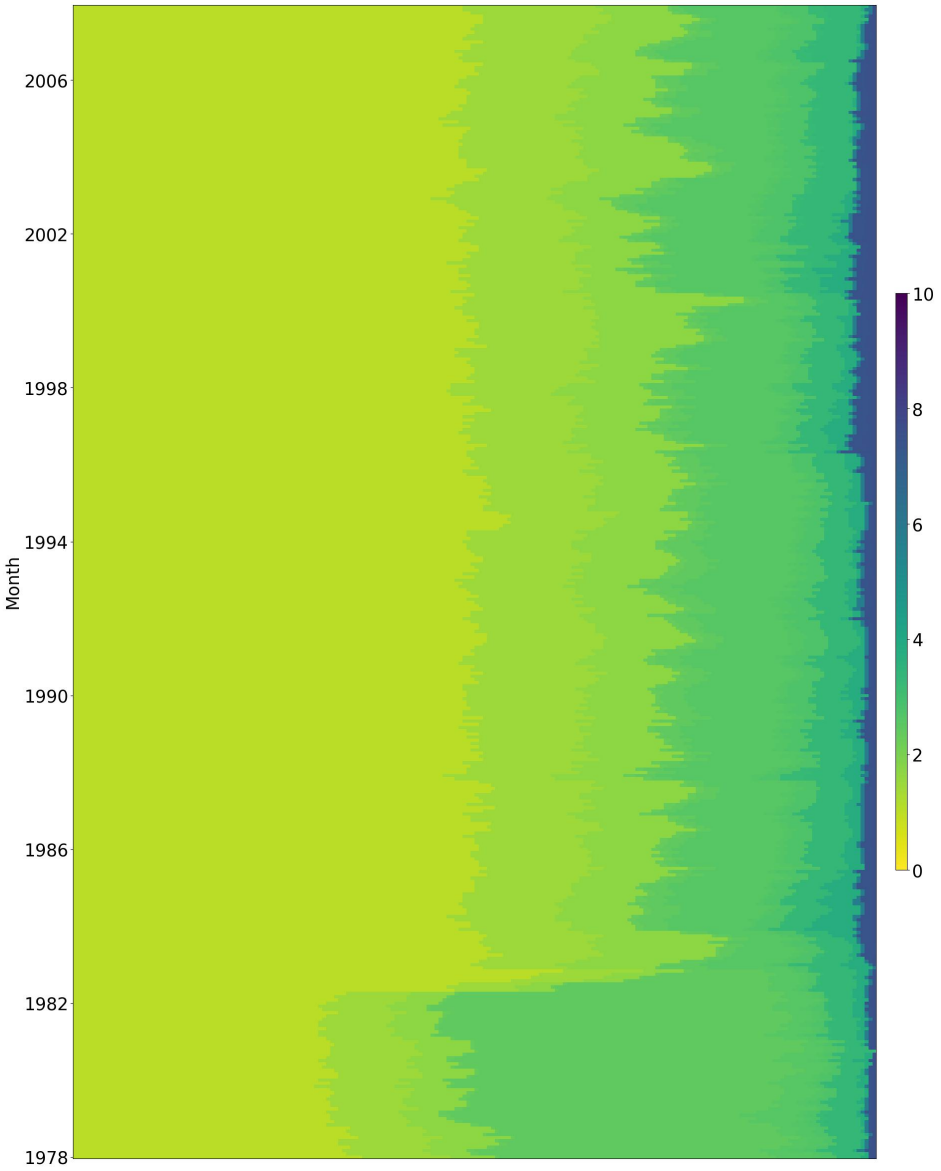


Figure A.5: **Tree-Based Cluster (Cross Section, 1983 - 2012)**

This figure shows the cross-sectional split tree-based clustering structure using monthly data from 1983 to 2012. The tree splits the panel of individual stock returns based on monthly cross-sectional standardization of firm characteristic ranks in [0,1]. The terminal leaves correspond to clusters identified by the interaction of firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. All nodes are labeled with cluster-wise model R^2.

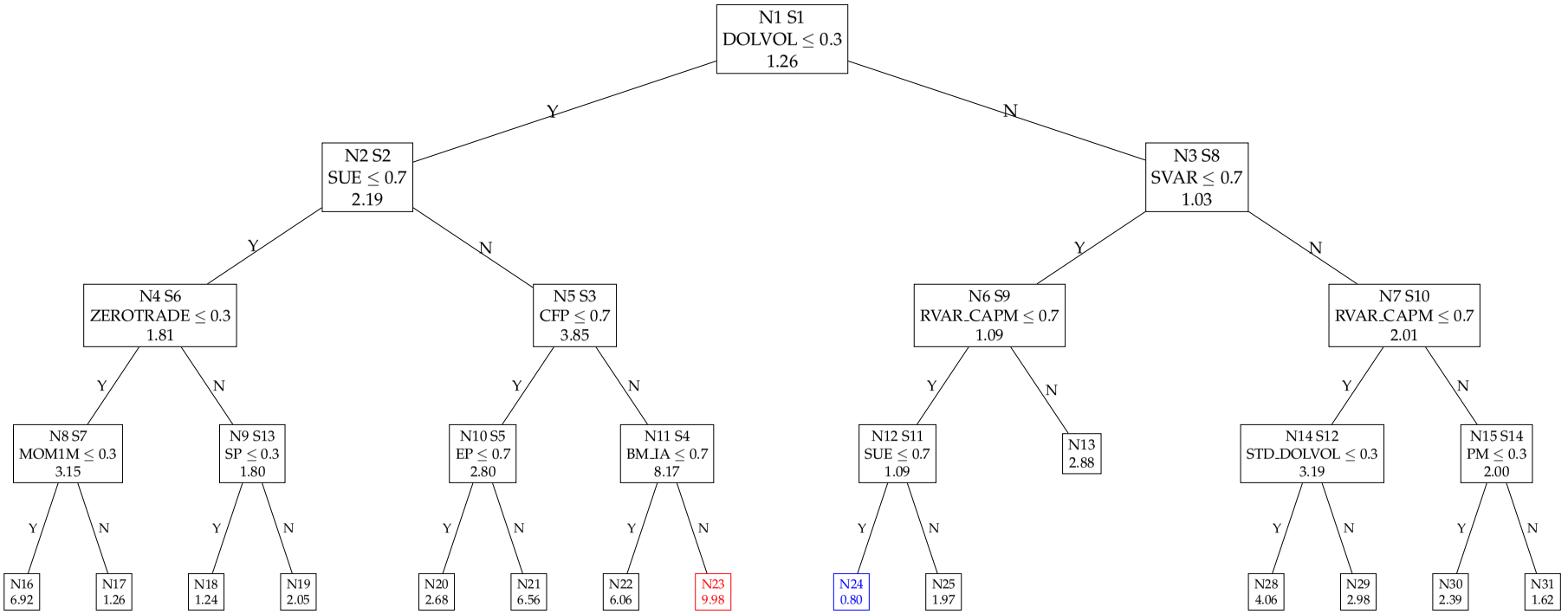


Figure A.6: **Mosaics of Predictability by Predictors (Cross Section, 1983 - 2012)**

We present four heat maps to summarize the average return predictability, R^2 values (% in the color bar), for the panel of individual stock returns corresponding to the tree-based clustering results from Figure A.5. The first three illustrate the average R^2 values for groups categorized by various years and deciles based on different characteristics (dollar trading volumes, earning surprise, and market equity value). The last one displays the average R^2 values for the 10×10 groups by bivariate-sorted deciles for the top two characteristics.

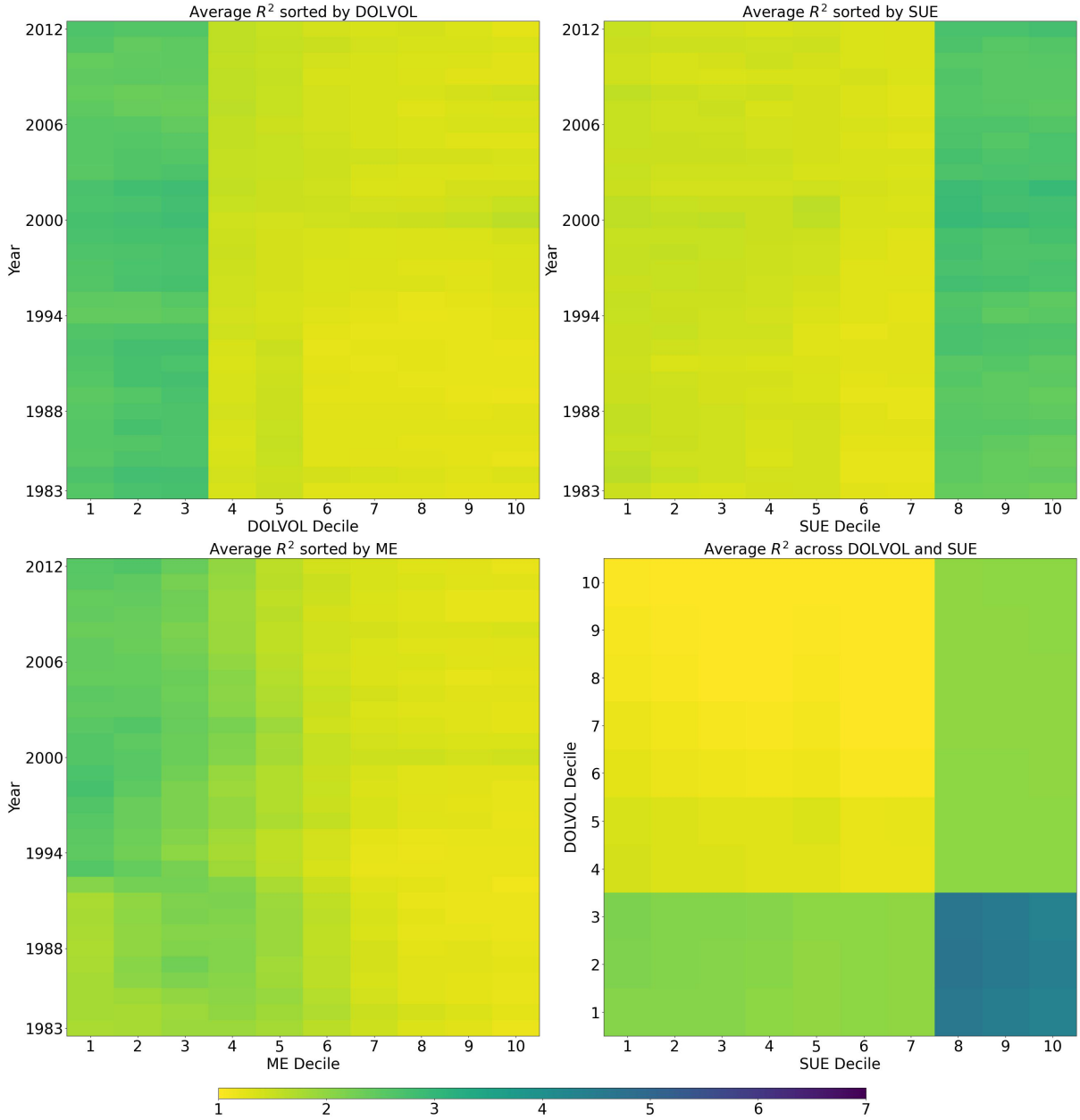


Figure A.7: **Mosaics of Predictability (Cross Section, 1983 - 2012)**

This heat map summarizes the predictability, R^2 values (% in the color bar), for the panel of individual stock returns by the tree-based clustering in Figure A.5. The vertical axis represents months, and colors from light to dark indicate ascending levels of return predictability of each cluster within each month. Horizontally, the length of each color bin corresponds to the proportion of observations for each cluster.

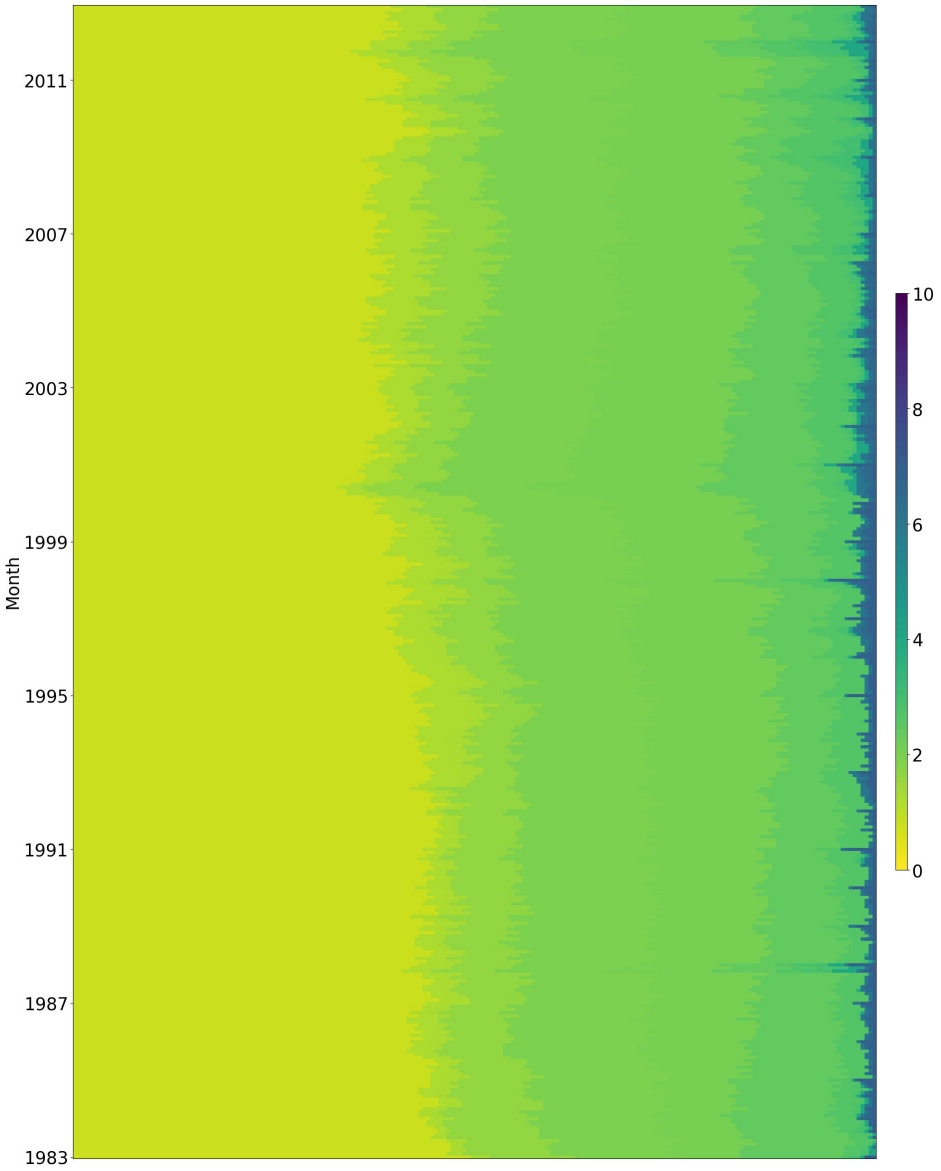


Figure A.8: **Tree-Based Cluster (Cross Section, 1988 - 2017)**

This figure shows the cross-sectional split tree-based clustering structure using monthly data from 1988 to 2017. The tree splits the panel of individual stock returns based on monthly cross-sectional standardization of firm characteristic ranks in [0,1]. The terminal leaves correspond to clusters identified by the interaction of firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. All nodes are labeled with cluster-wise model R^2.

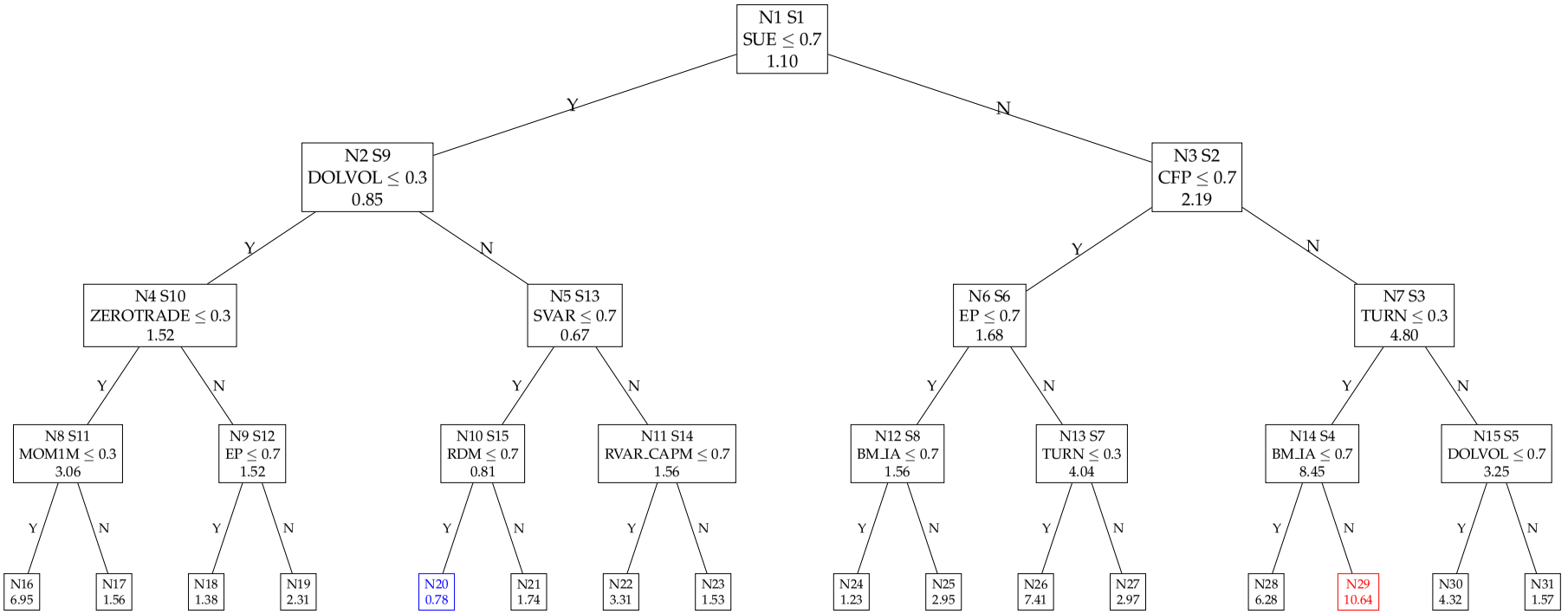


Figure A.9: **Mosaics of Predictability by Predictors (Cross Section, 1988 - 2017)**

We present four heat maps to summarize the average return predictability, R^2 values (% in the color bar), for the panel of individual stock returns corresponding to the tree-based clustering results from Figure A.8. The first three illustrate the average R^2 values for groups categorized by various years and deciles based on different characteristics (earning surprise, cashflow-to-price, and market equity value). The last one displays the average R^2 values for the 10×10 groups by bivariate-sorted deciles for the top two characteristics.

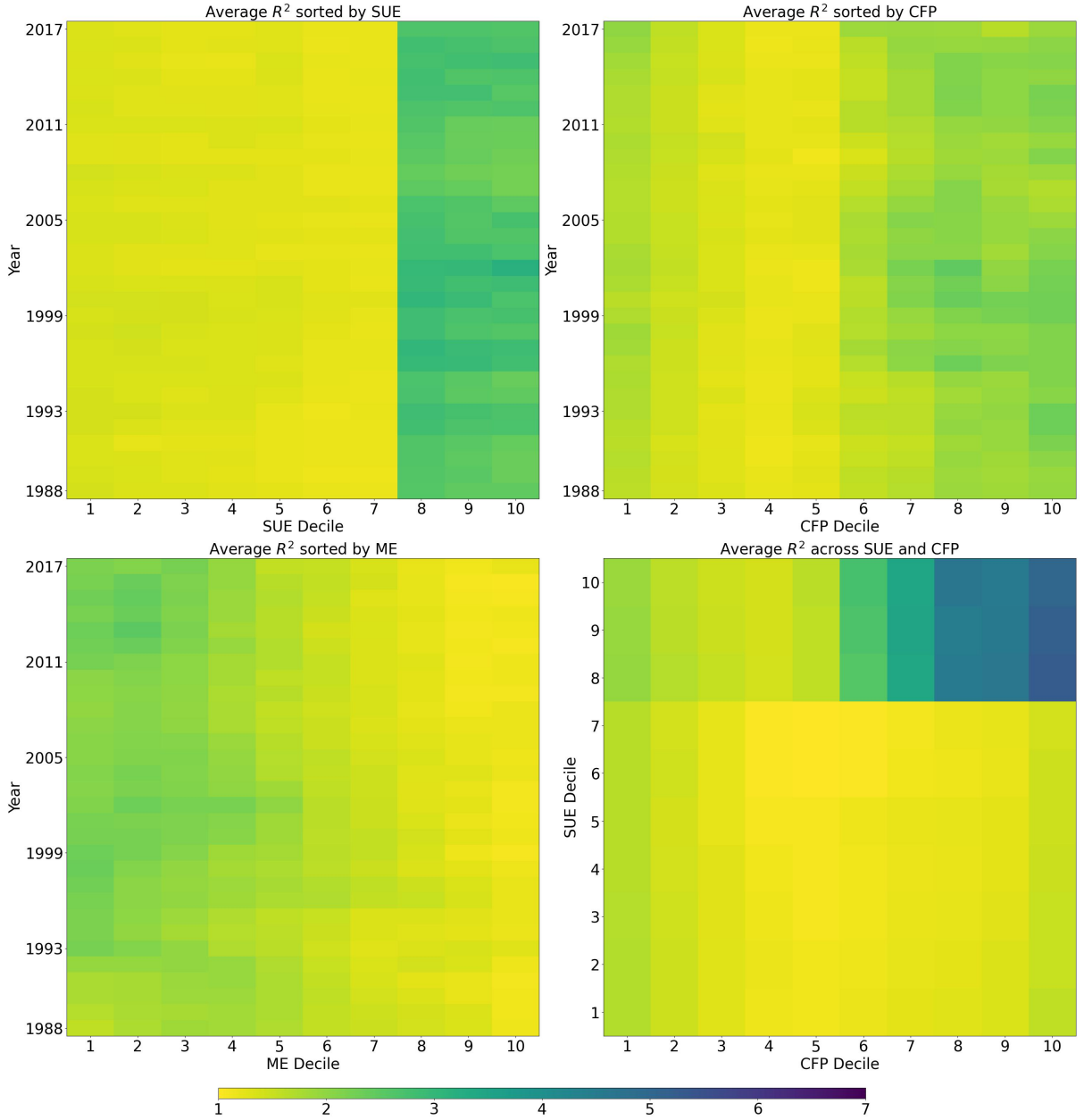
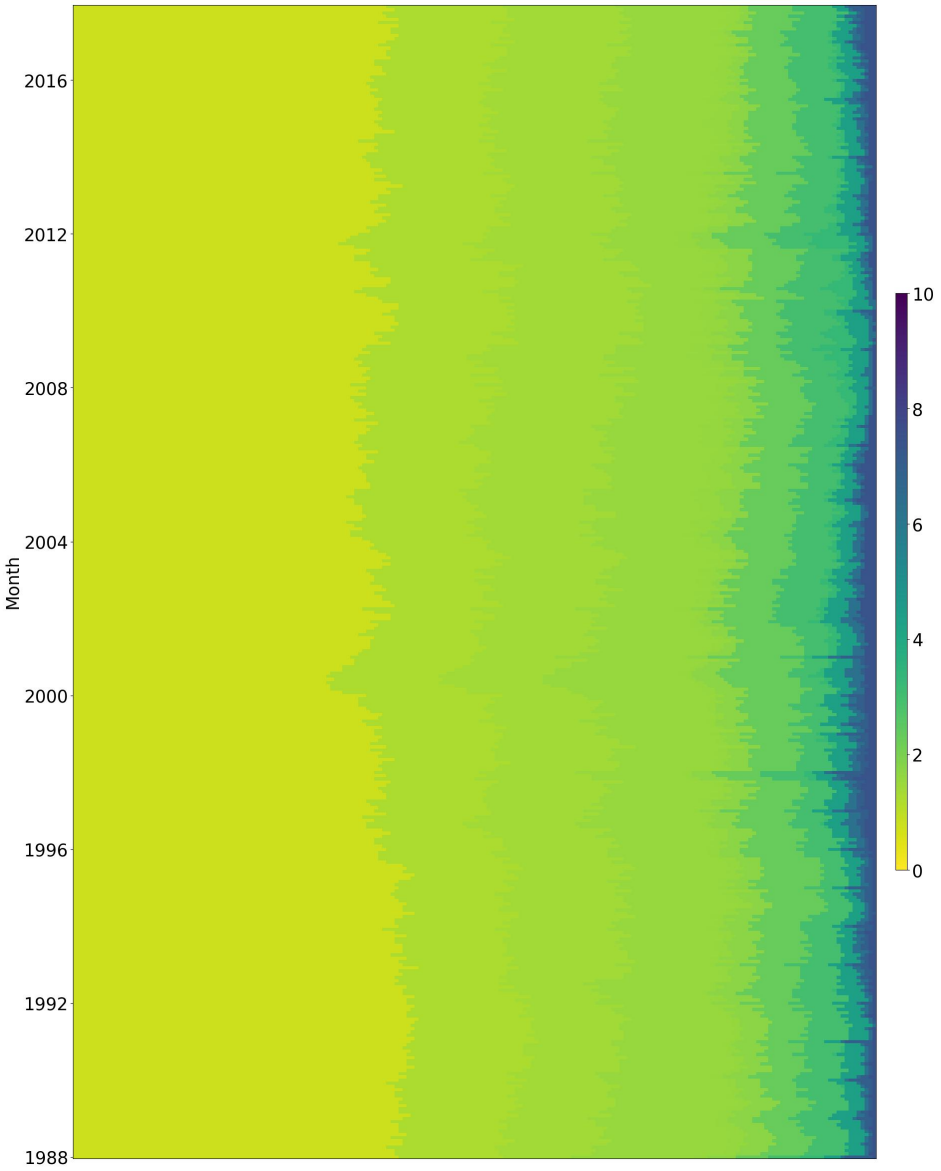


Figure A.10: **Mosaics of Predictability (Cross Section, 1988 - 2017)**

This heat map summarizes the predictability, R^2 values (% in the color bar), for the panel of individual stock returns by the tree-based clustering in Figure A.8. The vertical axis represents months, and colors from light to dark indicate ascending levels of return predictability of each cluster within each month. Horizontally, the length of each color bin corresponds to the proportion of observations for each cluster.



II..2 Structural Break

Table A.1: **Cluster-Wise Performance (+ Structural Break)**

This table shows the performance of the tree-based cluster model based on time series and cross-sectional splits. The in-sample return predictability, R^2 (in %), are calculated by Eq. (6). "# obs" represents the stock returns count for each cluster. "Avg" and "SR" denote the monthly average return (in %) and annualized Sharpe ratio for both cluster-wise equal-weighted and value-weighted forecast-implied portfolios, respectively. Each regime of values is arranged in the descending order of R^2 from left to right.

197301-197808	N7	N13	N24	N4	N10	N23	N25	N22						
# obs	12,456	10,468	11,341	19,069	16,687	16,022	11,907	148,017						
R^2	17.80	12.17	11.03	10.78	10.33	9.80	7.85	5.37						
Avg_{EW}	4.49	4.89	5.15	3.50	3.27	2.89	3.68	2.65						
SR_{EW}	1.85	2.42	2.86	2.00	2.33	1.91	2.15	1.76						
Avg_{VW}	3.82	3.97	4.75	2.65	2.61	2.42	3.40	1.67						
SR_{VW}	1.76	2.29	3.02	1.70	2.28	1.89	2.09	1.70						
197809-198308	N15	N14	N19	N10	N11	N16	N18	N6	N17					
# obs	13,683	15,824	11,955	12,017	10,097	21,130	33,132	14,132	107,707					
R^2	16.28	13.25	12.05	11.41	9.76	9.52	9.20	8.34	6.21					
Avg_{EW}	4.40	4.05	3.46	3.86	3.94	2.50	3.53	2.78	2.98					
SR_{EW}	2.74	2.65	2.43	2.55	2.14	2.45	2.38	2.31	2.08					
Avg_{VW}	2.97	2.31	3.45	3.00	2.38	2.06	3.53	1.91	1.86					
SR_{VW}	2.15	1.82	2.30	2.20	1.65	2.20	2.29	1.69	1.83					
198309-198811	N30	N13	N28	N31	N24	N17	N19	N16	N29	N25	N18	N5		
# obs	20,934	11,204	12,120	104,173	11,733	10,950	10,502	10,490	21,443	27,032	46,428	12,370		
R^2	10.08	8.07	7.51	7.11	6.80	6.68	5.69	5.60	4.72	3.17	2.07	-7.77		
Avg_{EW}	2.82	2.03	3.05	2.22	2.98	2.78	2.70	2.46	3.19	1.68	1.86	1.00		
SR_{EW}	1.75	2.06	2.25	1.43	1.76	2.17	2.03	2.25	2.64	1.55	1.83	1.22		
Avg_{VW}	2.04	1.75	2.95	1.61	1.77	2.53	1.99	2.41	3.25	1.13	1.42	0.95		
SR_{VW}	1.33	1.60	2.18	1.12	1.24	2.10	1.42	2.30	2.53	1.05	1.31	1.04		
198812-199803	N14	N31	N12	N20	N18	N30	N26	N21	N19	N23	N17	N27	N22	N16
# obs	14,635	14,915	12,594	14,883	33,550	12,315	107,725	10,176	57,794	18,007	264,478	23,615	14,053	21,101
R^2	12.78	8.11	7.11	6.74	4.87	4.46	3.38	2.75	2.74	2.31	1.94	1.30	0.60	0.39
Avg_{EW}	3.38	3.12	2.10	4.02	1.16	1.75	1.60	2.74	1.59	2.03	2.04	1.54	2.23	0.91
SR_{EW}	3.50	2.66	1.89	2.28	1.77	1.56	2.13	1.50	1.92	1.84	3.02	2.25	2.41	1.97
Avg_{VW}	2.20	2.32	2.13	2.07	1.19	1.52	1.31	2.08	1.23	1.78	1.48	1.10	2.22	0.93
SR_{VW}	2.52	2.36	1.67	1.11	1.35	1.35	1.76	1.07	1.44	1.45	1.69	0.98	2.46	1.32
199804-200303	N6	N15	N10	N19	N14	N18	N11	N16	N17					
# obs	19,659	11,027	10,722	12,155	11,809	54,089	10,912	24,346	197,515					
R^2	12.94	10.97	9.93	9.61	7.50	6.50	5.84	5.28	2.83					
Avg_{EW}	9.61	6.31	7.47	7.02	5.88	5.77	4.49	3.41	1.63					
SR_{EW}	2.53	1.67	2.07	2.12	2.24	3.18	2.37	2.10	1.86					
Avg_{VW}	7.41	6.96	4.94	6.55	4.01	5.79	2.27	3.97	0.63					
SR_{VW}	2.12	1.87	1.38	1.90	1.59	2.69	0.96	2.01	0.64					
200304-200808	N11	N6	N20	N17	N19	N7	N21	N16	N18					
# obs	18,849	20,951	16,527	12,396	52,975	16,343	36,240	30,237	91,152					
R^2	10.75	9.39	7.37	7.18	4.67	4.49	4.29	3.88	2.51					
Avg_{EW}	2.85	3.18	1.75	1.96	1.65	2.85	1.90	2.20	1.27					
SR_{EW}	2.70	1.64	2.76	1.58	1.99	1.88	1.69	2.15	1.81					
Avg_{VW}	1.32	2.02	1.10	1.69	1.46	2.20	0.96	1.01	0.66					
SR_{VW}	1.34	1.02	1.82	1.34	1.91	1.48	1.34	1.21	1.06					
200809-201708	N14	N25	N11	N15	N21	N24	N26	N19	N27	N20	N17	N18	N16	
# obs	10,211	14,442	10,783	20,192	14,600	19,766	13,437	24,431	41,970	47,813	101,001	17,378	64,402	
R^2	11.40	9.41	7.40	7.29	6.84	6.60	6.10	5.37	4.51	4.18	3.44	3.13	1.87	
Avg_{EW}	2.98	2.15	3.94	2.51	3.44	1.53	2.51	1.65	2.29	2.83	1.88	1.48	1.64	
SR_{EW}	2.43	1.78	1.80	1.81	1.84	1.20	2.08	1.34	1.36	1.52	1.45	1.34	1.60	
Avg_{VW}	1.69	2.24	4.00	1.81	1.92	1.28	2.07	1.33	1.97	2.02	1.50	1.39	1.50	
SR_{VW}	1.19	1.44	1.74	1.72	1.33	1.20	1.65	1.20	1.18	1.16	1.34	1.36	1.47	
201709-202212	N13	N22	N30	N12	N31	N23	N18	N10	N14	N19	N17	N16		
# obs	16,828	11,479	14,059	12,671	11,678	13,599	15,872	11,654	12,834	81,762	13,429	11,740		
R^2	13.59	9.39	9.38	9.26	6.79	6.75	6.74	5.00	4.35	3.98	1.54	-6.82		
Avg_{EW}	1.94	4.34	2.72	2.11	2.96	3.77	2.47	3.47	3.00	2.64	1.91	2.11		
SR_{EW}	1.38	1.65	1.60	1.74	1.69	1.75	1.84	2.10	1.34	1.77	1.36	1.36		
Avg_{VW}	2.03	2.73	2.62	1.68	2.80	2.46	1.56	1.44	2.06	1.84	1.67	2.42		
SR_{VW}	1.27	1.15	1.58	1.50	1.57	1.59	1.39	1.34	0.92	1.50	1.26	1.44		

Table A.2: **Evaluating Return Predictability (+ Structural Break)**

This table reports the return predictability, R^2s (in %), based on different predictive methods under various regimes. We present the full-sample results based on the tree-based cluster model incorporating both structural breaks and cross-sectional splits. Under each regime, we present four samples: Overall, High, Medium, and Low, determined by the predictive rankings within the tree clusters.

	Sample A: All Stocks			Sample B: Large-Cap		
197301-197808	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	8.61	7.12	8.21	6.67	5.45	6.34
High	12.94	10.95	12.46	7.37	6.81	7.48
Medium	9.27	7.31	8.95	9.11	7.23	8.83
Low	5.45	4.41	5.11	5.92	4.74	5.51
197809-198308	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	8.55	7.11	8.21	11.05	9.46	10.60
High	13.37	11.75	12.80	13.10	11.20	12.44
Medium	9.32	7.83	8.98	9.08	7.39	8.71
Low	6.26	4.90	6.00	9.08	7.99	8.90
198309-198811	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	5.28	3.55	4.98	7.78	5.35	7.25
High	10.07	8.22	9.59	13.38	11.08	12.74
Medium	6.88	4.40	6.30	8.49	5.50	7.67
Low	2.51	1.70	2.57	1.41	1.30	2.14
198812-199803	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	2.86	2.37	2.79	3.66	2.74	3.52
High	12.94	11.86	12.86	11.25	10.78	11.40
Medium	7.26	5.95	7.19	9.81	6.60	9.36
Low	2.31	1.91	2.24	3.36	2.51	3.22
199804-200303	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	7.50	5.78	6.60	6.56	5.04	5.79
High	11.94	10.20	10.34	11.67	10.23	10.58
Medium	6.48	4.62	5.88	6.82	4.66	6.15
Low	3.03	1.60	2.60	2.54	1.22	1.97
200304-200808	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	6.71	4.98	6.25	6.27	4.76	5.82
High	11.40	10.32	11.03	9.89	8.95	9.40
Medium	8.58	6.40	8.20	8.40	7.45	8.42
Low	5.11	3.26	4.62	5.55	3.91	5.08
200809-201708	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	4.29	2.42	4.27	6.12	3.76	6.12
High	10.88	7.63	10.72	9.59	5.80	9.59
Medium	6.67	4.40	6.61	8.24	5.48	8.23
Low	3.30	1.61	3.30	5.08	2.95	5.08
201709-202212	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Overall	3.83	3.18	3.99	5.76	4.95	5.86
High	9.64	8.38	9.37	11.68	9.79	11.19
Medium	5.89	4.27	5.59	6.83	5.38	6.55
Low	1.91	1.64	2.27	3.16	2.98	3.61

III. Predictor Descriptions

Table A.3: **Macroeconomic Variables**

No.	Variable Name	Description
1	X_TBL	3-month treasury bill rate
2	X_INFL	Inflation
3	X_TMS	Term spread
4	X_DFY	Default yield
5	X_DY	Dividend yield of S&P 500
6	X_SVAR	Rolling 12-month market excess return volatility
7	X_NI	Net equity issuance of S&P 500
8	X_LIQ	Rolling 12-month Pastor-Stambaugh illiquidity

Table A.4: **Equity Characteristics**

No.	Acronym	Description	Category
1	abr	Cumulative abnormal returns around earnings announcement dates	Momentum
2	acc	Operating Accruals	Investment
3	adm	Advertising Expense-to-market	Intangibles
4	agr	Asset growth	Investment
5	alm	Quarterly Asset Liquidity	Intangibles
6	ato	Asset Turnover	Profitability
7	baspread	Bid-ask spread rolling 3m	Liquidity
8	beta	Beta rolling 3m	Volatility
9	bm	Book-to-market equity	Value
10	bm_ia	Industry-adjusted book to market	Value
11	cash	Cash holdings	Value
12	cashdebt	Cash to debt	Value
13	cfp	Cashflow-to-price	Value
14	chpm	Industry-adjusted change in profit margin	Profitability
15	chtx	Change in tax expense	Momentum
16	cinvest	Corporate investment	Investment
17	depr	Depreciation / PPandE	Momentum
18	dolvol	Dollar trading volume	Liquidity
19	dy	Dividend yield	Value
20	ep	Earnings-to-price	Value
21	gma	Gross profitability	Investment
22	grltnoa	Growth in long-term net operating assets	Investment
23	herf	Industry sales concentration	Intangibles

Table A.4: Equity Characteristics (Continued)

No.	Acronym	Description	Category
24	hire	Employee growth rate	Intangibles
25	ill	Illiquidity rolling 3m	Liquidity
26	lev	Leverage	Value
27	lgr	Growth in long-term debt	Investment
28	maxret	Maximum daily returns rolling 3m	Volatility
29	me	Market equity	Size
30	me_ia	Industry-adjusted size	Size
31	mom12m	Momentum rolling 12m	Momentum
32	mom1m	Momentum	Momentum
33	mom36m	Momentum rolling 36m	Momentum
34	mom60m	Momentum rolling 60m	Momentum
35	mom6m	Momentum rolling 6m	Momentum
36	ni	Net Stock Issues	Investment
37	nincr	Number of earnings increases	Momentum
38	noa	(Changes in) Net Operating Assets	Investment
39	op	Operating profitability	Profitability
40	pctacc	Percent operating accruals	Investment
41	pm	Profit margin	Profitability
42	pscore	Performance Score	Profitability
43	rd_sale	R&D to sales	Intangibles
44	rdm	R&D Expense-to-market	Intangibles
45	rna	Quarterly Return on Net Operating Assets, Quarterly Asset Turnover	Profitability
46	Roa1	Return on Assets	Profitability
47	roe	Return on Equity	Profitability
48	rsup	Revenue surprise	Momentum
49	rvar_capm	Residual variance - CAPM rolling 3m	Volatility
50	svar	Return variance rolling 3m	Volatility
51	seas1a	Seasonality	Intangibles
52	sgr	Sales growth	Value
53	sp	Sales-to-price	Value
54	std_dolvol	Std of dollar trading volume rolling 3m	Volatility
55	std_turn	Std. of Share turnover rolling 3m	Volatility
56	sue	Unexpected quarterly earnings	Momentum
57	turn	Shares turnover	Liquidity
58	zerotrade	Number of zero-trading days rolling 3m	Liquidity