Mosaics of Predictability*

Lin William Cong   Guanhao Feng   Jingyu He   Yuanzhi Wang

First draft: Feb 2024; this draft: April 2025

Abstract

We postulate that return predictability is an intrinsic and time-varying asset characteristic potentially related to the cross section of expected returns, instead of just an attribute of the chosen predictors or models. We develop a tree-based clustering method to gauge heterogeneous return predictability by grouping a panel of asset returns using high-dimensional asset characteristics and market-wide predictors. Our approach tells what types of assets exhibit greater return predictability under what market conditions, and empirically reveals substantial predictability heterogeneity in the U.S. equity market. Stocks with high earnings surprises, high earnings-to-price ratios, and low trading volumes exhibit the strongest predictability; predictability diminishes sharply with low market dividend yield but peaks with high dividend yield and low market liquidity. Out-of-sample, a new anomaly linked to investors' model misspecification easily generates monthly excess alphas exceeding 1%, and investing in highly predictable clusters significantly outperforms conventional benchmarks with Sharpe ratios approaching 2.

Keywords: Clustering, Generative AI, Grouped Heterogeneity, Panel Tree, Regime Switching, Return Anomalies.

JEL Classification: G12, C38, C53, C55.

*We are grateful to Chunrong Ai, Doron Avramov, Daniele Bianchi, Pietro Bini, Jian Chen (Discussant), Qihui Chen, Lauren Cohen, Victor DeMiguel, Fuwei Jiang, Zhanye Luo, Hao Ma, Semyon Malamud, Kuntara Pukthuanthong, David Rapach (Discussant), Michael Weber, Dacheng Xiu, Lei Yang (Discussant), Yi Zhang, Geoffery Zheng, and the seminar and conference participants at 2025 AFA Annual Meeting, Cornell Tech, CUHK Shenzhen, NYU Shanghai, 2024 EFMA Annual Meeting, 2024 FMA Asia/Pacific Conference, 2024 FinEML Conference, HKUST Guangzhou, Queen Mary University of London, 4th Big Data and Econometric Conference at Xiamen University, 2024 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 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 has long been a central question in finance. Previous studies have documented return predictability in many asset classes.[1] They primarily treat predictability as an attribute of the chosen predictors or models, being time-invariant and uniformly applicable to all asset returns analyzed in each given setting.[2] However, similar to other asset characteristics such as beta and volatility, return predictability is unobservable. It often lacks a clear, consistent definition.[3] More importantly, predictability can be time-varying and heterogeneous across assets, which extant homogeneous models fail to capture.

These challenges motivate a novel framework to systematically examine which types of assets exhibit greater return predictability and how this predictability varies across market regimes. We argue that, beyond predictor significance or predictive model fitness, return predictability is an intrinsic asset characteristic potentially linked to the cross-section of expected returns. Given this, we propose a clustering approach that groups asset-return observations using high-dimensional asset characteristics and market predictors and measures heterogeneous return predictability.

Our findings from the U.S. equity market indicate that return predictability is heterogeneous, varying significantly in the cross-section (partitions based on firm characteristics) and over time (regimes identified via market predictors and structural breaks), thus exhibiting rich "mosaic" patterns on the panel. Moreover, the observed predictability disagreement between heterogeneous and homogeneous measures sug-

[1]The empirical literature has documented return predictability in the aggregate market (Lewellen, 2004; Ang and Bekaert, 2007; Campbell and Thompson, 2008), individual stocks (Fama and French, 2008; Rapach et al., 2013; Lewellen, 2015), corporate bonds (Lin et al., 2018; Feng et al., 2025), treasury bonds (Bianchi et al., 2021; Huang and Shi, 2023), options (Bali et al., 2023), and mutual fund alphas (Kaniel et al., 2023; DeMiguel et al., 2023).

[2]Research has long linked stock market return predictability to business cycle indicators (e.g., Fama and French, 1988, 1989). For predicting cross-sectional returns, researchers have also identified a large number of firm-specific characteristics (e.g., size, value, and momentum from Fama and French, 1992; Jegadeesh and Titman, 1993). Various methods have been proposed to study return predictability, including predictive regressions (e.g., Stambaugh, 1999), security sorting (e.g., Jensen et al., 1972), cross-sectional regression models (e.g., Kelly and Pruitt, 2013; Han et al., 2024), and machine learning approaches (e.g., Gu et al., 2020; Freyberger et al., 2020; Kelly et al., 2024).

[3]Concepts related to predictability include anomaly average return, predictor significance, out-of-sample R^2, forecast-implied portfolio, etc.

gests potential risks in model misspecification resulting from neglecting predictive model heterogeneity. Most investors' reliance on homogeneous models affects how information gets impounded into asset prices and, thus, the cross-section of expected returns. This gives rise to new predictability-based anomalies and profitable strategies by investing in stocks with higher return predictability or greater predictability disagreement between heterogeneous and homogeneous models.

To our knowledge, this paper is the first to systematically measure and analyze the heterogeneous predictability of asset returns in a frame that is both economics-guided and data-driven. We introduce the concept of *mosaics of predictability*, where return predictability exhibits cross-sectional and time-series variation, forming patterns reminiscent of a mosaic. Following the panel tree (P-Tree) of Cong, Feng, He, and He (2025) and Cong, Feng, He, and Li (2023), the panel of asset returns can be analyzed as a goal-oriented clustering problem using a "divide-and-conquer" approach, partitioning them into sub-panels. Accordingly, we propose a customized P-Tree clustering algorithm that categorizes asset-return observations by predictability level, effectively distinguishing between highly predictable and less predictable data, and offering a cluster-specific measure of predictability.

To this end, we first measure the cluster-specific return predictability using the R^2 of a cluster-wise best-fitted predictive model. R^2, commonly used to assess model fitness, is a useful proxy for the signal-to-noise (S2N) ratio. When R^2 is low, even for a correctly specified model with fine-tuned hyperparameters, the data exhibit low predictability, indicating a weak signal relative to noise. Empirical studies typically estimate a pooled model, yielding an aggregate R^2 for all stock-return observations, which overlooks potential heterogeneity. However, conducting asset-specific regressions to obtain R_i^2 is often inefficient due to limited observations. We strike a balance by endogenously clustering stock returns into groups based on varying levels of predictability, then estimating a cluster-wise model to capture heterogeneous cluster-specific R^2.

Using this intuitive measure, we construct a generative P-Tree tailored to uncover the optimal clustering structure that captures cross-sectional grouped heterogeneity in predictability and regime-dependent time variation. The model groups stock-return observations with similar predictability levels, partitioning observations by maximizing the difference in R^2 across clusters using cluster-wise heterogeneous predictive models. These clusters then lead to mosaics on the panel, described by firm characteristics and market predictors for interpretable visualization. All clusters are explicitly defined by characteristics for cross-sectional partitions or market predictors for regime detections. It maintains economic interpretability, highlighting key variables associated with different predictability levels and capturing the complex, asymmetric interactions between market predictors and firm characteristics.

We conduct the empirical analysis on the panel of U.S. monthly individual stock returns from 1973 to 2022, incorporating 51 firm characteristics and eight market predictors. Our findings highlight substantial heterogeneity in the cross-sectional predictability, identifying 24 disjoint characteristics-based clusters. Stocks with high earnings surprises, high earnings-to-price ratios, low trading volumes, and high sales-to-price ratios exhibit the strongest return predictability, with R^2 exceeding 10%. These stocks belong to the long legs of the anomalies for earnings surprises (post-earnings announcement drift, Bernard and Thomas, 1989) and earnings (Reinganum, 1981), both showing persistent abnormal returns. Conversely, stocks with low earnings surprises but high bid-ask spread and momentum display no predictability, indicated by negative R^2 values, suggesting the return forecast is worse than the naive zero forecast. This contrasts with a 1.49% in-sample R^2 for return predictability under a homogeneous model. The out-of-sample (OOS) results confirm the persistent predictability gap between highly and less predictable clusters, underscoring the robust link between latent return predictability and high-dimensional firm characteristics.

Following the methodology for identifying grouped heterogeneity in Cong et al. (2023), especially the structural break analysis, we extend the analysis of U.S. equities to capture regime shifts and structural breaks. We identify two primary drivers of time-varying return predictability: market dividend yield (Fama and French, 1988) and market liquidity (Pástor and Stambaugh, 2003). Stock return predictability varies across business cycles, diminishing at low market dividend yield but strengthening during high market dividend yield periods coupled with low market liquidity, typically seen in recessions. This aligns with evidence on time-varying market return predictability (e.g., see Henkel et al., 2011; Dangl and Halling, 2012). Predictability peaks at its highest when the market dividend yield is high and market liquidity is low, indicating early recessions. The forward-looking nature of the stock market suggests that during recessions, the increase in dividend yield is mainly due to declining prices rather than rising dividends. There is a higher probability of a stock price recovery following stock fire sales during the early recession. We interpret this result by offering a theoretical explanation for the increased return predictability using the present value framework from Campbell and Shiller (1988). We also estimate a structural break model within our framework for a robustness check, finding consistent evidence of increased return predictability during recessions.

Finally, given the cross-sectional grouped heterogeneity, the most predictable cluster achieves a 3.41% monthly average return and a 1.88 annualized Sharpe ratio, while the least predictable cluster performs poorly with a -0.06% return and a -0.02 Sharpe ratio. These findings, robust to considering transaction costs, establish a strong connection between return predictability and investment outcomes, despite the null findings in earlier attempts leaving out grouped heterogeneity (e.g., Rapach et al., 2010; Kelly et al., 2024).

We also document a positive link between predictability disagreement and cross-sectional expected returns, likely driven by model misspecification risk. Ranking clusters by predictability disagreement between cluster-specific heterogeneous and global homogeneous models produces long-short portfolios with unexplained monthly alphas exceeding 1% across various benchmark factor models over the past two decades. Moreover, our approach generates cluster-wise predictive models that consistently outperform the global homogeneous model, with the most predictable clusters achieving an OOS annualized Sharpe ratio of about 1.86. These results point to not only the inefficiencies of naive homogeneous approaches in prior studies (e.g., Feng and He, 2022; Evgeniou et al., 2023) but also promising new directions for practitioners.

Literature. Our study contributes to the extensive literature on return predictability as the first to systematically examine its heterogeneity. Previous research documents heterogeneous return predictability ad hoc or only for specific cases. For example, Avramov (2002), Green et al. (2017), and Avramov et al. (2023) document that return predictability is concentrated in micro-cap stocks, distressed stocks, and during periods of high market volatility. Earlier research shows that stock return predictability is linked to business cycles (e.g., Fama and French, 1988, 1989), and models with regime-switching or time-varying coefficients indicate stronger predictability during recessions (e.g., Henkel et al., 2011; Dangl and Halling, 2012). Lewellen (2004) examines the predictive power of aggregate market predictors across sub-samples, finding time-varying predictability in some valuation ratios, like the book-to-market and earnings-price ratio. Recent studies have progressed in detecting time-series variation and structural breaks in forecasting market returns, identifying pockets of predictability (e.g., Farmer et al., 2023, 2024; Cakici et al., 2024). However, Cakici et al. (2025) highlight the sensitivity of these findings and demonstrate a decline in return predictability over time. Our findings better rationalize the instability and inconsistencies of prior empirical findings in out-of-sample evaluations (e.g., Pesaran and Timmermann, 1995; Lettau and Van Nieuwerburgh, 2008; Welch and Goyal, 2008; Li et al., 2025).

We contribute to the literature conceptually by treating return predictability as a time-varying characteristic of individual assets. We thus go beyond traditional time-varying coefficient models by integrating high-dimensional firm characteristics and market predictors to endogenously find clusters in both the cross section and time series. Closely related to our study, Feng and He (2022) show that incorporating cross-asset information in return prediction models significantly enhances informational efficiency via a Bayesian hierarchical framework. Instead of focusing on exogenously given portfolios as they do, our key innovation lies in the seamless combination of data-driven clustering (which endogenizes asset group formation) and cluster-wise model fitting within a unified framework.

In terms of methodological innovations, along with Cong et al. (2025) and Cong et al. (2023), we are among the first to develop economically guided P-Trees, adapting the efficient algorithm of He and Hahn (2023). As such, we add to the emerging AI literature on goal-oriented search/generative modeling by proposing a data-driven approach to optimizing an economic goal in a large, flexible modeling space (e.g., Cong et al., 2020; Feng et al., 2024; Cao et al., 2024; Didisheim et al., 2024; Cong, 2025). The divide-and-conquer approach in P-Trees emulates human problem-solving by iteratively partitioning observation panels to optimize given economic objectives while maintaining interpretability. Cong et al. (2025) generate test assets and construct the SDF by decomposing the cross section and forming leaf-based portfolios. We ask and answer a different question by constructing a customized P-Tree to partition stock return panels optimally based on heterogeneous levels of return predictability.

Closely related by analyzing endogenous grouped heterogeneity in stock return panels, Ahn et al. (2009) employ unsupervised clustering based on return correlations; Patton and Weller (2022) adopt K-means to group assets by within-group slopes and averages, uncovering significant risk-price heterogeneity. More recently, Evgeniou et al. (2023) utilize K-means to cluster firms by characteristics and estimate post-cluster heterogeneous predictive models. In contrast to these approaches, which assign stocks to fixed clusters across all periods, ours allows assets to transition among characteristics-based clusters over time. These transitions naturally capture the evolving relationship between characteristics and return predictability.

The remainder of the article is structured as follows: Section 2 introduces the tree-based clustering framework. Section 3 describes the data and evaluation procedures. Section 4 examines cross-sectional grouped heterogeneity in predictability and the anomaly based on predictability disagreement. Section 5 explores time-varying predictability and market regime shifts. Section 6 evaluates investment performance gains from our cluster-based models. Section 7 concludes, and the appendices contain algorithms, data descriptions, and supplementary results. The internet appendix includes robustness checks and additional empirical discussions.

2 A P-Tree Framework for Analyzing Heterogeneous Predictability

Our framework aims to generate (i) data-driven clusters separating asset-return observations based on their predictability, which are described by a P-Tree structure, (ii) cluster-specific heterogeneous predictive models, and (iii) a measure of return predictability for each cluster. We begin by motivating the predictability metrics and detailing our heterogeneous prediction and clustering methodologies.

2.1 Measuring Return Predictability

As noted earlier, return predictability is unobservable and lacks a unified definition in the literature. We argue that the in-sample R^2 of a predictive model serves as a natural metric for predictability. Traditionally, statisticians interpret R^2 as the proportion of return variation a specific model explains, emphasizing its model-centric nature. Here, we take a data-centric view, where R^2 measures the machine learning (ML) model's return forecast outperformance over a naive zero-return forecast.

Partitioning the stock universe into clusters allows for fitting predictive models with optimized parameters, enhancing performance within each cluster. The R^2 values may vary across clusters, reflecting the differential challenges in forecasting returns. A lower R^2 intuitively suggests less predictability in returns. Let us revisit the calculation of R^2. A predictive model is typically represented as:

$$r_{i,t} = \mathbb{E}_{t-1}[r_{i,t}] + \varepsilon_{i,t}, \tag{1}$$

with the assumption $\mathbb{E}[\varepsilon_{i,t}] = 0$ such that the prediction of expected return is unbiased. The information regarding the heterogeneous predictable difficulty can be represented by the S2N ratio in $R^2_{i,t} = 1 - \mathbb{E}(r_{i,t} - \widehat{r}_{i,t})^2 / \mathbb{E}(r_{i,t} - \bar{r})^2$, where $\widehat{r}_{i,t}$ is the best return forecast and $\bar{r}$ is the average return for all stock i in all periods t. Intuitively, a high $R^2_{i,t}$ indicates that $\widehat{r}_{i,t}$ can capture the conditional expectation $\mathbb{E}_{t-1}[r_{i,t}]$ over the unconditional expectation $\bar{r}$. Conversely, if the noise variation $\mathbb{E}(r_{i,t} - \bar{r})^2$ is high, even knowing the true $\mathbb{E}_{t-1}[r_{i,t}]$ would yield a low $R^2_{i,t}$, let alone the estimation from noisy data. Therefore, in-sample $R^2_{i,t}$ offers a reasonable proxy for the S2N ratio, capturing return predictability across assets and periods. For a given dataset, adjusting model specifications or tuning hyperparameters may improve the R^2; however, such improvements are bounded. The R^2 of the best-fitted ML model with properly tuned hyperparameters reflects the intrinsic difficulty of return prediction for that dataset.

However, estimating $R^2_{i,t}$ for each asset i at time t is infeasible. As a compromise, the literature often imposes implicit homogeneity assumptions. For example, studies such as Gu et al. (2020); Freyberger et al. (2020) estimate a pooled model to compute a single R^2 across all assets and time periods, thereby assuming that return predictability is homogeneous both cross-sectionally and temporally. Our measurement strikes a balance by partitioning asset return observations into disjoint clusters, calculating a cluster-specific R^2 to reflect cluster-level predictability. The process, as detailed later, is designed to group asset-return observations with similar predictability levels into clusters and maximize predictability differences between clusters.

Why not use OOS R^2 (R^2_{OOS}) to guide the clustering procedure? The primary reason is that our clustering method relies on this measure to determine the clustering structure of asset returns. Incorporating the OOS data would introduce test sample information into the clustering process, resulting in data-snooping bias. Instead, we treat the clustering approach and the cluster-wise predictive model as a whole, determined using in-sample data. The performance of this system is then independently evaluated using the OOS data.

Finally, once our clustering system is fitted, we evaluate its OOS performance by examining the detected clustering patterns and the fitted cluster-wise predictive models. This standard OOS evaluation mitigates concerns about data snooping and further validates the effectiveness of our clustering algorithm.

2.2 Cluster-Wise Predictive Modeling

Clustering and predictive modeling are essential tasks in ML, typically addressed using separate models. Our methodology integrates these tasks within a unified framework: a single P-Tree is tailored for model-based clustering, while ML models (e.g., Ridge regression) are employed for prediction. In this subsection, we assume the clusters are predetermined and concentrate on cluster-wise analysis.

We denote the data as $\mathcal{D} = \{(r_{i,t}, \mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}); i = 1, \ldots, N; t = 1, \ldots, T_i\}$, where $r_{i,t}$ represents the excess return of stock i at time t. Predictors commonly used in the stock return prediction literature include $\mathbf{z}_{i,t-1}$, a C-dimensional vector of firm characteristics, and $\mathbf{x}_{t-1}$, an M-dimensional vector of market predictors.

Ideally, heterogeneous and time-varying expected excess returns are modeled as: $\mathbb{E}_{t-1}[r_{i,t}] = g_{i,t}(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$, which is a function that varies across assets and time periods. However, limited observations for individual asset returns make estimating each $g_{i,t}(\cdot)$ separately infeasible. Consequently, many studies (e.g., Gu et al., 2020) adopt a homogeneous predictive function, $g_t(\cdot)$, with time-varying coefficients updated via an expanding/rolling-window scheme. Yet, Feng and He (2022) and Evgeniou et al. (2023) highlight that this approach neglects heterogeneity in predictive power across assets. Moreover, it assumes uniform return predictability, or identical R^2 across stocks, which is contrary to empirical evidence (e.g., Avramov et al., 2023).

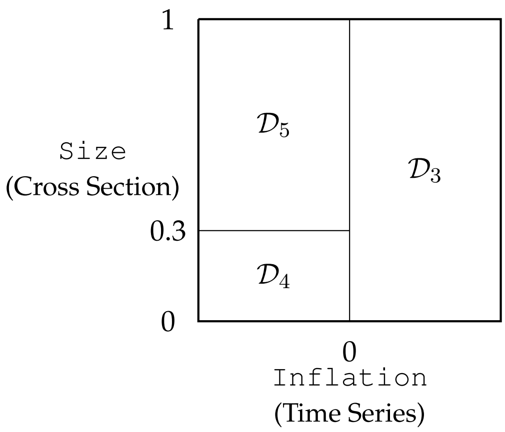
We propose a cluster-wise predictive modeling approach that effectively balances individual and pooled modeling by fitting group-specific models for endogenous clustered groups within a unified framework. Unlike the two-step clustering and estimation approach in Evgeniou et al. (2023), we adopt the P-Tree framework introduced by Cong et al. (2023), customizing it to cluster assets based on return predictability. This clustering method partitions the stock return panel into multiple clusters (leaf nodes) through characteristics for cross-sectional and market predictors for time-series dimensions. Figure 1 illustrates an example output of our approach, where the panel of stock returns is split into three non-overlapping clusters, described by characteristics or market predictors.[4]

Rather than fitting individual predictive models, $g_{i,t}(\cdot)$, for each asset i in period t, or a single pooled model, we estimate a small number of cluster-wise homogeneous

[4] A company's cluster membership may evolve as its characteristic values change. For example, if a company transitions from a small-cap to a large-cap firm, as reflected by market equity values, its cluster assignment may shift based on the partitioning outcome: $\mathcal{D}_3$: high inflation; $\mathcal{D}_4$: low inflation and small-cap; $\mathcal{D}_5$: low inflation and non-small-cap.

Figure 1: **Clustering Illustration via Partitions**

This figure separates the whole panel of stock returns into three rectangular $\mathcal{D}_3$, $\mathcal{D}_4$, and $\mathcal{D}_5$. The first partition is inflation at 0, and the second is size at 0.3 when inflation is low.



models that vary across clusters. As a result, the cluster-wise predictive model is:

$$\mathbb{E}_{t-1}[r_{i,t}] = g_j(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}), \tag{2}$$

where stock-return observations in the j-th cluster follow the same predictive model $g_j(\cdot)$. Our approach simultaneously clusters observations and estimates local predictive models, grouping stock-return observations with similar return predictability into the same cluster. This shares the spirit of Cong et al. (2023) but contrasts with the two-step approach of Evgeniou et al. (2023), which separates based on firm IDs $\{r_i \in j\}$.

Remarkably, our clustering method offers flexibility in choosing the predictive model $g_j(\cdot)$. For illustration, we use Ridge regression for the ML representative model, noted for its robustness in weak signal scenarios by Shen and Xiu (2024). Thus, we compute the in-sample R_j^2 using stock returns from the j-th cluster. We proceed to detail our tree-based clustering method step-by-step.

2.3 P-Tree Origins: The First Split

Our clustering approach aims to group asset-return observations with similar predictability levels into clusters. As outlined earlier, predictability is quantified using the in-sample R^2 of a cluster-wise predictive model. Standard clustering techniques, such as K-means, are unsuitable for this purpose. These methods typically aim to minimize within-cluster variation, which is measured by the distance between data points in a high-dimensional space. However, K-means does not account for predictive models or provide the interpretability of resulting clusters. Our tree-based approach instead explicitly incorporates a cluster-wise predictive model and ensures that the resulting clusters are interpretable.

We propose to partition the panel of stock returns based on unobservable predictability, distinguishing between more and less predictable observations. This method optimizes the partition of samples into two subgroups by maximizing the difference in their R^2 values.[5] Our approach is implemented iteratively to partition the entire panel of stock-return observations, adding one cluster at a time. The results are visualized using a P-Tree structure. Unlike the Random Forest algorithm, which constructs multiple trees for prediction only, our method uses conditioning variables to iteratively partition the sample, resulting in a deterministic clustering output.

The initial step tries to find the first split point (split predictor and cut-point value) to divide the data into two sets. This process includes fitting cluster-wise predictive models and calculating the absolute R^2 difference for each split candidate, ultimately selecting the one that maximizes the predictability difference.

Figure 2 illustrates a *candidate* for the first split in a tree, where the root node, $\mathcal{D}_1$, containing the entire dataset, is divided into two clusters, or *leaf nodes* $\mathcal{D}_2$ and $\mathcal{D}_3$. Following the terminology in Cong et al. (2025), leaf nodes refer to the nodes at the bottom of the tree without subsequent branches. The split is based on the rule "$\text{var}_p(\texttt{Inflation}) \leq c_k(0)$," where the p-th variable (firm characteristic, aggregate predictor, or calendar month) is inflation, and the k-th cut-point value is 0.[6] Observations of asset returns that satisfy this split rule are assigned to leaf node $\mathcal{D}_2$, while those that do not are placed in $\mathcal{D}_3$. To assess the quality of this split candidate, we propose a goal-oriented criterion that gauges its ability to separate returns with high versus low predictability.

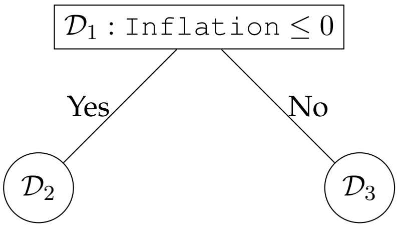
The first split candidate divides the entire return sample $\mathcal{D}_1$ into two clusters, $\mathcal{D}_2$

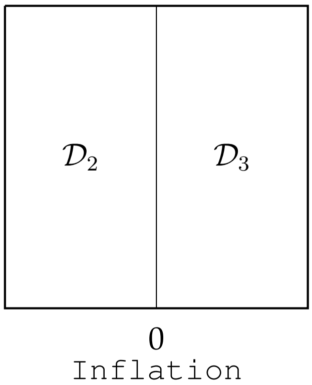
[5]Alternative clustering objectives achieve similar goals, but are often less interpretable and practical.

[6]Demeaning macro variable compared with 0 is equivalent to examine the raw macro condition corresponding to rolling 10-year average. See details in Section 3.

Figure 2: **Illustration for the First Split**

This figure illustrates one of the first split candidates, such as `Inflation`≤ 0. The left figure shows a panel tree that divides the sample, and the right figure shows the corresponding partition plot that only partitions over time.





and $\mathcal{D}_3$. For each cluster, we fit a separate predictive model using the training data specific to that cluster, denoted as $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$, respectively. Generally, for the j-th leaf node, we fit a specific predictive model $g_j(\cdot)$ in Eq. (3), and return forecasts are denoted as $\widehat{r}_{i,t} = \widehat{g}_j(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$, where $\mathbf{z}_{i,t-1}$ and $\mathbf{x}_{t-1}$ are lagged equity characteristics and aggregate macro predictors, respectively.

Notably, Fama and French (2008) criticize the dominance of small-cap stocks with high return variance in panel regressions, and Hou et al. (2020) highlight that many anomalies are driven by micro-caps (approximately 60% of firms) in cross-sectional regressions. To mitigate the influence of small-cap stocks, we employ volatility-weighted Ridge regression for cluster-wise predictive regressions and split criterion calculations:

$$g_j(\cdot) = \beta_0 + \beta^\intercal \mathbf{s}_{i,t-1} + \varepsilon_{i,t}, \tag{3}$$

where $\widehat{\beta}_j = \arg\min_{\beta_0,\beta} \left\{ \frac{1}{N_{\text{leaf}_j}} \sum_{\text{leaf}_j} w_{i,t-1} \left(r_{i,t} - \beta_0 - \beta^\intercal \mathbf{s}_{i,t-1} \right)^2 + \lambda ||\beta||_2^2 \right\}$ and $w_{i,t-1} = 1/\sigma^2_{i,t-1}$. In addition, $\mathbf{s}_{i,t-1} = \{\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}\}$ represents the set of lagged firm characteristics and market predictors, and $w_{i,t-1}$ is the inverse of idiosyncratic return variance. The volatility, $\sigma^2_{i,t-1}$, is estimated on a rolling-window basis, which helps to incorporate both cross-sectional and time-series variations for observation weights within the leaf cluster. The hyperparameter λ is determined through cross-validation.

Therefore, $\widehat{r}_{i,t} = \widehat{\beta}_{j,0} + \widehat{\beta}_j^\intercal \mathbf{s}_{i,t-1}$ is the heterogeneous return forecast for calculating the corresponding S2N ratios, R^2. Within the j-th leaf node:

$$R_j^2 = 1 - \frac{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t} - \widehat{r}_{i,t})^2}{\sum_{\{i,t\}\in \text{leaf}_j} r_{i,t}^2}. \tag{4}$$

The denominator differs from the cluster-wise observation variation in the standard R^2 definition. This evaluates returns by contrasting the ML return forecasts with a naive zero forecast. Since our goal is to separate returns with high predictability from those with lower predictability, it is natural to use the absolute value of the R^2 difference between the left and right clusters as the split criterion:

$$S_{\{\text{leaf}_l, \text{leaf}_r\}} \left(\text{var}_p, c_k\right) = \left| R^2_{\text{leaf}_l} - R^2_{\text{leaf}_r} \right|. \tag{5}$$

Intuitively, this criterion evaluates how effectively each split candidate differentiates the R^2 values between the two leaf nodes, regardless of which one is higher. A split candidate achieving a high value for this criterion indicates its success in partitioning stock returns into a highly predictable leaf node and a less predictable one.

We then evaluate the criterion in Eq. (5) across all split candidates, considering combinations of split variables and cut-point values. Each candidate pair, $\{\text{var}_p, c_k\}$ generates a different partition of the data into $\mathcal{D}_2$ and $\mathcal{D}_3$, resulting in distinct cluster-wise predictive models $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$. These, in turn, yield different values for the split criterion in Eq. (5). We select the candidate that maximizes this criterion as the optimal split. Given P candidate variables and K cut-point values, there are $P \times K$ possible splits for the initial partition. Each is evaluated to identify the optimal first split, growing the root node into two child clusters.

2.4 Subsequent Splits and Stopping Criteria

We essentially take a sequential approach to partition the panel into multiple clusters. Once the first split is determined, two leaf nodes are created, and subsequent splits can occur at either node to further split the clusters.

Specifically, the second split can occur within the left leaf node $\mathcal{D}_2$, partitioning it into $\mathcal{D}_4$ and $\mathcal{D}_5$, or within the right leaf node $\mathcal{D}_3$, dividing it into $\mathcal{D}_6$ and $\mathcal{D}_7$. Figure A.1 illustrates two potential candidates for the second split. Each side still evaluates $P \times K$ split candidates, resulting in $2 \times P \times K$ combinations. When evaluating split candidates for $\mathcal{D}_2$, cluster-wise predictive models $\widehat{g}_4(\cdot)$ and $\widehat{g}_5(\cdot)$ are fitted using the observations in $\mathcal{D}_4$ and $\mathcal{D}_5$, respectively, and the split criterion values are calculated. A similar procedure is applied for candidate splits within $\mathcal{D}_3$. Among all split candidates across $\mathcal{D}_2$ and $\mathcal{D}_3$, the one that maximizes the split criterion is selected as the second split. This procedure employs a local-global framework: it assesses the benefits of splitting each leaf node using local R^2 variations across sub-samples and globally selects the split that optimally enhances return predictability.

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

Stopping criteria. Stopping criteria are essential for regularizing in-sample model training, preventing overfitting. The clustering process stops when predetermined conditions are satisfied. We impose a minimum sample size requirement for each leaf cluster, eliminating split candidates whose resulting sub-samples fail to meet this threshold. This ensures the cluster-wise predictive model in each cluster can be fitted with sufficient observations. Additionally, we limit the tree structure's maximum depth and the number of terminal leaves to control its complexity. Finally, a node is no longer split if all candidates fail to improve predictability, specifically when the R^2 values of both child nodes are smaller than that of their parent node.

Cluster-wise forecasts and measurement of predictability. Once tree growth terminates, observations are partitioned into non-overlapping clusters based on interactions among variables, such as equity characteristics, market predictors, or calendar months. Within each cluster, we refit models using the corresponding data. As detailed in Section 2.2, Ridge regression is employed for the ML return forecasts. The final output of our methodology is: (i) a hierarchical tree structure illustrating the clustering pattern; (ii) cluster-wise ML models, yielding a heterogeneous predictive framework instead of a single global ML model applied to all stock-return observations; and (iii) the associated R_j^2 of the best-fitting model for each cluster, quantifying the predictability of stock returns within clusters and varying across them.

Notably, while we demonstrate our clustering approach using Ridge regression in Eq. (3) for simplicity, the framework is not restricted to any specific predictive model. Instead, it provides a flexible and generalizable framework for heterogeneous predictions that can incorporate a wide range of ML models.

3 Data and Empirical Evaluations

We apply our approach to the U.S. equity market to measure and analyze the heterogeneous predictability of individual stock returns.

3.1 Data and Variables

Data sample. The monthly sample spans from 1973 to 2022, with the first 30 years used for model training and the most recent 20 years for the OOS analysis.[7] We apply standard filters: (1) stocks listed on NYSE, AMEX, or NASDAQ for over one year; (2) firms with CRSP share codes 10 and 11; and (3) exclusion of stocks with negative book equity or lagged market equity. The average and median number of stocks in the training sample are 4,840 and 4,772, respectively; in the test sample, 3,909 and 3,694. Our algorithm accommodates unbalanced panel data.

Market predictors. Following Welch and Goyal (2008), we analyze eight market predictors to define and select market regimes characterized by time-varying return predictability. As detailed in Table A.1, these predictors include the 3-month Treasury bill rate, inflation, term spread, default yield, and market-level characteristics such as dividend yield, volatility, net equity issues, and liquidity.

To ensure comparability of these predictors over time, we standardize them by subtracting their respective 10-year rolling window averages. Additionally, we apply

[7]We begin in 1973 as CRSP expanded its data coverage in 1987 to include NASDAQ daily and monthly stock data, with information on domestic common stocks and ADRs traded on the NASDAQ Stock Market starting December 14, 1972.

a 12-month weighted moving average, assigning greater weights to more recent observations, to further smooth the data and mitigate volatility. A positive value of a macro indicator indicates that its current level exceeds its 10-year average, signaling a high regime in the economy. This standardized approach enhances the effectiveness of the tree clustering method in detecting market regimes based on recent data.

Characteristics. As detailed in Table A.2, our dataset comprises 51 firm-level characteristics categorized into eight major categories: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. Each characteristic is standardized cross-sectionally and scaled uniformly to the range [0, 1] for every month. To mimic the "top-middle-bottom" sorting approach, we define two cut-point values, 0.3 and 0.7, as split-value candidates for each characteristic. These characteristics are utilized for forecasts to construct tree-based clustering and cluster-wise predictive models.

3.2 Training and Evaluation

Model fitting design. The baseline analysis for cross-sectional partitions uses the first 30 years of data for tree-based clustering and model estimation, with the most recent 20 years as testing samples. To enhance the predictive performance, we update the tree-based clustering and cluster-wise models every five years, employing a 30-year rolling window of in-sample data to retrain the P-Tree structure. This process is repeated four times over the 20-year OOS data.

We perform a full-sample analysis complemented by extended investigations employing time-series splits to accommodate the prolonged and overlapping nature of business cycles. Furthermore, we implement cross-validation to optimize hyperparameters during the post-clustering predictive model training phase.

Performance evaluation. As mentioned in Section 2.1, we use in-sample R^2 to measure predictability. In addition to the in-sample R^2, we evaluate the R^2_{OOS} to check whether the predictability gap is consistent. This approach aligns with the standard practice in recent studies. Following Gu et al. (2020), we define the R^2_{OOS} using zero forecasts as the benchmark,

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

where subscript j represents the specific OOS predictions by the related in-sample cluster-wise predictive model of the j-th cluster.

4 Cross-Sectional Heterogeneity in Return Predictability

Our baseline setup investigates cross-sectional grouped heterogeneity in return predictability, focusing on the question: Does such heterogeneity exist, and what types of stock returns exhibit greater predictability? We refer to the clustering model in this section, which focuses solely on the cross-sectional partitions, as "CS clusters" to distinguish it from the "TS + CS clusters" model discussed in Section 5, which integrates regime-dependent time variations with cross-sectional partitions.

4.1 Cluster and Heterogeneous Predictability

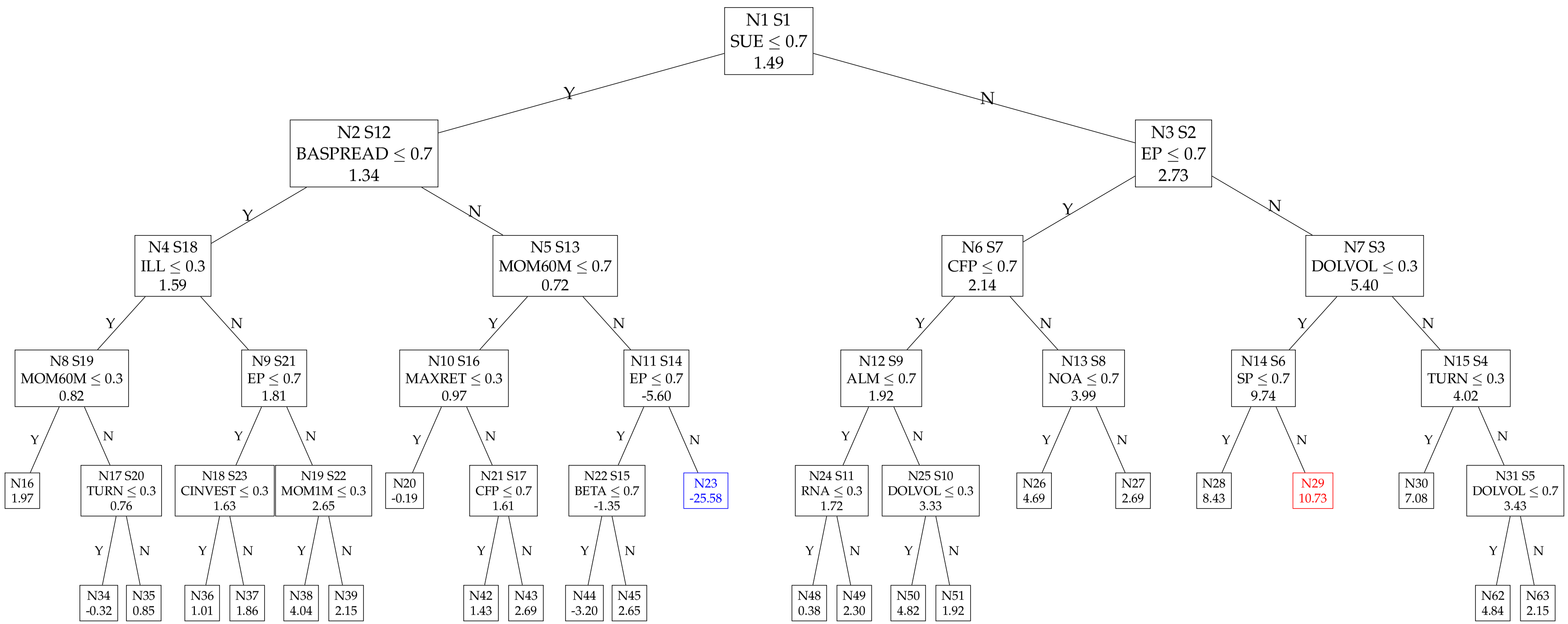
We present cross-sectional clustering results in Figure 3, based on monthly stock returns from 1973 to 2002. Following our predetermined stopping criteria, this tree stops growing after reaching 24 terminal leaves.[8] Each leaf node contains two or three rows of information. The top row identifies the leaf number and split order. For instance, the root node, representing the initial split, is labeled as leaf node N1 and split S1. The second row specifies the optimal split rule, directing stock-return observations satisfying the condition to the left child node, while others proceed to the right. Terminal leaves, which undergo no further splits, are denoted only by their node index in the first row. The bottom row reports the R^2 value, capturing cluster-wise return predictability at each node. This structure offers a clear visualization of clustering processes and interactions among characteristics.

The aggregate return predictability (R^2) of the homogeneous model is 1.49% (root node N1). Following the first split, the R^2 value improves markedly for stocks with high unexpected earnings ($\texttt{SUE} > 0.7$, top 30% by $\texttt{SUE}$), rising to 2.73% at N3, while

[8] A moderately deep P-Tree with large leaves balances robustness and interpretability. We cap the depth at 6 (up to 32 leaves) and impose a minimum leaf size of 30 monthly stock-return observations to ensure robust training. Under these settings, the algorithm completes in 23 splits.

Figure 3: **Tree-Based Cluster (CS Cluster)**

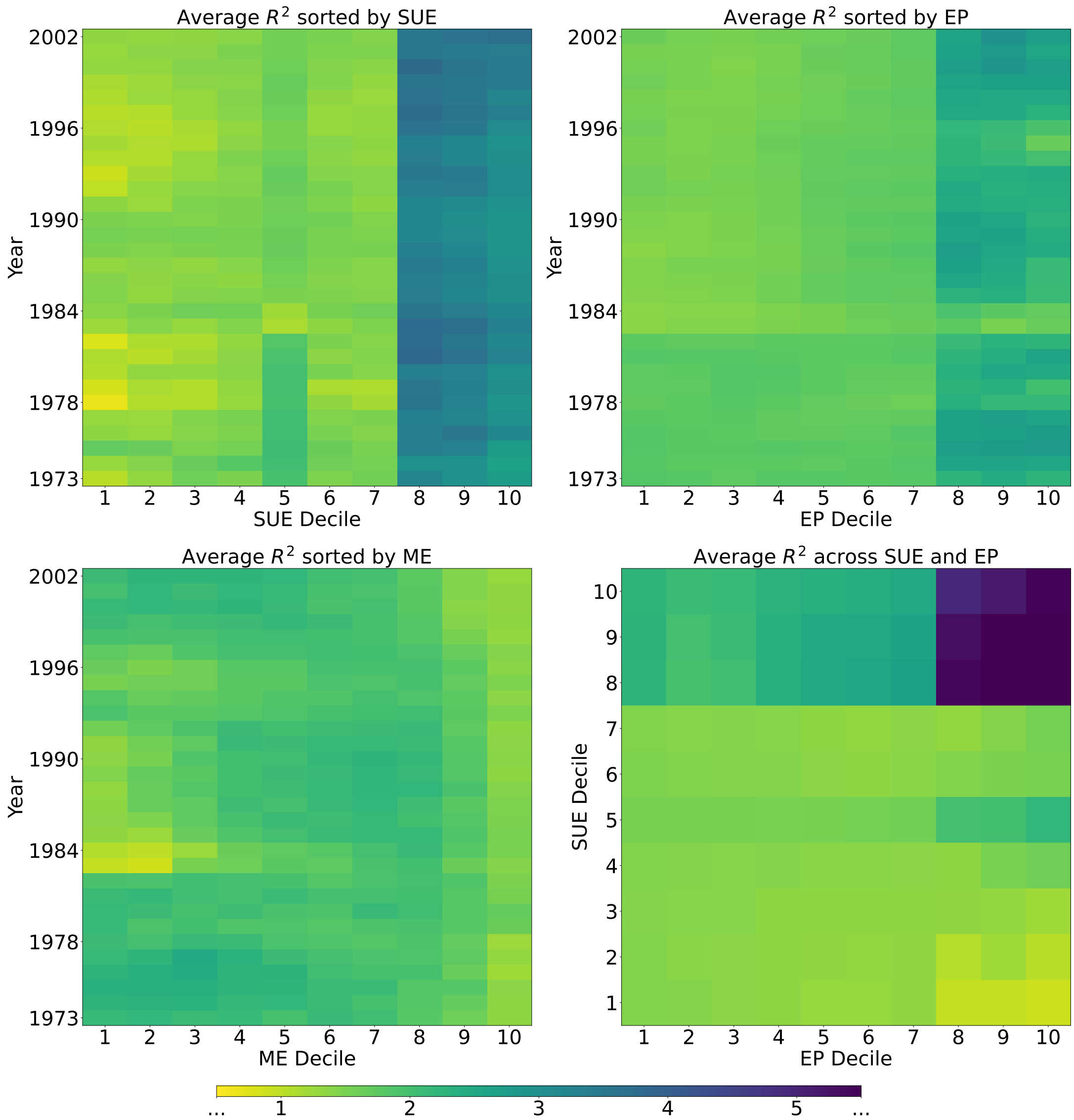
This figure illustrates the cross-sectional tree-based clustering derived from monthly data (1973–2002). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. Terminal leaves represent clusters defined by specific characteristic intervals. Each node, including intermediate and terminal leaves, is labeled with an ID (N#) where Ni-th node's child nodes have ID N($2i$) and N($2i+1$) respectively. S# indicates the sequence of splitting order. Cluster-specific R^2 values, indicating return predictability, are provided for each node.



declining slightly for the complement set, falling to 1.34% at N2. The R^2 difference (2.73% - 1.34%) represents the maximum value among all split candidates (51×2) identified by the algorithm. Higher earnings surprises are associated with stronger return predictability, aligning with the post-earnings announcement drift (e.g., Ball and Brown, 1968) that persists for weeks following earnings announcements.

Figure 4: **Mosaics of Predictability by Deciles**

These heat maps summarize return predictability (R^2 values, % in the color bar) for results presented in Figure 3. The first three heat maps show average R^2 values for groups sorted by years and deciles of earnings surprises, earnings-to-price, and market equity values. The fourth heat map displays average R^2 values for 10×10 groups formed by bivariate decile sorting of the top two characteristics. Colors range from light to dark, indicating increasing levels of return predictability.



The second split identifies high earnings-to-price ($\texttt{EP} > 0.7$) stocks, yielding a higher R^2 of 5.40% at N7 among high-SUE stocks. These high-SUE, high-EP stocks exhibit stronger return predictability, while others show weaker predictability ($R^2 = 2.14\%$ at N6). The maximum R^2 difference ($5.40\% - 2.14\%$) arises from an exhaustive search between cluster N2 ($\texttt{SUE} \leq 0.7$) and leaf N3 ($\texttt{SUE} > 0.7$).[9] Figure 4 presents marginal insights from the tree in Figure 3. The bottom-right panel highlights the interaction of these two characteristics (SUE and EP), with return predictability peaking in the top-right corner. These stocks belong to the long legs of the anomalies for earnings surprises (post-earnings announcement drift, Bernard and Thomas, 1989) and earnings (Reinganum, 1981), both showing persistent abnormal returns. These mosaic-like patterns emphasize the differences in return predictability arising from the interactions among various predictors.

Figure 4 also reports average decile R^2 values by year and decile clusters, sorted by the top two characteristics (SUE and EP). A clear pattern emerges: stocks with higher earnings surprises or earnings-to-price ratios exhibit stronger return predictability, which has persisted over time. The bottom-left panel shows that predictability differences across size-sorted clusters are modest, with small-cap stocks exhibiting slightly higher predictability.

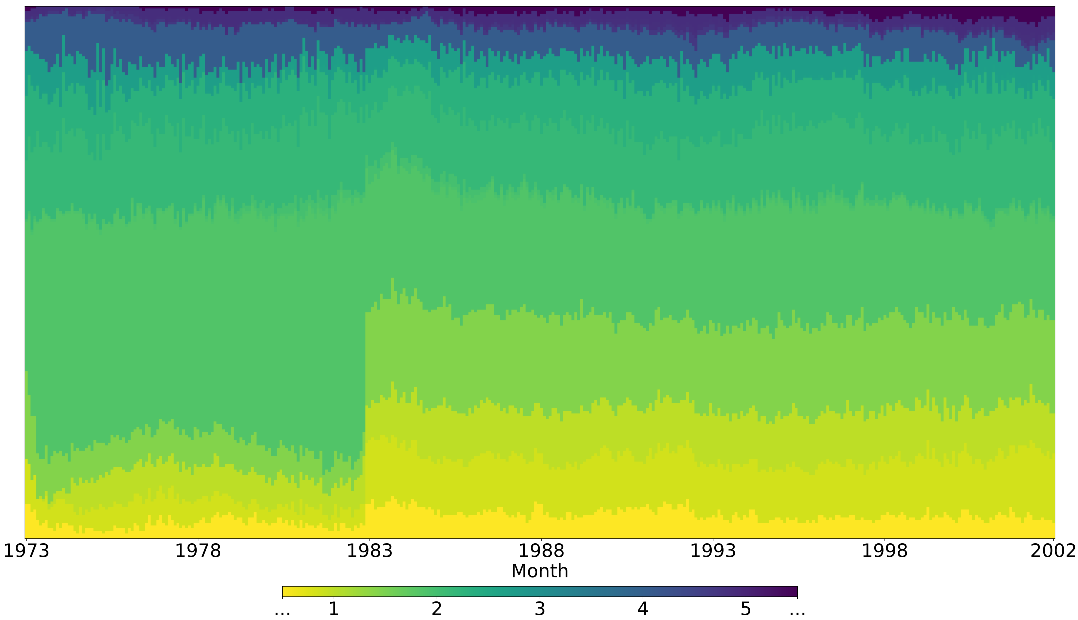
Figure 5 illustrates a mosaic-like visualization of return predictability based on tree-based clustering. The sample spans 360 months, aligned with the tree structure in Figure 3, and includes approximately 4,000 to 5,500 monthly observations. Each month, clusters are ordered by ascending predictability in the vertical bar, with segment length representing the proportion of stocks in each cluster.[10] Only a small proportion of stocks exhibit high predictability (top darker shades) over time, highlighting the difficulty in forecasting stock returns. Almost half of the observations cluster has near-zero R^2, represented by lighter shades. This horizontal, cascading 'mountain-like' pattern across the cross section, time series, and color dimensions underscores the heterogeneity in stock return predictability.

[9]The largest R^2 difference ($1.59\% - 0.72\%$) across all split candidates in cluster N2 is insufficiently pronounced, delaying its partition until later S12.

[10]The monthly proportions' denominators are derived from the total observations per month. The gap in 1982 reflects a surge in IPOs, coinciding with the April 1982 launch of the NASDAQ National Market, which featured larger, actively traded NASDAQ securities. This gap may also be related to the Volcker Shock, during which the Fed significantly raised interest rates, altering market dynamics.

Figure 5: **Mosaics of Predictability by Months**

This heat map summarizes monthly return predictability (R^2 values, % in the color bar) for results presented in Figure 3. The vertical axis indicates the proportion of observations per cluster, while the horizontal axis represents months. Colors transition from light (bottom) to dark (top) every month, indicating increasing levels of return predictability.



Furthermore, all clusters are determined by the cross-sectional standardized firm characteristics — the relative ranking of each firm among all stocks for each month. As a result, a stock can move from one cluster to another over time. We calculate the transition probability matrix to assess the persistence of such clustering patterns, counting the proportion of stocks moving from one cluster to all other clusters in the next month. Figure A.2 in Appendix IV.1 summarizes this transition probability matrix and illustrates that the clustering pattern is highly persistent over time.

4.2 Predictive Performance of Models with Grouped Heterogeneity

Cluster-wise predictive performance. Notably, clustering results in the previous subsection yield heterogeneous predictive models, which are the fitted Ridge regression model for each cluster. We now examine the performance gains of this heterogeneous model, comparing it to a homogeneous global model fitting Eq. (3) to all stock return observations, regardless of clusters.

Table 1 summarizes the results, with Panel A providing summary statistics for

Table 1: **Performance of Asset Clusters**

This table summarizes key metrics for each cluster from the cross-sectional tree structure in Figure 3. Panel A reports the number of observations (# obs), return predictability (R^2 in %) based on cluster-wise (R^2_{C}) and global (R^2_{G}) models, along with their differences ($R^2_{\mathrm{CMG}} = R^2_{\mathrm{C}} - R^2_{\mathrm{G}}$). Panel B presents the monthly average return (Avg, %) and annualized Sharpe ratio (SR) for equal- and value-weighted (EW/VW) portfolios. Leaf nodes are ordered by descending R^2_{C}.

	Panel A: Summary Statistics				Panel B: Profitability			
Leaf	# obs	R^2_{C}	R^2_{G}	R^2_{CMG}	$\mathrm{Avg}_{\mathrm{EW}}$	$\mathrm{SR}_{\mathrm{EW}}$	$\mathrm{Avg}_{\mathrm{VW}}$	$\mathrm{SR}_{\mathrm{VW}}$
N29	10,805	10.73	6.89	3.84	4.30	2.10	3.41	1.88
N28	11,917	8.43	6.38	2.05	3.04	1.79	2.45	1.54
N30	17,999	7.08	6.11	0.97	2.33	1.69	1.84	1.34
N62	29,415	4.84	4.55	0.29	2.34	1.35	1.82	1.14
N50	15,229	4.82	2.01	2.81	3.32	1.44	2.43	1.23
N26	14,714	4.69	3.52	1.17	3.05	1.59	2.30	1.16
N38	93,577	4.04	3.20	0.84	1.78	0.93	1.51	0.83
N27	11,525	2.69	2.55	0.14	1.51	0.91	0.91	0.58
N43	70,484	2.69	1.67	1.02	0.72	0.34	0.04	0.02
N45	11,101	2.65	1.54	1.12	-1.51	-0.56	-1.78	-0.63
N49	141,158	2.30	1.90	0.40	1.21	0.73	0.60	0.40
N63	31,443	2.15	1.44	0.71	1.27	0.76	1.05	0.66
N39	194,731	2.15	2.04	0.11	0.70	0.49	0.67	0.49
N16	17,204	1.97	1.19	0.78	0.90	0.48	0.99	0.54
N51	13,167	1.92	1.61	0.31	2.24	1.03	1.37	0.59
N37	443,144	1.86	1.56	0.30	0.30	0.16	0.17	0.10
N42	237,398	1.43	0.94	0.49	-0.34	-0.14	-0.65	-0.27
N36	140,194	1.01	0.92	0.08	0.12	0.06	0.11	0.06
N35	148,437	0.85	0.64	0.21	0.35	0.22	0.24	0.17
N48	31,799	0.38	0.62	-0.24	0.74	0.33	0.36	0.17
N20	20,560	-0.19	0.95	-1.14	0.45	0.31	-0.17	-0.09
N34	11,700	-0.32	0.40	-0.72	0.29	0.23	0.19	0.15
N44	13,697	-3.20	1.29	-4.49	-0.76	-0.41	-0.58	-0.28
N23	11,089	-25.58	1.92	-27.51	0.31	0.16	-0.06	-0.02

each cluster, including the number of observations and the cluster-specific R^2_{C} for the cluster-wise model. Additionally, we illustrate the R^2_{G} calculated by applying the global model (fitted on the entire dataset, irrespective of clusters) to predict within each cluster. Ordered by descending predictability of cluster-wise model forecasts, the table reveals substantial heterogeneity across clusters, with R^2_{C} values ranging from 10.73% to -25.58%, reflecting a difference exceeding 35%. The return predictability based on homogeneous models follows a similar downward trend (R^2_{G} values between 0.40% and 6.89%).

These cross-sectional variations in R^2 highlight that predictability varies significantly across clusters, highlighting the finding of characteristics-based stock return predictability heterogeneity. Notably, the bottom eight least predictable clusters (R^2 below the root node, 1.49%) represent around 35% of observations, while the top three most predictable clusters (R^2 exceeding 5%) comprise only about 2.3%. This sharp difference underscores that only a small fraction of stock returns are predictable.

We examine the link between return predictability and investment performance by constructing equal- and value-weighted portfolios for stocks within each cluster over time. Panel B of Table 1 reports monthly average returns and annualized Sharpe ratios for these portfolios. The most predictable cluster (N29, $R^2_{\mathrm{C}} = 10.73\%$) delivers an average return of 3.41% and an annualized Sharpe ratio of 1.88 for value-weighted strategies, highlighting strong profitability and a favorable risk-return tradeoff. In contrast, the least predictable cluster (N23, $R^2_{\mathrm{C}} = -25.58\%$) exhibits significantly weaker performance, with an average return of -0.06% and a Sharpe ratio of -0.02. These findings establish a strong connection between return predictability and investment outcomes, complementing Rapach et al. (2010) and Kelly et al. (2024) by exploiting cross-sectional grouped heterogeneity in return predictability. These findings also suggest the potential to identify predictability-related anomalies (see Section 4.3).

Out-of-sample robustness. Thus far, the easy- and hard-to-predict clusters have been identified using in-sample information. This naturally raises a question: do these patterns persist out of the sample? Importantly, our goal is not to engage in a horse race of R^2_{OOS} metrics among various ML models. Instead, the primary objective is to evaluate the persistence of the heterogeneous predictability pattern out of the sample, which cannot be solved by standard ML models or clustering models. To evaluate the OOS performance, we adopt a five-year rolling clustering setup with a two-fold cross-validation strategy for hyperparameter optimization.[11]

We classify leaves into three groups by return predictability: high, medium, and low, using the tree structure and cluster-specific performance in Table 1. The top six clusters form the high predictability group, while the bottom five clusters comprise

[11]The in-sample data is split into two contiguous periods. The model is trained on one period and validated on the other. Optimal hyperparameters are determined based on the quadratic loss, after which the model is retrained on the full in-sample dataset. The final coefficients are then used to forecast the OOS values for the subsequent five years.

the set with the lowest predictability. The remaining leaves constitute the medium predictability group.[12] We use R^2_{OOS} values, as defined in Eq. (6), to validate the cross-sectional patterns of stock return predictability.

Table 2: **Comparing Global and Cluster-Wise Predictive Models**

This table reports return predictability (R^2 in %) across predictive methods, with both in-sample and out-of-sample results from the cross-sectional tree structure in Figure 3. Panel A summarizes results for global models, while Panel B focuses on cluster-wise models. Five samples are analyzed: Global (homogeneous predictions without clustering), Aggregate (aggregated cluster-wise forecasts), and High, Medium, and Low, based on sub-sample predictability rankings across forecasts.

	In-Sample (1973 - 2002)			Out-of-Sample (2003 - 2022)		
Sample	OLS	Lasso	Ridge	OLS	Lasso	Ridge
			Panel A: Global Forecasts			
Global	1.49	0.52	0.54	0.27	0.40	0.35
High	4.12	2.45	2.28	3.16	2.14	2.06
Medium	1.40	0.44	0.47	0.12	0.31	0.27
Low	0.89	0.30	0.31	0.09	0.24	0.17
			Panel B: Cluster-Wise Forecasts			
Aggregate	1.75	0.83	0.74	-0.28	0.42	0.53
High	6.03	5.07	4.69	2.97	3.79	3.84
Medium	1.81	0.78	0.59	-0.24	0.40	0.39
Low	-2.57	-1.78	0.06	-1.75	-0.74	0.11

Table 2 reports the R^2 statistics for the homogeneous model (Panel A) and cluster-wise models (Panel B). The "Global Forecasts" employs a homogeneous model without clustering stock returns and is assessed on high-, medium-, and low-predictability sub-samples identified by our method. The "Aggregate" model in Panel B aggregates the prediction results from all cluster-wise models. For instance, the global Ridge model in Panel A only achieves 2.06% R^2 out of the sample for highly predictable clusters, while the cluster-wise model achieves 3.84% in Panel B. The aggregate model demonstrates improvements in R^2 values, highlighting the benefits of the cluster-wise heterogeneous model over the homogeneous predictive model.

When comparing clusters with different levels of predictability, highly predictable

[12] In contexts with varied dataset time spans and sample sizes, using fixed R^2 thresholds to distinguish predictability levels proves challenging. We classify predictability based on in-sample observation proportions. For cross-sectional clustering, a 10% cutoff is used, selecting leaves with high or low predictability based on sample point proportions meeting this threshold. This cutoff is also 10% in Section 5 under time-series partitioning schemes, but utilizing regime- or period-specific observations to ensure adequate data for investment analysis. All these cases ensure that clusters are not separated and that the sum of their relative percentages is at least over the cutoff value.

clusters consistently outperform their less predictable counterparts, regardless of the model used. For instance, when comparing the values in each row for the highly predictable cluster against the low predictability cluster, every metric — whether in-sample or out-of-sample, OLS or Ridge regression, global or cluster-wise model — shows a higher R^2 for the more predictable cluster. These findings reaffirm the persistent predictability disparity between high- and low-predictability clusters, highlighting the strong association between latent return predictability and their characteristics-based groupings. Moreover, the out-of-sample robustness tests validate the efficacy of our tree-based clustering approach in capturing this heterogeneity.

4.3 Disagreement over Return Predictability

Most researchers and investors rely on the global homogeneous predictive model, as stock clustering methods are uncommon in the literature. As such, our approach arguably provides the first genuine attempt at systematically measuring the unobservable and heterogeneous return predictability. Panel A of Table 1 demonstrates that cluster-wise heterogeneous models outperform the global homogeneous model for most stocks with predictable returns. This implies that most investors' models are misspecified, and an average investor faces the risk that information in the market is not efficiently impounded into asset prices due to trades all relying on misspecified global homogeneous models, such as those tested in Gu et al. (2020).

Recent studies, including Smith and Timmermann (2022) and Feng, He, Li, Sarno, and Zhang (2024), show in regime-switching models of stock and currency returns how homogeneous models underperform their regime-dependent heterogeneous counterparts. In the same spirit, we measure the disagreement over return predictability as the R^2 difference between cluster-wise and global predictive models for each cluster j, which remains constant across all stock return observations within the same cluster:

$$R^2_{\text{CMG},j} = R^2_{\text{C},j} - R^2_{\text{G},j}. \tag{7}$$

As reported in Table 1, this metric measures the R^2 difference between heterogeneous models (R^2_{C} from cluster-wise forecasts) and the global model (R^2_{G} from the global model without clustering, but evaluated on the current cluster). A higher value of R^2_{CMG} indicates a greater model misspecification when the majority of investors in the market ignore the heterogeneity in return prediction. This, in turn, implies that information may take longer times to be impounded into asset prices, creating risks of mispricing.[13] Stocks that face this extra risk command greater returns in equilibrium, revealing a new aspect of the risk-return relationship.

We construct value- and equal-weighted portfolios for each cluster and examine the relationship between average returns and the disagreement-sorted portfolios. Panel B of Table 1 highlights the spreads of average returns, showing a positive relationship: clusters with larger return predictability disagreement tend to yield higher average returns and Sharpe ratios. These results motivate novel long-short strategies to exploit anomalies in predictability disagreement. Our tree-based clustering approach demonstrates that these asset clusters display heterogeneous return predictability, and systematic predictability disagreement correlates with average returns.

We analyze the value-weighted portfolios detailed in Table 1. Specifically, we project the estimated predictability disagreement (R^2_{CMG}), computed at the cluster level using in-sample data, onto individual stock observations across both in-sample and out-of-sample periods.[14] For instance, the long-short strategy ("T3 - B3") takes long positions in the top 3 clusters and short positions in the bottom 3 clusters, rebalanced monthly. Within-cluster weights are value-weighted, while between-cluster weights are equally weighted. We estimate spanning regressions of portfolio returns on standard benchmark factor models to assess model-adjusted abnormal returns associated with these predictability disagreement anomalies.

Panel A of Table 3 summarizes the performance of four representative long-short anomalies.[15] Strategies with larger cluster-specific risk differentials (e.g., "T3 - B3") consistently outperform those with smaller gaps (e.g., "T5 - B5"), achieving higher

[13]Defined in Eq. (4), it is possible that R^2_{C} derived from cluster-wise forecasts may be lower than R^2_{G} obtained from global forecasts that potentially utilize more cross-sectional information.

[14]All stock-return observations with the same cluster label, whether in-sample or out-of-sample, are assigned the same value. However, the value for stock i may differ across months t.

[15]We primarily focus on the "T3 - B3" strategy. Those subsequent issues involve incorporating additional observations and conducting robustness checks.

Table 3: **Anomaly of Predictability Disagreement**

This table presents summary statistics (Panel A) and abnormal returns (Panel B) for the anomaly based on cross-sectional return predictability differences (R^2_{CMG} in Table 1) across clusters. The numbers following "T" and "B" indicate the number of top and bottom clusters, ranked by R^2_{CMG}, used to construct long and short portfolios. Combinations like "T-B" represent long-short strategies. Panel A reports average return (Avg, %), standard deviation (Std, %), and annualized Sharpe ratio (SR). Panel B provides alpha estimates (% with significance by "*") and t-statistics (in parentheses) from various factor models.

	In-Sample (1973 - 2002)				Out-of-Sample (2003 - 2022)			
	T1 - B1	T3 - B3	T5 - B5	T7 - B7	T1 - B1	T3 - B3	T5 - B5	T7 - B7
			Panel A: Summary Statistics					
Avg	3.47	3.03	1.82	1.46	2.79	1.74	1.32	1.10
Std	7.42	4.61	3.43	2.53	6.41	4.09	3.64	2.64
SR	1.62	2.28	1.84	2.00	1.51	1.48	1.25	1.44
			Panel B: Abnormal Returns					
CAPM	3.57***	3.07***	1.78***	1.43***	3.04***	1.81***	1.28***	1.10***
	(9.19)	(12.57)	(9.86)	(10.74)	(7.39)	(6.75)	(5.38)	(6.32)
FF3	3.20***	2.83***	1.64***	1.30***	3.06***	1.81***	1.30***	1.11***
	(8.27)	(12.14)	(10.36)	(10.87)	(7.47)	(6.92)	(5.84)	(6.74)
FF5	2.92***	2.72***	1.59***	1.26***	3.03***	2.00***	1.39***	1.17***
	(7.40)	(11.33)	(9.76)	(10.20)	(7.17)	(7.52)	(6.05)	(6.89)
FF5+MOM+IVOL	2.48***	2.53***	1.54***	1.27***	3.01***	2.05***	1.45***	1.22***
	(6.28)	(10.38)	(9.21)	(9.99)	(7.08)	(7.82)	(6.52)	(7.36)
Q5	2.48***	2.41***	1.50***	1.22***	3.10***	2.09***	1.61***	1.31***
	(5.43)	(8.68)	(7.88)	(8.44)	(7.24)	(7.73)	(7.19)	(7.89)
BS6	2.37***	2.49***	1.52***	1.20***	2.90***	2.03***	1.48***	1.23***
	(5.64)	(9.66)	(8.60)	(8.98)	(6.87)	(7.69)	(6.68)	(7.58)
DHS3	2.90***	2.70***	1.79***	1.47***	2.93***	1.98***	1.43***	1.19***
	(6.46)	(9.58)	(8.59)	(9.49)	(6.93)	(7.36)	(5.90)	(6.76)
SY4	2.32***	2.46***	1.50***	1.20***	3.85***	2.78***	1.86***	1.58***
	(5.65)	(9.72)	(8.68)	(9.15)	(7.11)	(8.91)	(8.47)	(8.58)

average returns and Sharpe ratios across both sample periods. Even the strategy incorporating the largest number of clusters ("T7 - B7") delivers robust performance, with an average monthly return of 1.10% and an annualized Sharpe ratio of 1.44 for the OOS two decades.

Panel B of Table 3 reports the model-adjusted abnormal returns, including CAPM, FF3, FF5, FF5+MOM+IVOL, Q5 (Hou, Mo, Xue, and Zhang, 2021), BS6 (Barillas and Shanken, 2018), DHS3 (Daniel, Hirshleifer, and Sun, 2020), and SY4 (Stambaugh and Yuan, 2017).[16] The performance results are highly positive, with all strategies demonstrating statistically significant unexplained alphas both in-sample and out-of-sample. The OOS monthly alphas for the strategy with larger predictability gaps ("T1 - B1")

[16]Data from Stambaugh's website is updated only through 2016, so we report 2003-2016 for SY4.

are highly significant both statistically and economically, ranging from 2.90% to 3.85%. In contrast, strategies with smaller cluster-specific risk gaps ("T5 - B5") exhibit slightly weaker performance, though their alphas remain over 1.2% and statistically significant. These findings underscore the robustness of the strategy's unexplained alphas, even after adjusting for risk factors in standard asset pricing models.

We further examine whether abnormal returns stem mainly from the long-leg, short-leg, or both components of the anomaly portfolio. Table A.3 in the appendix reveals that long-leg portfolios with larger predictability disagreement consistently contribute the most to abnormal returns, whereas short-leg portfolios generally exhibit insignificant OOS alphas. Therefore, stocks with the largest model misspecification risks (long-leg) drive the documented abnormal returns.

In conclusion, we document an anomaly linked to return predictability disagreements. Strategies exploiting aggregate predictability differences between cluster-based and homogeneous predictive models yield significant return spreads. This anomaly delivers robust alpha even after controlling for established risk factor models. These findings highlight predictability disagreement as a model misspecification risk by ignoring the heterogeneous modeling. Consequently, predictability-based strategies deliver superior model-adjusted abnormal returns, in contrast to the null results in Rapach et al. (2010) and Kelly et al. (2024).

5 Regime-Dependent, Time-Varying Predictability

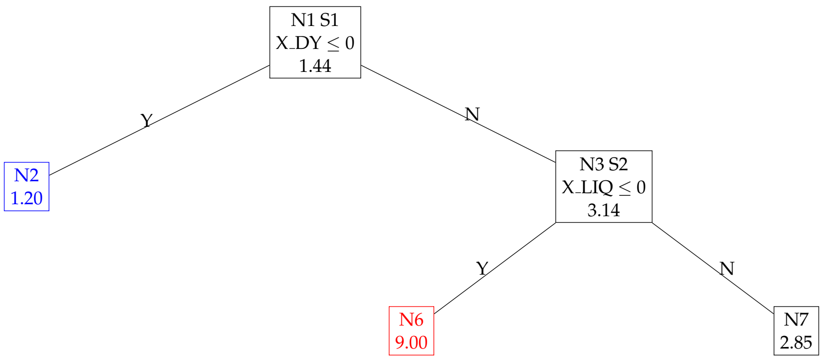
Building on Cong et al. (2023), we examine whether the heterogeneity in the cross section discussed earlier exhibits regime-switching by incorporating time-series information, i.e., splitting variables extend beyond characteristics ("CS Clusters," Section 4) to include market predictors. We investigate time-varying return predictability conditioned on aggregate macroeconomic or market states (Section 5.1). Unlike cross-sectional clustering, time-series heterogeneity is assessed by splitting the full sample period (1973–2022) to identify regime-dependent variations. We demonstrate that calendar months serve as alternative split variables, further facilitating the detection of structural breaks in time-varying predictability (Section 5.2).

5.1 Macro-Driven Regime Change

Like the regime detection in Feng et al. (2024) and Bie et al. (2024), we employ market predictors to split time-series horizons and identify disjoint market regimes for heterogeneous return predictability. These regimes are defined based on the market predictors, offering intuitive and economically meaningful interpretations. Furthermore, these regimes exhibit recurrent patterns aligned with business cycle dynamics. Specifically, we limit market predictors to time-series data for the initial two splits and employ firm characteristics for later cross-sectional partitions.[17] To our knowledge, this methodology is the first systematic framework that learns to cluster heterogeneous return predictability across time-series and cross-sectional dimensions.

Figure 6: **Tree-Based Clusters under Market Regimes**

This figure illustrates the tree-based clustering structure of stock return panels under market regimes, using monthly data from 1973 to 2022. The tree partitions returns based on market predictors, subtracting the prior 10-year rolling average and smoothing, dividing the sample into three regimes (market dividend yield and liquidity). Terminal leaves denote distinct regimes with nonconsecutive month indices in brackets. Each node, including intermediate and terminal ones, is assigned a unique ID (N#), with split order indicated by (S#). Cluster-specific R^2 values are reported for all nodes.



Clustering regime shifts. We fit the model using the 50-year sample for the macro-driven regimes, yielding a single tree-based clustering structure. Figure 6 presents the top part of the tree in the time dimension, while Figure A.3 shows further partitions

[17]We have experimented with relaxing these constraints, permitting both market predictors and firm characteristics as candidates for all splits. However, the trees prioritize market predictors for the first two partition layers. This indicates a strong preference for initial separation based on time-series data. To limit excessive regime changes, we impose these restrictions.

by firm characteristics across each regime. This framework captures time-series and cross-sectional grouped heterogeneity by creating 45 leaf clusters.[18]

Specifically, the in-sample R^2 for the global model across the entire sample is 1.44% before any partitions. Our approach identifies the aggregate market dividend yield and liquidity as the two most important market predictors for detecting regimes of return predictability, and creates the following three market regimes:

1. Regime I (377 months): $\texttt{X\_DY} \leq 0$, when the market dividend yield is low.
2. Regime II (57 months): $\texttt{X\_DY} > 0$ and $\texttt{X\_LIQ} \leq 0$, when the market dividend yield is high, and the market liquidity is low.
3. Regime III (166 months): $\texttt{X\_DY} > 0$ and $\texttt{X\_LIQ} > 0$, when both market dividend yield and market liquidity are high.

The dynamics of these two market predictors result in three disjoint regimes that transition throughout business cycles.[19] We find that the return predictability of stock returns is affected by the recurrence of these three market regimes. When the market dividend yield is low in Regime I, indicating high market valuation periods, the R^2 for return predictability is 1.20%, below the overall average of 1.44%. Predictability notably increases with high dividend yield and low market valuation.

In particular, the R^2 of predictability peaks at 9.00% when the market dividend yield is high and market liquidity is low, indicating early recessions. This aligns with evidence on time-varying market return predictability (e.g., see Henkel et al., 2011; Dangl and Halling, 2012). The forward-looking nature of the stock market suggests that during recessions, the increase in dividend yield is mainly due to declining prices rather than rising dividends. There is a higher probability of a stock price recovery following stock fire sales during the early recession. Regime II consistently exhibits superior return predictability compared to the other two regimes.

[18] We cap the maximum tree depth at 5 per regime (allowing up to 48 leaves) and impose a minimum leaf size of 30 monthly stock-return observations.

[19] We standardize market predictors by subtracting their 10-year rolling averages. Our standardized market dividend yield corresponds to the adjusted dividend-price ratio in Lettau and Van Nieuwerburgh (2008), who show that its stationary component predicts market returns. Their results attribute time-varying return predictability to structural breaks in expected returns and financial ratios.

The tree-based regime demonstration illustrates time-varying return predictability patterns. Further partitions based on cross-sectional characteristics reveal significant heterogeneity in stock return predictability across macro-driven regimes. For instance, leaf node N16 (red, Regime II in Figure A.3) achieves the highest R^2 of 35.51%, which includes low ILL, low STD_DOLVOL, low BETA, and low DEPR stocks. However, the weakest cluster in the same Regime II, node N25 (blue), which includes high ILL, low GMA, low ME, and high CASHDEBT stocks, records the weakest at -3.97%. Compared to the cross-sectional clustering results in Section 4, this wider range underscores the importance of incorporating regime-dependent time variation in enhancing return predictability assessments.

Figure 7: **Time-Varying Predictability across Market Regimes**

This figure illustrates market regime shifts for market predictors from 1973 to 2022. Three regimes are color-coded: low dividend yield (X_DY ≤ 0, orange), high dividend yield with low liquidity (X_DY > 0 & X_LIQ ≤ 0, purple), and high dividend yield with high liquidity (X_DY > 0 & X_LIQ > 0, green). These regimes capture varying predictability levels, as indicated by the color bar. Shaded regions represent NBER recessions, while major global events are labeled. The vertical axis measures the cross-sectional predictability of clusters within each regime, ranging from highly predictable (black) to less predictable (white) compared with the market predictability (background).

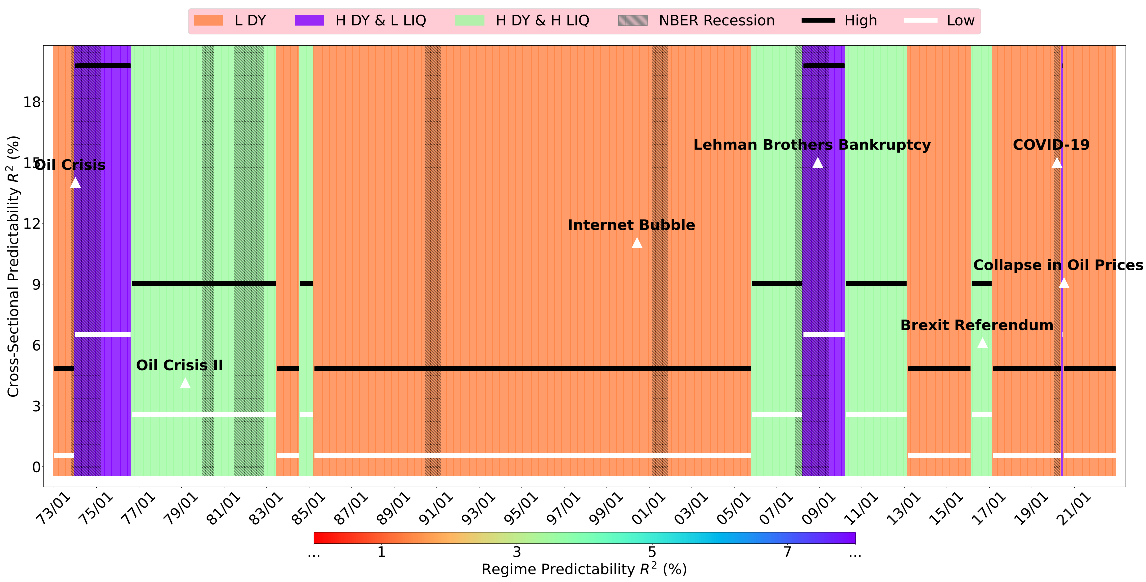
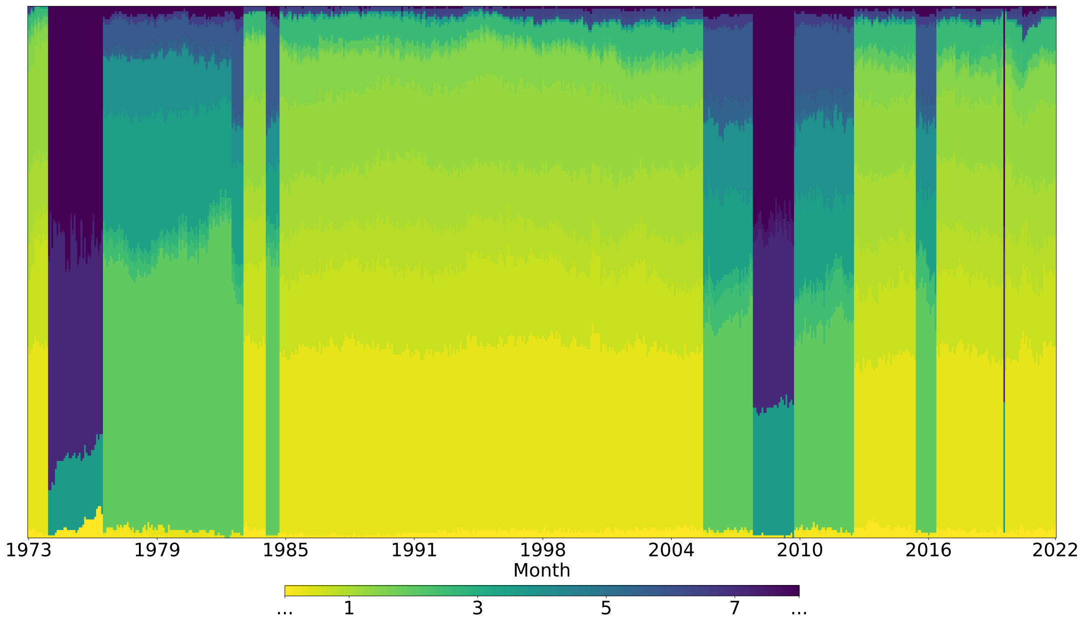


Figure 7 presents regime-dependent time variation from an alternative perspective. The color transitions denote the market regime switches identified by the tree structure in Figure 6. The range spans beyond the cross-sectional case, from approximately 1% (orange) to over 7% (purple). Subsequently, cross-sectional partitions of stock returns by firm characteristics amplify interval disparities across sub-samples.

We also illustrate the large cross-sectional variation in return predictability across regimes by presenting results for the top and bottom clusters (black and white bars). Furthermore, regime transitions align with recessions (e.g., 1974, 2008, 2020) and major global events, such as the 1973 Oil Crisis, the Lehman Brothers' collapse (2008), the Brexit referendum (2016), and the 2020 oil price crash.

Figure 8: **Mosaics of Predictability by Months (TS+CS Cluster)**

This heat map summarizes monthly return predictability (R^2 values, % in the color bar) for results presented in Figures 6 and A.3. The horizontal axis represents months, while the vertical axis indicates the proportion of observations per cluster. Colors transition from light (bottom) to dark (top) every month, indicating increasing levels of return predictability.



Compared to the CS clusters illustrated in Figure 5, the mosaic patterns of stock return predictability across months (as depicted in Figure 8) exhibit more regime-dependent time variations. Sorting the cluster-wise model R^2 values in ascending order from bottom to top and aggregating monthly reveals layered heterogeneity with three hierarchical types. Most clusters exhibit lower predictability (light green or even yellow). In contrast, specific regimes — such as around 1975 and 2009, characterized by high market dividend yields and low market liquidity during recessions — demonstrate significantly higher predictability across nearly all clusters than other periods. The frequent color shifts across clusters highlight time variation of market regime switches, empirically confirming the advantages of our P-Tree approach.

Cluster-specific performance. Next, we analyze cluster-specific performance to evaluate return predictability across regimes. Table 4 translates the tree structures in Figure 6 and Figure A.3 into a descending ranking of return predictability by market regime. These results highlight the significance of understanding heterogeneity in predictability patterns within various market regimes.

Table 4: **Cluster-Specific Performance under Market Regimes**

This table summarizes the cluster-based information under each regime within three panels corresponding to the time-series and cross-sectional tree structure in Figure 6 and A.3. Each panel counts the number of observations for each cluster ("# obs") and displays the return predictability (R^2 values in %) for each cluster. Besides, "Avg" and "Std" denote the monthly average return and standard deviation (both in %) for equal/value-weighted (EW/VW) portfolios based on all observations. Each regime of values is arranged in the descending order of R^2 from left to right.

			Regime I: $\mathbb{1}\{\texttt{X\_DY} \le 0\}$												
Leaf	N29	N28	N30	N26	N9	N31	N25	N17	N27	N21	N24	N16	N22	N20	N23
# obs	11,759	17,188	21,476	16,572	14,139	73,719	34,334	118,546	12,424	258,031	194,770	256,625	135,414	603,575	27,625
R^2	11.00	6.74	6.58	4.32	3.39	2.89	2.72	2.27	2.19	1.74	1.45	1.34	1.22	0.82	-12.79
Avg_{EW}	3.66	2.43	1.90	2.84	-0.77	1.46	2.49	-1.68	1.30	0.75	1.00	0.06	0.41	0.08	-0.28
Std_{EW}	5.65	5.02	4.54	6.35	7.12	5.59	7.29	8.46	5.71	5.07	5.57	5.21	5.00	6.58	6.75
Avg_{VW}	3.04	1.93	1.30	1.86	-0.29	0.90	1.19	-1.24	0.81	0.79	0.72	0.35	0.56	0.42	-0.51
Std_{VW}	5.41	4.80	4.32	6.32	7.23	5.10	7.53	9.94	5.01	4.48	4.82	4.71	4.73	4.88	7.90
			Regime II: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \le 0\}$												
Leaf	N16	N26	N29	N17	N19	N10	N31	N23	N28	N27	N18	N22	N30	N24	N25
# obs	1,946	2,411	16,065	1,857	18,921	2,434	9,977	1,738	25,867	8,011	6,217	4,585	78,607	39,164	3,831
R^2	35.51	19.80	19.51	18.63	17.85	16.73	15.29	15.23	12.78	12.77	12.05	10.14	9.65	6.99	-3.97
Avg_{EW}	0.01	0.56	5.00	-0.49	0.65	0.24	1.61	0.78	2.03	0.06	0.43	0.15	0.83	1.19	1.59
Std_{EW}	6.70	11.16	12.78	7.20	7.86	6.54	10.78	8.45	9.87	7.70	7.33	8.08	8.39	9.66	7.26
Avg_{VW}	0.24	0.37	3.49	-0.26	0.25	0.16	1.07	0.82	1.35	-0.10	-0.16	-0.25	0.40	0.76	0.80
Std_{VW}	6.47	11.26	12.46	6.59	6.65	6.43	9.41	8.71	9.26	7.73	7.97	7.66	7.70	8.72	7.00
			Regime III: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} > 0\}$												
Leaf	N14	N31	N27	N21	N20	N19	N30	N26	N23	N25	N16	N18	N22	N24	N17
# obs	8,010	6,477	12,087	5,013	5,658	60,434	19,067	74,436	108,106	9,159	16,811	28,896	296,448	6,427	5,849
R^2	12.39	9.78	7.66	6.84	6.41	6.40	6.05	4.69	4.22	4.12	3.24	3.15	2.65	-0.33	-0.49
Avg_{EW}	4.27	2.38	2.87	1.21	0.80	1.26	2.23	1.80	1.50	2.34	0.61	0.48	0.86	0.73	0.22
Std_{EW}	5.60	4.72	5.86	6.24	5.93	3.70	5.33	5.24	6.31	6.04	3.63	3.52	5.82	6.58	3.92
Avg_{VW}	3.55	1.21	1.52	0.78	0.60	0.66	1.46	0.97	1.38	1.05	0.44	0.50	0.62	0.30	-0.10
Std_{VW}	5.31	4.14	4.99	6.24	6.07	3.63	4.67	4.53	6.27	6.38	4.39	3.54	5.76	6.25	4.62

For example, the differences between the highest and lowest R^2 values are larger than in the cross-sectional setting: 11.00% vs. -12.79% in Regime I, 35.51% vs. -3.97% in Regime II, and 12.39% vs. -0.49% in Regime III. Sorting monthly stock returns by R^2 values in descending order highlights mosaic patterns of predictability across clusters. Similar to the cross-sectional analysis in Table 1, where modest positive correlations emerge between predictability and returns in Regimes I and III. Regime II, lasting 57 months, is the most predictable period but exhibits a near-zero correlation between R^2 and average returns. However, the higher volatility or return standard deviations confirm our explanation for the higher return predictability, possibly due to the stock price recovery following stock fire sales during the early recession.

Table 5: **Further Comparing Global and Cluster-Wise Predictive Models**

Similar to Table 2, this table reports return predictability (R^2 in %) across predictive methods under market regimes, with all and large-cap stocks. For each regime, two panels are presented: global and cluster-wise forecasts. Results include five samples: Global (homogeneous forecasts), Aggregate (cluster-aggregated predictions), and High, Medium, and Low, based on predictive rankings within tree clusters. "-" represents that the cluster is empty due to `ME` being selected as the split variable.

	Sample A: All Stocks			Sample B: Large-Cap		
	Regime I: $\mathbb{1}\{\texttt{X\_DY} \leq 0\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Panel A: Global Forecasts						
Global	0.99	0.24	0.84	0.97	0.39	0.87
High	3.20	1.70	3.00	2.20	1.53	2.15
Medium	1.11	0.25	0.96	0.94	0.36	0.85
Low	0.64	0.07	0.51	0.69	0.15	0.56
Panel B: Cluster-Wise Forecasts						
Aggregate	1.05	0.49	0.98	0.98	0.45	0.94
High	4.54	3.67	4.46	2.59	2.03	2.54
Medium	1.56	0.82	1.45	1.26	0.58	1.17
Low	0.18	-0.16	0.16	-0.20	-0.33	-0.10
	Regime II: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \leq 0\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Panel A: Global Forecasts						
Global	9.14	4.15	8.68	13.00	6.57	12.66
High	11.24	5.00	10.58	16.46	8.37	16.07
Medium	9.72	4.43	9.26	12.75	6.44	12.41
Low	6.67	3.06	6.29	-	-	-
Panel B: Cluster-Wise Forecasts						
Aggregate	11.29	8.99	9.28	15.33	11.29	12.91
High	19.17	16.32	17.35	23.59	21.20	20.44
Medium	11.51	9.11	9.24	14.73	10.57	12.36
Low	6.47	4.69	4.95	-	-	-
	Regime III: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} > 0\}$					
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Panel A: Global Forecasts						
Global	2.56	1.27	2.39	3.07	1.81	2.88
High	6.86	4.01	6.53	4.37	2.89	4.16
Medium	3.37	1.99	3.18	3.38	2.23	3.22
Low	1.80	0.67	1.65	2.24	0.81	2.01
Panel B: Cluster-Wise Forecasts						
Aggregate	3.19	2.07	3.00	3.54	2.33	3.35
High	8.72	7.03	8.43	6.47	3.64	5.95
Medium	4.19	2.83	3.92	3.99	2.63	3.74
Low	2.24	1.29	2.11	2.13	1.53	2.13

Aggregate evaluation. We further conduct aggregate evaluations to assess improvements in heterogeneous return predictability with regime-dependent time variation modeling. Clusters are aggregated into sub-samples, with Table 5 reporting the resulting predictability metrics. Under all regimes, heterogeneous models (Panel B) exhibit greater predictability differences in return predictability relative to the global homogeneous model (Panel A). We also analyze a large-cap sub-sample for robustness, which typically exhibits lower return predictability and transaction costs than small-caps. Solely focusing on large-cap stocks highlights heterogeneous predictability patterns across varying market conditions.

Among the three regimes, Regime II (characterized by high dividend yield and low liquidity) consistently shows greater predictability than the other two, a finding robust across all three predictive methods (OLS, Lasso, and Ridge). The downward trends in the final rows (high, medium, and low) underscore significant cross-sectional predictability variations within this regime, which are robust across all methods. However, the low market dividend yield regime exhibits the lowest level of predictability, with the large-cap segment demonstrating an in-sample R^2 of no more than 2.6%, suggesting negligible return predictability. Despite varying magnitudes, the improvement and decline patterns are consistent across methods, especially for return forecasts using cluster-wise models. These findings shed light on the mosaics of return predictability and stock heterogeneity across characteristics and market conditions.

Theoretical explanation. Our analysis identifies the market dividend yield as the primary driver for time-varying return predictability. Periods with high dividend yields show high return predictability. We next interpret this result through a theoretical framework built on Campbell and Shiller (1988) present value model. While it may not capture the cross-sectional returns, we aim to show the link between market dividend yield and market return predictability.

Let P_{t+1}, D_{t+1}, and R_{t+1} denote the price, dividend, and return for period $t+1$, respectively. Define $p_t = \log(P_t)$, $d_t = \log(D_t)$, and $R_{t+1} = \frac{D_{t+1}+P_{t+1}}{P_t} = \frac{1+P_{t+1}/D_{t+1}}{P_t/D_t}\frac{D_{t+1}}{D_t}$. The log return $r_{t+1} = \log(1+R_{t+1})$ adheres to the Campbell-Shiller decomposition:

$$
\begin{aligned}
r_{t+1} &= \log(1 + R_{t+1}) \\
&= \log\left(1 + \frac{P_{t+1}}{D_{t+1}}\right) - \log\left(\frac{P_t}{D_t}\right) + \log\left(\frac{D_{t+1}}{D_t}\right) \\
&= \log(1 + e^{\delta_{t+1}}) - \delta_t + \Delta d_{t+1},
\end{aligned}
$$

where $\delta_t = \log(P_t/D_t)$ represents the log price-dividend ratio and Δd_{t+1} denotes the log dividend growth rate. The standard Campbell-Shiller decomposition expands this function to the first order. Here, we additionally compute the second-order term: the logarithm of the Taylor series.

$$
r_{t+1} \approx \log(1 + \bar{\mathcal{D}}) + \underbrace{\frac{\bar{\mathcal{D}}}{1 + \bar{\mathcal{D}}}(\delta_{t+1} - \bar{\delta})}_{\text{first order}} + \underbrace{\frac{\bar{\mathcal{D}}}{2\left(1 + \bar{\mathcal{D}}\right)^2}(\delta_{t+1} - \bar{\delta})^2}_{\text{second order}} - \delta_t + \Delta d_{t+1}, \tag{8}
$$

where $\bar{\mathcal{D}}$ denotes the average price-dividend ratio and $\bar{\delta} = \log(\bar{\mathcal{D}})$. Notably, when explaining the log-return as a linear function of δ, the second-order term is absorbed by the regression residuals. When dividend-price ratios are high, the log price-dividend ratio tends to be lower, as it is the inverse of the dividend-price ratio. Consequently, the second-order term is smaller than the first-order term, reducing the residuals. This results in a higher R^2 in the linear regression model with only the first-order term, indicating increased market return predictability during high dividend-price ratio periods. Thus, individual stock return predictability likely increases with market return predictability as the dividend yield increases.

5.2 Structural Breaks in Calendar Time

Splitting by market predictors induces recurring yet non-continuous regimes over time. Another approach in modeling time variation is assessing structural breaks in the return dynamics (e.g., Smith and Timmermann, 2021; Cui et al., 2024). This approach can also be implemented using our P-Tree framework, splitting by calendar months rather than market predictors. For example, Feng et al. (2024) explore tree-based break detection methods in currency markets. One benefit of this approach is that calendar months progress over time, creating single, continuous regimes akin to structural breaks. This approach evaluates whether segmenting continuous periods effectively captures stock return predictability heterogeneity.

Figure 9: **Time-Varying Predictability over Structural Breaks**

This figure illustrates structural breaks by calendar month from 1973 to 2022. Distinct colors denote regimes with varying predictability, as the color bar shows. Shaded areas represent NBER recessions, while annotations highlight key global economic events. The vertical axis captures the cross-sectional predictability of clusters, ranging from the highest (black) to the lowest (white) within each regime.

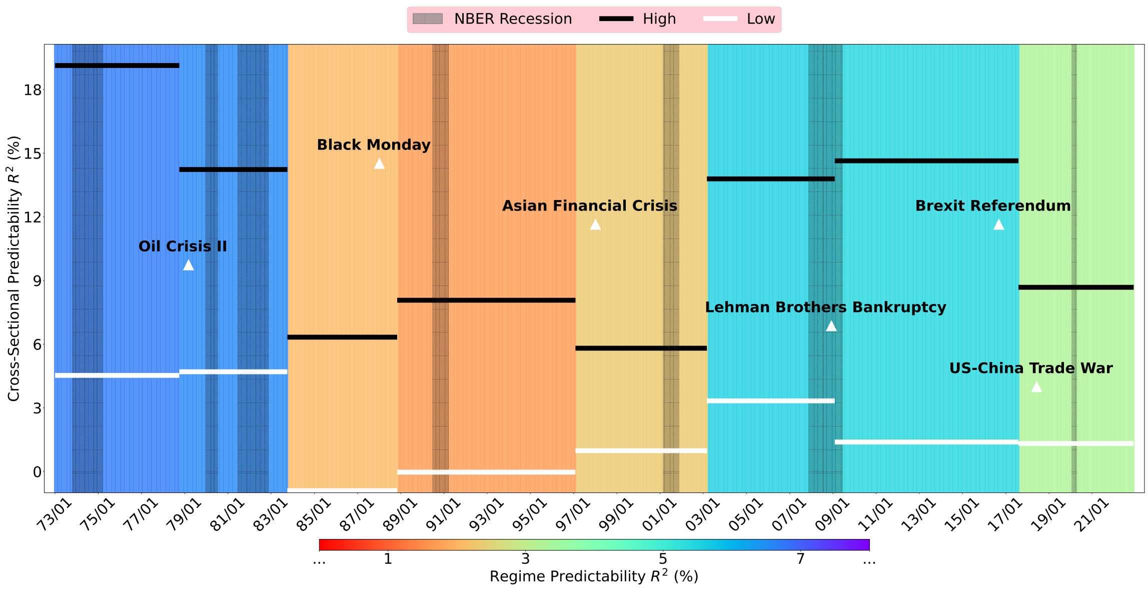


Figure 9 illustrates fluctuations in stock return predictability across structural breaks in calendar months, albeit smaller than in regime-based models using market predictors. Predictability peaks at 6.7% (Jan. 1973 – Oct. 1978, first purple) but declines to 1.67% (Nov. 1988 – Feb. 1997, orange). However, time-varying return predictability, analyzed via structural break models, intensifies during recessions, particularly before 1983, consistent with our findings in Section 5.1.

Moreover, events like Black Monday (1987) and the Brexit Referendum (2016) likely induce structural breaks in return predictability, which remain obscured in macro-driven regime analysis due to the sluggish nature of market predictors. Similarly, structural breaks may underlie other notable financial disruptions, such as the Oil Crisis (1979), the Asian Financial Crisis (1998), the Lehman Brothers' collapse (2008), and the US-China Trade War (2018).

Splitting further within the cross section of each regime between breaks reveals greater variation and heterogeneity in predictability. Table A.4 shows that return predictability and average returns across regimes vary between high (black) and low (white) predictable sub-samples, with trends exhibiting consistency across periods. These large spreads of heterogeneous predictability are robust for the large-cap sub-sample, as shown in Table A.5. Overall, all these positive results confirm the mosaic-like nature of stock return predictability in the time-series dimension.

6 Investment Performance of Cluster-Wise Models

The preceding sections highlight the heterogeneity in return predictability across assets and over time, employing our tree-based clustering approach. Beyond the risk anomaly linked to disagreement over return predictability, a key outcome for investment professionals is estimating cluster-specific predictive models. Assets are grouped into clusters, each with its own model, forming a heterogeneous framework. This stands in contrast to global models (e.g., Gu et al., 2020) that uniformly apply a single homogeneous model across all assets. We now evaluate the investment gains of the cluster-specific models by constructing a novel forecast-implied portfolio, which assesses predictive performance through portfolio returns.

We construct three types of portfolio weights based on model predictions. In equal-weighted and value-weighted portfolios, stock weights are based on the absolute values of the weighting schemes, with signs determined by model return forecasts. In addition, we construct forecast-weighted portfolios based on the ML return forecasts, controlling for leverage.[20] Under this framework, stocks predicted to have higher returns receive bigger weights in the long positions, and those with predicted lower (or even negative) returns receive smaller weights (or larger weights in the short positions).

Specifically, let the portfolio weights $w_{i,t}$ be equal-weighted or value-weighted, where i denotes the stock dimension and t represents time. These three portfolio weighting schemes are as follows:

[20] We follow the approach of Guijarro-Ordonez et al. (2024), which normalizes stock weights to sum to one in absolute terms, thereby implicitly imposing a leverage constraint.

$$\text{Sign-adjusted Equal/Value-weighted: } \widehat{w}_{i,t} = \begin{cases} w_{i,t}, & \text{if } \widehat{r}_{i,t} \geq 0 \\ -w_{i,t}, & \text{if } \widehat{r}_{i,t} < 0 \end{cases} \tag{9}$$

$$\text{Forecast-weighted: } \widehat{w}_{i,t} = \frac{\widehat{r}_{i,t}}{\sum_i |\widehat{r}_{i,t}|}.$$

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

$$R_{j,t} = \sum_{\{i,t\} \in \text{leaf}_j} \widehat{w}_{i,t} r_{i,t}. \tag{10}$$

These forecast-implied portfolios differ from the cluster portfolios in Section 4.3, where only value-weighted portfolios are constructed for each cluster, without incorporating return forecasts. In contrast, the forecast-implied portfolios are constructed as long-short portfolios, where the long and short positions are determined by the ML return forecasts. Intuitively, if the predictive model exhibits greater accuracy, the corresponding portfolio is expected to deliver higher investment returns.

Table 6 presents the results generated from the cross-sectional clustering model in Section 4. First, our results confirm that highly predictable stocks remain easier to predict out of the sample, and as a result, the corresponding trading strategies generate significant investment gains. The sign-adjusted value-weighted portfolio also performs well, indicating that the results are not dominated by small stocks. For example, in Panel A, the highly predictable sub-sample shows an OOS Sharpe ratio of 0.84 and a significant alpha of 0.46%. In contrast, investing in stocks with the lowest predictability achieves only about half of these values, with an OOS Sharpe ratio of 0.51 and a negative alpha with no significance.

Similarly, in Panels B and C, highly predictable stocks generate OOS Sharpe ratios of 1.66 and 1.86, and highly significant alphas of 1.84% and 2.30%, respectively. These figures significantly outperform the portfolio of medium- and low-predictability stocks. Similar patterns are observed for average returns and maximum drawdowns, especially the contrasts of high significance for alphas in the top two panels.

Second, cluster-wise predictive models ("Aggregate") perform better than the homogeneous predictive model ("Global"), demonstrating the economic gains of adopt-

Table 6: **Forecast-Implied Investment Performance (CS Cluster)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 10). Panels A and B present sign-adjusted value-/equal-weighted portfolios, while Panel C reports forecast-weighted portfolios, all constructed by observations in specific samples (Equation 9). We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap sub-samples of Global and Aggregate). Each panel includes five metrics: monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), Jensen's alpha (%), and maximum drawdown (MDD, %).

	In-Sample (1973 - 2002)					Out-of-Sample (2003 - 2022)				
	Panel A: Sign-adjusted Value-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	0.64	3.98	0.55	0.30^{***}	21.38	0.72	4.17	0.60	-0.03	17.21
Aggregate	0.56	4.03	0.48	0.21^{***}	21.14	0.79	4.08	0.67	0.05^{**}	16.58
High	1.93	4.83	1.38	1.58^{***}	21.84	1.21	5.01	0.84	0.46^{**}	24.17
Medium	0.56	4.08	0.48	0.21^{***}	21.44	0.81	4.15	0.67	0.06^{*}	16.92
Low	0.15	4.05	0.13	-0.14	15.49	0.61	4.12	0.51	-0.09	15.23
Global-Large	0.62	3.99	0.53	0.28^{***}	21.45	0.72	4.13	0.61	-0.02	16.98
Aggregate-Large	0.52	4.09	0.44	0.17^{***}	21.55	0.79	4.09	0.67	0.05^{**}	16.29
	Panel B: Sign-adjusted Equal-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.15	4.02	0.99	0.88^{***}	16.78	0.87	5.14	0.58	0.04	22.72
Aggregate	1.17	3.21	1.26	0.97^{***}	13.63	0.93	4.17	0.77	0.26^{**}	20.72
High	2.88	5.73	1.74	2.49^{***}	25.84	2.70	5.64	1.66	1.84^{***}	24.03
Medium	1.10	3.17	1.20	0.91^{***}	13.46	0.86	3.93	0.75	0.26^{*}	21.34
Low	0.51	3.80	0.47	0.26^{**}	14.23	0.67	5.06	0.46	-0.15	19.20
Global-Large	0.78	4.34	0.62	0.44^{***}	24.02	0.78	4.78	0.57	-0.06	20.56
Aggregate-Large	0.80	3.97	0.70	0.48^{***}	22.37	0.86	4.54	0.66	0.06	19.80
	Panel C: Forecast-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.58	4.69	1.17	1.26^{***}	20.52	1.12	5.32	0.73	0.25^{*}	23.04
Aggregate	2.01	4.26	1.64	1.74^{***}	16.52	1.56	4.77	1.13	0.79^{***}	21.31
High	3.42	6.08	1.95	3.01^{***}	26.23	3.22	5.99	1.86	2.30^{***}	24.82
Medium	1.74	4.05	1.48	1.50^{***}	16.10	1.25	4.57	0.95	0.53^{***}	21.96
Low	0.88	4.93	0.62	0.53^{***}	17.95	0.88	5.60	0.54	-0.03	19.98
Global-Large	1.04	4.77	0.76	0.66^{***}	24.81	0.92	4.93	0.64	0.05	20.79
Aggregate-Large	1.14	4.37	0.90	0.80^{***}	23.48	1.00	4.82	0.72	0.16^{*}	20.61

ing cluster-wise predictive models. For example, in Panel C, the OOS alpha for the aggregate model is 0.79%, higher than the 0.25% achieved by the global model, with Sharpe ratios of 1.13 and 0.73, respectively. This gap is larger in the forecast-weighted portfolio, primarily because the weights of the value- and equal-weighted portfolios in Panels A and B are identical, with only the sign adjusted by predictive models.

Overall, while both aggregate and global models exhibit comparable accuracy in predicting return signs (up or down) over the next month (Panels A and B), the aggregate model outperforms in forecasting *magnitude* of returns (Panel C). Large-cap sub-samples (top 30% by market equity values) outperform their global counterparts out-of-sample across all weighting strategies, although the improvement is smaller than the full sample. This may be due to the lower return predictability of large-cap stocks. In Table A.6 of the Appendix, we have also accounted for the impact of transaction costs. While the investment performance experiences a marginal decline, the documented pattern of decreasing economic gains across various return predictability levels remains robust. For example, the OOS Sharpe ratios for the highly predictable sub-sample in Panels B and C decrease from 1.66 and 1.86 to 1.59 and 1.80.[21]

All empirical evidence corroborates that stock return observations exhibiting high predictability — identified through our tree-based clustering approach — yield significantly superior investment returns than other stocks. Failing to include these highly predictable stocks significantly diminishes possible investment gains. Extending the findings from the stock return predictability mosaics to the contributions of investment strategies, it is clear that investors can achieve higher economic benefits by allocating funds to stocks with high earnings surprises, high valuations, or low trading volumes (see Figure 3). Finally, as mentioned in the introduction, the cluster-wise predictive model consistently outperforms the global homogeneous model, highlighting the limitations of naive homogeneous approaches.

7 Conclusion

Previous studies often conceptualize predictability as an attribute of the chosen predictors or models, examining it in aggregate within a homogeneous framework.

[21] Furthermore, in Section IA.II of the Internet Appendix, investment gains are amplified under regime-dependent temporal variations in heterogeneous predictive models. The results remain robust when applied to large-cap portfolios and after accounting for transaction costs.

We present a new perspective that treats return predictability as an intrinsic and time-varying characteristic of assets, which is unobservable but potentially related to the cross section of expected returns. We develop a tree-based clustering method to measure heterogeneous return predictability by grouping a panel of asset returns using firm characteristics and market predictors. We apply the framework to show what types of assets exhibit greater return predictability and how this predictability shifts across market regimes, documenting the "mosaics of predictability" in the U.S. equity market and offering rich insights for researchers and practitioners.

Specifically, we develop a novel clustering method based on P-Trees (Cong, Feng, He, and Li, 2023), to group asset-return observations with similar predictability levels into clusters and maximize predictability differences between clusters. We find substantial cross-sectional grouped heterogeneity and regime-dependent time variation in return predictability of U.S. equities: (i) Stocks characterized by high earnings surprises, high earnings-to-price ratios, and low dollar trading volumes exhibit the strongest return predictability; (ii) Return predictability diminishes at low market dividend yield but peaks during high market dividend yield periods coupled with low market liquidity; (iii) A new anomaly, potentially linked to predictability disagreement and model misspecification, generates monthly alphas easily exceeding 1%; and (iv) Investing in highly predictable stocks significantly outperforms popular benchmarks, achieving OOS annualized Sharpe ratios approaching 2 over the last two decades.

References

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

Ang, A. and G. Bekaert (2007). Stock Return Predictability: Is It There? *Review of Financial Studies 20*(3), 651–707.

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.

Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert (2023). Option Return Predictability with Machine Learning and Big Data. *Review of Financial Studies 36*(9), 3548–3602.

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

Barillas, F. and J. Shanken (2018). Comparing Asset Pricing Models. *Journal of Finance 73*(2), 715–754.

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.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond Risk Premiums with Machine Learning. *Review of Financial Studies 34*(2), 1046–1089.

Bie, S., F. X. Diebold, J. He, and J. Li (2024). Machine Learning and the Yield Curve: Tree-Based Macroeconomic Regime Switching. Technical report, City University of Hong Kong.

Cakici, N., C. Fieberg, T. Neumaier, T. Poddig, and A. Zaremba (2024). Pockets of Predictability: A Replication. *Journal of Finance, Forthcoming*.

Cakici, N., C. Fieberg, T. Neumaier, T. Poddig, and A. Zaremba (2025). The Devil in the Details: How Sensitive Are "Pockets of Predictability" to Methodological Choices? *Critical Finance Review, 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.

Campbell, J. Y. and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies 21*(4), 1509–1531.

Cao, S., W. Jiang, J. Wang, and B. Yang (2024). From Man vs. Machine to Man+Machine: The Art and AI of Stock Analyses. *Journal of Financial Economics 160*, 103910.

Cong, L., G. Feng, J. He, and X. He (2025). Growing the Efficient Frontier on Panel Trees. *Journal of Financial Economics 167*, 104024.

Cong, L. W. (2025). AI for Social Sciences: Generative Modeling Beyond LLMs. Technical report, Cornell University.

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. Technical report, Cornell University.

Cui, L., G. Feng, Y. Hong, and J. Yang (2024). Time-Varying Factor Selection: A Sparse Fused GMM Approach. Technical report, City University of Hong Kong.

Dangl, T. and M. Halling (2012). Predictive Regressions with Time-Varying Coefficients. *Journal of Financial Economics 106*(1), 157–181.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and Long-Horizon Behavioral Factors. *Review of Financial Studies 33*(4), 1673–1736.

DeMiguel, V., J. Gil-Bazo, F. J. Nogales, and A. A. Santos (2023). Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha. *Journal of Financial Economics 150*(3), 103737.

Didisheim, A., S. Ke, B. Kelly, and S. Malamud (2024). APT or "AIPT"? The Surprising Dominance of Large Factor Models. Technical report, Yale University.

Evgeniou, T., A. Guecioueur, 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 (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics 22*(1), 3–25.

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 (2008). Dissecting Anomalies. *Journal of Finance 63*(4), 1653–1678.

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

Farmer, L. E., L. Schmidt, and A. Timmermann (2024). Comment on Cakici, Fieberg, Neumaier, Poddig, and Zaremba: Pockets of Predictability: A Replication. Technical report, University of Viginia.

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.

Feng, G., J. He, N. Polson, and J. Xu (2024). Deep Learning of Characteristics-Sorted Factor Models. *Journal of Financial and Quantitative Analysis 59*(7), 3001–3036.

Feng, G., X. He, Y. Wang, and C. Wu (2025). Predicting Individual Corporate Bond Returns. *Journal of Banking and Finance 171*, 107372.

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 (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.

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti (2024). Deep Learning Statistical Arbitrage. *Management Science, Forthcoming*.

Han, Y., A. He, D. E. Rapach, and G. Zhou (2024). Cross-Sectional Expected Returns: New Fama–MacBeth Regressions in the Era of Machine Learning. *Review of Finance 28*(6), 1807–1831.

He, J. and P. R. Hahn (2023). Stochastic Tree Ensembles for Regularized Nonlinear Regression. *Journal of the American Statistical Association 118*(541), 551–570.

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., H. Mo, C. Xue, and L. Zhang (2021). An Augmented Q-Factor Model with Expected Growth. *Review of Finance 25*(1), 1–41.

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

Huang, J.-Z. and Z. Shi (2023). Machine-Learning-Based Return Predictors and the Spanning Controversy in Macro-Finance. *Management Science 69*(3), 1780–1804.

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.

Jensen, M. C., F. Black, and M. S. Scholes (1972). The Capital Asset Pricing Model: Some Empirical Tests.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023). Machine-Learning the Skill of Mutual Fund Managers. *Journal of Financial Economics 150*(1), 94–138.

Kelly, B., S. Malamud, and K. Zhou (2024). The Virtue of Complexity in Return Prediction. *Journal of Finance 79*(1), 459–503.

Kelly, B. and S. Pruitt (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance 68*(5), 1721–1756.

Lettau, M. and S. Van Nieuwerburgh (2008). Reconciling the Return Predictability Evidence. *Review of Financial Studies 21*(4), 1607–1652.

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.

Li, K., Y. Li, C. Lyu, and J. Yu (2025). How to Dominate the Historical Average. *Review of Financial Studies, Forthcoming*.

Lin, H., C. Wu, and G. Zhou (2018). Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach. *Management Science 64*(9), 4218–4238.

Pástor, L. and R. F. Stambaugh (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy 111*(3), 642–685.

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.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance 68*(4), 1633–1662.

Reinganum, M. R. (1981). Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values. *Journal of Financial Economics 9*(1), 19–46.

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.

Smith, S. C. and A. Timmermann (2022). Have Risk Premia Vanished? *Journal of Financial Economics 145*(2), 553–576.

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

Stambaugh, R. F. and Y. Yuan (2017). Mispricing Factors. *Review of Financial Studies 30*(4), 1270–1315.

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 Predictor Descriptions

Table A.1: **Market Predictors**

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 market liquidity

Table A.2: **Equity Characteristics**

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

Table A.2: Equity Characteristics (Continued)

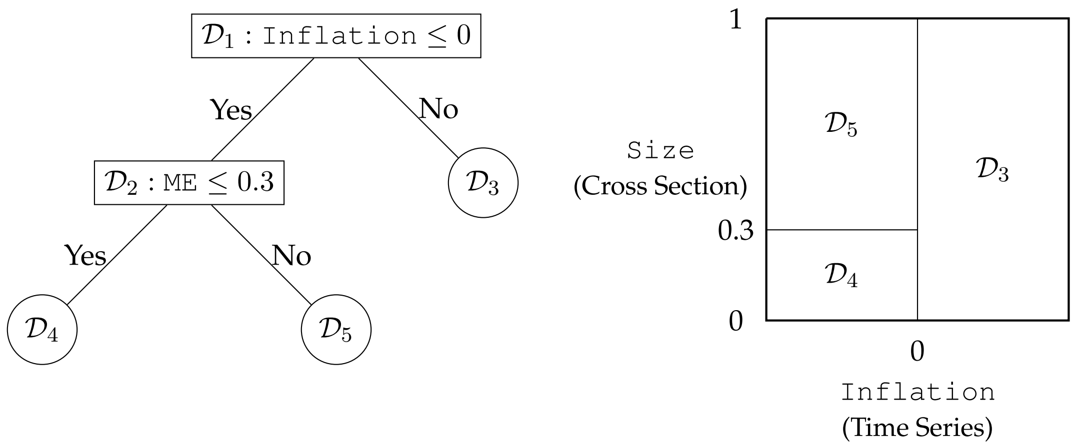
No.	Acronym	Description	Category
22	ill	Illiquidity rolling 3m	Liquidity
23	lev	Leverage	Value
24	lgr	Growth in long-term debt	Investment
25	maxret	Maximum daily returns rolling 3m	Volatility
26	me	Market equity	Size
27	me_ia	Industry-adjusted size	Size
28	mom12m	Momentum rolling 12m	Momentum
29	mom1m	Momentum	Momentum
30	mom36m	Momentum rolling 36m	Momentum
31	mom60m	Momentum rolling 60m	Momentum
32	mom6m	Momentum rolling 6m	Momentum
33	ni	Net stock issues	Investment
34	noa	(Changes in) Net operating assets	Investment
35	op	Operating profitability	Profitability
36	pctacc	Percent operating accruals	Investment
37	pm	Profit margin	Profitability
38	rna	Quarterly return on net operating assets	Profitability
39	roa	Return on assets	Profitability
40	roe	Return on equity	Profitability
41	rsup	Revenue surprise	Momentum
42	rvar_capm	Residual variance - CAPM rolling 3m	Volatility
43	svar	Return variance rolling 3m	Volatility
44	seas1a	Seasonality	Intangibles
45	sgr	Sales growth	Value
46	sp	Sales-to-price	Value
47	std_dolvol	Std of dollar trading volume rolling 3m	Volatility
48	std_turn	Std. of share turnover rolling 3m	Volatility
49	sue	Unexpected quarterly earnings	Momentum
50	turn	Shares turnover	Liquidity
51	zerotrade	Number of zero-trading days rolling 3m	Liquidity

II Illustration for the Second Split

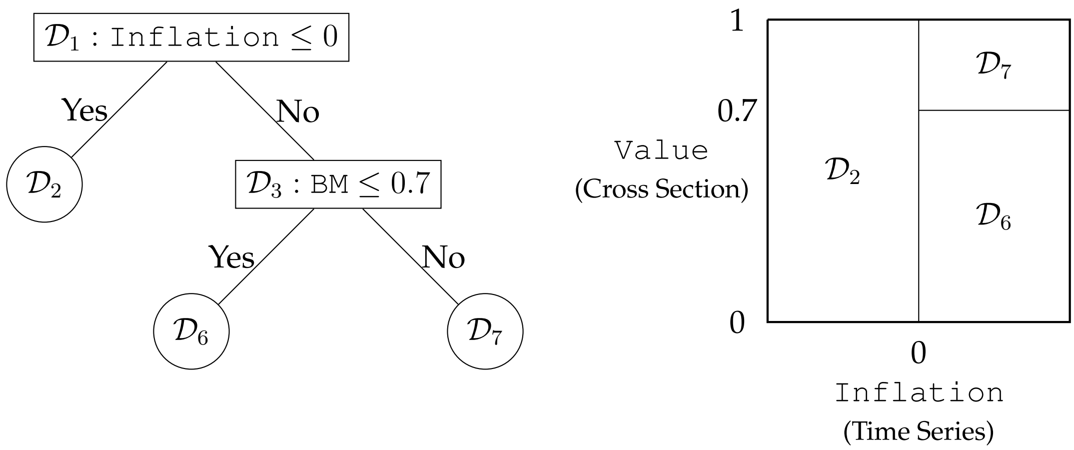
The initial split is determined based on the condition between the demeaning inflation and 0 (see Figure 2). Subsequently, we iteratively explore the optimal choice for the second split, as two candidates in Figure A.1 below. This split can occur either on the left based on size ($\texttt{ME} \leq 0.3$), or on the right corresponding to book-to-market ($\texttt{BM} > 0.7$). Evaluating and comparing all possible combinations of these conditions is necessary to identify the optimal choice. All further partitions are determined similarly, following more details in Section 2.4.

Figure A.1: **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 $\texttt{ME} \leq 0.3$



(b) If splitting node $\mathcal{D}_3$ at $\texttt{BM} \leq 0.7$

III Panel Regression Tree Algorithm

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

Algorithm Panel Regression Tree

procedure PANEL REGRESSION TREE
Input: Stock returns $r_{i,t}$, firm characteristics $z_{i,t-1}$, market 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 $\geq d_{\max}$ **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 the minimal leaf size **then**
            continue.
          else
            Obtain cluster-wise return forecasts as in (1).
            Calculate the cluster-based R_j^2 by (4).
          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 leaf node N (Figure 2).

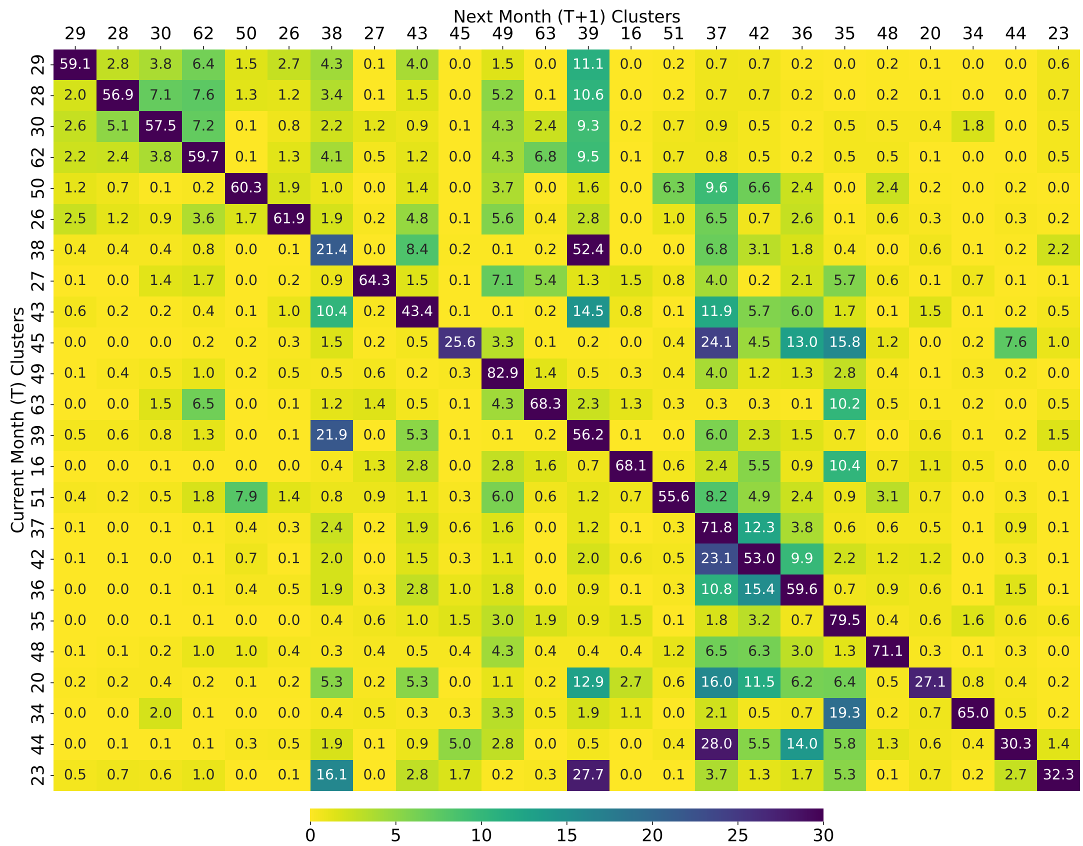
IV Additional Empirical Results

IV.1 Persistence of Clusters

This subsection analyzes the persistence of cross-sectional clustering in Section 4.1. We compute the transition matrix for each period, capturing the likelihood of stocks transitioning between clusters determined by the tree structure in Figure 3. We then average this transition matrix across all months, with Figure A.2 displaying a heatmap of the results. Large diagonal values highlight that many firms consistently remain in the same cluster, indicating stable membership. Off-diagonal values, though fewer, suggest potential shifts in return dynamics, such as momentum (N38 and N39) or bid-ask spread with earnings-to-price ratios (N44 and N37) in Figure 3.

Figure A.2: **Persistence of Clusters**

This figure presents the monthly average transition matrix of clustering results shown in Figure 3. Each row represents the probabilities of transitioning from the current cluster (time T) to other clusters in the next month (time $T+1$). Clusters in the rows and columns are sorted according to their predictability.



IV.2 Predictability Disagreement Anomaly Testing

This subsection provides a detailed breakdown of the long- and short-leg portfolios for the predictability disagreement anomaly in Table 3. Regardless of the number of highly predictable clusters included, the long-leg portfolios in Table A.3 consistently generate significant abnormal returns across both in-sample and out-of-sample periods. In contrast, the short-leg portfolios generally yield insignificant alphas out of the sample. This reinforces the conclusion that our discovered anomaly is primarily driven by stocks with the highest model misspecification risks (long-leg), as highlighted in Section 4.3.

Table A.3: **Testing the Predictability Disagreement Anomaly**

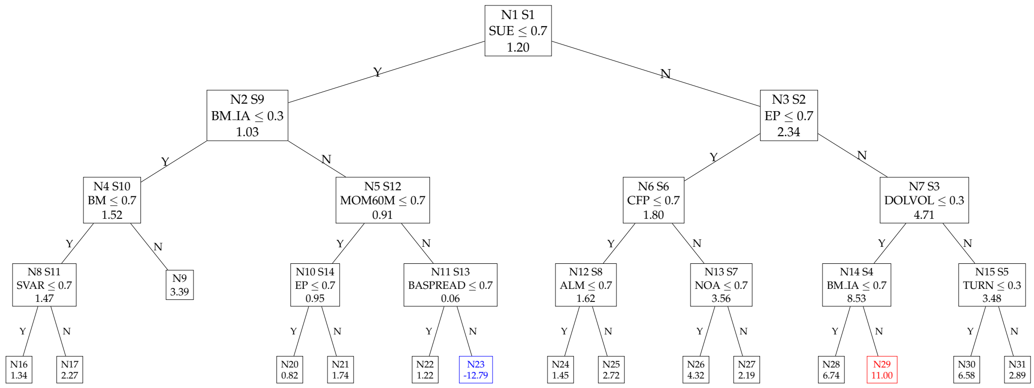
This table reports summary statistics (Panel A) and abnormal returns (Panel B) for the long- and short-leg portfolios corresponding to the long-short strategies in Table 3. The numbers following "T" and "B" denote the number of top and bottom clusters, ranked by R^2_{CMG}, used to construct long and short portfolios. Panel A reports average return (Avg, %), standard deviation (Std, %), and annualized Sharpe ratio (SR). Panel B provides alpha estimates (% with significance by "*") and t-statistics (in parentheses) from various factor models.

	In-Sample (1973 - 2002)						Out-of-Sample (2003 - 2022)					
	T1	T3	T5	B1	B3	B5	T1	T3	T5	B1	B3	B5
					Panel A: Summary Statistics							
Avg	3.41	2.76	1.76	-0.06	-0.26	-0.04	3.25	2.26	2.01	0.46	0.52	0.69
Std	6.28	5.67	5.94	8.38	5.82	5.05	7.03	5.83	6.33	7.18	5.58	5.28
SR	1.88	1.69	1.03	-0.02	-0.16	-0.03	1.60	1.35	1.10	0.22	0.32	0.46
					Panel B: Abnormal Returns							
CAPM	3.05***	2.40***	1.33***	-0.52	-0.66***	-0.44***	2.44***	1.42***	1.08***	-0.60**	-0.39**	-0.20*
	(12.19)	(11.58)	(7.49)	(-1.51)	(-3.32)	(-3.48)	(6.84)	(6.14)	(4.43)	(-2.17)	(-2.39)	(-1.68)
FF3	2.54***	1.97***	1.04***	-0.66*	-0.86***	-0.60***	2.49***	1.45***	1.12***	-0.58**	-0.36**	-0.18*
	(13.08)	(13.90)	(9.02)	(-1.88)	(-4.45)	(-4.89)	(7.28)	(7.08)	(5.58)	(-2.15)	(-2.45)	(-1.70)
FF5	2.37***	1.82***	0.98***	-0.56	-0.90***	-0.62***	2.51***	1.62***	1.21***	-0.52*	-0.38**	-0.18
	(12.49)	(13.03)	(8.32)	(-1.56)	(-4.59)	(-4.90)	(7.14)	(7.82)	(5.88)	(-1.88)	(-2.50)	(-1.62)
FF5+MOM+IVOL	2.47***	1.96***	1.11***	-0.02	-0.58***	-0.43***	2.54***	1.68***	1.28***	-0.47*	-0.37**	-0.17
	(12.71)	(13.96)	(9.54)	(-0.05)	(-3.07)	(-3.50)	(7.20)	(8.43)	(6.54)	(-1.71)	(-2.43)	(-1.54)
Q5	2.42***	1.88***	1.10***	-0.06	-0.53**	-0.40***	2.62***	1.75***	1.48***	-0.48*	-0.34**	-0.13
	(10.03)	(10.52)	(7.69)	(-0.14)	(-2.24)	(-2.70)	(7.27)	(8.58)	(7.84)	(-1.71)	(-2.15)	(-1.12)
BS6	2.10***	1.74***	0.98***	-0.27	-0.76***	-0.55***	2.45***	1.64***	1.30***	-0.45*	-0.38**	-0.18*
	(10.19)	(11.54)	(7.91)	(-0.74)	(-3.74)	(-4.16)	(6.96)	(8.23)	(6.85)	(-1.69)	(-2.54)	(-1.69)
DHS3	3.20***	2.58***	1.65***	0.31	-0.12	-0.14	2.57***	1.67***	1.31***	-0.36	-0.32*	-0.11
	(11.10)	(10.74)	(8.06)	(0.78)	(-0.55)	(-0.95)	(7.02)	(7.31)	(5.43)	(-1.30)	(-1.92)	(-0.94)
SY4	2.40***	1.88***	1.06***	0.08	-0.58***	-0.43***	3.30***	2.38***	1.71***	-0.56	-0.40**	-0.15
	(10.75)	(11.57)	(8.18)	(0.21)	(-2.75)	(-3.31)	(7.47)	(10.64)	(9.33)	(-1.60)	(-2.01)	(-1.08)

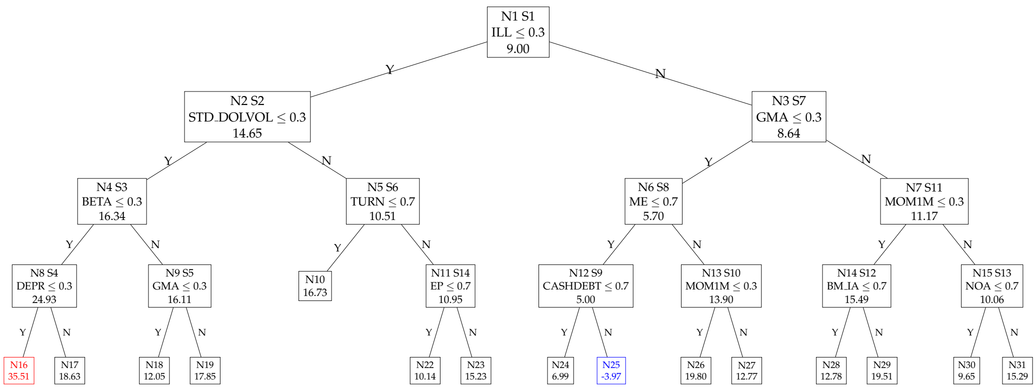
IV.3 Further Cross-Sectional Tree Structures across Market Regimes

Figure A.3: **Tree-Based Cluster (CS Clusters across Market Regimes)**

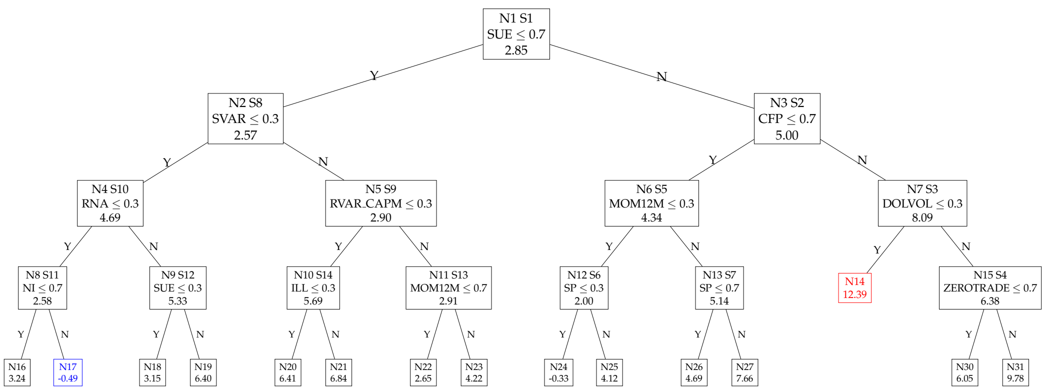
This figure shows the further cross-sectional split tree-based clustering structure under each regime determined by Figure 6 using monthly data from 1973 to 2022. Each tree splits stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. The terminal leaves correspond to clusters identified by the interaction of market predictors and 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.



(a) Regime I: $\mathbb{1}\{\texttt{X\_DY} \leq 0\}$



(b) Regime II: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \leq 0\}$



(c) Regime III: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} > 0\}$

IV.4 Structural Breaks

Table A.4: **Cluster-Wise Performance (+ Structural Break)**

This table summarizes the cluster-based information under each continuous period within eight panels corresponding to the structural breaks and further cross-sectional partitions. Each panel counts the number of observations for each cluster ("# obs") and displays the return predictability (R^2 values in %) for each cluster. Besides, "Avg" and "SR" denote the monthly average return (in %) and annualized Sharpe ratio for equal/value-weighted (EW/VW) portfolios based on all observations. Each regime of values is arranged in the descending order of R^2 from left to right.

197301-197810	N30	N13	N23	N31	N29	N25	N17	N22	N21	N18	N16	N28	N20	N24	N19	
# obs	2,289	3,533	3,040	3,115	4,556	8,776	2,958	6,011	13,432	7,561	10,595	2,836	9,611	39,735	135,631	
R^2	41.22	20.97	20.69	19.42	17.85	17.40	17.02	15.02	14.37	13.43	11.24	11.11	10.76	10.57	8.09	
Avg_{EW}	4.76	1.20	-0.39	0.38	0.24	1.72	0.09	-0.16	0.00	0.04	0.20	-0.24	0.18	0.94	0.66	
SR_{EW}	1.21	0.46	-0.17	0.12	0.08	0.78	0.04	-0.08	0.00	0.02	0.07	-0.08	0.10	0.44	0.30	
Avg_{VW}	3.27	0.94	-0.77	0.52	0.10	1.06	-0.51	-0.76	-0.46	-0.13	-0.09	-0.33	-0.31	-0.08	-0.06	
SR_{VW}	0.93	0.37	-0.38	0.17	0.03	0.50	-0.22	-0.42	-0.28	-0.09	-0.04	-0.11	-0.21	-0.04	-0.04	
197811-198310	N7	N26	N10	N25	N19	N23	N27	N22	N17	N18	N24	N16				
# obs	2,574	14,494	2,683	2,647	2,787	8,803	6,487	21,523	16,138	17,055	12,538	133,187				
R^2	16.62	14.59	14.24	13.77	13.27	12.84	9.32	8.36	8.22	7.27	7.01	4.93				
Avg_{EW}	3.00	3.06	1.15	2.21	1.95	1.57	2.01	0.73	0.86	1.23	1.77	1.35				
SR_{EW}	1.71	1.79	0.69	1.00	1.04	0.65	1.13	0.34	0.42	0.79	1.22	0.86				
Avg_{VW}	1.83	1.16	0.52	1.31	1.27	0.79	1.09	0.45	-0.09	0.77	1.02	0.66				
SR_{VW}	1.17	0.83	0.32	0.59	0.79	0.34	0.65	0.22	-0.05	0.61	0.80	0.57				
198311-198811	N31	N29	N30	N13	N11	N24	N28	N20	N25	N21	N4					
# obs	23,452	2,405	43,886	2,211	3,672	4,119	7,902	173,028	2,790	25,031	2,204					
R^2	18.48	15.66	14.14	13.55	13.33	10.37	9.12	4.76	3.75	-4.03	-5.57					
Avg_{EW}	1.03	0.04	0.55	0.54	-0.68	0.40	-0.16	-0.47	0.41	0.31	-0.58					
SR_{EW}	0.59	0.02	0.34	0.34	-0.29	0.31	-0.08	-0.28	0.29	0.19	-0.43					
Avg_{VW}	0.79	0.36	0.65	0.65	-0.49	0.44	0.26	-0.24	0.14	0.48	-0.99					
SR_{VW}	0.49	0.16	0.44	0.39	-0.21	0.35	0.15	-0.13	0.10	0.29	-0.45					
198812-199702	N29	N30	N28	N22	N25	N18	N31	N21	N23	N24	N27	N20	N19	N17	N16	N26
# obs	3,020	5,867	3,061	32,217	7,607	4,005	18,437	4,465	17,705	50,167	4,883	53,761	75,829	223,379	15,109	12,943
R^2	12.20	10.04	9.02	6.46	5.87	5.59	4.93	4.51	3.19	2.72	2.53	2.28	2.03	1.05	-0.22	-0.50
Avg_{EW}	4.68	2.43	3.32	1.09	3.01	1.89	1.83	-1.12	0.82	1.59	3.21	0.56	0.89	0.20	0.37	1.59
SR_{EW}	2.63	2.56	2.79	1.11	2.20	1.10	1.47	-0.48	0.76	1.29	1.88	0.48	0.81	0.14	0.50	1.00
Avg_{VW}	3.60	1.61	2.77	0.96	1.67	0.92	1.22	-0.62	0.90	1.16	1.56	0.66	0.77	0.35	0.55	0.34
SR_{VW}	2.60	1.57	2.47	0.96	1.36	0.51	1.01	-0.26	0.94	1.04	0.93	0.66	0.66	0.23	0.54	0.20
199703-200303	N29	N9	N23	N28	N16	N21	N22	N31	N27	N17	N20	N30	N24	N26	N25	
# obs	2,764	3,722	2,357	5,237	4,965	7,454	3,330	27,056	14,671	4,908	100,615	68,455	23,830	19,648	150,580	
R^2	18.56	16.10	12.06	10.96	10.57	8.37	8.34	7.95	6.84	5.52	4.87	4.83	4.64	4.27	2.71	
Avg_{EW}	2.13	-2.44	-2.72	2.51	-0.33	-1.33	-0.87	2.00	0.44	-0.32	-0.39	0.76	-0.25	0.31	0.36	
SR_{EW}	0.87	-0.54	-0.76	1.03	-0.10	-0.32	-0.33	1.55	0.13	-0.13	-0.13	0.61	-0.09	0.14	0.21	
Avg_{VW}	2.74	-3.40	-2.75	2.31	-0.91	-1.04	-1.13	0.75	0.17	-0.75	-1.08	0.41	0.04	0.42	-0.02	
SR_{VW}	0.83	-0.79	-0.75	0.80	-0.28	-0.26	-0.43	0.53	0.05	-0.31	-0.39	0.28	0.02	0.24	-0.02	
200304-200902	N14	N31	N26	N22	N25	N30	N27	N23	N20	N17	N24	N19	N21	N16	N18	
# obs	2,263	2,510	5,248	2,196	4,985	2,893	18,704	8,653	2,399	19,465	56,880	13,973	43,241	36,851	100,889	
R^2	18.15	16.18	15.90	13.93	11.19	10.74	10.10	9.31	8.92	8.20	6.98	6.47	5.93	5.68	3.80	
Avg_{EW}	0.08	0.07	0.04	3.57	0.32	-0.63	0.44	2.35	-0.79	0.29	-0.01	0.03	0.21	-0.02	0.01	
SR_{EW}	0.03	0.03	0.02	1.82	0.18	-0.20	0.31	1.41	-0.49	0.14	-0.01	0.03	0.13	-0.01	0.01	
Avg_{VW}	0.99	-0.40	0.12	1.96	0.08	-1.41	0.51	1.66	-0.86	-0.17	-0.14	0.14	0.26	-0.26	0.13	
SR_{VW}	0.34	-0.16	0.08	1.00	0.04	-0.51	0.40	0.93	-0.39	-0.09	-0.12	0.10	0.15	-0.14	0.09	
200903-201708	N28	N25	N20	N27	N24	N23	N29	N16	N30	N21	N26	N17	N31	N19	N22	N18
# obs	17,530	3,675	15,114	9,697	34,825	6,396	3,082	20,735	3,689	23,205	35,480	35,948	4,407	32,360	23,885	104,833
R^2	16.22	16.01	15.40	11.82	11.61	9.48	9.27	8.69	8.42	7.17	6.23	5.25	5.16	4.77	4.26	1.94
Avg_{EW}	1.78	1.87	2.35	3.24	1.81	4.23	1.08	1.59	1.43	2.01	1.37	1.41	1.69	1.99	1.08	0.89
SR_{EW}	0.84	1.41	0.83	2.56	1.71	2.14	0.54	1.38	0.74	1.07	1.02	0.98	0.91	1.44	0.60	0.62
Avg_{VW}	1.75	2.07	2.25	2.13	1.35	2.82	1.36	1.36	1.23	1.73	1.60	1.42	1.37	1.69	0.94	1.27
SR_{VW}	0.86	1.46	0.91	1.44	1.36	1.32	0.78	1.36	0.64	1.15	1.11	1.20	0.78	1.46	0.52	1.02
201709-202212	N24	N26	N7	N19	N11	N25	N20	N18	N27	N21	N17	N16				
# obs	4,772	1,963	3,447	2,685	2,244	25,418	18,294	17,411	29,653	24,949	75,469	21,048				
R^2	12.93	10.91	10.15	9.20	9.03	8.29	7.65	5.39	5.33	4.34	2.79	-1.29				
Avg_{EW}	1.15	1.07	0.66	0.84	0.97	0.90	0.67	0.64	0.74	0.96	-0.09	-0.05				
SR_{EW}	0.37	0.33	0.38	0.42	0.33	0.51	0.25	0.35	0.38	0.40	-0.05	-0.02				
Avg_{VW}	0.00	1.30	1.27	0.80	0.77	0.77	0.65	0.70	0.60	0.86	0.33	0.58				
SR_{VW}	0.00	0.47	0.66	0.40	0.30	0.52	0.26	0.37	0.37	0.38	0.17	0.28				

Table A.5: **Evaluating Return Predictability (+ Structural Break)**

This table reports return predictability (R^2, %) for different predictive methods under various periods. We present full-sample results from the tree-based cluster model, incorporating structural breaks and cross-sectional partitions. For each regime, we show four samples: Aggregate, High, Medium, and Low, based on the predictive rankings within the tree clusters. "-" represents empty samples due to `ME` being selected as the split criterion.

	Sample A: All Stocks			Sample B: Large-Cap		
197301-197810	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	11.22	5.37	6.70	12.35	5.25	6.64
High	25.66	16.35	19.13	20.39	11.61	13.53
Medium	12.24	5.11	6.69	12.30	5.13	6.65
Low	8.12	3.61	4.53	11.14	4.47	5.50
197811-198310	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	6.92	5.35	6.59	9.05	6.93	8.33
High	14.88	12.29	14.23	12.26	9.98	11.74
Medium	8.97	7.13	8.52	9.67	7.31	8.96
Low	4.92	3.61	4.70	6.70	5.04	5.89
198311-198811	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	5.76	1.54	2.11	14.68	7.80	7.10
High	19.00	11.65	6.33	19.00	11.65	6.33
Medium	6.02	1.54	2.54	13.28	6.55	7.35
Low	-4.02	-3.36	-5.67	-	-	-
198812-199702	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	1.71	1.06	1.67	2.61	1.55	2.45
High	8.47	6.37	8.07	7.58	4.27	6.57
Medium	1.63	0.94	1.57	2.28	1.36	2.16
Low	-0.41	0.34	-0.03	-1.28	-0.34	-0.57
199703-200303	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	4.94	1.33	2.29	5.64	1.84	2.58
High	11.19	4.00	5.81	13.83	5.92	7.73
Medium	5.11	1.35	2.41	6.28	1.88	2.75
Low	2.73	0.53	0.98	2.73	0.54	0.87
200304-200902	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	6.30	3.74	5.54	8.47	4.69	7.04
High	15.34	11.40	13.79	16.62	12.40	15.13
Medium	6.89	4.11	6.04	8.13	3.82	6.40
Low	3.79	1.85	3.33	4.25	2.21	3.73
200903-201708	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	6.23	3.49	5.43	9.31	5.41	7.92
High	16.22	9.81	14.64	20.53	12.49	18.28
Medium	7.60	4.59	6.74	8.96	5.32	7.60
Low	1.96	0.35	1.39	3.62	1.31	2.69
201709-202212	OLS	Lasso	Ridge	OLS	Lasso	Ridge
Aggregate	3.98	1.89	3.27	7.17	2.33	4.73
High	11.27	6.09	8.68	10.84	5.60	8.20
Medium	4.43	1.72	3.30	6.34	1.59	3.96
Low	-1.31	1.63	1.32	-	-	-

IV.5 Investment Performance with Transaction Costs

This subsection re-examines the performance of forecast-implied portfolios based on cross-sectional clustering (Table 6) after accounting for transaction costs. While these costs are empirically unobservable and inherently difficult to forecast, we adopt a conservative approach by imposing a fixed cost of 10 basis points per trade, applied uniformly across all rebalancing periods. This adjustment allows us to assess the robustness of the documented investment gains under realistic trading conditions. Table A.6 below presents performance metrics for transaction costs adjusted forecast-implied portfolios.

First, the results show that stocks with high predictability generate less significant out-of-sample (OOS) investment gains compared with Table 6. For instance, in Panel A, the high-predictability subsample yields an OOS Sharpe ratio of 0.77 and a marginally significant alpha of 0.36%, while the low-predictability subsample achieves only a Sharpe ratio of 0.43 and a negative alpha. Similar patterns emerge in Panels B and C, where high-predictability stocks deliver OOS Sharpe ratios of 1.59 and 1.80, with alphas of 1.74% and 2.20%, respectively—substantially outperforming medium- and low-predictability portfolios.

Moreover, the aggregate model outperforms the homogeneous model, highlighting the economic value of clustering. In Panel C, the Aggregate model achieves a highly significant OOS alpha of 0.69% and a Sharpe ratio of 1.06, compared to 0.15% (no significance) and 0.67 for the global model. This gap is particularly pronounced in forecast-weighted portfolios, where the aggregate model's superior return magnitude forecasts drive performance. Large-cap sub-samples stably outperform their global counterparts across all weighting schemes. However, the margin is narrower than the full sample, likely due to lower return predictability among large stocks.

Table A.6: **Forecast-Implied Investment Performance (CS Cluster)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 10). Panels A and B present sign-adjusted value-/equal-weighted portfolios, while Panel C reports forecast-weighted portfolios, all constructed by observations in specific samples (Equation 9). We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap sub-samples of Global and Aggregate). Each panel includes five metrics: monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), Jensen's alpha (%), and maximum drawdown (MDD, %).

	In-Sample (1973 - 2002)					Out-of-Sample (2003 - 2022)				
	Panel A: Sign-adjusted Value-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	0.54	3.98	0.47	0.20^{***}	21.48	0.62	4.17	0.52	-0.13^{***}	17.21
Aggregate	0.46	4.03	0.39	0.11^{**}	21.24	0.69	4.08	0.59	-0.05^{**}	16.58
High	1.83	4.83	1.31	1.48^{***}	21.94	1.11	5.01	0.77	0.36^{*}	24.17
Medium	0.46	4.08	0.39	0.11^{**}	21.54	0.71	4.15	0.59	-0.04	16.92
Low	0.05	4.05	0.05	-0.24^{**}	15.59	0.51	4.12	0.43	-0.19^{*}	15.23
Global-Large	0.52	3.99	0.45	0.18^{***}	21.55	0.62	4.13	0.52	-0.12^{***}	16.98
Aggregate-Large	0.42	4.09	0.36	0.07	21.65	0.69	4.09	0.58	-0.05^{**}	16.29
	Panel B: Sign-adjusted Equal-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.05	4.02	0.91	0.78^{***}	16.88	0.77	5.14	0.52	-0.06	22.82
Aggregate	1.07	3.21	1.16	0.87^{***}	13.73	0.83	4.17	0.69	0.16	20.82
High	2.78	5.73	1.68	2.39^{***}	25.94	2.60	5.64	1.59	1.74^{***}	24.13
Medium	1.00	3.17	1.09	0.81^{***}	13.56	0.76	3.93	0.67	0.16	21.44
Low	0.41	3.80	0.38	0.16	14.33	0.57	5.06	0.39	-0.25	19.30
Global-Large	0.68	4.34	0.54	0.34^{***}	24.12	0.68	4.78	0.49	-0.16^{*}	20.66
Aggregate-Large	0.70	3.97	0.61	0.38^{***}	22.47	0.76	4.54	0.58	-0.04	19.90
	Panel C: Forecast-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.48	4.69	1.09	1.16^{***}	20.62	1.02	5.32	0.67	0.15	23.14
Aggregate	1.91	4.26	1.55	1.64^{***}	16.62	1.46	4.77	1.06	0.69^{***}	21.41
High	3.32	6.08	1.89	2.91^{***}	26.33	3.12	5.99	1.80	2.20^{***}	24.92
Medium	1.64	4.05	1.40	1.40^{***}	16.20	1.15	4.57	0.87	0.43^{***}	22.06
Low	0.78	4.93	0.55	0.43^{***}	18.05	0.78	5.60	0.48	-0.13	20.08
Global-Large	0.94	4.77	0.68	0.56^{***}	24.91	0.82	4.93	0.57	-0.05	20.89
Aggregate-Large	1.04	4.37	0.82	0.70^{***}	23.58	0.90	4.82	0.65	0.06	20.71

Internet Appendix

IA.I Robustness Check

Amid rising concerns about result stability in predictability research, as highlighted by Cakici et al. (2025) in the "pockets of predictability" framework (Farmer et al., 2023, 2024), we conduct rigorous robustness checks on our tree-based clustering method. Our framework includes adjustable parameters like sample periods, tree depth, minimum leaf size, and return winsorization thresholds. The results below confirm the robustness of our core findings through cross-sectional partitioning.

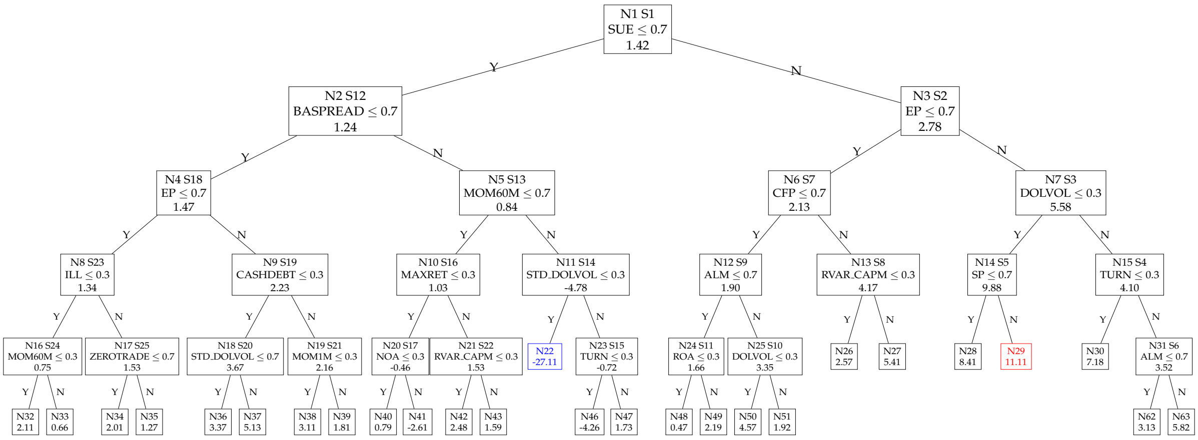
As detailed in Section 3.2, our cross-sectional analyses use a 30-year rolling window updated every five years. The optimal splitting variables show consistent robustness across sub-samples — the first two splits in tree structures by different periods (Figure IA.1 for 1978–2007, 1983–2012, and 1988–2007) still choose earnings surprises and value characteristics (`EP` and `CFP`) as Figure 3 — the interpretation remains stable. The high-predictability clusters consistently exhibit three key traits: high earnings surprises, high values (`EP`, `SP`, or `BM`), and low liquidity (`DOLVOL` or `TURN`). The order of these splits varies slightly without affecting cluster interpretability.

Moreover, the robustness extends to cluster-wise performance metrics. As shown in Table IA.1-IA.3 (similar as Table 1), predictability dispersion persists across specifications: the top cluster consistently delivers in-sample R^2 almost exceeding 10%, while the bottom cluster shows negative values. The relationship between predictability rankings and investment performance remains consistent, as shown in Section 4.

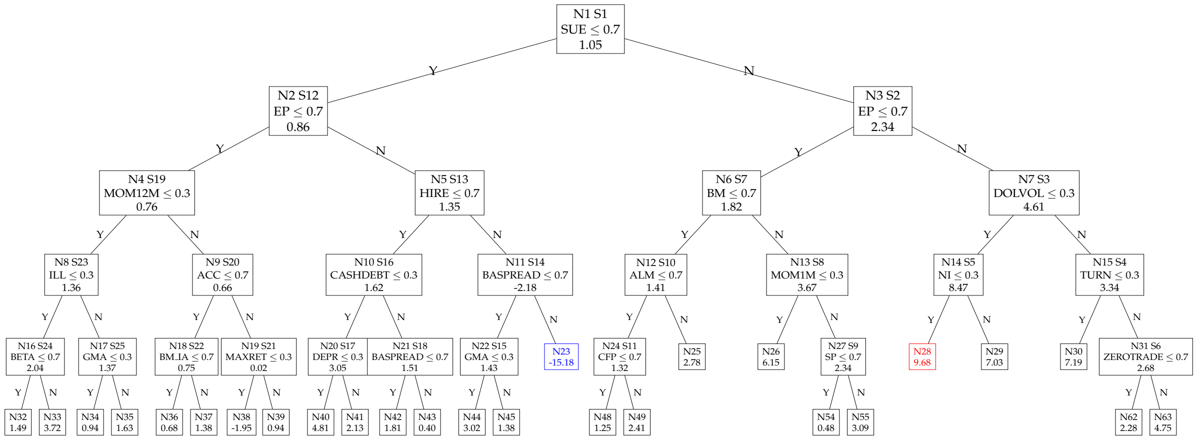
Additional sensitivity analyses, including variations in tree depth, minimum leaf observations, and winsorization levels, demonstrate consistent stability in selected firm characteristics. While splitting sequences differ slightly, the same variables reappear across specifications. These findings confirm the heterogeneous nature of return predictability and uphold the replicability of our cluster definitions. The investment performance of highly predictable clusters withstands rigorous robustness checks, mitigating concerns about data-mining biases.

Figure IA.1: **Tree-Based Cluster (CS Clusters across Other Sample Periods)**

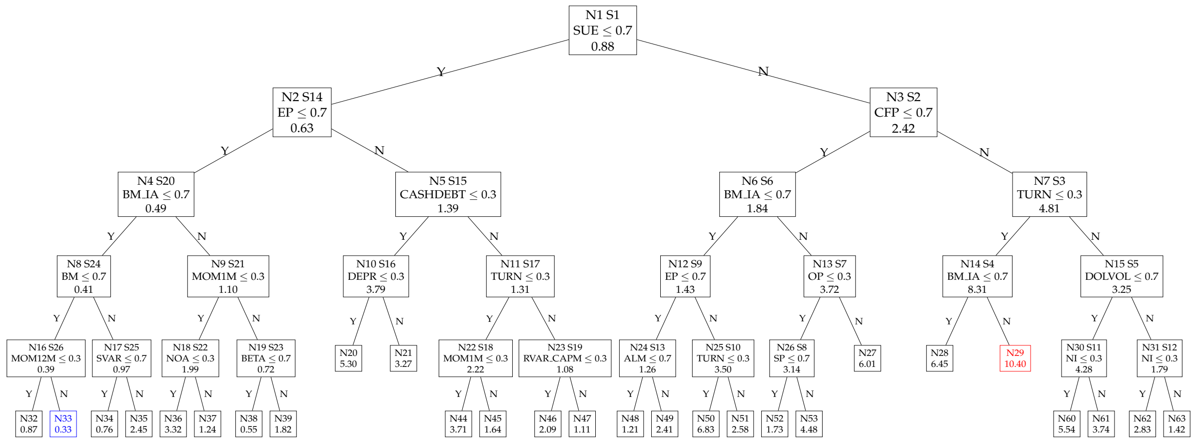
This figure illustrates the cross-sectional tree-based clustering derived from the other three monthly data sample periods (1978–2007, 1983-2012, 1988-2017). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. Terminal leaves represent clusters defined by specific characteristic intervals. Each node, including intermediate and terminal leaves, is labeled with an ID (N#) where Ni-th node's child nodes have ID N$(2i)$ and N$(2i+1)$ respectively. S# indicates the sequence of splitting order. Cluster-specific R^2 values, indicating return predictability, are provided for each node.



(a) Sample 1: 1978 - 2007



(b) Sample 2: 1983 - 2012



(c) Sample 3: 1988 - 2017

Table IA.1: **Performance of Asset Clusters (1978 - 2007)**

This table summarizes key metrics for each cluster from the cross-sectional tree structure (a) in Figure IA.1. Panel A reports the number of observations (# obs), return predictability (R^2 in %) based on cluster-wise (R^2_{C}) and global (R^2_{G}) models, along with their differences ($R^2_{\mathrm{CMG}} = R^2_{\mathrm{C}} - R^2_{\mathrm{G}}$). Panel B presents the monthly average return (Avg, %) and annualized Sharpe ratio (SR) for equal- and value-weighted (EW/VW) portfolios. Leaf nodes are ordered by descending R^2_{C}.

	Panel A: Summary Statistics				Panel B: Profitability			
Leaf	# obs	R^2_{C}	R^2_{G}	R^2_{CMG}	$\mathrm{Avg}_{\mathrm{EW}}$	$\mathrm{SR}_{\mathrm{EW}}$	$\mathrm{Avg}_{\mathrm{VW}}$	$\mathrm{SR}_{\mathrm{VW}}$
N29	11,186	11.11	7.13	3.98	4.64	2.60	3.68	2.37
N28	14,820	8.41	6.53	1.88	2.94	2.10	2.42	1.81
N30	20,250	7.18	6.29	0.89	2.21	1.94	1.67	1.48
N63	10,828	5.82	5.12	0.70	2.85	1.63	2.18	1.19
N27	17,539	5.41	3.60	1.82	2.99	1.71	1.97	0.93
N37	12,430	5.13	3.35	1.78	1.85	0.91	1.28	0.66
N50	16,710	4.57	2.40	2.16	3.53	1.67	2.61	1.44
N36	23,189	3.37	2.91	0.46	0.97	0.58	1.05	0.54
N62	60,075	3.13	2.90	0.23	1.61	1.11	1.10	0.80
N38	96,666	3.11	2.64	0.46	1.58	0.96	1.21	0.74
N26	10,877	2.57	3.26	-0.70	1.48	1.19	0.64	0.46
N42	10,897	2.48	1.74	0.74	0.37	0.29	0.47	0.25
N49	155,137	2.19	1.86	0.33	1.43	0.91	0.75	0.54
N32	17,257	2.11	1.20	0.91	0.92	0.53	0.99	0.60
N34	272,689	2.01	1.36	0.65	0.35	0.17	0.40	0.21
N51	14,288	1.92	1.74	0.18	2.35	1.13	1.46	0.65
N39	219,773	1.81	1.74	0.08	0.80	0.68	0.67	0.59
N47	19,614	1.73	1.28	0.45	-0.44	-0.20	-0.36	-0.14
N43	324,644	1.59	0.99	0.60	0.08	0.04	-0.32	-0.14
N35	285,072	1.27	1.24	0.03	0.42	0.30	0.37	0.30
N40	11,843	0.79	0.75	0.04	0.26	0.14	0.44	0.18
N33	111,851	0.66	0.56	0.10	0.40	0.26	0.37	0.27
N48	31,569	0.47	0.64	-0.17	0.77	0.35	0.88	0.40
N41	10,799	-2.61	1.72	-4.34	0.46	0.34	-0.03	-0.02
N46	11,603	-4.26	1.56	-5.82	0.11	0.08	0.02	0.01
N22	11,122	-27.11	1.22	-28.32	-0.80	-0.32	-0.67	-0.25

Table IA.2: **Performance of Asset Clusters (1983 - 2012)**

This table summarizes key metrics for each cluster from the cross-sectional tree structure (b) in Figure IA.1. Panel A reports the number of observations (# obs), return predictability (R^2 in %) based on cluster-wise (R^2_{C}) and global (R^2_{G}) models, along with their differences ($R^2_{\mathrm{CMG}} = R^2_{\mathrm{C}} - R^2_{\mathrm{G}}$). Panel B presents the monthly average return (Avg, %) and annualized Sharpe ratio (SR) for equal- and value-weighted (EW/VW) portfolios. Leaf nodes are ordered by descending R^2_{C}.

	Panel A: Summary Statistics				Panel B: Profitability			
Leaf	# obs	R^2_{C}	R^2_{G}	R^2_{CMG}	$\mathrm{Avg}_{\mathrm{EW}}$	$\mathrm{SR}_{\mathrm{EW}}$	$\mathrm{Avg}_{\mathrm{VW}}$	$\mathrm{SR}_{\mathrm{VW}}$
N28	13,950	9.68	6.49	3.19	3.47	2.47	2.78	2.09
N30	19,632	7.19	6.24	0.95	2.34	1.84	1.84	1.53
N29	13,023	7.03	4.96	2.07	3.27	2.14	2.51	1.82
N26	14,351	6.15	2.68	3.47	4.01	1.67	1.65	0.63
N40	12,694	4.81	3.02	1.79	0.87	0.51	0.57	0.30
N63	20,640	4.75	4.35	0.39	1.83	1.37	1.06	0.76
N33	14,614	3.72	0.95	2.77	-0.28	-0.10	-0.43	-0.17
N55	17,049	3.09	2.10	0.99	2.54	1.34	1.78	0.79
N44	15,506	3.02	2.12	0.90	0.83	0.59	0.74	0.40
N25	14,563	2.78	1.76	1.02	2.84	1.30	1.88	0.88
N49	16,233	2.41	2.35	0.06	1.77	1.14	0.77	0.52
N62	53,955	2.28	1.72	0.56	1.59	0.98	1.15	0.80
N41	18,264	2.13	2.36	-0.23	1.29	0.60	1.27	0.44
N42	249,510	1.81	1.71	0.10	0.98	0.72	0.77	0.61
N35	214,544	1.63	0.93	0.70	0.06	0.03	-0.17	-0.07
N32	13,763	1.49	0.13	1.36	0.74	0.36	0.34	0.17
N45	52,046	1.38	1.14	0.25	0.69	0.39	0.91	0.52
N37	108,011	1.38	1.27	0.11	1.01	0.67	0.85	0.60
N48	188,053	1.25	0.91	0.34	1.18	0.74	0.68	0.50
N39	141,082	0.94	0.54	0.39	-0.06	-0.03	0.03	0.01
N34	114,791	0.94	0.69	0.25	-0.41	-0.16	-0.91	-0.37
N36	373,551	0.68	0.65	0.03	0.15	0.10	0.41	0.31
N54	11,821	0.48	1.42	-0.94	1.40	0.85	1.05	0.62
N43	53,455	0.40	1.01	-0.60	0.75	0.45	0.18	0.09
N38	20,075	-1.95	0.95	-2.90	0.01	0.01	0.08	0.05
N23	17,782	-15.18	0.96	-16.14	0.15	0.08	0.04	0.01

Table IA.3: **Performance of Asset Clusters (1988 - 2017)**

This table summarizes key metrics for each cluster from the cross-sectional tree structure (c) in Figure IA.1. Panel A reports the number of observations (# obs), return predictability (R^2 in %) based on cluster-wise (R^2_{C}) and global (R^2_{G}) models, along with their differences ($R^2_{\mathrm{CMG}} = R^2_{\mathrm{C}} - R^2_{\mathrm{G}}$). Panel B presents the monthly average return (Avg, %) and annualized Sharpe ratio (SR) for equal- and value-weighted (EW/VW) portfolios. Leaf nodes are ordered by descending R^2_{C}.

	Panel A: Summary Statistics				Panel B: Profitability			
Leaf	# obs	R^2_{C}	R^2_{G}	R^2_{CMG}	$\mathrm{Avg}_{\mathrm{EW}}$	$\mathrm{SR}_{\mathrm{EW}}$	$\mathrm{Avg}_{\mathrm{VW}}$	$\mathrm{SR}_{\mathrm{VW}}$
N29	14,133	10.40	6.38	4.02	4.16	2.73	3.04	1.96
N50	15,267	6.83	5.19	1.64	1.72	1.58	1.34	1.09
N28	18,956	6.45	4.80	1.65	2.74	2.22	1.98	1.52
N27	15,560	6.01	4.33	1.68	2.91	1.87	1.38	0.82
N60	13,085	5.54	4.16	1.37	3.05	1.93	2.12	1.42
N20	15,457	5.30	2.77	2.53	0.92	0.60	0.63	0.35
N53	13,754	4.48	2.34	2.14	3.56	1.55	2.07	0.77
N61	20,718	3.74	3.21	0.53	2.74	1.61	2.05	1.34
N44	34,249	3.71	2.57	1.14	1.82	1.29	1.41	0.93
N36	46,729	3.32	0.95	2.36	2.98	1.19	2.25	0.87
N21	19,769	3.27	2.38	0.90	1.33	0.66	1.37	0.48
N62	14,817	2.83	1.82	1.01	1.41	1.03	0.94	0.74
N51	27,003	2.58	2.22	0.36	1.43	1.03	1.19	0.85
N35	12,927	2.45	0.30	2.15	-1.30	-0.56	-1.92	-0.63
N49	12,255	2.41	1.45	0.96	2.20	1.06	1.53	0.67
N46	95,412	2.09	2.02	0.07	1.02	0.84	0.82	0.68
N39	41,412	1.82	0.63	1.19	0.26	0.11	0.76	0.34
N52	13,413	1.73	1.44	0.29	2.37	1.09	1.27	0.56
N45	87,795	1.64	1.22	0.42	0.63	0.62	0.55	0.48
N63	20,776	1.42	1.22	0.20	1.24	0.78	0.91	0.68
N37	69,495	1.24	0.73	0.51	1.01	0.42	0.56	0.25
N48	186,935	1.21	1.02	0.18	1.26	0.86	0.80	0.66
N47	147,541	1.11	0.74	0.37	0.70	0.38	0.93	0.51
N32	169,088	0.87	0.43	0.43	-0.65	-0.30	0.07	0.03
N34	28,500	0.76	0.59	0.17	0.22	0.16	0.18	0.10
N38	124,152	0.55	0.58	-0.03	0.48	0.33	0.65	0.56
N33	457,756	0.33	0.38	-0.05	0.31	0.20	0.53	0.43

IA.II Investment Gains within Time-Varying Modeling

This section examines the investment performance of forecast-implied portfolios by considering the regime-dependent time variation modeling. The empirical results are robust to transaction cost and are consistent with the cross-sectional analysis (Section 6 and Appendix IV.5): highly predictable clusters outperform their medium- and low-predictability counterparts, while the aggregated heterogeneous forecast model demonstrates superior performance to the globally homogeneous model.

Table IA.4: **Forecast-Implied Investment Performance (TS+CS Cluster)**

This table reports full-sample and large-cap sub-sample baseline long-short investment performance for forecast-implied portfolios (Equation 10). The first two panels show sign-adjusted value- and equal-weighted portfolios, while Panel C reports forecast-weighted portfolios, all constructed from observations in specific samples (Equation 9). We show five samples: Global (no clustering), Aggregate (aggregation of clustering results), and High, Medium, and Low, based on predictive rankings within the tree clusters. We show results of monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), Jensen's alpha (in %), and monthly maximum drawdown (MDD, %).

	Sample A: All Stocks					Sample B: Large-Cap				
	Panel A: Sign-adjusted Value-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	0.81	2.88	0.97	0.57***	13.61	0.78	2.88	0.94	0.54***	13.49
Aggregate	1.16	2.90	1.39	0.99***	6.94	1.13	2.91	1.34	0.95***	7.34
High	2.57	4.65	1.91	2.34***	20.68	2.51	7.88	1.10	2.41***	7.78
Medium	1.17	3.34	1.21	0.95***	11.33	0.90	2.72	1.14	0.63***	8.26
Low	0.77	3.05	0.87	0.70***	14.92	0.75	3.16	0.83	0.67***	20.27
	Panel B: Sign-adjusted Equal-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.45	3.34	1.50	1.25***	15.16	0.98	3.21	1.06	0.73***	17.17
Aggregate	1.83	3.41	1.86	1.74***	11.61	1.34	3.22	1.44	1.17***	8.67
High	3.71	5.18	2.48	3.35***	21.22	2.98	7.97	1.29	2.77***	11.89
Medium	1.85	3.43	1.87	1.71***	11.92	1.09	2.82	1.34	0.84***	9.18
Low	1.08	3.53	1.06	1.15***	24.61	0.85	3.15	0.93	0.80***	17.60
	Panel C: Forecast-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	2.05	4.02	1.77	1.83***	18.41	1.31	3.82	1.19	1.04***	19.94
Aggregate	2.76	4.30	2.23	2.68***	16.24	1.83	3.87	1.63	1.67***	16.52
High	4.79	6.11	2.72	4.36***	22.84	3.36	8.30	1.40	3.10***	15.43
Medium	2.87	4.35	2.28	2.75***	16.67	1.60	3.52	1.58	1.39***	12.84
Low	1.45	4.34	1.16	1.54***	28.41	1.04	4.40	0.82	1.06***	27.62

Table IA.5: **Forecast-Implied Investment Performance (TS+CS Cluster)**

This table reports full-sample and large-cap sub-sample baseline long-short investment performance for forecast-implied portfolios (Equation 10). The first two panels show sign-adjusted value- and equal-weighted portfolios, while Panel C reports forecast-weighted portfolios, all constructed from observations in specific samples (Equation 9). We show five samples: Global (no clustering), Aggregate (aggregation of clustering results), and High, Medium, and Low, based on predictive rankings within the tree clusters. We show results of monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), Jensen's alpha (in %), and monthly maximum drawdown (MDD, %).

	Sample A: All Stocks					Sample B: Large-Cap				
	Panel A: Sign-adjusted Value-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	0.71	2.88	0.85	0.47***	13.71	0.68	2.88	0.82	0.44***	13.59
Aggregate	1.06	2.90	1.27	0.89***	7.04	1.03	2.91	1.22	0.85***	7.44
High	2.47	4.65	1.84	2.24***	20.78	2.41	7.88	1.06	2.31***	7.88
Medium	1.07	3.34	1.11	0.85***	11.43	0.80	2.72	1.01	0.53***	8.36
Low	0.67	3.05	0.76	0.60***	15.02	0.65	3.16	0.72	0.57***	20.37
	Panel B: Sign-adjusted Equal-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.35	3.34	1.40	1.15***	15.26	0.88	3.21	0.95	0.63***	17.27
Aggregate	1.73	3.41	1.76	1.64***	11.71	1.24	3.22	1.34	1.07***	8.77
High	3.61	5.18	2.41	3.25***	21.32	2.88	7.97	1.25	2.67***	11.99
Medium	1.75	3.43	1.77	1.61***	12.02	0.99	2.82	1.21	0.74***	9.28
Low	0.98	3.53	0.96	1.05***	24.71	0.75	3.15	0.82	0.70***	17.70
	Panel C: Forecast-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.95	4.02	1.68	1.73***	18.51	1.21	3.82	1.10	0.94***	20.04
Aggregate	2.66	4.30	2.15	2.58***	16.34	1.73	3.87	1.54	1.57***	16.62
High	4.69	6.11	2.66	4.26***	22.94	3.26	8.30	1.36	3.00***	15.53
Medium	2.77	4.35	2.20	2.65***	16.77	1.50	3.52	1.48	1.29***	12.94
Low	1.35	4.34	1.08	1.44***	28.51	0.94	4.40	0.74	0.96***	27.72