NBER WORKING PAPER SERIES

MOSAICS OF PREDICTABILITY

Lin William Cong
Guanhao Feng
Jingyu He
Yuanzhi Wang

Working Paper 35158
http://www.nber.org/papers/w35158

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We thank Chunrong Ai, Doron Avramov, Daniele Bianchi, Pietro Bini, Svetlana Bryzgalova, 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), Hong Zhang, Xiaoyan Zhang, Yi Zhang, Geoffery Zheng, Yifeng Zhu (Discussant), and participants at the 2026 Shanghai Forum, 2025 SFS Cavalcade Asia-Pacific Conference, 2025 CICF, 2025 Global AI Finance Research Conference, 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Mosaics of Predictability
Lin William Cong, Guanhao Feng, Jingyu He, and Yuanzhi Wang
NBER Working Paper No. 35158
April 2026
JEL No. C38, C53, C55, G12

ABSTRACT

We argue that return predictability is a latent, asset-specific, and state-dependent characteristic. We develop an interpretable Panel Tree that endogenously partitions the U.S. equity panel into out-of-sample and persistent "mosaic" patterns, and estimate cluster-specific forecasting models. Predictability concentrates in stocks with large earnings surprises, high earnings–price ratios, and low trading volume. It is countercyclical, stronger when market dividend yields are high and liquidity is low. Accounting for predictability heterogeneity, which conventional models ignore, improves forecasts and yields portfolios with out-of-sample Sharpe ratios around 2. Across 50 years of data, the mosaic map shows where signals arise and where noise dominates.

Lin William Cong
Cornell University
and NBER
will.cong@cornell.edu

Guanhao Feng
City University of Hong Kong
gavin.feng@cityu.edu.hk

Jingyu He
City University of Hong Kong
jingyuhe@cityu.edu.hk

Yuanzhi Wang
City University of Hong Kong
yuanzwang5-c@my.cityu.edu.hk

1 Introduction

Forecasting asset returns has long been central to asset pricing, portfolio choice, and market efficiency. A large and influential literature documents statistically and economically meaningful return predictability across assets and horizons,[1] and proposes econometric tests for it.[2] Yet most standard empirical designs impose homogeneity, treating predictability as similar across assets, states, or both. Predictive relationships are typically evaluated through pooled regressions, aggregate forecasting models, or representative portfolios, implicitly assuming that the inherent difficulty of forecasting returns is uniform across the return panel. By "predictability," we refer to the forecastability of conditional expected returns, rather than ex post realized premiums or structural estimates of risk premia levels.

Economic intuition suggests that predictability varies systematically across assets and market regimes. In other words, information frictions, limits to arbitrage, and state dependence in expected returns imply non-uniform predictability. Predictability differs because assets face different trading costs, limited attention, information frictions, and macro exposures, and market conditions can amplify these gaps. Consistent with this view, predictability often concentrates in particular regimes, such as recessions (e.g., Henkel et al., 2011; Farmer et al., 2023), and in specific groups of firms, such as micro-cap or distressed stocks (e.g., Avramov et al., 2023).

We argue that return predictability is a latent, asset-specific, state-dependent characteristic, analogous to beta or volatility. Predictability governs the signal-to-noise ratio in conditional expected returns and, in turn, the attainable gains relative to a naive benchmark. When predictability varies across assets and states, a single homogeneous predictive model can be misspecified. Such misspecification affects not only statistical fit but also asset prices, portfolio allocations, and the cross section of expected returns.

[1] Market return predictability often aligns with business cycle indicators (e.g., Fama and French, 1989; Campbell and Thompson, 2008). Cross-sectional return variations are widely linked to firm characteristics (e.g., Fama and French, 1992; Lewellen, 2015).

[2] These include predictive regressions to assess the significance of predictors (e.g., Stambaugh, 1999), security sorting methods (e.g., Jensen et al., 1972), and cross-sectional models (e.g., Kelly and Pruitt, 2013; Han et al., 2024) for the evaluation of long-short portfolios, along with machine learning techniques for forecast-implied investment performance (e.g., Gu et al., 2020; Freyberger et al., 2020).

Viewing predictability as an economic object has implications. If variation in predictability is tied to state-dependent movements in risk premia (rather than their level), it should co-move with business-cycle conditions and aggregate liquidity. If it is driven by information frictions and limits to arbitrage, it should be concentrated in assets where information is costly and trading is thin (e.g., low-volume stocks). Finally, if investors rely on forecasting structures that impose homogeneity across assets and states, then cross-sectional differences in predictability relative to such benchmarks may be associated with systematic return differentials. The challenge is that the population degree of predictability is unobservable, and there is no single empirical measure that applies uniformly across assets and states.[3]

In that spirit, this paper develops a framework to systematically measure and analyze heterogeneous return predictability across assets in the cross section and over identified market regimes. Rather than focusing on new predictors or aggregate forecast gains, we examine which assets are more predictable, under which market conditions, and what that implies economically. Our approach treats predictability as an object of interest in its own right and aims to characterize its cross-sectional and regime-dependent structure in a unified way. Doing so requires a method that allows predictive relationships to vary across firms and states, while remaining disciplined enough to deliver stable, out-of-sample estimates.

To operationalize this idea, we build on the Panel Tree (P-Tree) framework of Cong, Feng, He, and He (2025), but adapt it to target predictability heterogeneity. Specifically, we develop a customized, goal-oriented clustering algorithm that partitions the asset-return panel by predictability level. The key insight is that, while predictability is unobservable at the individual asset level, it can be inferred at the group level by clustering assets that share similar signal-to-noise properties. We interpret the cluster-level R^2 from properly regularized predictive models as a reduced-form measure of return predictability, capturing the intrinsic difficulty of forecasting returns

[3]In the literature, predictability is often linked to empirical concepts such as anomaly returns (Hou et al., 2020), out-of-sample R^2 (e.g., Welch and Goyal, 2008; Campbell and Thompson, 2008), and the performance of forecast-implied portfolios (e.g., Lewellen, 2015; Gu et al., 2020).

within each group.[4] The clustering algorithm focuses on maximizing the difference in predictability across different clusters. Unlike pooled models that report a single aggregate measure or firm-level regressions that suffer from limited data, our approach strikes a balance by endogenously forming economically interpretable clusters and estimating cluster-specific predictive models.

Applying this framework to U.S. monthly stock returns (1973–2022) with 51 firm characteristics and eight market predictors, we uncover significant heterogeneity in return predictability across the panel. We use the resulting mosaics to: (i) identify where predictability concentrates in the cross-section, (ii) examine how it evolves across market regimes, and (iii) assess the economic value of accounting for these differences.

Return predictability varies sharply across assets and over market regimes, forming non-uniform patterns that resemble mosaics defined by interactions between firm characteristics and market conditions. In the cross section, stocks with large earnings surprises, high earnings-to-price ratios, low trading volume, and high sales-to-price ratios exhibit the strongest predictability, with in-sample R^2 exceeding 10%. These stocks correspond to the long legs of well-known anomalies such as post-earnings announcement drift (Bernard and Thomas, 1989) and earnings yield effects (Reinganum, 1981), both characterized by persistent abnormal returns. In contrast, stocks with low earnings surprises, high bid-ask spreads, and strong momentum exhibit little or no predictability, with negative R^2 values indicating forecasting performance worse than a naive zero-return benchmark. For comparison, a homogeneous pooled model estimated on the full sample yields an in-sample R^2 of only 1.49%, illustrating how averaging across heterogeneous predictive structures can mask substantial variation in predictability. Out-of-sample results confirm the persistent gap between highly and weakly predictable clusters.

Moreover, predictability also varies substantially over time and is strongly state-dependent. Two market-level variables, the market dividend yield (Fama and French, 1988) and aggregate liquidity levels (Pástor and Stambaugh, 2003), emerge as key

[4]We evaluate alternative measures like the Sharpe ratio and confirm the robustness of our findings.

drivers of regime shifts in predictability, which aligns with previous evidence on state-dependent market return predictability (e.g., Henkel et al., 2011). Predictability is countercyclical, weakening during periods of low market dividend yields and abundant liquidity, and peaking when market dividend yields are high but liquidity is constrained, conditions typically observed in the early stages of recessions. This pattern is consistent with the present-value framework of Campbell and Shiller (1988), in which elevated expected returns during recessions reflect heightened risk premia associated with liquidity and distress risks, rather than purely behavioral mispricing.

Finally, the economic consequences of heterogeneity are substantial. Forecast-implied portfolios based on cluster-specific models outperform the corresponding homogeneous benchmark. The gains are especially pronounced when portfolio weights depend on forecast magnitude, indicating that heterogeneous models add value not only by improving forecast direction, but also by determining how aggressively capital should be allocated across stocks. The most highly predictable cluster earns an average monthly return of 3.41% with an annualized Sharpe ratio of 1.88, whereas the least predictable cluster exhibits sharply negative numbers (-0.06% and -0.02).

We further show in the appendix that gaps between leaf-specific and pooled-model predictability map into sizable long-short return differentials, linking statistical misspecification to economically relevant portfolio outcomes. These results are robust to transaction costs and demonstrate that heterogeneity in return predictability can be directly translated into economically meaningful value, extending earlier studies that focus on aggregate or homogeneous predictability (e.g., Rapach et al., 2010; Kelly et al., 2024). Homogeneous models can miss economically relevant heterogeneity in predictability (e.g., Feng and He, 2022; Evgeniou et al., 2023) so that exploiting grouped heterogeneity has substantial practical value.

Literature. Our study contributes to the return predictability literature by analyzing its heterogeneity systematically in a data-driven manner for the first time. Prior studies document heterogeneity in ad hoc settings (e.g., Avramov, 2002; Green et al., 2017) or under specific market conditions (e.g., Henkel et al., 2011). Recent work on "pockets

of predictability" (Farmer et al., 2023, 2024) is sensitive to human specifications (Cakici et al., 2025). In contrast, our framework reconciles these inconsistencies by connecting predictability to asset characteristics and regimes, offering a coherent explanation for the instability and inconsistencies observed in prior OOS evaluations (e.g., Pesaran and Timmermann, 1995; Lettau and Van Nieuwerburgh, 2008; Welch and Goyal, 2008).

Conceptually, we contribute to the literature by treating return predictability as a state-dependent, asset-specific characteristic. Our approach extends traditional time-varying coefficient models by endogenously identifying clusters via high-dimensional characteristics and market predictors. Building on the P-Tree framework of Cong et al. (2025), we adapt it to optimize for predictability heterogeneity. Unlike approaches relying on K-means clustering (e.g., Evgeniou et al., 2023) or fixed portfolios, our framework allows assets to transition dynamically across clusters as characteristics and regimes evolve. While closely related to Feng and He (2022), who show that incorporating cross-asset information enhances informational efficiency via a Bayesian hierarchical framework, our primary innovation departs from the reliance on exogenously given portfolios by integrating data-driven clustering and cluster-specific model predictions within a unified framework.[5]

In terms of methodological innovation, we extend Cong et al. (2025) and Cong et al. (2023) by developing an economically guided P-Tree framework that builds upon the efficient algorithm presented in He and Hahn (2023). Our study thus contributes to 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 intelligence 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

[5]In the high-frequency domain, Aït-Sahalia et al. (2025) find returns and trade direction are more predictable for stocks with lower prices, liquidity, volatility, and weaker correlation with the market.

heterogeneous levels of return predictability.

Finally, 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. Unlike fixed-cluster approaches, our generative framework enables assets to dynamically transition across clusters as their characteristics evolve. 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. Section 5 reports out-of-sample validation and the main portfolio implications of heterogeneous predictability. Section 6 concludes. The Appendix provides algorithmic details, data descriptions, additional strategies, and supplementary results, while the Online Appendix contains additional robustness checks and extended empirical discussions.

2 Predictability and P-Tree Framework for Heterogeneity

We first describe how we measure unobservable return predictability, then provide a detailed implementation of the P-Tree framework.

2.1 Measurement of Latent Return Predictability

A challenge in studying return predictability is that the (population) degree of predictability is not observable. Existing studies often assess predictability through hypothesis testing, out-of-sample forecast accuracies, or the performance of forecast-implied portfolios. While informative for specific empirical purposes, these approaches conflate two distinct objects: the presence of conditional structure in returns and the economic outcomes of exploiting such structure through portfolio choice.

We instead conceptualize return predictability as a latent, asset-specific, and state-dependent characteristic, analogous to beta or volatility, that captures how much of return variation is driven by time-varying conditional means that can be proxied from predictors. In this sense, predictability is a property of the return data-generating process (DGP) rather than an attribute of any particular forecasting model, predictor set, or trading strategy.

To be more specific, we interpret in-sample R^2 as a natural metric for return predictability.[6] Although R^2 is traditionally viewed as a measure of model fit, we adopt a data-centric interpretation: the population $\mathcal{R}^2$ captures the intrinsic signal-to-noise ratio of the return-generating process. When predictive models are properly trained, the corresponding empirical R^2 provides a reasonable approximation. To avoid confusion, we use $\mathcal{R}^2$ to denote the population predictability and R^2 to denote its sample analogue computed from fitted models. Specifically, consider the decomposition for asset returns:

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

where $\mathbb{E}[\varepsilon_{i,t}|I_{t-1}] = 0$. From this perspective, intrinsic predictability can be characterized by the signal-to-noise ratio:

$$\mathcal{R}^2_{i,t} \equiv 1 - \frac{\mathrm{Var}(\varepsilon_{i,t})}{\mathrm{Var}(r_{i,t})} = \frac{\mathrm{Var}(\mathbb{E}_{t-1}[r_{i,t}])}{\mathrm{Var}(r_{i,t})},$$

where variances are taken with respect to the unconditional distribution over information states. A higher $\mathcal{R}^2_{i,t}$ implies that a larger fraction of return variation is attributable to systematic conditional mean variation; thus, it is easier for a statistical or machine learning model to learn the conditional mean. On the contrary, a low $\mathcal{R}^2_{i,t}$ indicates that idiosyncratic noise dominates and conditional expected returns are weakly informative and hard to capture by forecasting models.

While $\mathcal{R}^2_{i,t}$ provides a conceptually clean definition, it is not directly observable. In practice, only a single realization of $r_{i,t}$ is observed for each asset-time pair, rendering point-wise estimation infeasible. As a result, empirical studies typically rely

[6] We also complement the framework by using the Sharpe ratio as an economic predictability measure for robustness. The results in the Appendix Section A. 5 are consistent with our main findings.

on pooled-sample analogues, often estimating a single predictive model on the full panel and reporting a single aggregate R^2 (e.g., Gu et al., 2020; Freyberger et al., 2020). This homogeneous approximation implicitly assumes that the DGP-level predictability, $\mathcal{R}^2_{i,t}$, is constant across assets and over time. We relax this restriction by partitioning asset-return observations into disjoint clusters and approximating the latent $\mathcal{R}^2_{i,t}$ using cluster-specific empirical measures (detailed explanations of R^2_j in the next subsection). These measures summarize the average signal-to-noise ratio within each subgroup, and the clustering procedure groups observations with similar latent predictability while maximizing the empirical dispersion of R^2 across clusters. With proper training, cluster-specific R^2 value can provide a stable approximation to the predictive content of returns, with improvements bounded by the intrinsic signal-to-noise ratio from the data.

Additionally, we do not use out-of-sample R^2 to guide clustering, as doing so would rely on test-period outcomes for model selection. Since the clustering procedure determines the partition endogenously, using test-sample information to guide split decisions would contaminate model selection and invalidate subsequent out-of-sample evaluations. Instead, we treat clustering and cluster-wise model estimation as a unified in-sample exercise, and subsequently assess the resulting heterogeneity using out-of-sample forecasts and portfolio performance, thereby ensuring a clean separation between measuring latent predictability and evaluating its economic relevance.

2.2 Cluster-Level Predictive Modeling and Predictability Score

We next specify the predictive model used within each cluster, along with its associated predictability score. In this subsection, we assume a cluster (or leaf) j is given. Then, we estimate a regularized predictive regression and compute a cluster-specific R^2_j, which serves as our empirical measure of predictability for return observations within this cluster.

Let the data be denoted by $\mathcal{D} = \{(r_{i,t}, \mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}); i = 1, \ldots, N; t = 1, \ldots, T_i\}$, where T_i indexes the (possibly unbalanced) time periods available for asset i. $r_{i,t}$ 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. Let $\mathbf{s}_{i,t-1} = \{\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}\}$ denote the combined predictor set.

In general, expected excess returns for each asset i at time t may satisfy

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

allowing the predictive relation to vary across assets and over time. In practice, limited observations for individual assets render the direct estimation of each $g_{i,t}(\cdot)$ infeasible, motivating the use of pooled forecasting models that impose a common predictive relationship across assets (e.g., Gu et al., 2020). Our approach instead estimates a cluster-wise predictive function that is homogeneous within each cluster but allowed to differ across clusters:

$$\mathbb{E}_{t-1}[r_{i,t}] = g_j(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}), \tag{3}$$

where all stock-month observations assigned to cluster j share the same predictive function $g_j(\cdot)$.

Inverse-variance-weighted Ridge regression. To perform predictive modeling in a high-dimensional setting, we use Ridge regression estimated with inverse-volatility weights, where each observation is weighted by the inverse of an estimate of the stock's conditional return variance. This choice mitigates multicollinearity in the predictor set and reduces the undue influence of high-variance (often small-cap) stocks in pooled estimation, as emphasized by Fama and French (2008) and Hou et al. (2020). Meanwhile, Ridge regression performs more robustly than Lasso in weak-signal environments (Shen and Xiu, 2024). Therefore, for the j-th cluster, we estimate the predictive regression

$$g_j(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}) = \beta_{j,0} + \beta_{j,1}^{\intercal}\mathbf{s}_{i,t-1}$$

by solving

$$\widehat{\beta}_j = \arg\min_{\beta_{j,0},\beta_{j,1}^{\intercal}} \left\{ \frac{1}{N_{\text{leaf}_j}} \sum_{\{i,t\}\in \text{leaf}_j} w_{i,t-1} \left(r_{i,t} - \beta_{j,0} - \beta_{j,1}^{\intercal}\mathbf{s}_{i,t-1}\right)^2 + \lambda\|\beta\|_2^2 \right\}, \tag{4}$$

where $w_{i,t-1} = 1/\sigma^2_{i,t-1}$ is the inverse-variance weight, and $\sigma^2_{i,t-1}$ is a rolling-window estimate of stock i's return variance available at $t-1$. For others, $\mathbf{s}_{i,t-1} = \{\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}\}$ represents the set of predictors, $\beta_j = (\beta_{j,0}, \beta_{j,1}^\intercal)$ are corresponding coefficients, and λ is the regularization hyperparameter. Importantly, we select λ by cross-validation within each cluster, thereby limiting both underfitting and overfitting, and ensuring a stable mapping from data to predictability scores.

Cluster-level predictability score. Given the fitted forecast $\widehat{r}_{i,t} = \widehat{\beta}_{j,0} + \widehat{\beta}_{j,1}^\intercal \mathbf{s}_{i,t-1}$ for all observations in leaf j (i.e., $\{i,t\} \in \text{leaf}_j$), we define the cluster-specific in-sample R^2 as

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

This in-sample score evaluates predictive content relative to a zero-return benchmark, which is analogous to the Campbell and Thompson (2008) out-of-sample R^2 definition. We treat R^2_j as the empirical predictability score for each cluster j, reflecting the best attainable performance within the cluster-specific regularized predictive model. Subsequently, the P-Tree algorithm, described in the next subsection, uses dispersion in $\{R^2_j\}$ across candidate partitions to construct clusters that differ sharply in latent predictability.

2.3 P-Tree Construction: The First Split

Given the cluster-level predictability score defined in the previous subsection, we now describe how the Panel Tree (P-Tree) algorithm constructs clusters endogenously through a sequence of splits on observable variables.

The P-Tree represents heterogeneity in return predictability using a decision-tree structure that recursively partitions the stock return panel into economically interpretable clusters. Each split condition is based on either firm characteristics or aggregate predictors. Splits on firm characteristics capture cross-sectional heterogeneity in predictability across stocks, while splits on aggregate predictors or calendar variables isolate market regimes in which predictability differs over time. By combining both dimensions within a single tree, the P-Tree provides a unified representation of

predictability heterogeneity across assets and market states.

Our approach iteratively partitions the panel by splitting one existing leaf at each iteration, thereby increasing the number of terminal clusters by one. At each node, the algorithm selects a split that maximizes separation in predictability between the resulting subsamples, where predictability is measured by the cluster-level score R_j^2. This goal-oriented design distinguishes the P-Tree from unsupervised clustering methods, such as K-means, which group observations based on distance metrics, and from approaches like Evgeniou et al. (2023), which rely on pre-specified characteristic-based clusters. In contrast, the P-Tree endogenously selects splitting variables and cut-points to maximize predictability heterogeneity, yielding a single, deterministic, and economically interpretable partition of the return panel.

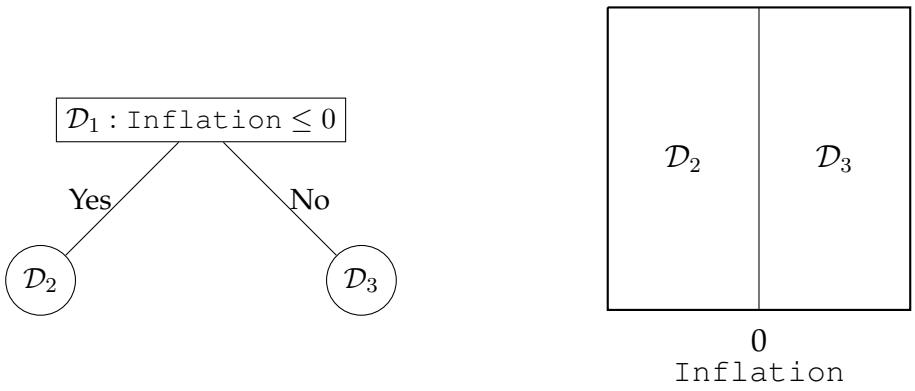
Figure 1 illustrates a *split candidate* for the first split in a tree, where the root node, $\mathcal{D}_1$, representing the full dataset, is divided into two clusters (*leaf nodes* $\mathcal{D}_2$ and $\mathcal{D}_3$). Following the terminology in Cong et al. (2025), leaf nodes refer to the nodes 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. Observations of asset returns that satisfy this split rule are assigned to the left leaf node $\mathcal{D}_2$, while those that do not are placed in the right-hand side $\mathcal{D}_3$. To assess the quality of this split candidate, we propose a goal-oriented split criterion to measure its ability to separate returns into high- and low-predictability groups.

The first split candidate divides the entire return sample $\mathcal{D}_1$ into two clusters, $\mathcal{D}_2$ and $\mathcal{D}_3$. For these two resulting subsamples, we separately fit predictive models using only the observations in each subsample, denoted as $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$. Generally, for the j-th leaf node, we fit a specific predictive model $g_j(\cdot)$ according to Eq. (4), 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 predictors, respectively.

Split criterion. Since our goal is to separate returns into high- and low-predictability groups, we need a specific split criterion to implement this. For each candidate par-

Figure 1: **Illustration for the First Split**

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



tition, the sample is separated into two subsamples (left and right leaves). We fit a cluster-wise predictive model based on the setup in Eq. (4) and calculate the R^2 values for each cluster by Eq. (5). We score a candidate split by the absolute difference in predictability between these two child leaves as:

$$S_{\{\text{leaf}_l,\text{leaf}_r\}}\left(\text{var}_p, c_k\right) = \left|R^2_{\text{leaf}_l} - R^2_{\text{leaf}_r}\right|. \tag{6}$$

This criterion rewards splits that create a large separation in predictability, regardless of which side is more predictable. A larger value of the criterion indicates that the split successfully separates stock-return observations into a highly predictable group and a less predictable one based on the predictability level.

Reviewing the first split in Figure 1, we evaluate the criterion values by Eq. (6) 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. (6), completing the first split of the tree.

2.4 Subsequent Splits and Stopping Criteria

We use a sequential, greedy procedure to partition the panel into multiple clusters. Following the initial partition, we recursively split leaf nodes to refine the "mosaics" of the return panel.

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 heterogeneity in 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. Our goal-oriented approach endogenously partitions the panel into clusters that optimize for predictability heterogeneity before fitting local models.

Stopping criteria. To preserve interpretability and regularize training, we terminate tree growth based on minimum leaf size, maximum depth, and R^2 improvement thresholds. First, we impose a minimum sample size requirement at each leaf node, excluding split candidates whose resulting subsamples do not meet this threshold. This ensures each cluster-wise predictive model is estimated with sufficient observations. Second, we restrict the tree's maximum depth and terminal leaf count to control model complexity. Finally, we do not split a node if no candidate split increases predictability relative to the parent node. Specifically, we require at least one child cluster to have

R^2 exceeding that of its parent.

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.[7] The outputs of our methodology are: (i) a hierarchical tree structure showing the clustering pattern based on split variables and cut-point values; (ii) cluster-specific ML models, providing a heterogeneous predictive framework rather than a single global model; and (iii) the R_j^2 for each terminal cluster, quantifying within-cluster predictability, validated via out-of-sample forecasts.

The resulting P-Tree structure provides a parsimonious and interpretable mosaic of predictability. We now proceed to detail the data and the empirical evaluation procedures used to validate the economic significance of these clusters.

3 Data and Evaluations

3.1 Data and Variables

Data sample. We apply our approach to the U.S. equity market to measure and analyze the heterogeneous predictability of individual stock returns. Our monthly sample spans from 1973 to 2022.[8] We implement 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. To account for cross-sectional heterogeneity, we split the data into the first 30 years for model training and the most recent 20 years for OOS analysis. 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. On the other hand, we use the entire 50-year sample for the time-series analyses. Our algorithm accommodates

[7] Notably, while we choose a relatively reasonable linear predictive model to speed up the calculations, our framework is not limited to any particular machine learning algorithms.

[8] 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.

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.

We demean each market predictor using its 10-year rolling average and apply a 12-month weighted moving average to reduce high-frequency noise. Therefore, a positive value of a standardized indicator signals that its current level exceeds its historical 10-year average, indicating a high-regime state of the economy. This transformation makes regime definitions depend on deviations from recent historical norms.

Characteristics. As detailed in Table A.2, our dataset comprises 51 firm-level characteristics categorized into eight major categories: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. Each characteristic is standardized cross-sectionally and uniformly scaled to the range [0, 1] for each month. To mimic the "top-middle-bottom" sorting approach, we use 0.3 and 0.7 as candidate split points 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. To accommodate gradual changes in the predictability environment while maintaining parameter stability, we implement a rolling-window re-estimation scheme for both the tree-based clustering framework and the cluster-wise predictive models to analyze cross-sectional heterogeneity. Specifically, the P-Tree structure and associated predictive models are estimated using a 30-year rolling window of in-sample data and updated every five years. We repeat this re-estimation procedure four times over the 20-year out-of-sample period.

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. Fur-

thermore, 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 in-sample R^2 to measure predictability. In addition to the in-sample R^2, we evaluate the R^2_{OOS} to check whether the predictability gap is consistent. This approach aligns with the standard practice in recent studies. Following Gu et al. (2020), we define the $R^2_{OOS,j}$ using naive zero-return forecast as the benchmark,

$$R^2_{OOS,j} = 1 - \frac{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t+1} - \widehat{r}_{i,t+1})^2}{\sum_{\{i,t\}\in \text{leaf}_j} (r_{i,t+1} - 0)^2}, \tag{7}$$

where subscript j represents the specific OOS predictions by the related in-sample 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 higher predictability? We begin with cross-sectional partitions, referred to as "CS clusters" in Section 4.1, and then extend the analyses 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, along with further cross-sectional partitions.

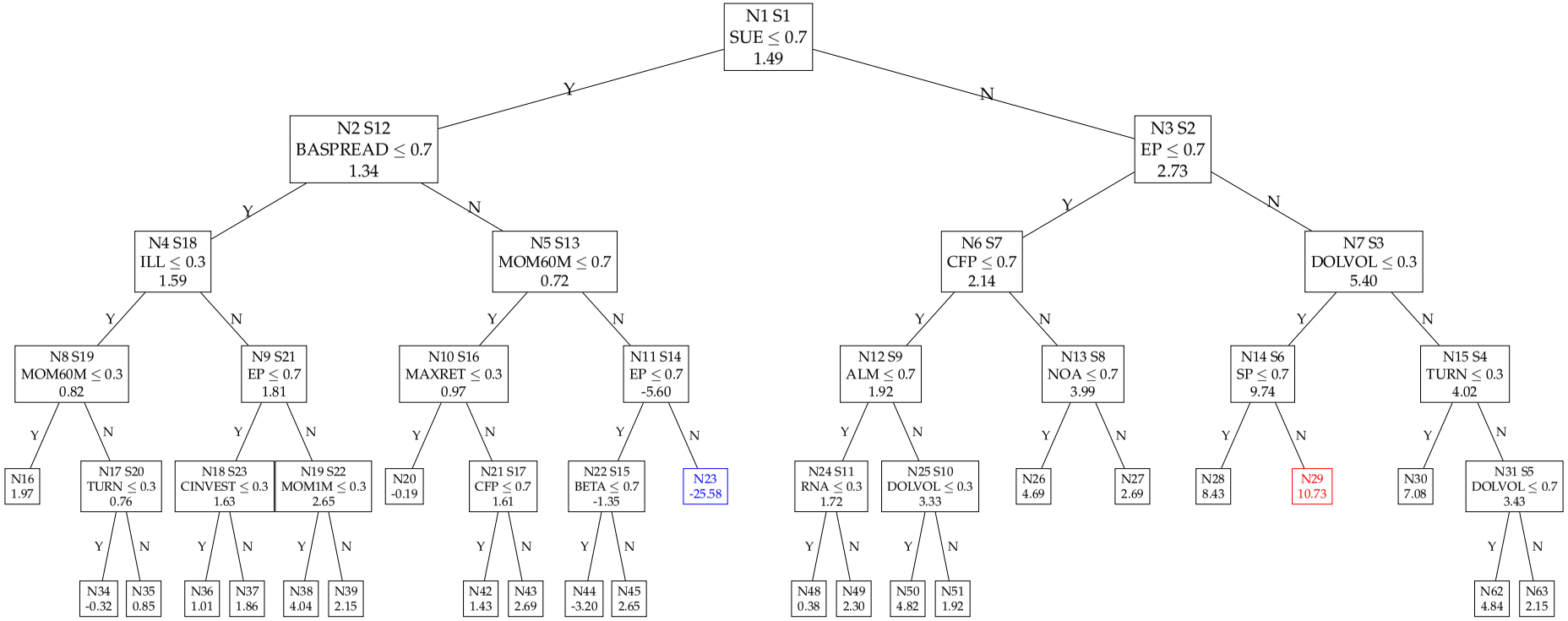
4.1 Cross-Sectional Heterogeneity

Figure 2 reports the cross-sectional tree estimated on monthly data from 1973 to 2002. Per our predetermined stopping criteria, this tree stops growing once it reaches 24 terminal leaves.[9] 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 the algorithm's iterations,

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

Figure 2: **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.



which directs stock-return observations that satisfy the condition to the left child node and those that do not to the right child node. Terminal leaves, which undergo no further splits, are denoted only by their leaf node index in the first row. Figure 2 shows a clear hierarchy in predictability across clusters. This structure provides 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%, obtained by fitting a Ridge model to predict returns across the dataset without clustering. Following the first split, the R^2 improves markedly for stocks with high unexpected earnings (SUE $>$ 0.7, monthly top 30% by SUE), rising to 2.73% at N3, while declining slightly to 1.34% for the complement set at N2. The R^2 difference ($2.73\% - 1.34\% = 1.39\%$) represents the maximum among all split candidates (51×2) identified by the algorithm.

The second split (S2) identifies high earnings-to-price (EP $>$ 0.7) stocks, yielding a higher R^2 of 5.40% at N7 among high-SUE stocks. These high-SUE, high-EP stocks exhibit stronger return predictability, while others show weaker predictability ($R^2 =$ 2.14% at N6). The maximum R^2 difference ($5.40\% - 2.14\%$) arises from an exhaustive search between cluster N2 (SUE $\leq$ 0.7) and leaf N3 (SUE $>$ 0.7).[10] The bottom-right panel highlights the interaction of these two characteristics (SUE and EP), with return predictability peaking in the top-right corner.

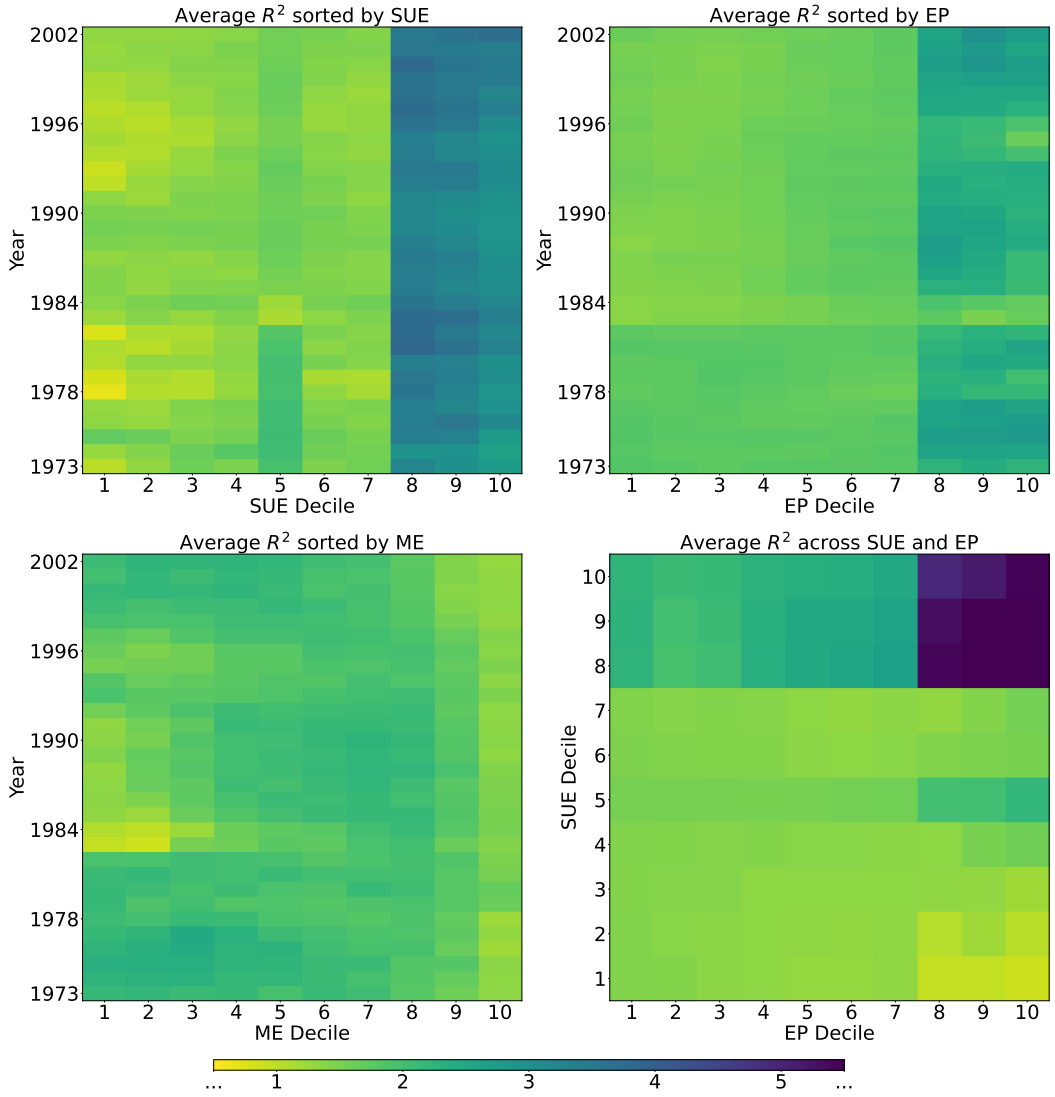
Mosaic pattern visualization. Figure 3 provides a decile-based visualization of the mosaic patterns in Figure 2. Figure 3 also reports average decile R^2 values by year and decile clusters, sorted by the top two characteristics (SUE and EP). A clear pattern emerges: stocks with higher earnings surprises or higher earnings-price ratios exhibit stronger return predictability, a pattern that has persisted over decades. The bottom-left panel shows that predictability differences across size-sorted clusters are modest, with small-cap stocks exhibiting slightly higher predictability.

These stocks belong to the long legs of the anomalies for earnings surprises (post-earnings announcement drift, Bernard and Thomas, 1989) and earnings (Reinganum,

[10]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.

Figure 3: **Mosaics of Predictability by Deciles**

These heat maps summarize return predictability (R^2 values, % in the color bar) for results presented in Figure 2. The first three heat maps display average R^2 values for groups sorted by years and deciles of earnings surprises, earnings-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.



1981), both showing persistent abnormal returns. These patterns suggest that predictability is largely driven by interactions among predictors rather than by single predictors. Economically, this pattern is driven by information-processing delays and limited arbitrage, particularly in stocks where fundamental signals are slow to incorporate into prices due to frictions (e.g., Ball and Brown, 1968; Engelberg et al., 2018). These results suggest that pooling across the full panel can mask predictability that is concentrated in characteristic-defined states.

Table 1: **Performance of Asset Clusters**

This table summarizes key metrics for each cluster from the cross-sectional tree structure in Figure 2. 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	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

Heterogeneity magnitude. 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. (4) using all stock-return observations jointly, without accounting for clusters.

As summarized in Table 1, the predictability measure R^2_C estimated from the heterogeneous model varies substantially across clusters. Importantly, this variation arises from the interaction of firm characteristics rather than individual effects alone. For example, cluster N3, characterized by high earnings surprises (SUE > 0.7), achieves an R^2 of 2.73%. However, when high SUE combines with high earnings-to-price (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. The characteristics driving predictability (high SUE, high EP, low volume) correspond to stocks exhibiting persistent return behavior, specifically long legs of well-documented, predictable patterns. These returns exhibit higher predictability because information-processing delays slow transmission of fundamental signals into prices, while limited arbitrage prevents rapid corrections.

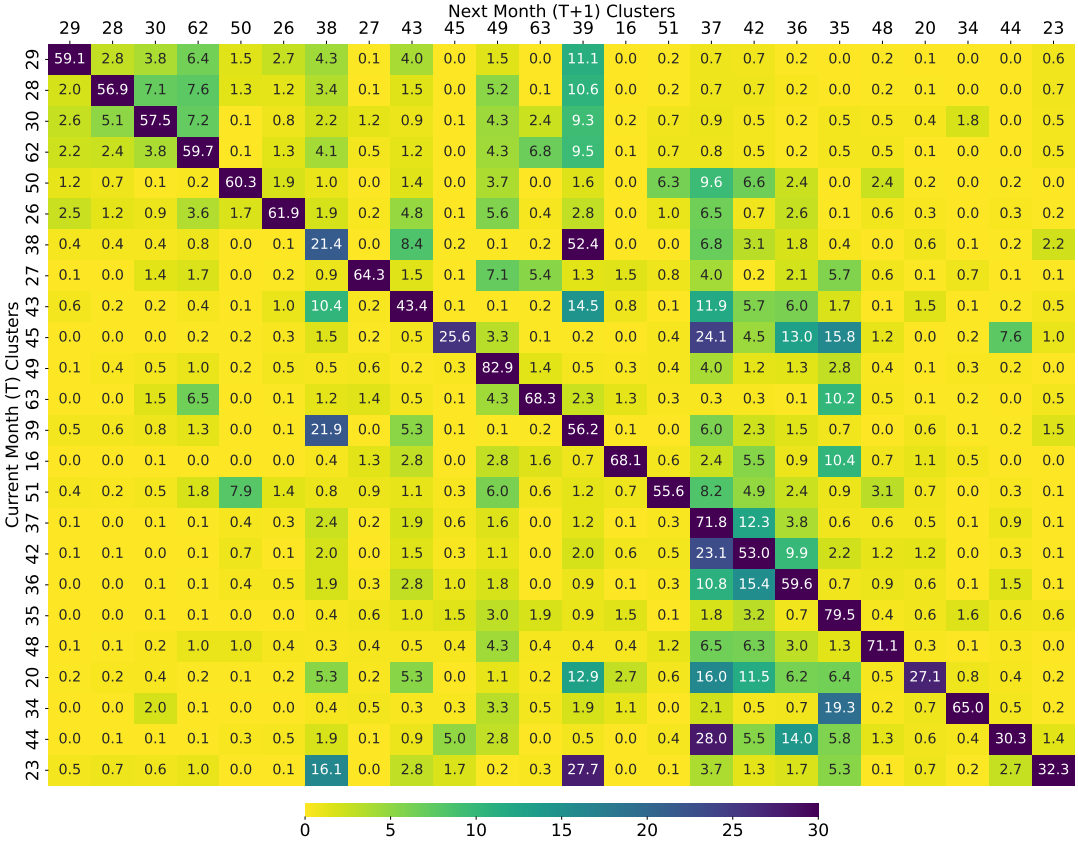
When we apply the global model, which estimates on the full dataset without accounting for clusters, to each cluster and compute its within-cluster R^2, denoted R^2_{G}, we also observe considerable variation across clusters. This highlights the cross-sectional differences in return predictability that 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 the 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_{C} below the root node, 1.49%) represent around 35% of observations, while the top three most predictable clusters (R^2_{C} 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., monthly top 30% SUE), the tree prioritizes splits that maximize predictability differences. For example, the tree uncovers that low volume only boosts predictability when paired with high SUE and EP, but not in isolation.

Figure 4: **Persistence of Clusters**

This figure presents the monthly average transition matrix of clustering results shown in Figure 2. 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.



State persistence. Because the clusters are determined by cross-sectional standardized firm characteristics, a stock may transition between clusters over time as its characteristics evolve. Figure 4 reports the monthly average transition matrix. High diagonal values confirm stable cluster membership, while the remaining off-diagonal entries capture economically interpretable transitions, such as momentum states (N38 and N39) or bid-ask-spread and earnings-price states (N44 and N37) in Figure 2. This evidence clarifies the "pockets of predictability" interpretation: firms move across states as characteristics change, but the characteristic-defined states remain persistent sources of alpha, suggesting that predictability is a stable feature of the economic regime rather than a fleeting anomaly of specific assets.

Economic links. 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, $R^2_{\mathrm{C}} = 10.73\%$) delivers an average return of 3.41% and an annualized Sharpe ratio of 1.88 under value weighting, highlighting strong profitability and a favorable risk-return trade-off. In contrast, the least predictable cluster (N23, $R^2_{\mathrm{C}} =$ -25.58%) exhibits significantly weaker performance, with an average return of -0.06% and a Sharpe ratio of -0.02. These findings establish a strong connection between return predictability and investment outcomes, complementing Rapach et al. (2010) and Kelly et al. (2024) by exploiting cross-sectional grouped heterogeneity in return predictability. They also suggest the potential to identify predictability-related strategies (see Section 5).

4.2 Time-Series Analyses and Macroeconomic Regimes

We further investigate whether the cross-sectional heterogeneity documented above varies systematically with time-series conditions. The aim is not to deliver a sharp identification between behavioral and rational channels, but to analyze the economic state dependence of the model's performance. We therefore extend the set of splitting variables beyond firm characteristics to include aggregate market states. Calendar months are considered additionally in the Appendix Section A. 4.5, where they provide a complementary way to capture the time-varying heterogeneity in return predictability based on structural breaks.

Macro-driven regime changes. Similar to the regime detection in Feng et al. (2024) and Bie et al. (2024), we use aggregate predictors to segment time-series horizons and identify market regimes, thereby capturing heterogeneous return predictability. This approach offers intuitive and economically meaningful interpretations of these regimes. Furthermore, these regimes exhibit recurrent patterns aligned with business cycle dynamics. We restrict the first two splits by market predictors and employ firm characteristics for subsequent cross-sectional partitions.[11]

[11] 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 suggests a strong preference for initial separation, as indicated by time-series data. To limit excessive regime changes, we impose these restrictions.

Figure 5: **Tree-Based Clusters under Market Regimes**

This figure illustrates tree-based clusters of stock returns across market regimes, using monthly data from 1973 to 2022. The tree structure is constructed by two standardized and smoothed aggregate predictors (market dividend yield and liquidity), thereby dividing the sample into three regimes. Each node, including intermediate and terminal nodes, is assigned a unique ID (N#), and the split order is indicated by (S#). Cluster-specific R^2 values are reported for all nodes.

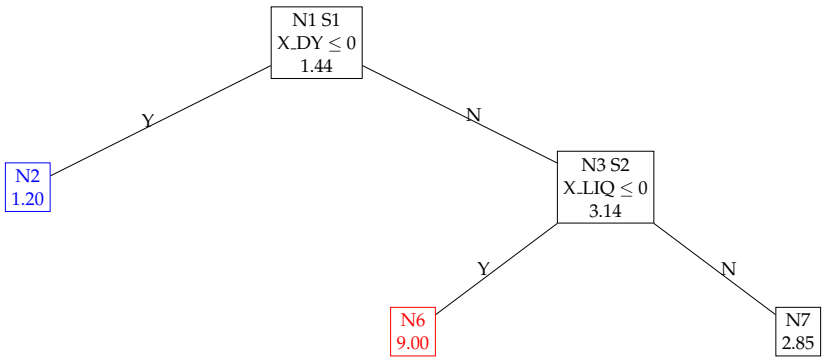


Figure 5 displays the upper portion of the tree, which captures the time-series splits, while Figure A.3 illustrates subsequent partitions by firm characteristics within each regime. We find three regimes and 45 terminal clusters in total.[12] The in-sample R^2 of the global model across the sample is 1.44% without 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:

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

Interactions between market dividend yield and liquidity endogenously define three discontinuous regimes closely aligned with underlying business cycles.[13]

[12] 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.

[13] 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.

Table 2: **Cluster-Specific Performance under Market Regimes**

This table summarizes the cluster-based information for each regime across three panels, corresponding to the time-series and cross-sectional tree structures in Figure 5 and A.3. Each panel displays the number of observations ("# obs") and the return predictability level (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 in subscripts) portfolios based on all observations within every cluster. 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

Additionally, further splits by firm characteristics reveal significant cross-sectional heterogeneity within macro-driven regimes. Table 2 translates the tree structures in Figure 5 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. For example, leaf N16 (red, Regime II in Figure A.3) exhibits the highest R^2 of 35.51%, formed by stocks with low ILL, low STD_DOLVOL, low BETA, and low DEPR. However, the weakest cluster in Regime II, leaf N25 (blue), characterized by high ILL, low GMA, low ME, and high CASHDEBT stocks, exhibits the poorest performance at -3.97%. Similar gaps can be seen in Regime I (11.00% versus -12.79%) and Regime III (12.39% versus -0.49%). 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.

Similar to the cross-sectional analysis in Table 1, modest positive correlations emerge between predictability and returns in Regimes I and III. In contrast, Regime

II, the most predictable period, 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 also conduct evaluations to assess improvements in heterogeneous return predictability enabled by regime-dependent leaf-specific predictive models. Clusters are aggregated into subsamples, as shown in Table 3. Under all regimes, heterogeneous models (Panel B) exhibit greater differences in return predictability than 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 these three regimes, Regime II consistently exhibits greater predictability than the other two, a finding that is robust across all three predictive models (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.

Furthermore, we also design Figure 6 to present the regime-dependent time variation from an alternative perspective. The color transitions denote the market regime switches identified by the tree-based clusters above. 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

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

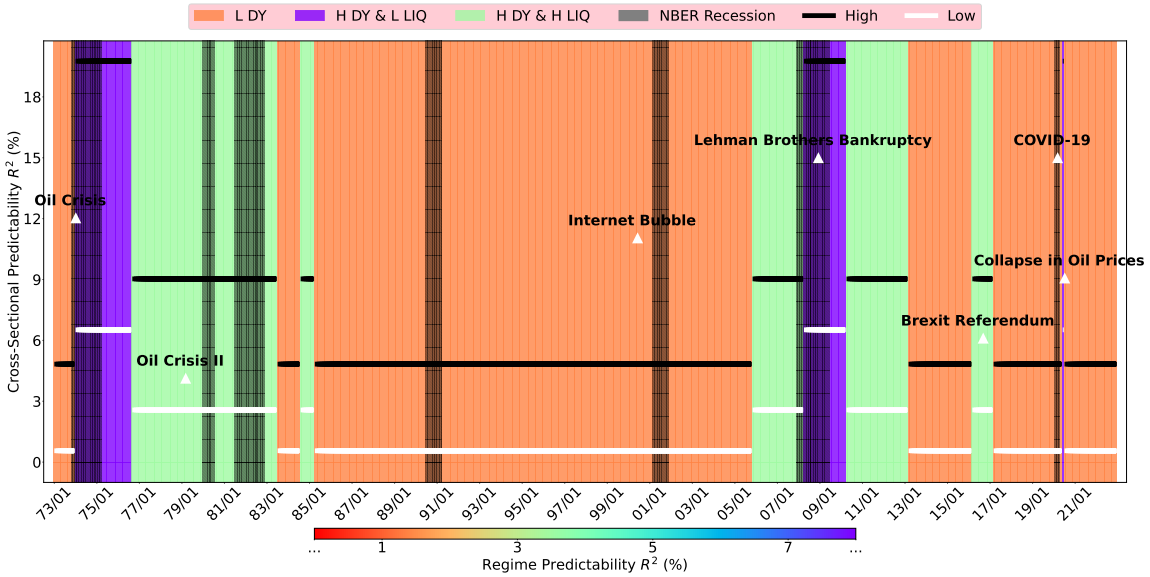
This table reports return predictability levels (R^2 values in %) across models under market regimes, using all (Sample A) and large-cap (Sample B) stocks. For each regime, two panels are presented: global and cluster-wise forecasts. Results include five samples: Global (homogeneous forecasts), Aggregate (cluster-wise predictions), and High, Medium, and Low, based on predictive rankings within tree clusters. Symbol "-" 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

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 2019 COVID-19 pandemic. These findings shed light on

Figure 6: **Time-Varying Predictability across Market Regimes**

This figure illustrates market regime shifts corresponding to aggregate 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 levels of predictability, as indicated by the color bar. Shaded regions represent NBER recessions, while major global events are labeled. The vertical axis measures the further cross-sectional predictability of clusters within each regime, ranging from highly predictable (black) to less predictable (white) groups, relative to the regime predictability (color bar).



the mosaics of return predictability across characteristics and market conditions.

Economic interpretation. The regime evidence is informative because it suggests that predictability is not merely a cross-sectional feature of certain stocks, but also a state-dependent feature of the environment in which those stocks trade. In particular, predictability is strongest when dividend yields are high and liquidity is low, that is, when discount-rate variation should be more pronounced and arbitrage capital more constrained. This is the setting in which a given cross section of characteristics is most likely to generate economically meaningful differences in the signal-to-noise ratio of expected returns. This pattern is consistent with the present-value intuition of Campbell and Shiller (1988), under which recessionary price declines raise dividend yields and sharpen forward-looking return signals. A formal derivation is reported in the Appendix Section A. 4.4.

5 Out-of-Sample Facts and Portfolio Implications

To validate the persistence and economic relevance of heterogeneous predictability, we conduct out-of-sample (OOS) tests and then evaluate the main portfolio implications of the clustered forecasts.[14]

5.1 Grouped Heterogeneity Out of Sample

We have identified easy- and hard-to-predict clusters using in-sample information. This naturally raises an important question: Do these mosaic patterns persist out of sample? We do not focus on comparing forecasting models; instead, we test whether the predictability ranking implied by the tree persists out of sample. To assess OOS performance, we implement a five-year rolling clustering approach with two-fold cross-validation for hyperparameter tuning. This rigorous OOS testing mitigates concerns about in-sample artifacts and ensures replicability.[15]

Similarly, we classify leaves into three distinct levels by return predictability: high, medium, and low, using the tree structure and empirical cluster-specific performance in Table 1. The top six clusters form the high-predictability group, while the bottom five form the low-predictability group. The remaining leaves collectively constitute the medium level.[16] We use R^2_{OOS} values, as defined in Eq. (7), to validate the cross-sectional patterns of stock return predictability.

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

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

[15]The in-sample data is split into two contiguous periods. The model is trained on one period and validated on the other. Optimal hyperparameters are determined based on the quadratic loss, and the model is then retrained on the full in-sample dataset. The final coefficients are then used to forecast OOS values for the next 5 years.

[16]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 in-sample 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.

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 2. 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.

	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

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 an out-of-sample R^2 of 2.06% for highly predictable clusters, while the cluster-wise model achieves 3.84% in Panel B. Cluster-wise forecasts improve out-of-sample R^2 in the high-predictability subsample.

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 in-sample or OOS, OLS or Ridge regression, and global or cluster-wise model, shows a higher R^2 for the more predictable cluster. The high-predictability clusters remain easier to forecast out of sample than the low-predictability clusters. 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, one with high predictability and the other with low predictabil-

ity, based on the relative proportions of in-sample and OOS observations.[17] For each subsample, we re-estimate a homogeneous predictive model using only the in-sample 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 after removing highly predictable stocks from the analysis.

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 2. We split the entire sample into subsamples based on different partitions of sorted clusters and refit homogeneous predictive models locally. The "True %" column summarizes the accurate percentages of subsample observations relative to 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

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 high-predictability stocks. Some of the OOS R^2 values 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

[17] 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.

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 Forecast-Implied Strategies

Thus far, our analysis has primarily focused on uncovering heterogeneity in stock return predictability using clustering methods. However, beyond identifying clusters, our framework also produces cluster-specific predictive models, each estimated via cluster-wise Ridge regressions. This contrasts with global models (e.g., Gu et al., 2020), which impose a single predictive relation across all assets, thereby overlooking localized patterns of predictability.

We now turn to the main portfolio implication by evaluating trading portfolios constructed 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 (equal-weighted) or by market capitalization (value-weighted), 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.[18]

Specifically, let the portfolio weights $w_{i,t-1}$ 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:

[18] Guijarro-Ordonez et al. (2025) also normalize absolute portfolio weights to sum to one, thereby implicitly imposing a leverage constraint.

$$\text{Sign-adjusted Equal/Value-weighted: } \widehat{w}_{i,t-1} = \begin{cases} w_{i,t-1}, & \text{if } \widehat{r}_{i,t} \geq 0 \\ -w_{i,t-1}, & \text{if } \widehat{r}_{i,t} < 0 \end{cases} \tag{8}$$

$$\text{Forecast-weighted: } \widehat{w}_{i,t-1} = \frac{\widehat{r}_{i,t}}{\sum_i |\widehat{r}_{i,t}|}.$$

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

$$R_{j,t} = \sum_{\{i,t\} \in \text{leaf}_j} \widehat{w}_{i,t-1} r_{i,t}. \tag{9}$$

Forecast-implied strategies directly incorporate individual return forecasts into position signs and weights. The forecast-implied portfolios are explicitly active, long–short portfolios whose positions are determined by the signs and magnitudes of the underlying machine learning return forecasts. Greater predictive accuracy should mechanically translate into consistently superior realized portfolio performance.

Table 6 presents the results of the cross-sectional clustering model in Section 4.1. First, our empirical results confirm that highly predictable stocks remain consistently easier to predict out of sample, and as a result, the corresponding trading strategies generate significant investment gains. The sign-adjusted value-weighted portfolio also performs well, indicating that the results are not driven solely by small caps. For example, in Panel A, the highly predictable subsample shows an OOS Sharpe ratio of 0.84 and a significant alpha of 0.46%. In contrast, investing in the least predictable stocks achieves only about half of these values, with an OOS Sharpe ratio of 0.51 and a statistically insignificant negative alpha.

Panels B and C of Table 6 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 2.31%, 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 them. For example, in Panel C, the OOS alpha for the aggregate model is 0.79%, higher

Table 6: **Forecast-Implied Investment Performance**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 9). 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 8). We present seven samples: Global (no clustering), Aggregate (cluster aggregation), High/Medium/Low (predictive rankings within clusters), and Global-Large/Aggregate-Large (large-cap stocks in 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

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.

Both aggregate and global models show similar accuracy in predicting the direc-

tion of returns (Panels A and B), but the aggregate model excels in capturing return *magnitudes* (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.8 in the Appendix) slightly reduces absolute performance levels under conservative trading-friction assumptions throughout, but leaves 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.[19]

Overall, across all specifications, stocks robustly identified by our tree-based clustering as highly predictable out of sample deliver markedly superior investment performance relative to others. Excluding these stocks significantly reduces potential portfolio gains, underscoring their outsized contribution to overall profitability. Extending the mosaic evidence of return predictability to practical investment applications, investors can systematically enhance economic gains by allocating capital toward stocks with high earnings surprises, strong valuation ratios, or low trading volume (see Figure 2).

Consistent with our central argument, the cluster-wise predictive model persistently outperforms the standard global homogeneous benchmark, highlighting the inherent limitations of naive homogeneous approaches. More broadly, this result helps position the paper relative to the recent machine-learning literature. Much of that work asks whether a richer global model can, on average, improve forecasts. Our evidence points to a different margin: even within a common predictive architecture, substantial gains arise from allocating model flexibility unevenly across the panel. The economic value of heterogeneity, therefore, comes less from inventing new signals than from using existing signals where they are most informative.

A related asset-pricing implication concerns the gap between local and global model fit. If grouped heterogeneity captures information omitted by homogeneous

[19]Section IA.2 of the Internet Appendix shows that investment gains remain robust—and even strengthen—when allowing for regime-dependent time variation and large-cap constraints.

benchmarks, assets with larger local-global fit gaps should be more exposed to model misspecification risk. Appendix Section A. 4.6 develops this idea and shows that portfolios sorted on this predictability differential earn large and persistent alphas.

6 Conclusion

We show that return predictability is a latent, asset-specific, and state-dependent characteristic rather than a uniform feature of the return panel. We interpret our models and findings as evidence of heterogeneity in the predictability of conditional expected returns. In practice, a single pooled/universal forecasting model can mask where signals appear consistently and where predictability is hard to distinguish from noise, motivating our endogenous "mosaics" that separate high- and low-predictability segments.

Methodologically, P-Tree departs from pooled approaches by producing an interpretable partition that captures heterogeneity in predictability. Unlike fixed-cluster methods, the framework allows assets to move across groups as characteristics and regimes change, helping track how the linkages among characteristics, market conditions, and market efficiency evolve over time.

Empirically, we find substantial heterogeneity driven by interactions between firm characteristics and macroeconomic conditions; these differences persist out of sample and deliver investment gains, including monthly alphas above 1%. The mosaics yield granular insights into market efficiency, highlighting information delays and limits to arbitrage that sustain pockets of predictability in the cross section. Ignoring grouped heterogeneity can therefore lead to misspecification. Return predictability strengthens in adverse macro-financial environments, when distress and liquidity risks are elevated, providing a benchmark for macro-finance interpretations of state-dependent signal strength. We conclude that integrating heterogeneous predictability is critical for advancing asset pricing theory and enhancing portfolio construction.

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 X. Zhu (2025). How and when are high-frequency stock returns predictable? *Management Science, Forthcoming*.

Avramov, D. (2002). Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D., S. Cheng, and L. Metzker (2023). Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability. *Management Science 69*(5), 2587–2619.

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

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.

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 (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.

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

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 Virginia.

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 (2025). 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.

Jensen, M. C., F. Black, and M. S. Scholes (1972). The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Market*.

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

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.

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? Technical report, University of Chicago, Becker Friedman Institute for Economics Working Paper.

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

A. 1 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)

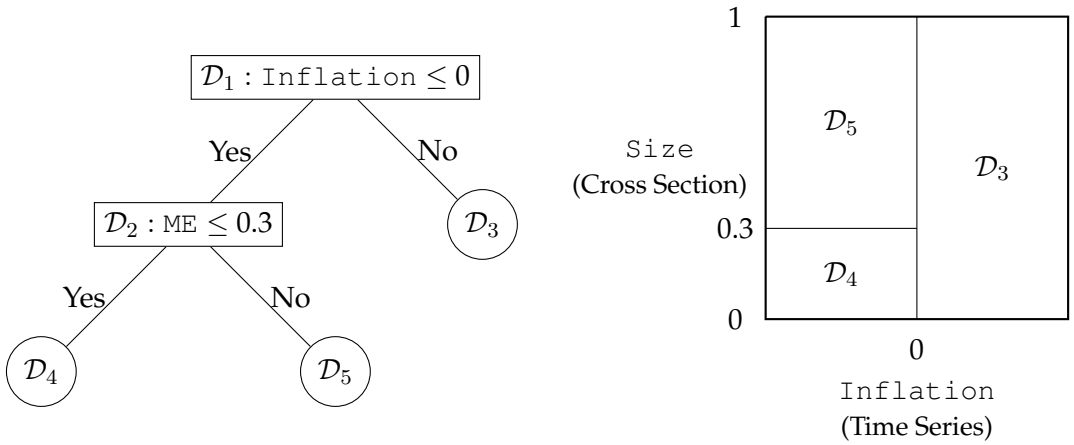
No.	Acronym	Description	Category
22	ill	Illiquidity Rolling 3M	Liquidity
23	lev	Leverage	Value
24	lgr	Growth in long-term debt	Investment
25	maxret	Maximum daily returns Rolling 3M	Volatility
26	me	Market equity	Size
27	me_ia	Industry-adjusted size	Size
28	mom12m	Momentum rolling 12m	Momentum
29	mom1m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

A. 2 Illustration for the Second Split

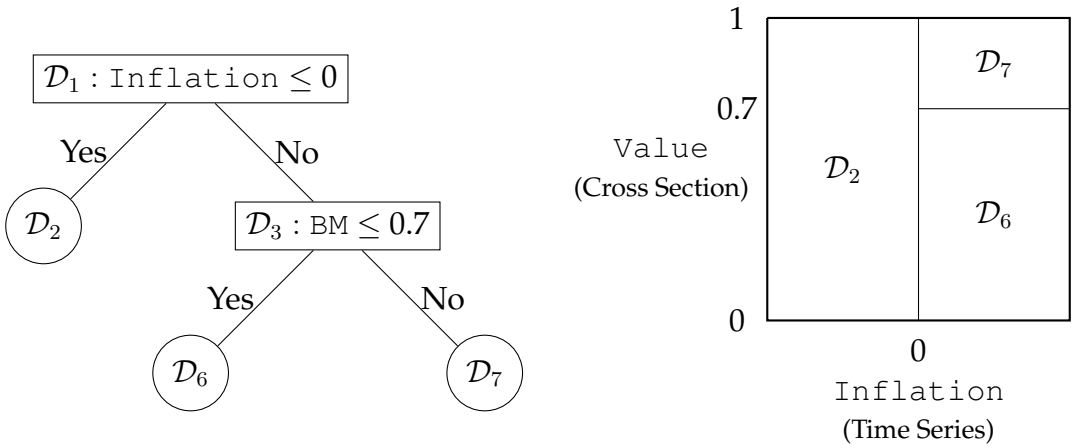
The initial split is determined based on demeaned inflation relative to the threshold 0 (see Figure 1). Subsequently, we iteratively explore the optimal choice for the second split using the two candidates in Figure A.1 below. This split can occur either on the left based on size (ME ≤ 0.3), or on the right based on book-to-market (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, as detailed in Section 2.4.

Figure A.1: **Illustration for the Second Split**

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



(a) If splitting node $\mathcal{D}_2$ at ME ≤ 0.3



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

A. 3 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 (4).
            Calculate the cluster-based R_j^2 by (5).
          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 1).

A. 4 Additional Empirical Results

A. 4.1 Additional Cross-Sectional Visualizations

Figure A.2: **Mosaics of Predictability by Months**

This heat map summarizes monthly return predictability (R^2 values, % in the color bar) for results presented in Figure 2. The vertical axis shows the proportion of observations per cluster, and the horizontal axis shows months. Colors transition from light (bottom) to dark (top) each month, indicating an increase in return predictability.

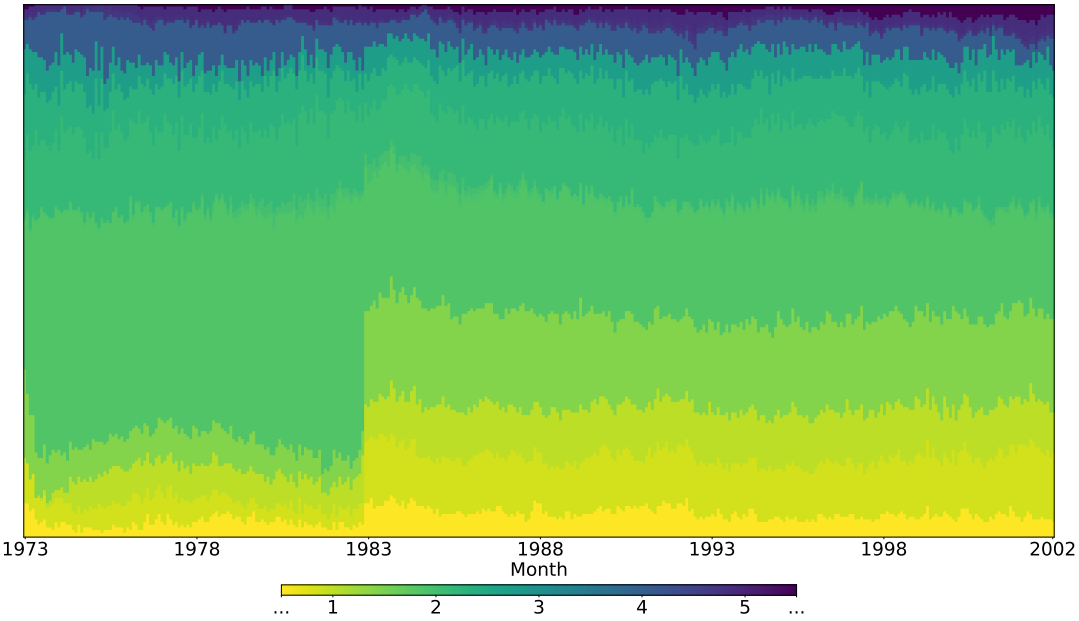
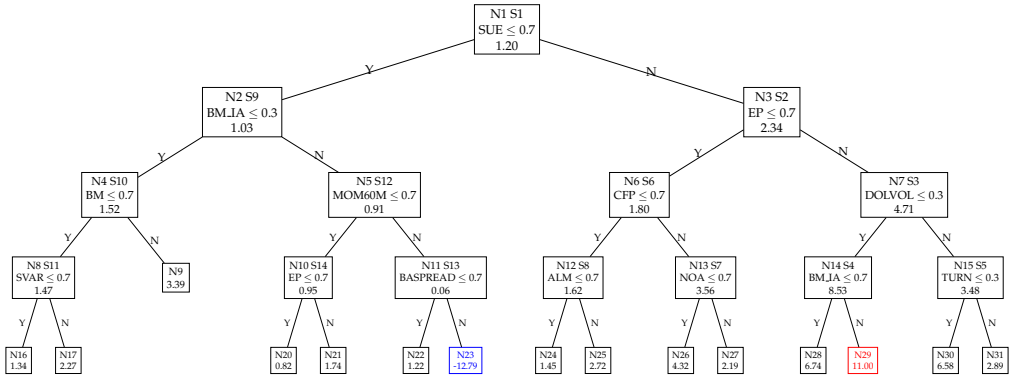


Figure A.2 illustrates a month-by-month mosaic of the cross-sectional clusters. The sample spans 360 months, aligned with the tree structure in Figure 2, and includes approximately 4,000 to 5,500 monthly observations. Because all clusters are determined by cross-sectional standardized firm characteristics, each firm's cluster membership can evolve over time even when the predictability ranking remains stable.

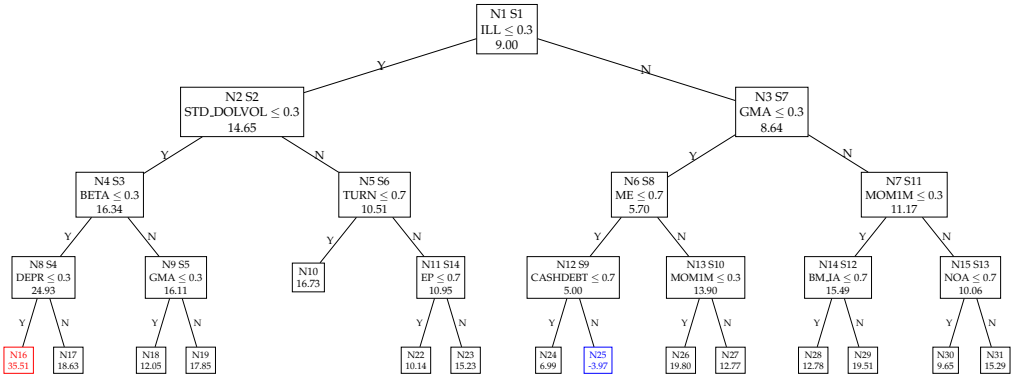
A. 4.2 Further Cross-Sectional Tree Structures across Market Regimes

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

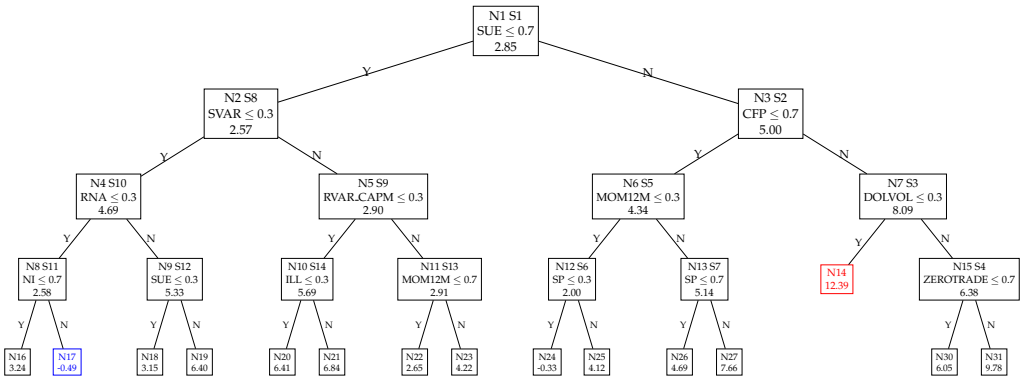
This figure illustrates the further cross-sectional split tree-based clustering structure under each regime, as determined by Figure 5, using monthly data from 1973 to 2022. Each tree splits stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. The terminal 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: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \leq 0\}$

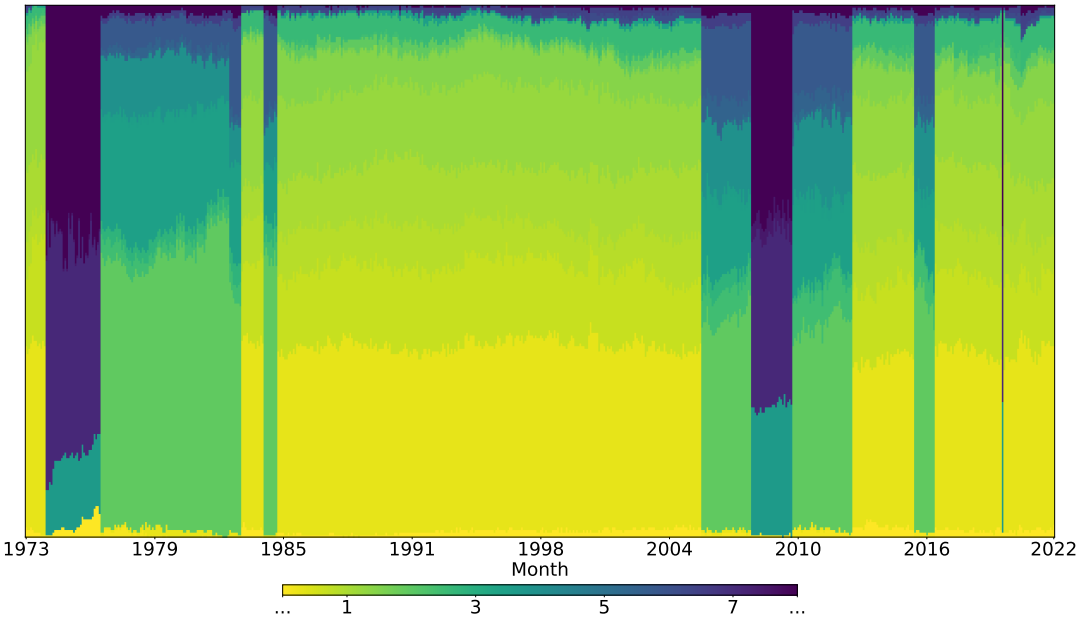


(c) Regime III: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} > 0\}$

A. 4.3 Additional Visualizations of Regime-Dependent Predictability

Figure A.4: **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 5 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 A.2, the patterns of stock return predictability across months (Figure A.4) 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, characterized by high market dividend yields and constrained liquidity, highlight the countercyclical nature of return predictability during recessionary periods. 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.

A. 4.4 Present-Value Intuition for Regime Dependence

Our analysis identifies the market dividend yield as the key driver of time-varying return predictability, with higher yields associated with stronger predictability. Within the present-value framework of Campbell and Shiller (1988), we interpret this finding as reflecting systematic shifts in expected returns. Although their model focuses on aggregate returns rather than cross-sectional variation, it offers a theoretical basis for understanding how dividend yield fluctuations influence market-level predictability, particularly during recessions when price declines enhance forward-looking return signals. This evidence links time-varying predictability to core asset-pricing dynamics driven by economic conditions.

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, conse-

quently, a higher R^2 in linear predictive regressions, reflecting increased predictability of market returns 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.

A. 4.5 Structural Breaks in Calendar Month

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 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 A.5: **Time-Varying Predictability over Structural Breaks**

This figure depicts structural breaks by calendar month from 1973 to 2022. Colors represent regimes with varying predictability, as shown in the color bar. Shaded regions indicate NBER recessions, while annotations mark key global economic events. The vertical axis measures cross-sectional predictability, ranging from highest (black) to lowest (white) within each regime.

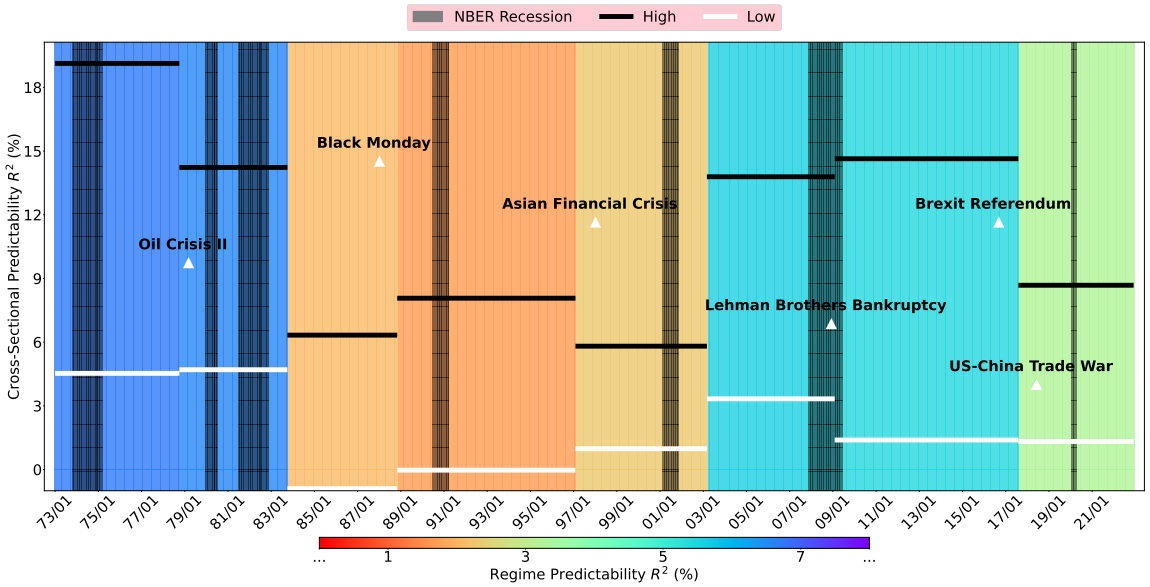


Figure A.5 illustrates fluctuations in predictability across these calendar-time structural breaks. Although the magnitude of variation is generally smaller than in regime-based models using market predictors, predictability peaks at 6.7% during January 1973 to October 1978 (the first purple segment) and declines to 1.67% between Novem-

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				

ber 1988 and February 1997 (the orange segment). Notably, predictability intensifies during recessions, especially prior to 1983, consistent with patterns in Figure 6.

Major events, such as Black Monday (1987) and the Brexit Referendum (2016), likely induced structural breaks in return predictability, which macro-driven regime

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	-	-	-

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.

A. 4.6 Predictability Differential 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 homogeneous 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 homogeneous model for most stocks with predictable returns. This finding suggests that conventional pooled models may be misleading about predictive accuracy, exposing average investors to the risk that cluster-specific relevant information is not efficiently incorporated into asset prices when all trades are based on homogeneous benchmarks (e.g., Gu et al., 2020).

Model misspecification occurs when homogeneous benchmarks fail to capture cluster-specific signals, leading to mispricing. 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 advantage of the heterogeneous model relative to the global one:[20]

$$R^2_{\text{CMG},j} = R^2_{\text{C},j} - R^2_{\text{G},j}. \tag{10}$$

Specifically, a higher R^2_{CMG} indicates that cluster-specific 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 heterogeneous model underperforms the global one, possibly reflecting weaker or

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

Table A.5: **Predictability Differential Portfolio**

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)

noisier information.[21]

To implement this strategy, 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 A.5 reports the comparative performance of four long-short predictability differential strategies, sorted by cluster-specific differentials.[22] Strategies targeting larger underlying predictability differentials (e.g., "T3 - B3") consistently

[21] As the global model incorporates broader cross-sectional data, R^2_{C} can be lower than R^2_{G}.
[22] We focus on the "T3 - B3" strategy, with additional clusters and robustness checks discussed later.

yield superior risk-adjusted returns in both in-sample and OOS evaluations. Even the strategy incorporating the largest number of clusters ("T7 - B7") delivers highly robust performance, with an average monthly return of 1.10% and an annualized Sharpe ratio of 1.44 during the entire OOS sample period.

Panel B of Table A.5 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 (FF5+MOM+IVOL), Q5 (Hou, Mo, Xue, and Zhang, 2021), BS6 (Barillas and Shanken, 2018), DHS3 (Daniel, Hirshleifer, and Sun, 2020), and SY4 (Stambaugh and Yuan, 2017).[23] All strategies exhibit statistically significant unexplained alphas in both in-sample and OOS settings. Additionally, the persistence of this premium is consistent with limited arbitrage. The strategy's alpha remains significant even after accounting for transaction costs (see Table A.7 in the Appendix). The premium is stable out-of-sample (e.g., the "T7 - B7" strategy can still exceed 0.9% per month), confirming that it is not a data-mined result.

To isolate the source of performance, we decompose the portfolio returns into contributions from the long and short legs. Appendix Table A.6 shows that the long-leg portfolios, corresponding to clusters with stronger predictability differentials, consistently generate most of the abnormal returns, whereas short-leg portfolios typically exhibit weak or insignificant OOS alphas. Hence, the documented profitability stems primarily from stocks with pronounced cluster-specific informational structures that global models fail to capture.

Overall, the results show economically meaningful return spreads associated with heterogeneity in return predictability. Portfolios constructed from the differential between cluster-wise and global models consistently produce sizable and persistent spreads, even after adjusting for established factor models. These results highlight that ignoring heterogeneity in return predictability constitutes a form of latent model misspecification risk, and that strategies recognizing heterogeneous predictive structures can deliver superior model-adjusted performance, contrasting sharply with the null find-

[23] Due to data availability limitations, SY4 results are reported for 2003–2016 only.

ings in Rapach et al. (2010) and Kelly et al. (2024).

Additionally, we provide a detailed breakdown of the long- and short-leg portfolios for the predictability differential strategy, as shown in Table A.5. Regardless of the number of highly predictable clusters included, the long-leg portfolios in Table A.6 consistently generate significant abnormal returns across both in-sample and OOS periods. In contrast, the short-leg portfolios generally yield insignificant alphas out of sample. This reinforces the conclusion that the strategy we identify is primarily driven by stocks with the highest model misspecification risk (long-leg), as highlighted in Section A. 4.6 above.

Table A.6: **Decomposition of 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 A.5. 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)

A. 4.7 Investment Performance with Transaction Costs

This subsection re-examines the forecast-implied strategy in Section 5 and the predictability-differential strategy in Section A. 4.6. 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.7: **Predictability Differential Portfolio with Transaction Costs**

This table presents the results after considering transaction costs, as shown in Table A.5. 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.7 above reports the performance of the predictability differential strategy after accounting for transaction costs. The results show that investment profitability remains robust: average returns and Sharpe ratios decline only modestly (all com-

binations exhibit an OOS Sharpe ratio greater than 1), and all factor-adjusted alphas remain economically and statistically significant across models (at least 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.

Similarly, Table A.8 presents performance metrics for transaction-cost-adjusted forecast-implied portfolios. First, the results show that stocks with high predictability generate weaker OOS investment gains than those reported in Table 6. For instance, in Panel A, the high-predictability subsample yields an OOS Sharpe ratio of 0.77 and a marginally significant alpha of 0.36%, while the low-predictability subsample yields 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. These results substantially outperform portfolios with medium- and low-predictability.

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 consistently outperform their global counterparts across all weighting schemes. However, the margin is narrower than the full sample, likely due to lower return predictability among large stocks.

Table A.8: **Forecast-Implied Investment Performance (CS Cluster, with Transaction Costs)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios after accounting for transaction costs (Equation 9). 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 8). 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

A. 5 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. (6) with Sharpe ratios. For each split, we construct forecast-weighted portfolios based on all observations within the left and right clusters following Eq. (8), and then compute the corresponding Sharpe ratios. This new split criterion can also yield a tree-based clustering structure similar to that shown in Figure 2. Although the selected characteristics may differ, we can still observe clear heterogeneity gaps in return predictability across clusters.

Table A.9: **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.

	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 Sections 4 and 5, we present two of the most important empirical results to test the mosaic patterns of return predictability. Table A.9 reports both in-sample 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 per-

formance across all predictive methods. In both panels, the high-predictability clusters yield the highest R^2 values, 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, as defined in economic terms, yields similar cross-sectional rankings and confirms the robustness of our main findings.

Table A.10: **Forecast-Implied Investment Performance (SR Partitions)**

This table reports in-sample and out-of-sample performance of forecast-implied portfolios (Equation 9). 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 8). 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

Furthermore, Table A.10 also examines the forecast-implied investment performance based on Sharpe ratios as the split criterion. The overall results closely resem-

ble those in Tables 6, showing that clusters identified as highly predictable continue to deliver superior economic performance in both in-sample 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, this evidence demonstrates 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 power into stronger, more persistent investment performance.

Internet Appendix for "Mosaics of Predictability"

IA.1 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.

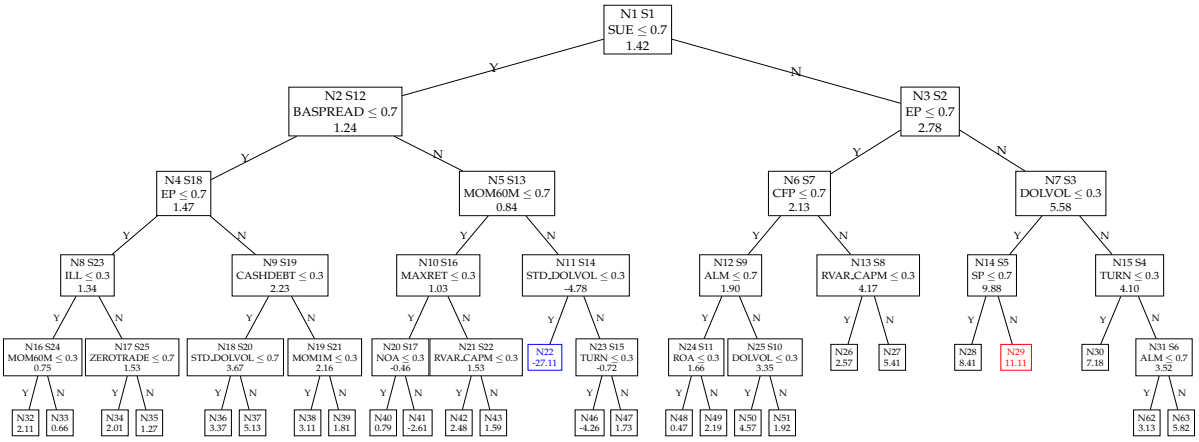
As detailed in Section 3.2, our cross-sectional analyses use a 30-year rolling window updated every five years. The optimal splitting variables show consistent robustness across subsamples. The first two splits in tree structures by different periods (Figure IA.1 for 1978–2007, 1983–2012, and 1988–2017) still choose earnings surprises and value characteristics (`EP` and `CFP`), as in Figure 2. 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 an in-sample R^2 close to or exceeding 10%, while the bottom cluster shows negative values. The relationship between predictability rankings and investment performance remains consistent, as shown in Section 4.

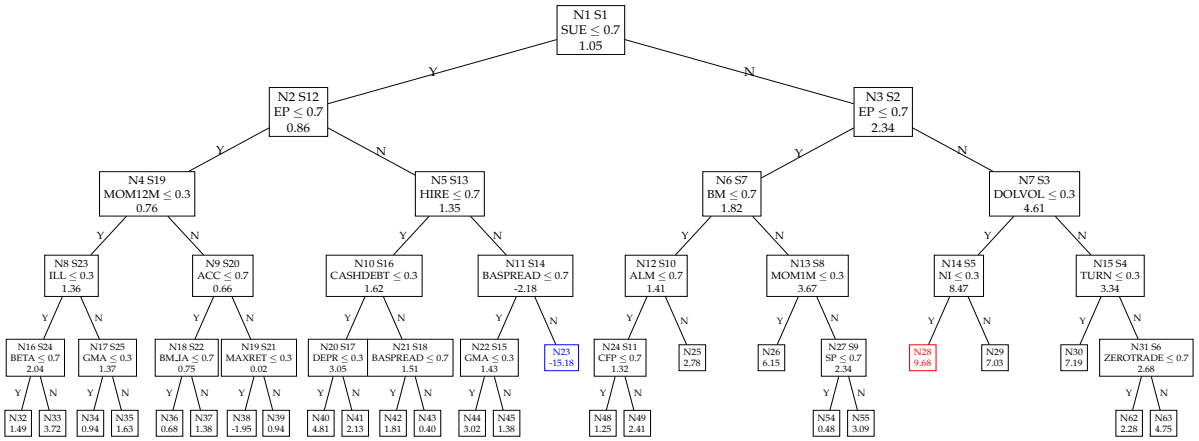
Additional sensitivity analyses, including variations in tree depth, minimum leaf observations, and winsorization levels, demonstrate consistent stability in selected firm characteristics. While splitting sequences differ slightly, the same variables reappear across specifications. These findings confirm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

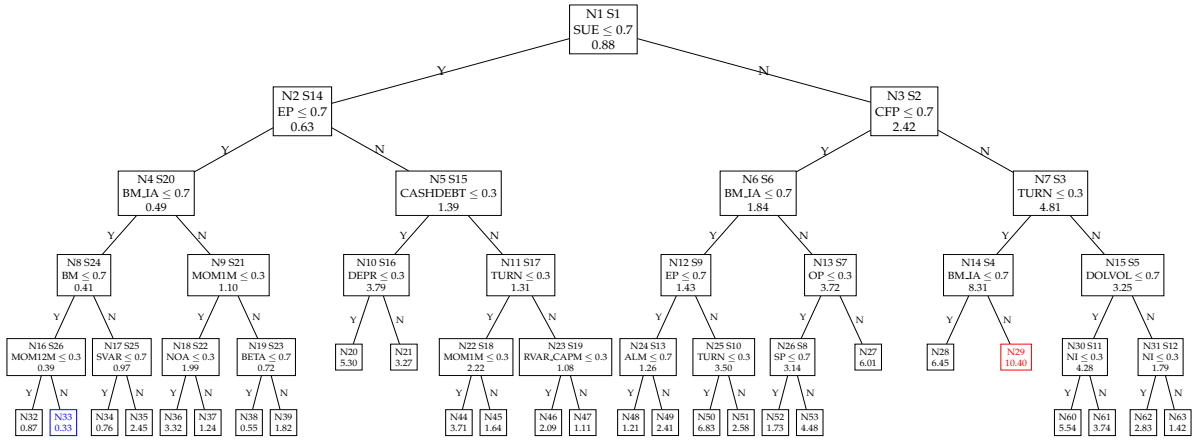
This figure illustrates the cross-sectional tree-based clustering derived from the other three monthly data sample periods (1978–2007, 1983–2012, 1988–2017). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the range of [0, 1]. Terminal leaves represent 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.2 Investment Gains within Time-Varying Modeling

This section examines the investment performance of forecast-implied portfolios by considering regime-dependent time-variation modeling. The empirical results are robust to transaction costs and are consistent with the cross-sectional analysis (Section 5 and Appendix A. 4.7): 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 9). 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 8). 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, with Transaction Costs)**

This table reports full-sample and large-cap subsample transaction-cost-adjusted long-short investment performance for forecast-implied portfolios (Equation 9). 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 8). 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