Mosaics of Predictability*

Lin William Cong Guanhao Feng Jingyu He Yuanzhi Wang

First draft: Feb. 2024; this draft: Nov. 2025

Abstract

We document that return predictability is an asset-specific characteristic linked to the cross section of expected returns, rather than a uniform feature of selected predictors or models. Utilizing a novel tree-based clustering framework, we partition the U.S. equity return panel based on rich firm characteristics and market-wide variables to capture heterogeneity in predictability across assets and market states. We find that stocks with large earnings surprises, high earnings-to-price ratios, and low trading volumes exhibit the strongest predictability. This predictability is time-varying, weakening when dividend yields are low and strengthening when dividend yields are high, particularly in markets with limited liquidity. Out-of-sample, a predictability differential strategy that leverages investors' model misspecification generates monthly alphas exceeding 1%, and investing in highly predictable clusters significantly outperforms conventional benchmarks with Sharpe ratios approaching 2.

Keywords: Return Predictability, Tree-Based Clustering, Asset Heterogeneity, Regime Switching.

JEL Classification: G12, C38, C53, C55.

*We thank Chunrong Ai, Doron Avramov, Daniele Bianchi, Pietro Bini, Jian Chen (Discussant), Qihui Chen, Lauren Cohen, Victor DeMiguel, Yi Ding (Discussant), Bing Han, Zhiguo He, Kewei Hou, Fuwei Jiang (Discussant), Sophia Li, Yu-Jane Liu, Zhanye Luo, Hao Ma, Semyon Malamud, Kuntara Pukthuanthong, David Rapach (Discussant), Allan Timmermann (Discussant), Michael Weber, Dacheng Xiu, Lei Yang (Discussant), Xiaoyan Zhang, Yi Zhang, Geoffery Zheng, and participants at the AI and Big Data in Finance Research Webinar, 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, Xiamen University Big Data and Econometrics Conference, 2024 SoFiE Meeting, PKU-NUS Quantitative Finance and Economics Conference, Summer Institute of Finance 2024, and other conferences for valuable feedback. Cong (will.cong@cornell.edu) is with Cornell SC Johnson College of Business (Johnson), ABFER, and NBER. Feng (gavin.feng@cityu.edu.hk), He (jingyuhe@cityu.edu.hk), and Wang (yuanzwang5-c@my.cityu.edu.hk) are with the City University of Hong Kong.

1 Introduction

Forecasting asset returns has long been a central question in finance. Previous literature has documented return predictability in many asset classes.[1] Most studies treat predictability as a property of the chosen predictors or models, assuming it is time-invariant and uniformly applicable to all asset returns analyzed in each setting. Various methods have been proposed to test return predictability, including predictive regressions (e.g., Stambaugh, 1999), security sorting (e.g., Jensen et al., 1972), cross-sectional models (e.g., Kelly and Pruitt, 2013; Han et al., 2024), and machine learning techniques (e.g., Gu et al., 2020; Freyberger et al., 2020).

However, other evidence suggests that predictability may itself be an asset-specific characteristic, concentrated, for example, among micro-cap or distressed firms. Like other latent characteristics, such as beta or volatility, predictability is unobservable and lacks a universally accepted definition. Related concepts include anomaly average return, predictor significance, out-of-sample R^2, and forecast-implied portfolios. Moreover, predictability is likely both time-varying and heterogeneous across assets, features that extant homogeneous models fail to capture. This uniform assumption poses significant challenges for asset pricing: if predictability varies across assets (e.g., concentrated in stocks with weak signals or high trading frictions) or over time (e.g., stronger during recessions), then homogeneous models risk misestimating expected returns and obscuring the true economic drivers of predictability. Consequently, this leaves a critical gap in our understanding of how predictability relates to asset characteristics and market regimes.

These challenges motivate the development of a new framework to systematically examine which assets exhibit stronger return predictability and how such pre-

[1]Return predictability has been documented across various asset classes. At the aggregate market level, predictability is often linked to business-cycle variables such as dividend yields and term spreads (e.g., Fama and French, 1988, 1989; Ang and Bekaert, 2007; Campbell and Thompson, 2008). In the cross section of individual stocks, firm characteristics such as size, value, and momentum serve as significant predictors of expected returns (e.g., Fama and French, 1992; Jegadeesh and Titman, 1993; Rapach et al., 2013; Lewellen, 2015). Predictability has also been found in corporate bonds (e.g., Lin et al., 2018), treasury bonds (e.g., Bianchi et al., 2021; Huang and Shi, 2023), options (e.g., Bali et al., 2023), and mutual fund alphas (e.g., Kaniel et al., 2023; DeMiguel et al., 2023).

dictability evolves across market regimes. We argue that, beyond the significance of individual predictors or the fitness of forecasting models, return predictability itself can be viewed as an asset-level characteristic that relates to the cross-section of expected returns. Building on this perspective, we introduce a clustering approach that groups asset-return observations according to high-dimensional firm characteristics and market predictors. This enables the measurement and comparison of heterogeneous patterns of return predictability across assets and over time.

Our analysis of the U.S. equity market uncovers a rich mosaic of predictability across the panel of individual stock returns. Predictability varies remarkably across the cross-section (partitions based on firm characteristics) and over time (regimes identified via market predictors and structural breaks), forming a mosaic-like pattern within the return panel. Moreover, the substantial discrepancy between heterogeneous and homogeneous predictability measures highlights the risk of model misspecification that arises when variation in predictive structures is ignored. Investors' reliance on uniform forecasting models shapes how information is incorporated into prices and, consequently, the cross-section of expected returns. Recognizing the advantages of local models, we design predictability-differential strategies that exploit cluster-specific information and construct forecast-implied portfolios that earn higher risk-adjusted returns by focusing on stocks with stronger return predictability.

To our knowledge, this paper is the first to systematically measure and analyze heterogeneous return predictability within an economics-guided, data-driven framework. Building on the Panel Tree (P-Tree) framework of Cong, Feng, He, and He (2025); Cong, Feng, He, and Li (2023), we develop a customized P-Tree clustering algorithm that partitions asset-return observations by predictability level—directly addressing the uniform predictability gap. Unlike homogeneous models or ad hoc sorting, the P-Tree endogenously groups stocks with similar predictability using firm characteristics and market variables, enabling us to measure cluster-specific predictability and its link to expected returns.

To quantify heterogeneous return predictability, we measure the cluster-specific

R^2 from a properly tuned and regularized predictive model estimated within each cluster. Although R^2 is conventionally interpreted as a measure of model fit, it also serves as an empirical proxy for the signal-to-noise (S2N) ratio in return forecasting. A higher R^2 indicates that the model captures a larger fraction of systematic variation relative to idiosyncratic noise, suggesting stronger predictability. Importantly, we obtain these R^2 values from models that have undergone full tuning and regularization, ensuring that they reflect intrinsic predictive content rather than overfitting. Traditional pooled regressions report a single aggregate R^2 that masks cross-sectional variation in predictability, whereas firm-by-firm regressions suffer from limited data. We strike a balance by endogenously clustering stocks according to their return and characteristics profiles, and estimating cluster-wise models that yield robust, cluster-specific R^2 measures capturing heterogeneous predictability across assets and over time.

Using this intuitive measure, we develop 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 partitions stock return observations by maximizing the cross-cluster difference in R^2 under heterogeneous predictive models. The resulting clusters form mosaics on the return panel, defined by firm characteristics for cross-section and market predictors for regime detection, yielding an interpretable visualization of dynamic predictive patterns. This framework preserves economic interpretability, highlights the firm and macro variables most associated with varying levels of predictability, and flexibly captures nonlinear and asymmetric interactions between firm characteristics and market conditions.

We apply this framework to monthly U.S. individual stock returns from 1973 to 2022, incorporating 51 firm characteristics and eight market predictors. Our analysis reveals substantial cross-sectional heterogeneity, 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 correspond 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. In contrast, stocks with low earnings surprises but high bid-ask spreads and momentum display no predictability. Their negative R^2 values imply forecasting performance worse than a naive zero forecast. For comparison, a homogeneous pooled model estimated on the full sample yields an in-sample (INS) R^2 of only 1.49%, suggesting that averaging across heterogeneous predictive structures can mask hard-to-predict returns. Out-of-sample (OOS) results confirm the persistent gap between highly and weakly predictable clusters, highlighting the robust link between latent return predictability and high-dimensional firm characteristics.

Building on the proposed method, we extend the analysis of U.S. equities to capture regime shifts and structural breaks. We identify two primary drivers of time variation in return predictability: the market dividend yield (Fama and French, 1988) and market liquidity (Pástor and Stambaugh, 2003). Predictability varies across business cycles, weak during periods of low market dividend yield and strong when market dividend yield is high and liquidity is low, conditions typically observed in recessions. This pattern aligns with prior evidence on time-varying market return predictability (e.g., see Henkel et al., 2011; Dangl and Halling, 2012). Predictability peaks when the market dividend yield is high and market liquidity is low, indicating early recessions. We interpret this result through the present value framework of Campbell and Shiller (1988).[2] A structural-break analysis within our framework provides consistent evidence of heightened return predictability during recessionary periods.

Finally, accounting for cross-sectional grouped heterogeneity reveals a strong link between predictability and performance. The most predictable cluster earns an average monthly return of 3.41% with an annualized Sharpe ratio of 1.88, whereas the least predictable cluster delivers a -0.06% return and a -0.02 Sharpe ratio. These results, robust to transaction costs, demonstrate that return predictability translates into economic value, contrasting with the null findings of earlier studies that ignored grouped heterogeneity (e.g., Rapach et al., 2010; Kelly et al., 2024).

[2]The forward-looking nature of the stock market suggests that elevated dividend yields during recessions primarily reflect declining prices. Subsequent price recoveries enhance return predictability.

We document a positive association between predictability differentials and cross-sectional expected returns, likely reflecting compensation for the risk of model misspecification. Sorting clusters by predictability gaps between cluster-specific heterogeneous models and the global homogeneous benchmark, we construct long-short portfolios with unexplained monthly alphas exceeding 1% across various factor models over the past two decades. Cluster-wise predictive models consistently outperform the global homogeneous approach, with the most predictable clusters achieving an out-of-sample annualized Sharpe ratio near 1.86. These results highlight the inefficiency of conventional homogeneous models (e.g., Feng and He, 2022; Evgeniou et al., 2023) and emphasize the practical value of exploiting grouped heterogeneity in return predictability. This evidence indicates that predictability varies systematically across asset clusters, reflecting economic or informational differences. Ignoring such heterogeneity can cause mispricing and suboptimal strategies. Our findings support asset pricing and forecasting methods that explicitly incorporate grouped heterogeneity, improving expected return estimates and the economic understanding of predictability.

Literature. Our study contributes to the extensive literature on return predictability, as it is the first to systematically examine the heterogeneity of this phenomenon. Previous research documents heterogeneous return predictability ad hoc or only for specific cases. For instance, Avramov (2002), Green et al. (2017), and Avramov et al. (2023) show that return predictability is concentrated among micro-cap and distressed stocks, particularly during periods of high market volatility. Other research links stock return predictability 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 subsamples, finding time-varying predictability in valuation ratios, such as book-to-market and earnings-to-price. More recent work uncovers time-series variation and structural breaks in market return predictability, so-called "pockets of predictability" (e.g., Farmer et al., 2023, 2024; Cakici et al., 2024). However, Cakici et al. (2025) show that such findings

are sensitive to specification and have weakened over time. Our findings offer a more rational explanation for the instability and inconsistencies in prior empirical findings on OOS 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, asset-specific characteristic. This perspective extends beyond traditional time-varying coefficient models by integrating high-dimensional firm characteristics and market predictors to endogenously identify clusters across 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.[3] Our primary innovation departs from the reliance on exogenously given portfolios by integrating data-driven clustering, which endogenously determines asset groupings, with cluster-specific model estimation within a unified framework.

In terms of methodological innovation, we build on Cong et al. (2025) and Cong et al. (2023) to develop an economically guided P-Tree framework that adapts the efficient algorithm of He and Hahn (2023). Our study thus contributes to the emerging AI literature on goal-oriented search and generative modeling by proposing a data-driven approach to optimizing an explicitly defined economic objective within a large and flexible model 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. While Cong et al. (2025) generate test assets and construct the SDF by decomposing the cross-section and forming leaf-based portfolios, our paper addresses a different question: we construct a customized P-Tree that partitions stock–return panels optimally based on heterogeneous levels of return predictability.

[3] In the high-frequency domain, Aït-Sahalia et al. (2022) document that returns and trade direction are more predictable for stocks with smaller nominal share prices, less liquidity, less volatility, and less correlation with the aggregate market.

Our work is also related to studies 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; and Evgeniou et al. (2023) apply 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, our framework enables assets to transition dynamically across characteristics-based clusters. These transitions naturally capture the evolving relationship between characteristics and return predictability.

The remainder of the paper proceeds as follows: Section 2 introduces the tree-based clustering framework. Section 3 describes the data and evaluation procedures. Section 4 examines the cross-sectional and time-series sources of heterogeneity in return predictability, with a focus on firm characteristics, macroeconomic conditions, and calendar effects. Section 5 examines the OOS robustness of these patterns and the performance of different investment strategies advantaged from our cluster-based models. Section 6 concludes. The appendices provide algorithmic details, data descriptions, and supplementary results, while the online appendix contains additional robustness checks and extended empirical discussions.

2 A P-Tree Framework for Analyzing Heterogeneous Predictability

Our framework aims to generate three key components: (i) data-driven clusters that partition asset–return observations by their level of predictability, represented through a P-Tree structure; (ii) cluster-specific heterogeneous predictive models; and (iii) a corresponding 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

Return predictability is inherently unobservable and lacks a universally accepted definition in the literature. We propose using the INS R^2 of a well-specified and reg-

ularized predictive model as a concise and intuitive measure of an asset's intrinsic predictability. This approach captures the underlying signal-to-noise ratio rather than merely reflecting model fit. By incorporating regularization, the measure achieves a balance between robustness against overfitting and the ability to identify meaningful cross-asset variation in predictability. This validates the concept of "predictability as an asset characteristic" and offers a practical framework for quantifying it.

While some studies advocate for economic performance measures, such as the Sharpe ratio, to assess predictability more meaningfully (e.g., Campbell and Thompson, 2008; Lewellen, 2015), we show in Appendix V that Sharpe ratio–based evaluations corroborate our main findings based on R^2.

Formally, consider the predictive model $r_{i,t} = \mathbb{E}_{t-1}[r_{i,t}] + \varepsilon_{i,t}$, where $\mathbb{E}[\varepsilon_{i,t}] = 0$ ensures unbiased forecasts. The INS R^2 for stock i at time t is defined as

$$R^2_{i,t} = 1 - \frac{\mathbb{E}(r_{i,t} - \widehat{r}_{i,t})^2}{\mathbb{E}(r_{i,t} - \bar{r})^2},$$

where $\widehat{r}_{i,t}$ denotes the fitted prediction and $\bar{r}$ the unconditional mean return across all stocks and periods.

Intuitively, a high $R^2_{i,t}$ indicates that the fitted model captures the conditional expectation $\mathbb{E}_{t-1}[r_{i,t}]$ beyond the unconditional mean. This reflects the asset's predictability of return. Conversely, when the noise variance $\mathbb{E}[(r_{i,t} - \bar{r})^2]$ is large, even perfect knowledge of the conditional mean yields a low R^2, highlighting the intrinsic difficulty of return prediction. Therefore, the INS R^2 can serve as a proxy for the data's signal-to-noise ratio and, by extension, the inherent predictability of the asset's returns.

We emphasize the use of INS estimation because the clustering algorithm depends on identifying characteristic-driven structures within the data. Incorporating OOS information during clustering risks contaminating this process and introducing data-snooping bias. Regularization in the predictive model further ensures that the R^2 statistic reflects genuine underlying characteristics rather than overfitting to noise.

From a classical perspective, the F-statistic in predictive regressions quantifies predictability. Consider the linear model: $r_{i,t} = \alpha + x'_{i,t-1}\beta + \varepsilon_{i,t}$. Let SSR and SSE

denote the regression's explained and residual sums of squares, respectively. The F-statistic for testing joint significance is:

$$F = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)},$$

where k is the number of predictors and n is the sample size. Since F increases monotonically with R^2, both metrics reflect the signal-to-noise tradeoff. Thus, the F-statistic aligns naturally with R^2 as a measure of predictability.

Finally, one might question why we do not use OOS R^2 to guide clustering. The clustering algorithm determines the INS structure of asset returns; incorporating OOS data during clustering would introduce test-sample information, resulting in data-snooping bias. Instead, we treat the clustering procedure and cluster-wise predictive model as a unified INS object, evaluating their performance exclusively OOS to ensure a valid assessment.

2.2 Cluster-Wise Predictive Modeling

Clustering and predictive modeling are two fundamental tasks in machine learning, typically handled by separate models. Our methodology unifies these tasks within a single framework: a machine learning model (e.g., Ridge regression) is employed for prediction, and a P-Tree is constructed based on the chosen predictive model to perform clustering. The clustering design ensures that stock–return observations with similar levels of predictability are grouped into the same cluster. In this subsection, we treat the clusters as predetermined and focus on the cluster-wise analysis first.

Let the data be denoted 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 $r_{i,t}$ is 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.

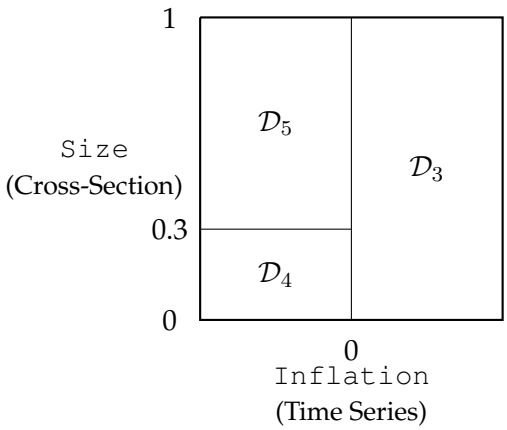
Ideally, heterogeneous and time-varying expected excess returns are modeled as: $\mathbb{E}_{t-1}[r_{i,t}] = g_{i,t}(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$, which is a function that varies across assets and time periods. However, limited observations for individual asset returns make it infeasible

to estimate each $g_{i,t}(\cdot)$ individually. Consequently, many studies (e.g., Gu et al., 2020) employ a homogeneous predictive function, $g_t(\cdot)$, with time-varying coefficients that are updated via an expanding/rolling-window scheme. Yet, as shown by Feng and He (2022), this approach overlooks heterogeneity in predictive strength across assets and implicitly assumes uniform return predictability, or identical R^2 values across stocks, which is contrary to empirical evidence (e.g., Avramov et al., 2023).

We propose a cluster-wise predictive modeling approach that strikes a balance between 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 the level of return predictability yielded by the chosen predictive model. This clustering method partitions the stock return panel into multiple clusters (leaf nodes) based on characteristics for cross-sectional and market predictors, as well as time-series dimensions. Figure 1 illustrates an example where the panel of stock returns is split into three non-overlapping clusters, described by characteristics or market predictors.[4]

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.



Rather than fitting individual predictive models, $g_{i,t}(\cdot)$, for each asset i in period

[4] A firm'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.

t, or estimating a single pooled model, we estimate a small number of cluster-wise models that remain homogeneous within clusters but vary across them. The resulting cluster-wise predictive model is:

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

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 employ Ridge regression as a representative machine learning model, given its robustness in weak-signal environments, as documented by Shen and Xiu (2024). Thus, we compute the INS R_j^2 using stock returns from the j-th cluster. We then proceed to describe our tree-based clustering methodology and provide empirical demonstrations of its implementation.

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 INS R^2 of a cluster-wise predictive model. Standard clustering techniques, such as K-means, are unsuitable for this purpose because they minimize within-cluster variation based on distances in feature space, ignoring predictive model performance. Moreover, their resulting clusters often lack economic interpretability. In contrast, our tree-based approach explicitly incorporates a cluster-wise predictive model, ensuring that the resulting clusters are both interpretable and predictive.

Our approach iteratively partitions the panel of stock–return observations, adding one cluster at a time. Among all candidate splits, the selected split is the one that divides the sample into two subgroups by maximizing the difference in their R^2 values.[5]

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

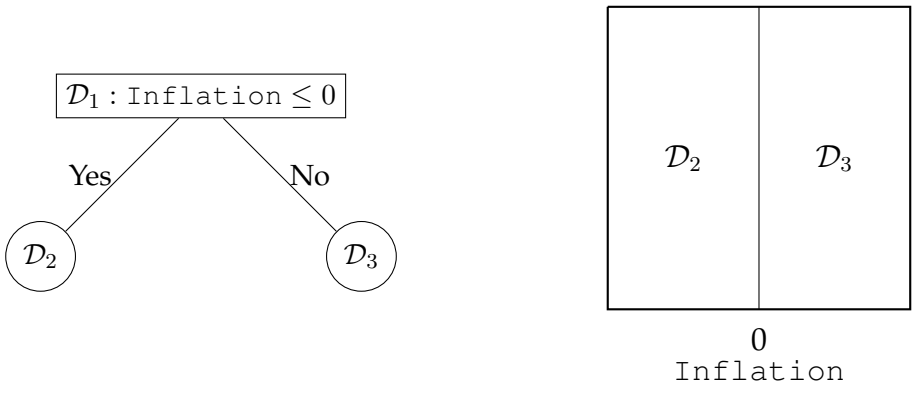
The resulting partitions are visualized in a P-Tree structure. Unlike Random Forests, which construct multiple trees solely for prediction, our method employs conditioning variables to sequentially partition the sample, yielding a deterministic and economically interpretable clustering outcome.

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, $\mathcal{D}_1$, representing the full 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.

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

This figure illustrates one of the first split candidates, such as $\texttt{Inflation} \leq 0$. The left figure shows a panel tree that divides the sample, and the right figure shows the corresponding partition plot, which partitions only over time.



[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.

The first split candidate divides the entire return sample $\mathcal{D}_1$ into two clusters, $\mathcal{D}_2$ 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. (2), 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:

$$g_j(\cdot) = \beta_0 + \beta^\intercal \mathbf{s}_{i,t-1} + \varepsilon_{i,t}, \tag{2}$$

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:

$$R^2_j = 1 - \frac{\sum_{\{i,t\} \in \text{leaf}_j} (r_{i,t} - \widehat{r}_{i,t})^2}{\sum_{\{i,t\} \in \text{leaf}_j} r^2_{i,t}}. \tag{3}$$

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:

$$S_{\{\text{leaf}_l, \text{leaf}_r\}}(\text{var}_p, c_k) = \left| R^2_{\text{leaf}_l} - R^2_{\text{leaf}_r} \right|. \tag{4}$$

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 high value of the criterion indicates that the split successfully separates stock–return observations into a highly predictable group and a less predictable one.

We evaluate the criterion in Eq. (4) across all candidate splits, defined by combinations of split variables and cut-point values. With P candidate variables and K cut-point values, there are $P \times K$ possible splits for the initial partition. 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 $\widehat{g}_3(\cdot)$, and consequently different criterion values. The optimal split is selected as the one that maximizes Eq. (4), completing the first split of the tree.

2.4 Subsequent Splits and Stopping Criteria

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

For instance, the second split may 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 follows a local–global principle: locally, each leaf node's benefit from splitting is assessed via within-subsample R^2 variation, while globally, the algorithm selects the split that most improves overall return predictability.

All subsequent splits are determined in the same manner. At each iteration, the algorithm examines all existing leaf nodes, searches for all possible split candidates, and selects the one that maximizes the global criterion. 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 fits post-cluster heterogeneous predictive models.

Stopping criteria. Stopping criteria are essential for regularizing INS model training and preventing overfitting. The clustering process terminates once predefined conditions are met. First, we impose a minimum sample-size requirement for each leaf node, excluding split candidates whose resulting subsamples do not meet this threshold. This ensures that each cluster-wise predictive model is estimated with sufficient observations. Second, we restrict the tree's maximum depth and the number of terminal leaves to control model complexity. Finally, a node is not further partitioned if all candidate splits fail to improve predictability—specifically, when the R^2 values of both child nodes fall below 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 corresponding 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 the clustering approach using Ridge regression in Eq. (2) for simplicity, the framework is not limited to any particular predictive model. The predictability of individual stock returns inherently depends on the choice of machine learning algorithm used for forecasting. Our framework accommodates this by providing a flexible and generalizable structure for modeling heterogeneous return predictability with a wide range of machine learning methods.

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 numbers of stocks in the training sample are 4,840 and 4,772, respectively. In the test sample, these values are 3,909 and 3,694, respectively. 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, including 3-month Treasury bill rate, inflation, term spread, default yield, and market-level characteristics such as dividend yield, volatility, net equity issues, and liquidity. Details are listed in Table A.1.

To ensure comparability over time, each predictor is standardized by subtracting its 10-year rolling average. We further apply a 12-month weighted moving average, assigning greater weight to more recent observations, to smooth short-term fluctuations and reduce volatility. A positive value of a standardized indicator, therefore, signals that its current level exceeds its historical 10-year average, corresponding to a high-regime state of 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,

[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.

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 to construct tree-based clustering and cluster-wise predictive models for forecasts.

3.2 Training and Evaluation

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

We complement the full-sample analysis with additional tests based on time-series splits to account for the long and overlapping nature of business cycles. Furthermore, hyperparameters in the post-clustering predictive models are optimized through cross-validation to ensure robust model performance.

Performance evaluation. As mentioned in Section 2.1, we use INS R^2 to measure predictability. In addition to the INS 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,

$$R^2_{OOS,j} = 1 - \frac{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t+1} - \widehat{r}_{i,t+1})^2}{\sum_{\{i,t\}\in \text{leaf}_j} r^2_{i,t+1}}, \tag{5}$$

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

4 Mosaics of Return Predictability

Our baseline analysis investigates two core questions: (1) Does heterogeneity exist? (2) Which stocks or periods exhibit greater predictability? We begin with cross-sectional partitions, referred to as "CS clusters" 4.1. We then extend the analysis to incorporate time-series dimensions through the "TS + CS clusters" model in Section 4.2, which integrates regime-dependent time variations based on macroeconomic information and calendar months, along with cross-sectional partitions.

4.1 Cross-Sectional Heterogeneity

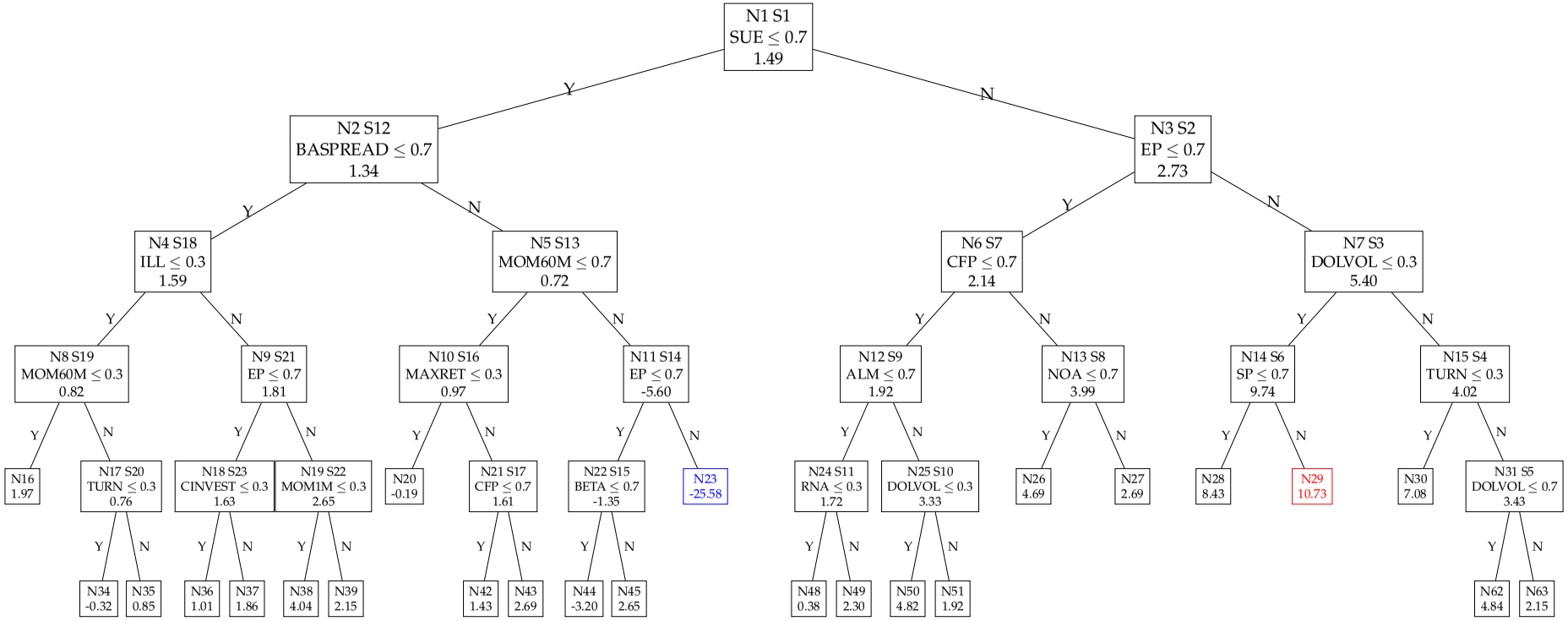
We present cross-sectional clustering results in Figure 3, based on monthly stock returns from 1973 to 2002. Following our predetermined stopping criteria, this tree stops growing after reaching 24 terminal leaves.[8] Each leaf node contains two or three rows of information. The top row identifies the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

The aggregate return predictability of the homogeneous model at the root node (N1) is 1.49%, where we fit a single ridge model to predict returns for all data without clustering. 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 declining slightly for the complement set, falling to 1.34% at N2. The R^2 difference (2.73% - 1.34% = 1.39%) represents the maximum value among all split candidates (51×2)

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

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

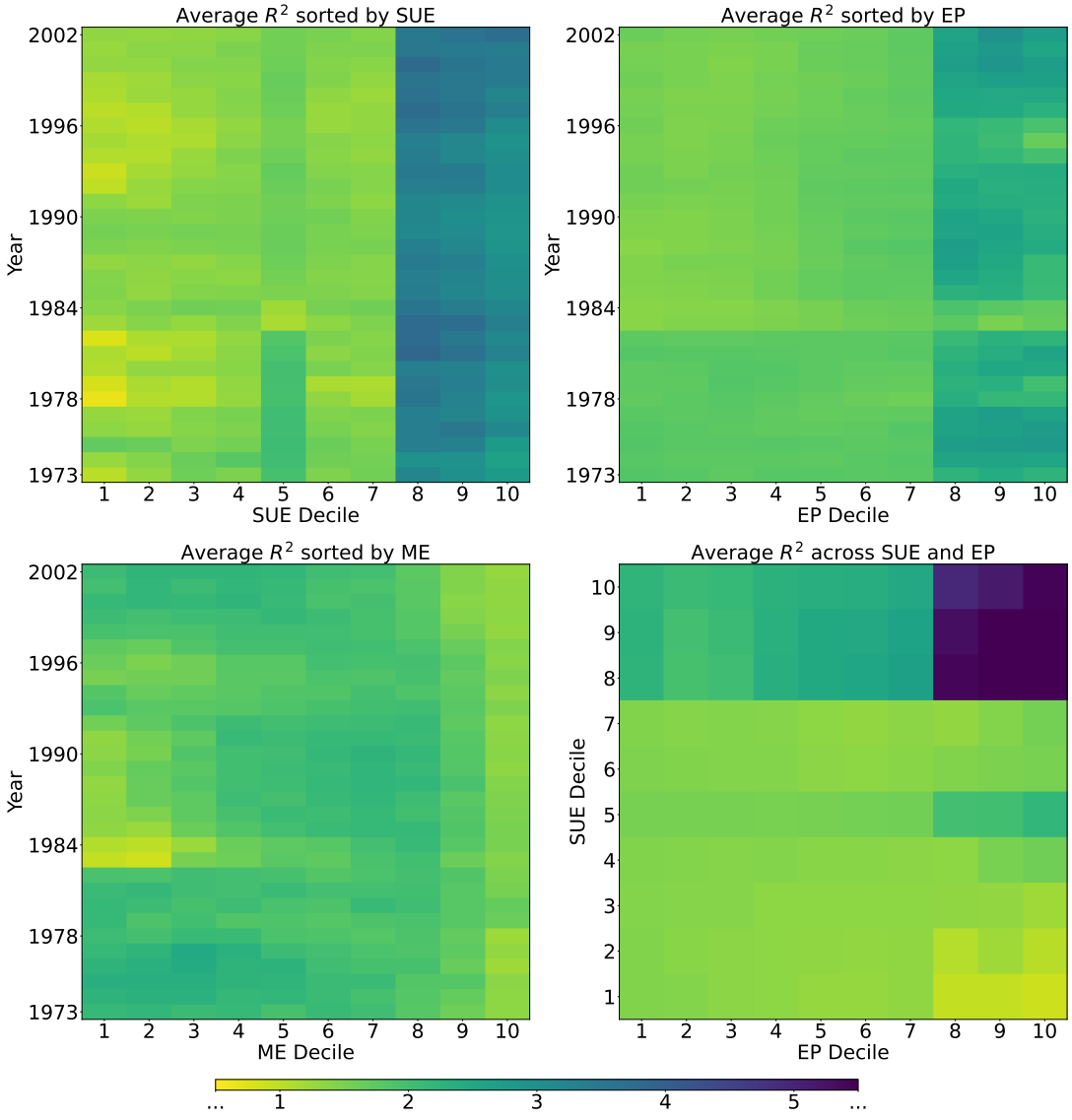
This figure illustrates the cross-sectional tree-based clustering derived from monthly data (1973–2002). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the range of [0,1]. Terminal leaves represent clusters defined by specific characteristic intervals. 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.



identified by the algorithm. The finding that stocks with higher earnings surprises exhibit stronger predictability is consistent with the post-earnings-announcement drift documented by Ball and Brown (1968). Engelberg et al. (2018) further show that anomaly returns are concentrated around earnings announcements and news events, reinforcing the role of earnings information in driving return predictability.

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 display average R^2 values for groups sorted by years and deciles of earnings surprises, earnings-to-price ratios, 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.



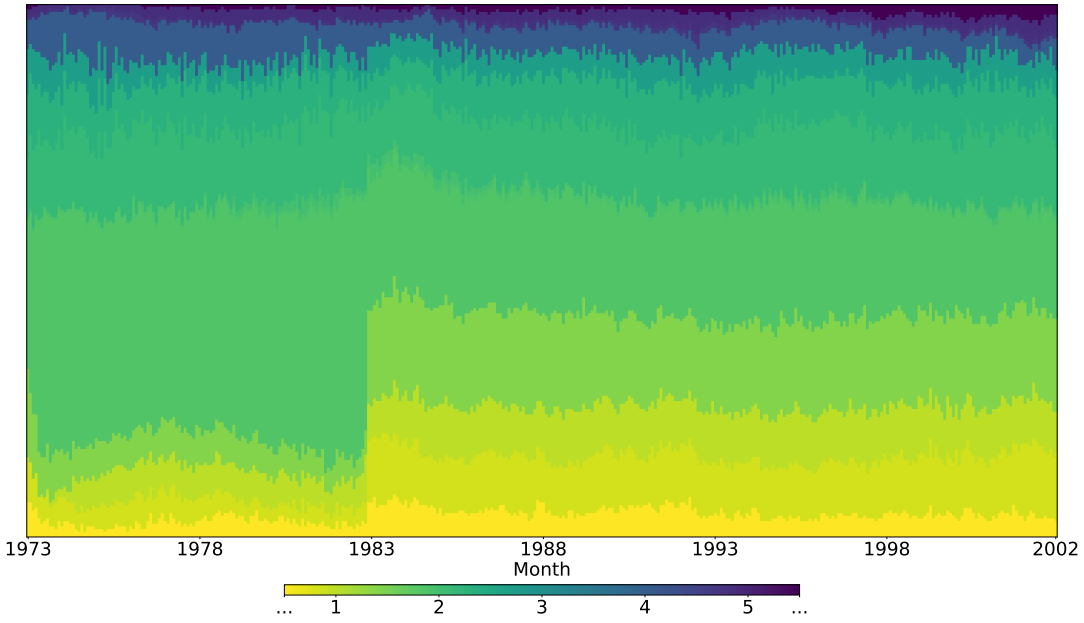
The second split (S2) identifies high earnings-to-price ($\texttt{EP} > 0.7$) stocks, yielding a higher R^2 of 5.40% at N7 among high-$\texttt{SUE}$ stocks. These high-$\texttt{SUE}$, high-$\texttt{EP}$

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 that differences in return predictability arise from the interactions among various predictors rather than from individual effects alone.

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 an increase in return predictability levels.



Clustering and predictability. 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 pat-

[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.

tern 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. The sample spans 360 months, aligned with the tree structure in Figure 3, and includes approximately 4,000 to 5,500 monthly observations. Furthermore, all clusters are determined by the cross-sectional standardized firm characteristics—that is, each firm's relative rank among all stocks in a given month. As a result, a stock may transition between clusters over time. We calculate the transition probability matrix to assess the persistence of these clustering patterns, counting the proportion of stocks that move from one cluster to all other clusters in the next month. Figure A.2 in Appendix IV.1 reports this matrix, showing that the clustering structure remains highly stable over time.

The clustering results yield heterogeneous predictive models, specifically Ridge regression models fitted within each cluster. We next evaluate the performance gains from this heterogeneous specification by comparing it with a homogeneous global model that estimates Eq. (2) using all stock–return observations jointly, without accounting for clusters. As summarized earlier, the predictability measure R^2_{C} estimated from the heterogeneous model varies substantially across clusters. Importantly, this variation arises from the **joint interaction of firm characteristics rather than individual effects alone**. For example, cluster N3, characterized by high earnings surprises ($\texttt{SUE} > 0.7$), achieves an R^2 of 2.73%. However, when high SUE combines with high earnings-to-price ($\texttt{EP} > 0.7$) and low trading volume in cluster N29, predictability surges to 10.73%. This interaction, uniquely identified by our tree-based clustering, explains why predictability varies across stocks: it depends on specific combinations of characteristics, rather than isolated features.

Economic links and performance. The characteristics driving predictability (high SUE, high EP, low volume) correspond to stocks exhibiting persistent return behavior, specifically, the long legs of well-documented and predictable patterns. The returns of

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 ($R^2_{\text{CMG}} = R^2_{\text{C}} - R^2_{\text{G}}$). 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}	Avg_{EW}	SR_{EW}	Avg_{VW}	SR_{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

these stocks are more predictable because their fundamental signals (earnings, valuation) are slower to be reflected in prices, or because their low liquidity limits arbitrage, thereby preserving predictability over time.

When we apply the global model, which estimates on the full dataset without accounting for clusters, to each cluster and compute its within-cluster R^2, denoted R^2_{G}, we also observe considerable variation across clusters. This highlights the cross-sectional differences in return predictability that our approach captures. The pattern remains robust even when evaluated ex post using the global model. However, because the global model assumes homogeneous predictability, it cannot uncover the underlying clustering structure.

Specifically, Panel A of Table 1 reports R^2_{C} and R^2_{G}. 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%). 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 contrast underscores that only a small fraction of stock returns are meaningfully predictable.

Our P-Tree framework surpasses standard characteristic sorting by endogenously identifying which characteristic combinations are most important. Unlike fixed splits (e.g., top 30% `SUE`), the tree prioritizes splits that maximize predictability differences. For example, uncovering that low volume only boosts predictability when paired with high `SUE` and `EP`, but not in isolation.

We further 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 the corresponding average monthly returns and annualized Sharpe ratios. The most predictable cluster (N29, $R^2_{\text{C}} = 10.73\%$) delivers an average return of 3.41% and an annualized Sharpe ratio of 1.88 for value weighting, highlighting strong profitability and a favorable risk-return tradeoff. In contrast, the least predictable cluster (N23, $R^2_{\text{C}} = -25.58\%$) 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. They also suggest the potential to identify predictability-related strategies (see Section 5).

4.2 Time Series Extensions

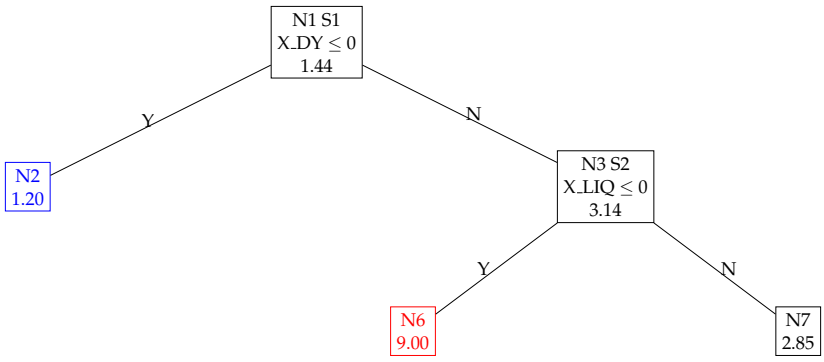
Building on Cong et al. (2025), we further investigate whether the cross-sectional heterogeneity documented above exhibits regime dependence by incorporating time-

series information. In this extension, splitting variables are expanded beyond firm characteristics ("CS Clusters" in the above Section 4.1) to include market situations. We examine time-varying return predictability conditional on aggregate macroeconomic and market states, and we also demonstrate that calendar months can serve as additional split variables, facilitating the detection of structural breaks in predictability over time. Unlike the cross-sectional clustering, the time-series heterogeneity analysis partitions the full sample period (1973–2022) to identify regime-dependent variations in return predictability.

Macro-driven regime changes. Similar to the regime detection in Feng et al. (2024) and Bie et al. (2024), we first utilize aggregate predictors to segment time-series horizons and identify distinct market regimes for heterogeneous return predictability. These regimes are defined based on the macroeconomic predictors, offering intuitive and economically meaningful interpretations. Furthermore, these regimes exhibit recurrent patterns aligned with business cycle dynamics. l two splits and employ firm characteristics for subsequently cross-sectional partitions.[10]

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 return 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.



[10]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.

We examine regime-shift patterns using the full 50-year sample. Figure 6 displays the upper portion of the tree, which captures the time-series splits, while Figure A.3 illustrates subsequent partitions by firm characteristics within each regime. We find three regimes and 45 terminal clusters in total.[11] The INS R^2 of the global model across the entire sample is 1.44% without 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 interaction between these two market predictors gives rise to three distinct regimes that evolve endogenously in response to business cycle dynamics.[12] We find that stock return predictability varies systematically across these recurring market regimes. During periods of low market dividend yield (Regime I), which are typically characterized by high market valuations and investor optimism, the R^2 is 1.20%, compared to the global model's 1.44%.

Return predictability reaches its peak at 9.00% when the market dividend yield is high and market liquidity is low. These conditions usually coincide with the early stages of recessions. This pattern aligns with prior evidence on time-varying market return predictability (e.g., see Henkel et al., 2011; Dangl and Halling, 2012) and with the well-documented positive relation between low liquidity and high predictability

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

[12] We standardize market predictors by subtracting their 10-year rolling averages. The 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 findings attribute time-varying return predictability to structural breaks in expected returns and financial ratios.

in the high-frequency domain (Aït-Sahalia et al., 2022). Because equity markets are forward-looking, elevated dividend yields during recessions primarily reflect falling prices rather than increasing dividends. These episodes typically precede price recoveries, as assets rebound from fire-sale levels during the onset of downturns. Overall, Regime II exhibits the highest level of return predictability among the three regimes.

The tree-based regime analysis identifies notable temporal shifts in return predictability. Further segmentation by firm characteristics reveals significant cross-sectional heterogeneity within macro-driven regimes. For example, leaf node N16 (red, Regime II in Figure A.3) exhibits the highest R^2 of 35.51%, characterized by stocks with low ILL, low STD_DOLVOL, low BETA, and low DEPR. However, the weakest cluster in Regime II, node N25 (blue), characterized by high ILL, low GMA, low ME, and high CASHDEBT stocks, exhibits the poorest performance at -3.97%. Compared with the cross-sectional clustering results in Section 4.1, this wider dispersion highlights the importance of incorporating regime-dependent time variation when assessing stock return predictability.

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).

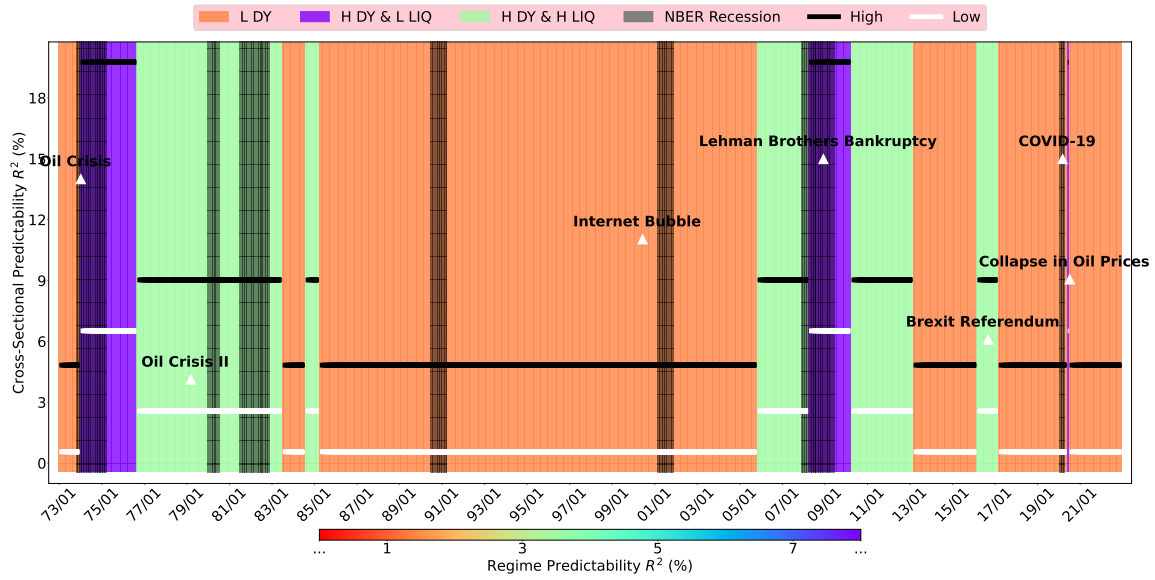
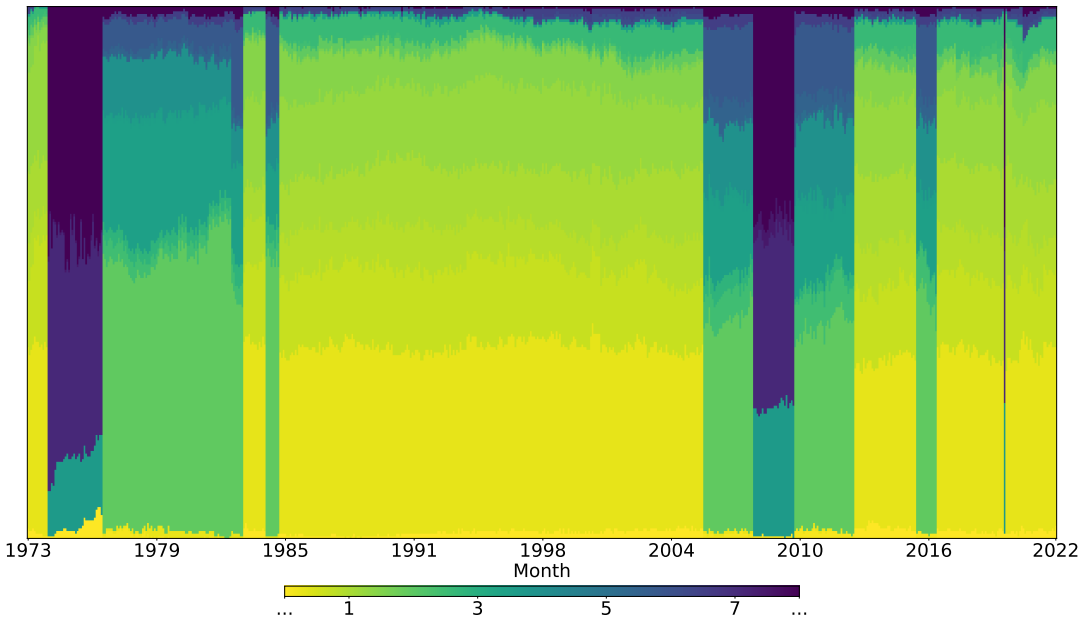


Figure 7 presents regime-dependent time variation from an alternative perspective. The color transitions denote the market regime switches identified by the tree structure in Figure 6. The range extends beyond the cross-sectional case, spanning from approximately 1% (orange) to over 7% (purple). Subsequent cross-sectional partitions by firm characteristics further amplify these differences, widening the dispersion of predictability across subsamples. To illustrate the extent of this cross-sectional variation, we plot the results for the most and least predictable clusters (black and white bars, respectively). Moreover, the timing of regime transitions closely aligns with U.S. recessions (e.g., 1974, 2008, 2020) and major global shocks, including the 1973 Oil Crisis, the 2008 Lehman Brothers collapse, the 2016 Brexit referendum, 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 with the cross-sectional mosaics in Figure 5, the patterns of stock return predictability across months (Figure 8) display stronger regime-dependent time variation. 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 relatively low predictability (light green to yellow shades), whereas certain periods, such as the mid-1970s and 2009, show markedly higher predictability across nearly all clusters. These episodes correspond to high market dividend yields and low market liquidity, characteristic of recessionary phases. The frequent color transitions across clusters visually capture the evolution of market regimes over time, providing empirical validation of the P-Tree's ability to uncover dynamic, regime-dependent patterns in return predictability.

Table 2: **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: $\mathbb{1}\{\texttt{X\_DY} \leq 0\}$												
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: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \leq 0\}$												
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

Next, we analyze cluster-specific performance to evaluate the predictability of returns across regimes. Table 2 translates the tree structures in Figure 6 and Figure A.3 into a descending ranking of return predictability by market regimes. The dispersion in predictability is much wider than in the purely cross-sectional setting: 11.00% versus -12.79% in Regime I, 35.51% versus -3.97% in Regime II, and 12.39% versus -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, the most predictable period, lasting 57 months, exhibits a near-zero correlation between R^2 and average returns. However, the higher volatility and return dispersion within this regime support our interpretation that heightened predictability reflects post-fire-sale price recoveries during early recessions.

We further conduct aggregate evaluations to assess improvements in heterogeneous return predictability with regime-dependent time variation modeling. Clusters are aggregated into subsamples, as shown in Table 3, which reports the resulting predictability metrics. Under all regimes, heterogeneous models (Panel B) exhibit greater differences in predictability of returns relative to the global homogeneous model (Panel A). We also perform a robustness check using a large-cap subsample—typically associated with lower return predictability and transaction costs than small-cap stocks. Even when focusing exclusively on large-cap stocks, distinct heterogeneity patterns persist across various market conditions.

Among the three regimes, Regime II 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 variations in cross-sectional predictability within this regime, which are robust across all methods. By contrast, the low market dividend yield regime exhibits the weakest predictability, with large-cap stocks showing R^2 values below 2.6%, suggesting negligible predictability of returns. Despite differences in magnitude, the improvement and decline patterns remain consistent across methods, particularly for cluster-wise forecasts. 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 of time-varying return predictability, with higher dividend yields associated with stronger predictability of returns. We interpret this empirical finding within the present-value framework of Campbell and Shiller (1988). Although their model is

Table 3: **Further Comparing Global and Cluster-Wise Predictive Models**

This table reports return predictability (R^2 in %) across predictive methods under market regimes, with all (Sample A) and large-cap (Sample B) 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

formulated for aggregate market returns rather than cross-sectional variation, it provides a valuable theoretical lens for understanding how fluctuations in the dividend yield relate to changes in market-level return predictability.

Let P_{t+1}, D_{t+1}, and R_{t+1} denote the price, dividend, and gross return in period $t+1$, respectively. Define $p_t = \log(P_t)$, $d_t = \log(D_t)$, and note that

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} = \frac{1 + \frac{P_{t+1}}{D_{t+1}}}{\frac{P_t}{D_t}} \cdot \frac{D_{t+1}}{D_t}.$$

Taking logarithms, the log return $r_{t+1} = \log(1 + R_{t+1})$ satisfies the Campbell-Shiller decomposition:

$$r_{t+1} = \log\left(1 + e^{\delta_{t+1}}\right) - \delta_t + \Delta d_{t+1},$$

where $\delta_t = \log(P_t/D_t)$ is the log price-dividend ratio and $\Delta d_{t+1} = d_{t+1} - d_t$ is the log dividend growth.

Expanding $\log(1 + e^{\delta_{t+1}})$ around the average log price-dividend ratio $\bar{\delta} = \log(\bar{\mathcal{D}})$ using a second-order Taylor approximation yields

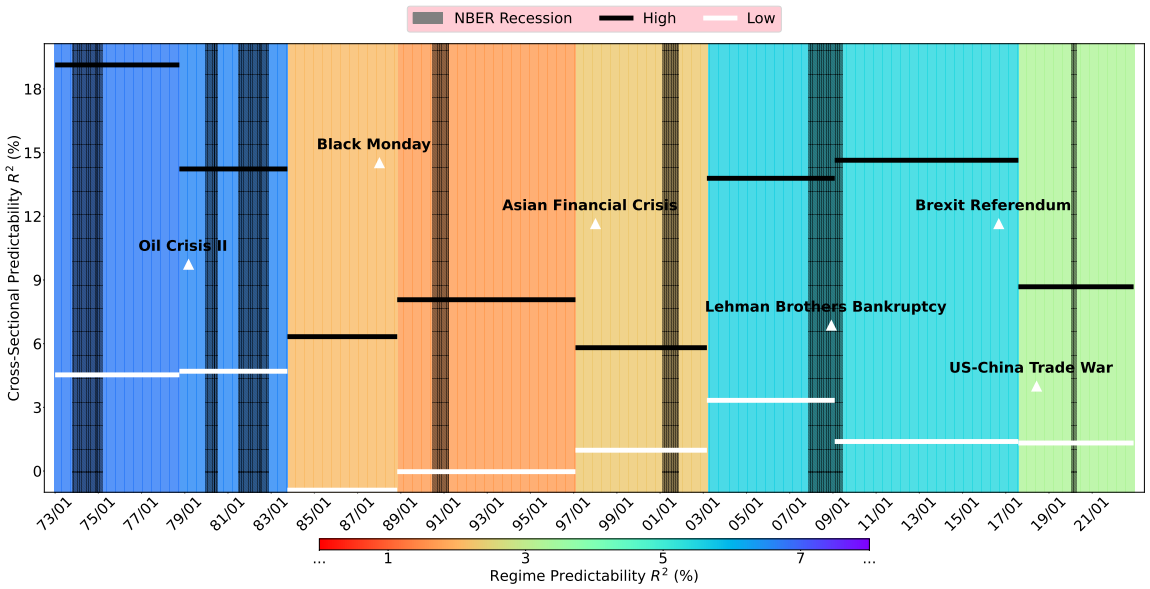
$$r_{t+1} \approx \log(1 + \bar{\mathcal{D}}) + \frac{\bar{\mathcal{D}}}{1 + \bar{\mathcal{D}}}(\delta_{t+1} - \bar{\delta}) + \frac{\bar{\mathcal{D}}}{2(1 + \bar{\mathcal{D}})^2}(\delta_{t+1} - \bar{\delta})^2 - \delta_t + \Delta d_{t+1}.$$

When modeling log returns as a linear function of δ, the second-order term is absorbed into the regression residuals. Notably, since the dividend yield is the inverse of the price-dividend ratio, periods of high dividend yields correspond to lower δ_t, which reduces the magnitude of the second-order term relative to the first-order term. This reduction in nonlinear components leads to smaller residual variance and, consequently, a higher R^2 in linear predictive regressions, reflecting increased market return predictability during high-dividend-yield regimes. By extension, as market return predictability strengthens with rising dividend yields, individual stock return predictability is also likely to increase, consistent with our empirical findings.

Structural breaks in calendar time. Traditional regime classification based on market predictors often produces recurring but non-continuous regimes over time. An alternative approach to capturing time variation in return predictability is to model structural breaks in return dynamics (e.g., Smith and Timmermann, 2021; Cui et al., 2024). This approach can be naturally implemented within our P-Tree framework by

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 indicated by the color bar. 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.



using calendar months as splitting variables instead of market predictors, as demonstrated in Feng et al. (2024) for currency markets.

A key advantage of this method is that calendar months progress sequentially, generating single, continuous regimes analogous to structural breaks. This enables us to assess whether segmenting continuous time periods can effectively capture heterogeneity in stock return predictability.

Figure 9 illustrates fluctuations in predictability across these calendar-time structural breaks. Although the magnitude of variation is generally smaller than in regime-based models using market predictors, predictability peaks at 6.7% during January 1973 to October 1978 (the first purple segment) and declines to 1.67% between November 1988 and February 1997 (the orange segment). Notably, predictability intensifies during recessions, especially prior to 1983, consistent with patterns in Figure 7.

Major events, such as Black Monday (1987) and the Brexit Referendum (2016), likely induced structural breaks in return predictability, which macro-driven regime analyses may overlook due to the slower adjustments of market predictors. Similar

discontinuities align with major crises, such as the 1979 Oil Crisis, the 1998 Asian Financial Crisis, the 2008 Lehman collapse, and the 2018 U.S.–China trade war.

Further partitioning within each regime by firm-specific characteristics reveals even greater heterogeneity in predictability. Table A.3 shows significant differences in return predictability and average returns between high- and low-predictability subsamples, with consistent trends across periods. These heterogeneous effects remain robust within the large-cap subsample (Table A.4).

Importantly, these findings suggest that abrupt changes in market conditions and investor sentiment around major economic and geopolitical events induce shifts in risk premia and return dynamics that cannot be fully captured by slowly evolving market predictors. This highlights the critical role of sudden structural breaks in shaping the mosaic-like and time-varying nature of stock return predictability.

5 Out-of-Sample Robustness and Economic Gains

The previous section focuses primarily on the heterogeneity of return predictability in the INS setting. In this section, we extend the analysis to the OOS domain. We first conduct robustness checks to examine whether the observed heterogeneity across different levels of predictability persists out of sample. We then investigate whether the identified mosaics of stock return predictability translate into economically meaningful gains through various investment strategies.[13]

5.1 Robustness Checks

Thus far, the easy- and hard-to-predict clusters have been identified using INS information. This naturally raises a question: Do these patterns persist out of the sample? Our objective is not to compare the R^2_{OOS} performance of competing machine learning models in a predictive "horse race," but rather to evaluate the persistence and stability of the heterogeneous predictability structure—something conventional ML or clustering models cannot directly address. To evaluate the OOS performance,

[13]Since all time-series results are based on the full-sample analysis, the OOS evaluation in this section primarily emphasizes the cross-sectional dimension.

we adopt a five-year rolling clustering setup with a two-fold cross-validation strategy for hyperparameter optimization.[14]

We classify leaves into three levels 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 the set with the lowest predictability. The remaining leaves constitute the medium level.[15] We use R^2_{OOS} values, as defined in Eq. (5), to validate the cross-sectional patterns of stock return predictability.

Table 4: **Out-of-Sample Predictability Robustness**

Similar to Table 3, 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 subsample 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 4 reports the OOS R^2 for two predictive models. The "Global Forecasts" (Panel A) applies a homogeneous model estimated over the entire sample without clustering and evaluates its performance on the high-, medium-, and low-predictability

[14]The INS 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 INS dataset. The final coefficients are then used to forecast the OOS values for the subsequent five years.

[15]Given varying dataset lengths and sample sizes, setting fixed R^2 thresholds to define predictability levels is impractical. Instead, we classify clusters based on their proportional representation in the INS data. For cross-sectional clustering, we apply a 10% cutoff: clusters with cumulative sample shares above (below) this threshold are categorized as high (low) predictability. The same 10% rule applies in Section 4.2 for time-series clustering, adjusted for regime- or period-specific observations to ensure sufficient sample sizes for investment analysis. This approach guarantees that clusters remain cohesive and that their combined representation exceeds the cutoff threshold.

subsamples identified by our method. The "Aggregate" model (Panel B) instead aggregates forecasts 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. This improvement demonstrates the benefit of the heterogeneous model in terms of OOS prediction accuracy.

Across all specifications, highly predictable clusters consistently outperform their less predictable counterparts, regardless of model type or estimation window. For instance, when comparing the values in each row for the highly predictable cluster against the low-predictability cluster, every metric—whether INS or OOS, OLS or Ridge regression, global or cluster-wise model—shows a higher R^2 for the more predictable cluster. These results reaffirm the persistence of the return-predictability gap between highly and weakly predictable clusters identified by our approach. Moreover, the OOS robustness tests validate the efficacy of our tree-based clustering approach in capturing this heterogeneity.

Beyond the aggregate cluster-level evaluation, we conduct an additional robustness test to further assess the persistence of OOS heterogeneity. While maintaining the predictability ranking in Table 1, this alternative approach divides the sample into two balanced groups—high- and low-predictability—based on the relative proportions of INS and OOS observations.[16] For each subsample, we re-estimate a homogeneous predictive model using only the INS observations within that group and evaluate its corresponding OOS forecasts. This "local-global" predictive modeling framework enables us to investigate whether the weak OOS predictability observed in the full sample will diminish or even disappear once highly predictable stocks are removed from the analysis.

Table 5 presents results that align closely with our expectations. The top row reports the predictive performance of the homogeneous model using the full sample. In contrast, each of the four subsequent groups shows a decline in overall predictive accuracy after removing stocks with high predictability. Some of the OOS R^2 values

[16] Each cluster's predictability is evaluated as a unit, and clusters are selected to form subsamples that are as balanced in size as possible at each threshold.

Table 5: **Further Local-Global Predictive Models Testing**

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. We separate the entire sample into subsamples by different propositions of sorted clusters and refit homogeneous predictive models locally. The 'True %' column summarizes the accurate percentages of subsamples' observations compared with the entire in-sample and out-of-sample datasets.

	In-Sample (1973 - 2002)				Out-of-Sample (2003 - 2022)			
Sample	OLS	Lasso	Ridge	True %	OLS	Lasso	Ridge	True %
Global	1.49	0.52	0.54		0.27	0.40	0.35	
High 10%	5.34	4.37	4.41	5.7	3.67	3.58	3.83	6.2
Non-High 10%	1.33	0.39	0.43	94.3	0.11	0.29	0.23	93.8
High 20%	3.61	2.29	2.29	15.8	1.68	1.61	1.77	18.7
Non-High 20%	1.19	0.28	0.36	84.2	0.03	0.22	0.16	81.3
High 30%	2.95	1.65	1.84	26.4	1.12	1.22	1.36	27.9
Non-High 30%	1.15	0.27	0.34	73.6	0.07	0.21	0.15	72.1
High 40%	2.61	1.34	1.50	37.5	0.58	0.84	0.91	38.5
Non-High 40%	1.09	0.25	0.23	62.5	0.18	0.23	0.16	61.5

even decrease by more than half (e.g., from 0.35% to 0.16%), in sharp contrast to the consistently strong predictability observed among the high-predictability subsample across all model specifications.

As the relative sample sizes of the high- and non-high-predictability groups become more balanced (from the top to bottom panels), the predictability gaps gradually narrow. These results highlight an interactive structure of stock return predictability: when data are pooled, a small subset of highly predictable stocks mechanically inflates the apparent predictability of the full sample. This aggregation effect can lead investors to overstate the true degree of return predictability and, consequently, misjudge the potential gains from predictive strategies. These findings reinforce the importance of recognizing heterogeneity in return predictability and motivate our subsequent analysis on whether the identified mosaics of predictability translate into tangible investment benefits.

5.2 Predictability Differential and Local Advantage Strategy

Many investors and researchers rely on global, homogeneous predictive models, while clustering stocks at the individual level remains rare in the literature. Our framework systematically quantifies the unobservable heterogeneity in return predictability that such global models tend to overlook.

Panel A of Table 1 shows that cluster-wise heterogeneous models—which tailor predictions to groups of similar stocks—outperform the global homogeneous model for most stocks with predictable returns. This finding suggests that conventional global models are likely misspecified, exposing average investors to the risk that locally relevant information is not efficiently incorporated into asset prices when all trades are based on homogeneous benchmarks (e.g., Gu et al., 2020).

In asset pricing terms, **model misspecification risk arises when global, homogeneous models fail to capture cluster-specific predictability—leading investors to underprice stocks where local signals matter most.** In equilibrium, these stocks command higher expected returns because investors bear the cost of ignoring local predictive structures: relying solely on global models means missing information that drives expected returns for clustered stocks.

Recent studies, such as Smith and Timmermann (2022) and Feng et al. (2024), demonstrate that homogeneous models of stock and currency returns underperform their regime-dependent, heterogeneous counterparts in regime-switching frameworks. Building on this insight, we define a cluster-level predictability differential that measures the local advantage of the heterogeneous model relative to the global one:[17]

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

A higher R^2_{CMG} indicates that local models explain returns better than the global benchmark, suggesting that investors relying solely on global models may overlook locally relevant signals. Conversely, a low or negative value implies the local model un-

[17] As reported in Table 1, this metric measures the difference in R^2 between cluster-wise forecasts (R^2_{C}) and the global model (R^2_{G}), evaluated on the same cluster.

derperforms the global one, possibly reflecting weaker or noisier local information.[18]

The **Local Advantage Strategy exploits the 'mosaics of predictability' identified earlier: by targeting clusters where local models outperform global ones (high R^2_{CMG}), it captures returns that homogeneous models overlook.** This strategy's alpha—exceeding 1% monthly across benchmarks—reflects the economic value of recognizing predictability's heterogeneous nature rather than assuming uniformity.

To examine these implications, we construct value- and equal-weighted portfolios within each cluster and analyze the relationship between average returns and predictability differentials. We then evaluate portfolio performance using standard factor-spanning regressions to estimate model-adjusted alphas, thereby quantifying the abnormal returns associated with heterogeneity in return predictability.

Panel A of Table 6 reports the performance of four long-short predictability differential strategies, sorted by cluster-specific differentials.[19] Strategies with larger risk differentials (e.g., "T3 - B3") consistently outperform those with smaller gaps (e.g., "T5 - B5"), yielding higher average returns and Sharpe ratios in both INS and OOS 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 during the OOS sample period.

Panel B of Table 6 presents risk-adjusted abnormal returns (alphas) estimated using benchmark models, including CAPM, Fama-French three-factor (FF3), five-factor (FF5), FF5 with momentum and idiosyncratic volatility (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).[20] All strategies exhibit statistically significant unexplained alphas in both INS and OOS settings. Additionally, this risk premium persists because arbitrage is limited. The strategy's alpha remains significant even after accounting for transaction costs (see Table A.6 in the Appendix). The premium is stable out of sample (e.g., can still be above 0.9% monthly return for

[18] As the global model incorporates broader cross-sectional data, R^2_{C} can be lower than R^2_{G}.
[19] We focus on the "T3 - B3" strategy, with additional clusters and robustness checks discussed later.
[20] Due to data availability limitations, SY4 results are reported for 2003–2016 only.

Table 6: **Predictability Differential Portfolio Performance**

This table presents summary statistics (Panel A) and abnormal returns (Panel B) for the strategy 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)

the "T7 - B7" strategy), confirming it is not a data-mined result.

Our findings highlight that ignoring heterogeneity in return predictability constitutes a form of model misspecification risk. Strategies recognizing local predictive structures can deliver superior model-adjusted performance, contrasting sharply with the null results reported in Rapach et al. (2010) and Kelly et al. (2024).

5.3 Forecast-Implied Strategies

Thus far, our analysis has primarily focused on uncovering heterogeneity in stock return predictability through the use of clustering mechanisms. However, beyond identifying clusters, our framework also produces cluster-specific predictive models, each estimated via cluster-wise Ridge regressions, which contrasts with global mod-

els (e.g., Gu et al., 2020), which impose a single predictive relation across all assets, thereby overlooking localized patterns of predictability.

We now evaluate the investment performance of these cluster-wise predictive models. Unlike the Local Advantage Strategy discussed in the previous section, which exploits differences between local and global models, the current analysis constructs trading portfolios directly from the cluster-wise model forecasts. At each rebalancing date, stock weights are determined in three ways: (1) Sign-adjusted equal-weighting and (2) sign-adjusted value-weighting, where the absolute base weights are assigned equally or by market capitalization, and the position sign (long or short) follows the model's return forecast; and (3) forecast-weighting, where the weight is directly proportional to the model-predicted return, subject to a leverage constraint following Guijarro-Ordonez et al. (2024).[21]

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

$$\begin{aligned} \text{Sign-adjusted Equal/Value-weighted: } \widehat{w}_{i,t} &= \begin{cases} w_{i,t}, & \text{if } \widehat{r}_{i,t} \geq 0 \\ -w_{i,t}, & \text{if } \widehat{r}_{i,t} < 0 \end{cases} \\ \text{Forecast-weighted: } \widehat{w}_{i,t} &= \frac{\widehat{r}_{i,t}}{\sum_i |\widehat{r}_{i,t}|}. \end{aligned} \tag{6}$$

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

$$R_{j,t} = \sum_{\{i,t\} \in \text{leaf}_j} \widehat{w}_{i,t} r_{i,t}. \tag{7}$$

These forecast-implied portfolios differ from the cluster portfolios analyzed in Section 5.2, which rely solely on value-weighted returns without incorporating forecasts. In contrast, the forecast-implied portfolios are explicitly long–short portfolios whose positions are determined by the signs and magnitudes of the machine-learning return forecasts. Intuitively, a model with greater predictive accuracy should generate

[21] Guijarro-Ordonez et al. (2024) normalize absolute portfolio weights to sum to one, thereby implicitly imposing a leverage constraint.

a portfolio with stronger realized performance.

Table 7 reports results for the cross-sectional clustering model introduced in Section 4. yield a Sharpe ratio of roughly half these values. Classified as highly predictable, these remain easier to forecast out of sample, and trading strategies based on these forecasts yield substantial investment gains. Sign-adjusted value-weighted portfolios also perform strongly, suggesting that the results are not driven by small-cap stocks. For instance, in Panel A, the highly predictable subsample achieves an OOS Sharpe ratio of 0.84 and a significant monthly alpha of 0.46%. In contrast, portfolios formed from the least predictable stocks yield a Sharpe ratio of roughly half this value (0.51) and a negative, statistically insignificant alpha.

Panels B and C of Table 7 present similar patterns. Highly predictable stocks yield OOS Sharpe ratios of 1.66 and 1.86, accompanied by statistically significant alphas of 1.84% and 2.30%, respectively. These substantially outperform portfolios formed from medium- and low-predictability stocks. Comparable contrasts emerge for average returns and maximum drawdowns, particularly for alphas in the top two panels.

Second, cluster-wise predictive models ("Aggregate") outperform the homogeneous predictive model ("Global"), demonstrating the economic benefits of adopting 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. The performance gap widens for the forecast-weighted portfolio, where both the magnitude and sign of the return forecasts drive the portfolio weights.

Overall, both aggregate and global models show similar accuracy in predicting the direction of returns (Panels A and B), but the aggregate model excels in capturing *magnitude* (Panel C). Large-cap subsamples (top 30% by market equity value) outperform their global counterparts OOS across all weighting strategies, although the improvement is smaller than that of the full sample. This may be due to the lower predictability of returns for large-cap stocks. Finally, accounting for transaction costs (Table A.7 in the Appendix) slightly reduces absolute performance levels but leaves

Table 7: **Forecast-Implied Investment Performance**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 7). 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 6). We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap subsamples 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	6.00	1.86	2.31^{***}	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

all qualitative patterns intact. For example, the OOS Sharpe ratios for the highly predictable subsample in Panels B and C decrease marginally, from 1.66 and 1.86 to 1.59 and 1.80.[22]

[22]Section IA.II of the Internet Appendix shows that investment gains remain robust—and even

Table 8: **Forecast-Sorted Long-Short Strategies**

This table reports in-sample and out-of-sample performances of forecast-sorted long-short value- (Panel A) and equal-weighted (Panel B) portfolios in deciles, respectively. We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap subsamples 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: Value-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	3.60	5.66	2.20	3.71***	20.89	0.99	4.98	0.69	1.31***	16.70
Aggregate	3.88	6.26	2.15	4.06***	29.94	1.77	5.26	1.16	1.98***	31.30
High	3.74	6.41	2.02	3.63***	15.09	3.13	7.15	1.52	2.69***	20.27
Medium	3.91	5.71	2.37	4.01***	19.98	1.36	5.42	0.87	1.53***	28.19
Low	1.92	8.06	0.83	1.88***	34.42	0.58	5.93	0.34	0.66*	32.30
Global-Large	2.23	4.59	1.68	2.32***	14.23	0.88	4.08	0.75	1.08***	14.85
Aggregate-Large	2.49	5.09	1.70	2.63***	18.56	0.62	3.74	0.57	0.79***	15.03
	Panel B: Equal-Weighted									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	4.51	5.22	2.99	4.60***	29.85	2.12	5.34	1.38	2.24***	22.27
Aggregate	4.58	5.10	3.11	4.68***	31.32	3.04	4.15	2.54	3.22***	20.24
High	4.95	5.55	3.09	4.81***	10.70	3.67	5.79	2.19	3.35***	14.50
Medium	4.26	4.77	3.09	4.33***	27.70	2.63	4.12	2.22	2.78***	18.63
Low	0.99	5.14	0.67	0.90***	18.22	0.76	4.90	0.54	0.75**	23.31
Global-Large	2.75	5.55	1.71	2.88***	38.21	0.68	3.86	0.61	0.75***	10.00
Aggregate-Large	2.58	5.19	1.72	2.72***	33.85	0.91	3.11	1.01	1.05***	12.14

Overall, conventional long–short strategies yield the strongest performance (Table 8). Across all specifications, stocks identified by our tree-based clustering approach as highly predictable deliver markedly superior investment performance relative to others. Excluding these stocks significantly reduces attainable portfolio gains, underscoring their outsized contribution to overall profitability. Extending the mosaic evidence of return predictability to practical investment applications, investors can enhance economic gains by allocating capital toward stocks with high earnings surprises, strong valuation ratios, or low trading volumes (see Figure 3). Consistent with our central argument, the cluster-wise predictive model persistently outperforms the

strengthen—when allowing for regime-dependent temporal variation and large-cap constraints.

global homogeneous benchmark, highlighting the limitations of naive homogeneous approaches.

6 Conclusion

This paper advances the understanding of return predictability by proposing a novel perspective that treats predictability as an intrinsic, time-varying characteristic of individual assets, rather than a uniform property of predictors or models examined in aggregate. To quantify this heterogeneous predictability, we develop a tree-based clustering framework that groups asset-return observations by firm characteristics and market conditions, thereby maximizing differences in predictability across clusters. Empirically, we uncover substantial cross-sectional heterogeneity and regime-dependent variation in the predictability of U.S. equity returns. Specifically, stocks with high earnings surprises, elevated earnings-to-price ratios, and low trading volumes exhibit the strongest predictability. Moreover, predictability varies systematically with market conditions, weakening during periods of low dividend yields and peaking when dividend yields are high and market liquidity is low.

Importantly, portfolios that exploit predictability differentials between heterogeneous and homogeneous models successfully capture model misspecification risk, generating monthly alphas exceeding 1%. Investing in highly predictable stocks also yields significant economic benefits, achieving out-of-sample annualized Sharpe ratios of approximately 2 over the past two decades and substantially outperforming standard benchmarks. Our findings have important implications for asset pricing theory and portfolio management. They highlight the value of incorporating heterogeneous and time-varying predictability into both empirical research and practical investment strategies, offering new insights into the "mosaics of predictability" that characterize the U.S. equity market.

References

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

Aït-Sahalia, Y., J. Fan, L. Xue, and Y. Zhou (2022). How and when are high-frequency stock returns predictable? Technical report, National Bureau of Economic Research.

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.

Engelberg, J., R. D. McLean, and J. Pontiff (2018). Anomalies and news. *Journal of Finance 73*(5), 1971–2001.

Evgeniou, T., A. Gueciouer, and R. Prieto (2023). Uncovering Sparsity and Heterogeneity in Firm-Level Return Predictability Using Machine Learning. *Journal of Financial and Quantitative Analysis 58*(8), 3384–3419.

Fama, E. F. and K. R. French (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.

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)

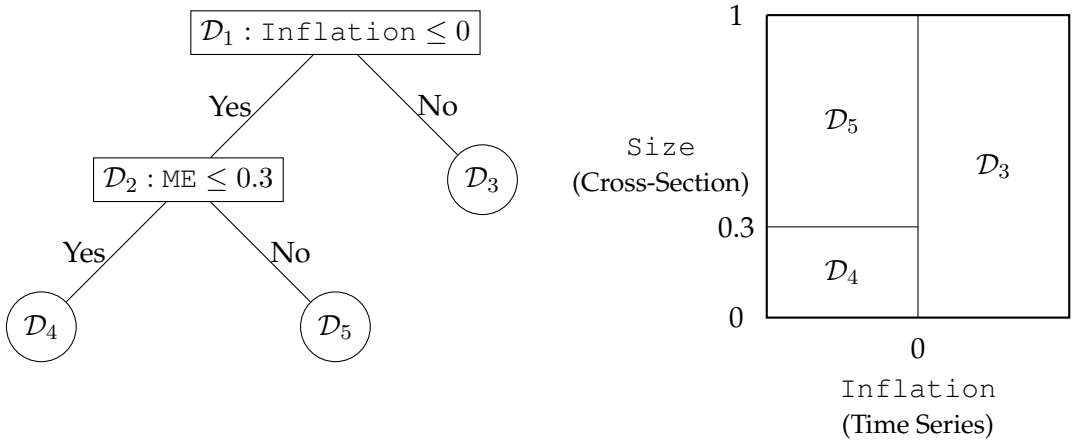
No.	Acronym	Description	Category
22	ill	Illiquidity rolling 3m	Liquidity
23	lev	Leverage	Value
24	lgr	Growth in long-term debt	Investment
25	maxret	Maximum daily returns rolling 3m	Volatility
26	me	Market equity	Size
27	me_ia	Industry-adjusted size	Size
28	mom12m	Momentum rolling 12m	Momentum
29	mom1m	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

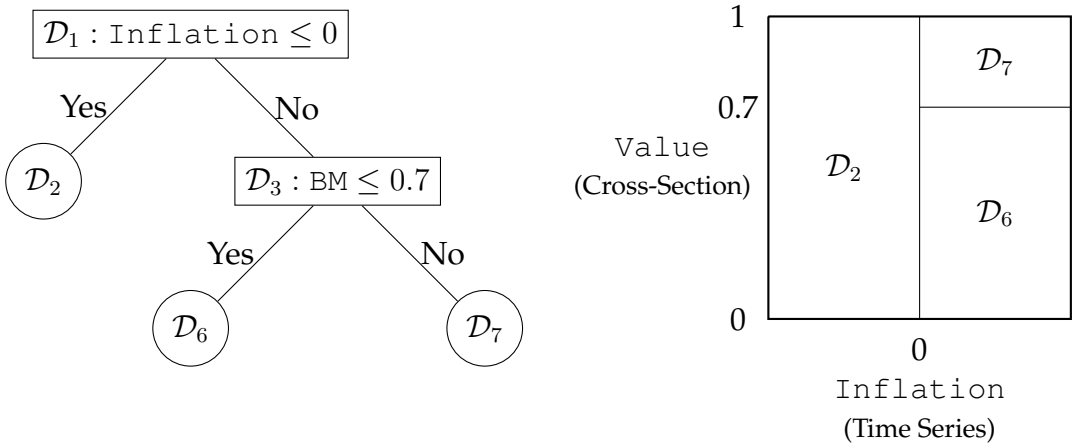
The initial split is determined based on the condition between the demeaning inflation and 0 (see Figure 2). Subsequently, we iteratively explore the optimal choice for the second split, as two candidates in Figure A.1 below. This split can occur either on the left based on size ($\texttt{ME} \leq 0.3$), 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 $\geq d_{\text{max}}$ **then**
      return.
    else
      Search the tree, find all potential leaf nodes $\mathcal{N}$
      for each leaf node N in $\mathcal{N}$ **do** ▷ Loop over all current leaf nodes
        for each split candidate $\tilde{c}_{p,k,N}$ in $\mathcal{C}_N$ **do**
          Partition data temporally in N according to $\tilde{c}_{p,k,N}$.
          if Left or right child node cannot satisfy the minimal leaf size **then**
            continue.
          else
            Obtain cluster-wise return forecasts as in (2).
            Calculate the cluster-based R_j^2 by (3).
          end if
        end for
      end for
      Find the best leaf node and split rule that maximizes the 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).

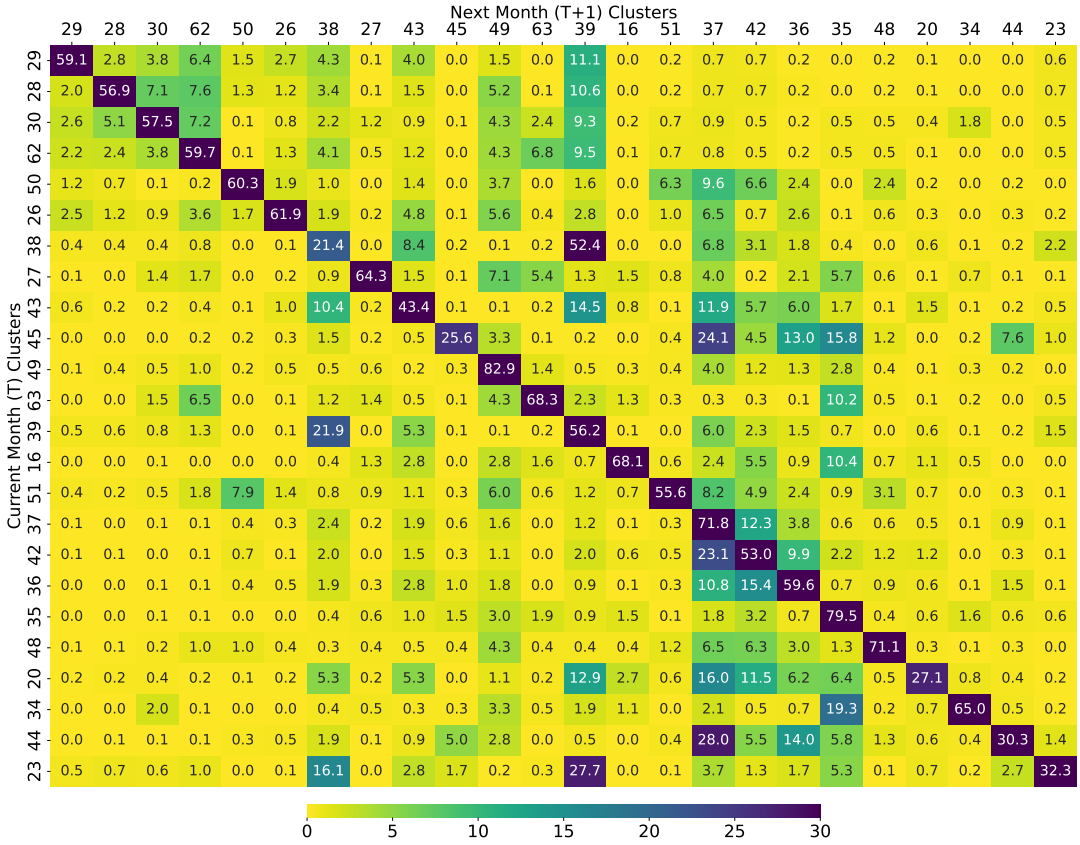
IV Additional Empirical Results

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**

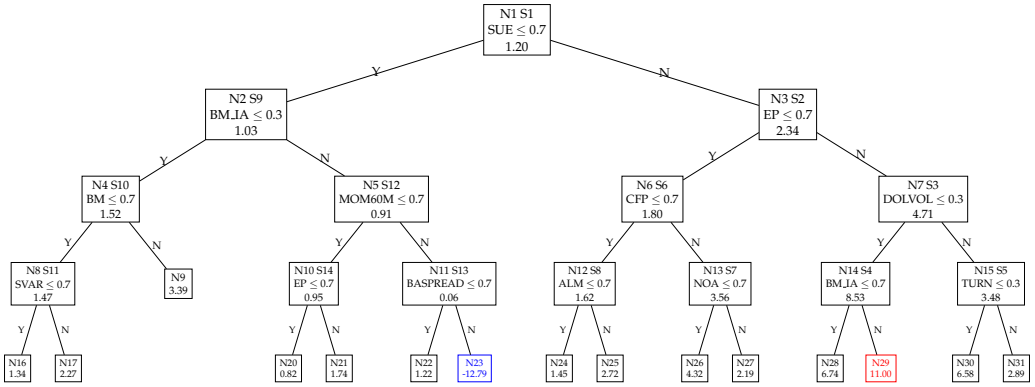
This figure presents the monthly average transition matrix of clustering results shown in Figure 3. Each row represents the probabilities of transitioning from the current cluster (time T) to other clusters in the next month (time $T+1$). Clusters in the rows and columns are sorted according to their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



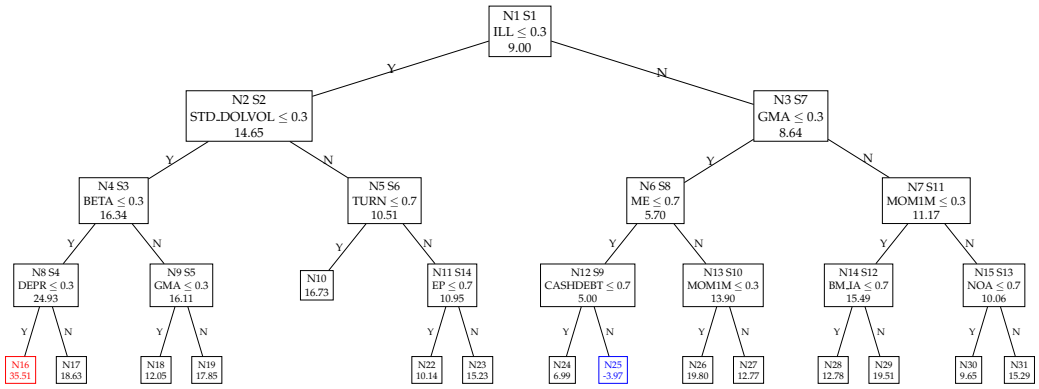
IV.2 Further Cross-Sectional Tree Structures across Market Regimes

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

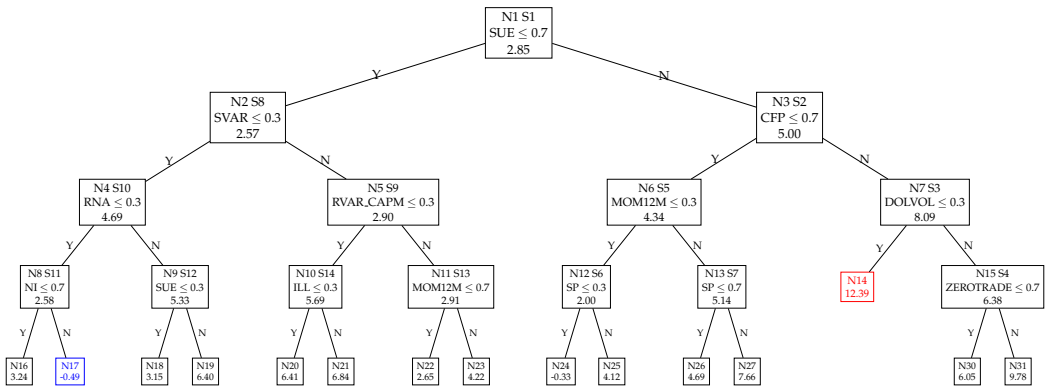
This figure illustrates the further cross-sectional split tree-based clustering structure under each regime, as determined by Figure 6, using monthly data from 1973 to 2022. Each tree splits stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. The terminal 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 the 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.3 Structural Breaks

Table A.3: **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.4: **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.4 Predictability Differential Strategy Testing

This subsection provides a detailed breakdown of the long- and short-leg portfolios for the predictability differential strategy, as shown in Table 6. Regardless of the number of highly predictable clusters included, the long-leg portfolios in Table A.5 consistently generate significant abnormal returns across both INS and OOS periods. In contrast, the short-leg portfolios generally yield insignificant alphas out of the sample. This reinforces the conclusion that our discovered strategy is primarily driven by stocks with the highest model misspecification risks (long-leg), as highlighted in Section 5.2.

Table A.5: **Testing the Predictability Differential Strategy**

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 6. 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.5 Investment Performance with Transaction Costs

This subsection re-examines both types of investment strategies introduced in the last two parts of Section 5. 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: **Predictability Differential Portfolio with Transaction Costs**

This table presents the results after considering transaction costs, as shown in Table 6. 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	3.27	2.83	1.62	1.26	2.59	1.54	1.12	0.90
Std	7.42	4.61	3.43	2.53	6.41	4.09	3.64	2.64
SR	1.53	2.13	1.64	1.72	1.40	1.31	1.06	1.18
			Panel B: Abnormal Returns					
CAPM	3.37***	2.87***	1.58***	1.23***	2.84***	1.61***	1.08***	0.90***
	(8.67)	(11.75)	(8.75)	(9.24)	(6.90)	(6.01)	(4.54)	(5.17)
FF3	3.00***	2.63***	1.44***	1.10***	2.86***	1.61***	1.10***	0.91***
	(7.75)	(11.28)	(9.10)	(9.20)	(6.98)	(6.15)	(4.94)	(5.53)
FF5	2.72***	2.52***	1.39***	1.06***	2.83***	1.80***	1.19***	0.97***
	(6.90)	(10.49)	(8.54)	(8.58)	(6.70)	(6.77)	(5.17)	(5.71)
FF5+MOM+IVOL	2.28***	2.33***	1.34***	1.07***	2.81***	1.85***	1.25***	1.02***
	(5.78)	(9.56)	(8.01)	(8.42)	(6.61)	(7.06)	(5.62)	(6.15)
Q5	2.28***	2.21***	1.30***	1.02***	2.90***	1.89***	1.41***	1.11***
	(4.99)	(7.96)	(6.83)	(7.05)	(6.77)	(6.99)	(6.29)	(6.68)
BS6	2.17***	2.29***	1.32***	1.00***	2.70***	1.83***	1.28***	1.03***
	(5.16)	(8.88)	(7.47)	(7.49)	(6.40)	(6.93)	(5.78)	(6.36)
DHS3	2.70***	2.50***	1.59***	1.27***	2.73***	1.78***	1.23***	0.99***
	(6.02)	(8.87)	(7.63)	(8.20)	(6.46)	(6.62)	(5.07)	(5.63)
SY4	2.12***	2.26***	1.30***	1.00***	3.65***	2.58***	1.66***	1.38***
	(5.16)	(8.93)	(7.52)	(7.63)	(6.74)	(8.27)	(7.56)	(7.49)

Table A.6 above reports the performance of the predictability differential strategy after accounting for transaction costs. The results show that the investment profitability remains robust: average returns and Sharpe ratios decline only modestly (all combinations exhibit an OOS Sharpe ratio larger than 1), and all factor-adjusted alphas

remain economically and statistically significant across models (at least over 0.9% for OOS). These findings confirm that the return premia associated with heterogeneity in return predictability are not driven by excessive trading or market frictions, underscoring the economic relevance of local predictability advantages.

Table A.7: **Forecast-Implied Investment Performance (CS Cluster)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 7). 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 6). We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap subsamples 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

Similarly, Table A.7 presents performance metrics for transaction-cost-adjusted forecast-implied portfolios. First, the results show that stocks with high predictability generate less significant OOS investment gains compared with Table 7. 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 magnitude of return forecasts drives performance. Large-cap subsamples 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.

V Sharpe Ratio Partitions

To complement the statistical measure of return predictability (R^2) discussed in the main text, we further evaluate an economic version by using Sharpe ratios to measure return predictability and also to complete the clustering procedure. All steps are identical to those described in Section 2, except that we replace R^2 in Eq. (4) with Sharpe ratios. For each split, we construct forecast-weighted portfolios based on all observations within the left and right clusters following Eq. (6), and then compute the corresponding Sharpe ratios. This new split criterion can also yield a tree-based clustering structure similar to Figure 3. Although the selected characteristics may differ, we can still observe clear heterogeneity gaps in return predictability across clusters.

Table A.8: **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 determined by Sharpe ratios. 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 subsample 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	1.62	0.54	0.52	0.86	0.46	0.46
Medium	1.47	0.52	0.54	0.08	0.37	0.30
Low	1.24	0.54	0.63	0.17	0.50	0.48
		Panel B: Cluster-Wise Forecasts				
Aggregate	2.14	1.19	0.94	0.04	0.62	0.63
High	3.01	1.96	1.70	0.69	0.98	1.04
Medium	1.94	1.03	0.76	-0.05	0.53	0.51
Low	1.65	0.51	0.60	-1.61	0.36	0.42

Following all previous analyses in Section 4 and 5, we present here two of the most important empirical results to test the previous mosaic patterns of return predictability. Table A.8 reports both INS and OOS results for predictive models when the clustering structure is determined by Sharpe ratios. The overall pattern closely mirrors that in Table 4, where clusters identified as highly predictable continue to exhibit stronger performance across all predictive methods. In both panels, the high-

predictability clusters deliver the highest values of R^2, while the low-predictability clusters perform poorly or even negatively in the OOS evaluation. These results indicate that using the Sharpe ratio as the clustering criterion yields partitions consistent with those obtained under the R^2-based approach. In other words, predictability defined in economic terms leads to similar cross-sectional rankings and confirms the robustness of our main findings.

Furthermore, Table A.9 also examines the forecast-implied investment performance based on Sharpe ratios as the split criterion. The overall results closely resemble those in Tables 7 and 8, showing that clusters identified as highly predictable continue to deliver superior economic performance in both INS and OOS periods. The high-predictability clusters consistently exhibit greater average returns, alphas, and Sharpe ratios, while the low-predictability clusters display weak or negligible performance. These patterns remain robust across portfolio weighting schemes, indicating that the economic relevance of predictability is not sensitive to model specification.

Overall, these evidences demonstrate that the heterogeneity in return predictability is both statistically and economically meaningful. Regardless of whether the clustering criterion is based on R^2 or the Sharpe ratio, we consistently observe that highly predictable clusters translate their predictive abilities into stronger and more persistent investment performance.

Table A.9: **Forecast-Implied Investment Performance (SR Partitions)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 7). 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 6). We show seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap subsamples 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	0.68	3.87	0.61	0.37***	21.70	0.77	4.27	0.63	−0.00	17.26
High	2.09	3.46	2.10	1.97***	8.29	1.23	3.82	1.11	0.80***	20.43
Medium	0.61	4.28	0.49	0.24***	20.57	0.77	4.41	0.60	−0.03	17.24
Low	0.76	4.46	0.59	0.52***	24.04	0.92	3.99	0.79	0.27**	17.70
	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	1.40	3.51	1.38	1.17***	15.32	1.05	4.43	0.82	0.33**	20.67
High	2.59	4.16	2.15	2.44***	8.45	1.64	4.09	1.39	1.19***	19.38
Medium	1.24	3.54	1.21	1.01***	16.07	0.96	4.58	0.72	0.20	20.71
Low	0.80	4.70	0.59	0.46***	28.28	0.93	4.91	0.65	0.12	24.05
	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	2.18	4.10	1.84	1.95***	17.21	1.61	4.71	1.19	0.88***	20.92
High	3.82	5.17	2.56	3.64***	14.33	2.57	4.71	1.89	2.04***	20.15
Medium	1.83	4.05	1.56	1.60***	17.65	1.39	4.75	1.01	0.63***	21.26
Low	1.00	4.97	0.70	0.63***	28.81	1.06	5.27	0.70	0.19	23.07
	Panel D: Value-Weighted Long-Short									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	3.60	5.66	2.20	3.71***	20.89	0.99	4.98	0.69	1.31***	16.70
Aggregate	4.87	6.17	2.73	5.01***	21.93	2.50	4.66	1.86	2.56***	13.20
High	7.58	4.59	5.71	7.56***	12.67	4.99	5.89	2.93	4.90***	19.77
Medium	4.58	5.77	2.75	4.75***	30.66	2.26	4.78	1.63	2.33***	13.13
Low	2.04	5.20	1.36	2.10***	24.09	0.95	6.70	0.49	1.04**	42.51
	Panel E: Equal-Weighted Long-Short									
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	4.51	5.22	2.99	4.60***	29.85	2.12	5.34	1.38	2.24***	22.27
Aggregate	5.83	4.35	4.64	5.93***	13.68	3.35	4.04	2.88	3.52***	15.17
High	9.07	4.71	6.66	9.04***	12.22	5.95	5.61	3.68	5.91***	14.65
Medium	5.00	5.17	3.35	5.16***	25.93	2.65	4.53	2.03	2.82***	18.73
Low	2.51	4.88	1.78	2.64***	23.40	0.78	6.20	0.44	0.95**	34.56

Internet Appendix

IA.I Robustness Check

Amid rising concerns about the stability of results 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, such as 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.

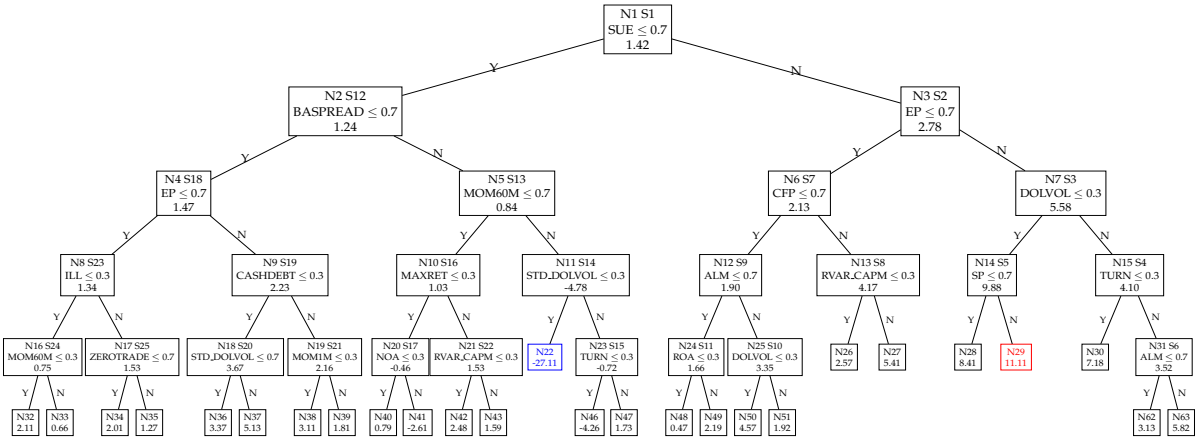
As detailed in Section 3.2, our cross-sectional analyses use a 30-year rolling window updated every five years. The optimal splitting variables show consistent robustness across subsamples—the first two splits in tree structures by different periods (Figure IA.1 for 1978–2007, 1983–2012, and 1988–2007) 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 characteristics: 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 to Table 1), predictability dispersion persists across specifications: the top cluster consistently delivers INS 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.

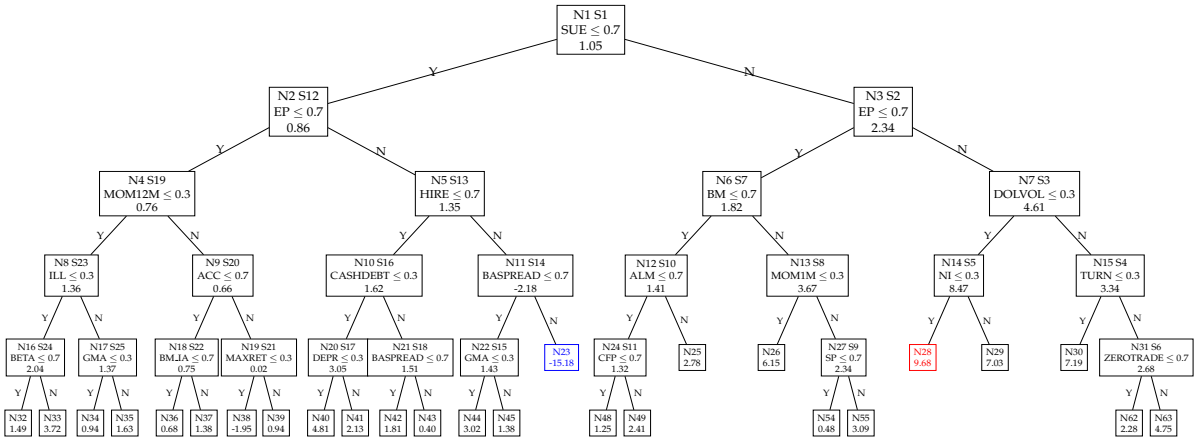
Additional sensitivity analyses, including variations in tree depth, minimum leaf observations, and winsorization levels, demonstrate consistent stability in selected firm characteristics. While splitting sequences differ slightly, the same variables reappear across specifications. These findings confirm 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)**

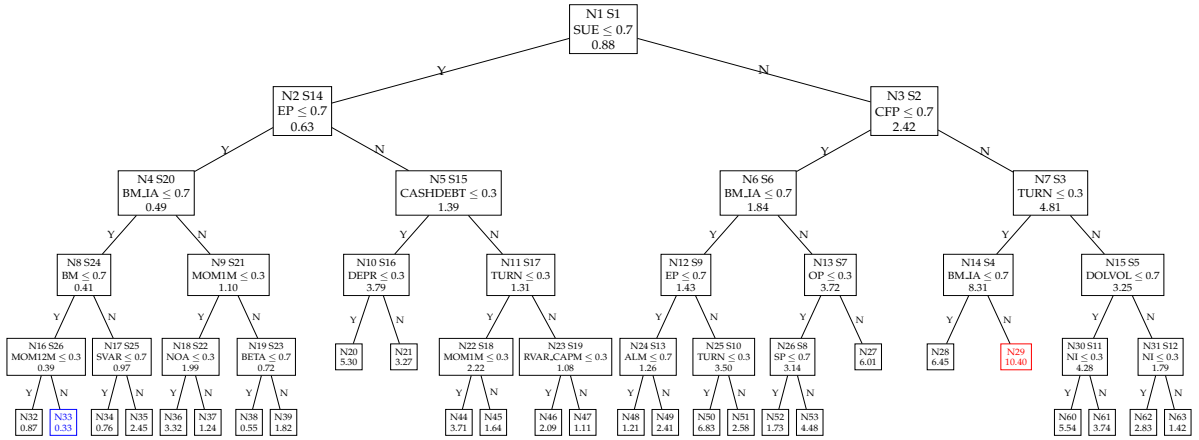
This figure illustrates the cross-sectional tree-based clustering derived from the other three monthly data sample periods (1978–2007, 1983-2012, 1988-2017). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the range of [0, 1]. Terminal leaves represent 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 ($R^2_{CMG} = R^2_C - R^2_G$). 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}	Avg_{EW}	SR_{EW}	Avg_{VW}	SR_{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 ($R^2_{CMG} = R^2_C - R^2_G$). 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}	Avg_{EW}	SR_{EW}	Avg_{VW}	SR_{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 ($R^2_{CMG} = R^2_C - R^2_G$). 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}	Avg_{EW}	SR_{EW}	Avg_{VW}	SR_{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 5 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 subsample baseline long-short investment performance for forecast-implied portfolios (Equation 7). 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 6). 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 subsample baseline long-short investment performance for forecast-implied portfolios (Equation 7). 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 6). 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