# Mosaics of Predictability

Lin William Cong Guanhao Feng Jingyu He Yuanzhi Wang*

First draft: Feb. 2024; this draft: March 20, 2025

**Abstract**

We postulate that return predictability is an unobservable, yet intrinsic and time-varying asset characteristic tied to expected returns, instead of merely an attribute of predictors or models. We introduce a novel tree-based clustering method to measure heterogeneous return predictability by grouping asset-return observations with similar predictability levels. The resulting clusters are described using high-dimensional asset characteristics and lagged macroeconomic variables. Our empirical findings reveal substantial heterogeneity in return predictability among U.S. stocks. Clusters with high earnings surprises, high earnings-to-price ratios, and low trading volumes exhibit the strongest predictability; predictability diminishes sharply when the aggregate dividend yield is low but peaks during periods of high aggregated dividend yield and low liquidity. Moreover, we document a new anomaly associated with traders' model misspecification risks: long-short portfolios based on variants of predictability-sorting generate unexplained monthly alphas exceeding 1%.

**Key Words:** Asset Pricing, GenAI, Goal-oriented Clustering, Grouped Heterogeneity, Machine Learning, Regime Switches, Trading Anomalies.

*We are grateful to Chunrong Ai, Doron Avramov, Daniele Bianchi, Jian Chen (Discussant), Qihui Chen, Lauren Cohen, Victor DeMiguel, Fuwei Jiang, Hao Ma, Semyon Malamud, Kuntara Pukthuanthong, David Rapach (Discussant), Dacheng Xiu, Lei Yang (Discussant), Yi Zhang, and the seminar and conference participants at 2025 AFA Annual Meeting, Cornell Tech, CUHK Shenzhen, 2024 EFMA Annual Meeting, 2024 FMA Asia/Pacific Conference, 2024 FinEML Conference, HKUST Guangzhou, Queen Mary University of London, 4th Big Data and Econometric Conference at Xiamen University, 2024 Annual SoFiE Meeting, 8th PKU-NUS Annual International Conference on Quantitative Finance and Economics, Summer Institute of Finance 2024, 2024 Academic Conference on Digital Economy Development and Governance, SIDF 2024, 17th ARMC 2024. Cong (E-mail: will.cong@cornell.edu) is at Cornell University SC Johnson College of Business (Johnson) and NBER; Feng (E-mail: gavin.feng@cityu.edu.hk), He (E-mail: jingyuhe@cityu.edu.hk), and Wang (E-mail: yuanzwang5-c@my.cityu.edu.hk) are at the City University of Hong Kong.

# 1 Introduction

Forecasting asset returns has long been a central topic in finance. Previous research has provided ample evidence of return predictability in various asset classes.[1] However, return predictability is typically treated as an attribute of the predictors or the models, being invariant and uniformly applicable to all assets analyzed in a given model.[2] However, return predictability not only is unobservable, but also can be time-varying and heterogeneous across individual assets or groups of assets.[3]

We recognize this challenge and develop a novel framework to address it. Specifically, we investigate what assets exhibit higher return predictability and how such predictability varies under different environments. We propose that, beyond predictors or models, return predictability reflects an unobserved yet intrinsic property of assets, potentially time-varying according to macroeconomic conditions and dependent on cross-sectional asset characteristics. We find return predictability to be extremely heterogeneous in the cross section (partitions based on asset characteristics as opposed to human-designated industry, country, or class groups) and over time (regimes delineated through lagged macroeconomic variables instead of calendar-time breaks), exhibiting rich data-driven "mosaic" patterns. Moreover, we show that heterogeneous predictability is closely linked to the cross-section of expected returns, likely because how traders incorporate it affects how information gets impounded into asset prices. We further explore new predictability-based anomalies and profitable strategies.

To our knowledge, this paper is the first to systematically investigate and measure

[1] For example, empirical studies have documented return predictability in the aggregate market (e.g., Welch and Goyal, 2008; Campbell and Thompson, 2008), individual stocks (e.g., Fama and French, 2008; Rapach et al., 2013; Lewellen, 2015), corporate bonds (e.g., Feng et al., 2025), treasury bonds (e.g., Bianchi et al., 2021), and mutual funds (e.g., Kaniel et al., 2023; DeMiguel et al., 2023).

[2] Researchers have identified aggregate or macroeconomic predictors for forecasting market returns, including dividend yield and default yield (e.g., Keim and Stambaugh, 1986; Fama and French, 1988); firm characteristics, such as size, value, and momentum, can predict cross-sectional variations in returns and be used for constructing factors or generating excess returns (e.g., Fama and French, 1992, 1993; Jegadeesh and Titman, 1993). Numerous methods have been developed to examine return predictability, including predictive regressions (e.g., Stambaugh, 1999), security sorting (e.g., Jensen et al., 1972), cross-sectional models (e.g., Fama and MacBeth, 1973; Han et al., 2024), and machine learning approaches (e.g., Kelly and Pruitt, 2013; Gu et al., 2020; Kelly et al., 2024).

[3] Many equity characteristics are similarly unobservable and require measurements, such as beta and volatility. Concepts related to predictability include anomaly average return, predictor significance, out-of-sample $R^2$, forecast-implied portfolio, etc.

the heterogeneous return predictability of individual assets. We propose the concept of *mosaics of predictability*, where return predictability varies in time and the cross-section, revealing mosaic patterns on the panel. As demonstrated by Cong et al. (2023), heterogeneous panel data analyses can be framed as a goal-oriented clustering problem through a strategy of "divide-and-conquer." The empirical solution, therefore, involves developing a similar clustering algorithm that separates highly predictable observations from less predictable ones in tandem with providing an effective group-level predictability measure.

To this end, we first propose measuring return predictability using the $R^2$ of the cluster-wise predictive model, assuming that observations within a cluster share similar predictability but vary across clusters. Note that because predictability is closely linked to the signal-to-noise (S2N) ratio or the regression $R^2$, we focus on estimating it as an inherent measure of return predictability. Traditional approaches either estimate a panel regression, producing a single $R^2$ homogeneously for all assets, or run time series regressions for each individual stock, which is often infeasible due to limited data. Instead, we adopt a balanced approach: clustering asset returns into multiple groups according to predictability, estimating a cluster-wise model $R^2$ to measure heterogeneous return predictability.

Given this intuitive measure, we next introduce a customized panel tree approach to identify the optimal clustering structure that captures variations in stock return predictability, grouping stocks with similar predictability levels. This partitioning is achieved by maximizing the difference in the S2N ratios, or $R^2$, across clusters using cluster-wise heterogeneous return forecasting models. The decision tree structure then organizes these clusters into mosaics on the panel, described by firm characteristics and aggregate predictors for easier visualization. Our clustering can be viewed as generalized multi-way security sorting in high-dimensional space. This approach preserves economic interpretability and highlights key variables associated with different levels of predictability. By leveraging the asymmetric interactions between aggregate predictors and firm characteristics, we measure heterogeneous return predictability

through cluster-wise predictive models.

Empirically, we test our framework using monthly returns of U.S. individual stocks from 1973 to 2022. The results reveal substantial heterogeneity in the cross-sectional predictability of stock returns. Our method identifies approximately 22 clusters (distinct leaf nodes in a decision tree), as illustrated in Figure 4. We find that clusters of stocks with high unexpected earnings, high earnings-to-price ratios, and low trading volumes exhibit the highest predictability. In contrast, stocks with modest earnings surprises, low return variance, and low liquidity are the least predictable. The gap in predictability between highly and less predictable clusters persists out-of-sample, confirming the robustness of the link between latent predictability and high-dimensional characteristics.

Building on Cong et al. (2023), we extend clustering analysis to capture regime shifts driven by macroeconomic predictors or calendar months (structural breaks). This method uncovers dynamic patterns in stock return predictability across cross-sectional and time-series dimensions. Our analysis identifies two key macroeconomic drivers: dividend yield and liquidity, both central to business cycles (Fama and French, 1988; Pástor and Stambaugh, 2003). Return predictability declines sharply during low dividend yield periods but peaks during high dividend yield phases, typically linked to recessions. These findings align with Henkel et al. (2011) and Dangl and Halling (2012). We find individual stock return predictability is highest when the market dividend yield is high, and market liquidity is low, conditions often signaling the end of recoveries or early recessions. Across three macroeconomic regimes, decision trees consistently highlight value characteristics as significant predictors for cross-sectional predictability difference. Similar patterns observed across calendar month structural breaks further validate these results.

Finally, while our primary goal is to cluster stock returns based on predictability, our approach also generates cluster-specific predictive models, linking heterogeneous predictability to expected returns. Although Rapach, Strauss, and Zhou (2010) and Kelly, Malamud, and Zhou (2024) suggest no direct link between OOS $R^2$ and invest-

ment gains, we identify a novel anomaly using our proposed measure of cluster-wise return predictability relative to the global homogeneous model. By ranking clusters based on the predictability difference between cluster-specific and global models -— effectively sorting stocks by predictability, which is uniform within clusters -— we uncover varying degrees of model misspecification risk. This aligns with recent research on regime-switching model misspecification risks in equities and currencies (e.g., Smith and Timmermann, 2022; Feng et al., 2024). Our findings reveal a positive relationship between heterogeneous predictability models, investor model misspecification risk, and expected returns. Long-short value-weighted portfolios sorted by predictability differences generate unexplained alphas exceeding 1% over multiple benchmark factor models over a two-decade OOS period. Additionally, our analysis highlights consistent improvements in investment metrics such as Sharpe ratios, market alphas, and maximum drawdowns.

**Literature.** Our study contributes to the extensive empirical literature on return predictability in general, and emerging studies on heterogeneous return predictability in particular. We are not the first to point out that return predictability is not homogeneous. For example, Avramov (2002) and Avramov et al. (2023) demonstrate that predictability is concentrated in micro-cap stocks, distressed stocks, and during periods of high market volatility; Green et al. (2017) show forecast-implied portfolios exploiting characteristics-based predictability are insignificant outside micro-caps since 2003. Long-established research also links stock market return predictability to business cycle indicators (e.g., Keim and Stambaugh, 1986; Fama and French, 1989), and advanced predictive models (e.g., with regime-switching VAR and time-varying coefficients) reveal stronger predictability during economic recessions (e.g., Henkel et al., 2011; Dangl and Halling, 2012).

Most closely related to our study and advancing our understanding of predictability heterogeneity, Feng and He (2022) develop a Bayesian framework to utilize cross-asset information to predict returns of exogenously given portfolios of assets, while a series of studies identify heterogeneous sets of predictors effective for various cal-

endar time windows defined by detected breakpoints (Smith and Timmermann, 2021; Farmer et al., 2023; Cakici et al., 2024; Farmer et al., 2024).

We contribute by conceptually treating return predictability as a characteristic of assets under particular environments, and systematically analyzing such heterogeneity for the first time. Importantly, we not only extend traditional time-varying coefficient models by incorporating high-dimensional asset-level and aggregate predictors, but also using asset characteristics and lagged macroeconomic regimes to endogenously generate the heterogeneity groups in both the cross-section and time series. In so doing, we can also better help rationalize the apparent instability and inconsistencies of prior empirical findings in OOS or post-publication evaluations (Pesaran and Timmermann, 1995; Welch and Goyal, 2008; Harvey et al., 2016; McLean and Pontiff, 2016). Another key distinction of our approach is the seamless integration of clustering and cluster-wise model fitting within a unified framework, rather than treating these processes as separate steps. Moreover, the framework is model-agnostic, allowing it to be combined with various machine learning (ML) methods, including Lasso, PCA, and others.

Speaking of methodological innovations, along with Cong et al. (2025) and Cong et al. (2023), we are among the first to develop economically guided panel trees, part of the emerging AI literature on goal-oriented search/generative modeling — a data-driven approach to optimizing an economic goal in a large, flexible modeling space (e.g., Cong et al., 2020; Feng et al., 2024; Cong, 2025). The panel trees' "divide-and-conquer" approach mimics how humans solve complex problems by completing constituent tasks. Cong et al. (2023) and Feng et al. (2024) extend the panel-tree framework to fit heterogeneous models by maximizing marginal likelihood. Bie et al. (2024) examine regime switching within the dynamic Nelson-Siegel model for yield curves. We differ in focusing on the economic objective of maximally separating assets of differential return predictabilities.

Closely related by analyzing endogenous grouped heterogeneity in financial markets, Ahn et al. (2009) use unsupervised clustering based on return correlations, and

Patton and Weller (2022) generalize $K$-means to group assets by within-group slopes and averages, finding pervasive risk-price heterogeneity. More recently, Evgeniou et al. (2023) apply unsupervised $K$-means to iteratively cluster firms to fit regression residuals and estimate post-cluster heterogeneous predictive models. Our tree-based clustering employs a goal-oriented approach, segmenting predictable and less predictable groups based on asset traits and lagged macroeconomic variables. This data-driven method generates "mosaics" of return predictability.

The remainder of the article is structured as follows: Section 2 introduces the $R^2$ predictability measure and the clustering framework. Section 3 describes the data and evaluation procedures. Section 4 examines cross-sectional heterogeneity in predictability and predictability-sorted trading strategies. Section 5 explores time-varying predictability and regime shifts. Section 6 evaluates investment performance gains from cluster-based models. Section 7 concludes. The appendices contain algorithms, data descriptions, and supplementary results.

## 2 A P-Tree Framework for Analyzing Heterogeneous Predictability

### 2.1 Measurement of Return Predictability for Clustering

As discussed in the introduction, return predictability is unobservable and lacks a consensus definition in the literature. We propose that the in-sample $R^2$ of a predictive model serves as a natural measure of data predictability. While statisticians traditionally interpret $R^2$ as a property of the model, representing the proportion of return variation explained by the model, we extend this interpretation from a data perspective. Specifically, we view $R^2$ as a measure of how much variation in a specific dataset can be explained by the best-fitted ML model. Equivalently, it serves as an indication of the difficulty in predicting the dataset.

Intuitively, if the stock universe is already partitioned into multiple clusters, fitting a predictive model (with the necessary turning of parameters) allows each model to achieve optimal predictive performance within its respective cluster. However, the resulting $R^2$ values for each cluster may still vary, reflecting the intrinsic difficulty of

predicting returns within that cluster. A cluster with a lower $R^2$ value suggests that its returns contain less signal to be explained by the predictors, making it more challenging to predict out of the sample.

Specifically, we revisit the calculation of $R^2$ to justify its usage as a measure of return predictability. First, in the literature (e.g., Fama and French, 2008; Lewellen, 2015; Gu et al., 2020), a predictive model is typically represented as:

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

with the assumption $\mathrm{E}[\varepsilon_{i,t}] = 0$ such that the prediction of expected return is unbiased. The information regarding the heterogeneous predictable difficulty can be represented by the S2N ratio in Equation (2):

$$R^2_{i,t} = 1 - \frac{\mathrm{Var}(\varepsilon_{i,t})}{\mathrm{Var}(r_{i,t})} := 1 - \frac{\sigma^2_{\varepsilon,i,t}}{\sigma^2_{i,t}}, \tag{2}$$

where $\sigma^2_{\varepsilon,i,t}$ and $\sigma^2_{i,t}$ denote the variances of $\varepsilon_{i,t}$ and $r_{i,t}$, respectively. Subscripts $i$ and $t$ of this metric emphasize its cross-sectional and time-series variability. Conceptually, when $R^2_{i,t}$ is high for asset $i$ at a specific period $t$, it is relatively easier for a predictive model to capture the conditional expectation. Conversely, if the noise for asset $i$ is large, even with knowledge of the true $\mathrm{E}_{t-1}[r_{i,t}]$, the resulting $R^2$ would still be low, let alone when learning the conditional expectation from noisy data. Thus, with appropriate model regularization, the in-sample $R^2_{i,t}$ provides a reasonable measure of the S2N ratio, reflecting the return predictability of various assets across different periods.

Second, estimating $R^2_{i,t}$ for each asset $i$ at time $t$ is challenging due to the lack of data. The literature often addresses this issue by fitting a pooled model and calculating a single $R^2$ for all assets across all periods (Gu et al., 2020). However, this approach combines all data under one homogeneous predictive model, neglecting the potential heterogeneity of return predictability on the panel. Our approach balances these extremes by dividing asset returns into a few subsets and calculating the $R^2$ for each group to measure cluster-wise predictability. This goal-oriented clustering method is

specifically designed to maximize differences in predictability across clusters. Details of the splitting process are discussed in the following subsections.

Third, readers may ask why we do not use OOS $R^2$ to guide our clustering approach. The primary reason is that our clustering method relies on this measure to determine the clustering structure of assets. Incorporating OOS data would introduce future information into the clustering process and result in data-snooping bias. Instead, we treat the clustering approach and the cluster-wise predictive model as a whole, both determined using in-sample data. The performance of this system is then independently evaluated on OOS data. Additionally, the expected OOS mean squared error has the well-known bias-variance decomposition:

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

When predicting stock returns, $\sigma^2_{\varepsilon,i,t+1}$ dominates due to the low S2N ratio. Robust prediction benchmarks, such as zero for individual stocks or the historical average for the market index, may not reduce predictive bias but typically result in almost zero predictive variance. Furthermore, as documented in the literature, the OOS $R^2$ generally remains below 1%, making it hard to produce robust clustering results based on OOS $R^2$ in environments with low S2N ratios.

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

### 2.2 Cluster-wise Predictive Modeling

Clustering and predictive modeling are two fundamental tasks in machine learning, typically addressed by different models. However, our approach integrates both tasks within a unified framework, decision trees are customized for clustering, and ML models (e.g., Ridge regression) are used for prediction.

In this section, we assume the clusters are predetermined and focus on introduc-

ing the cluster-wise analysis. The subsequent section details our approach for detecting clusters using cluster-wise predictive models.

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

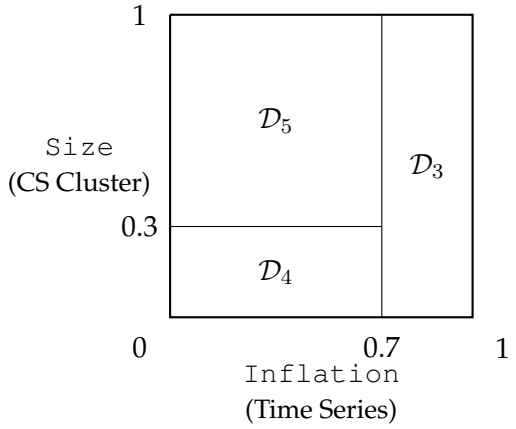
Ideally, heterogeneous and time-varying expected excess returns, $\mathrm{E}_{t-1}[r_{i,t}]$, would be modeled as $g_{i,t}(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$, a function that varies across assets and time periods. However, due to the limited observations available for individual stock returns, estimating each $g_{i,t}(\cdot)$ separately is infeasible. As a result, many studies (e.g., Gu et al., 2020) adopt a homogeneous predictive function, $g_t(\cdot)$, and update time-varying coefficients using a rolling-window scheme. However, Feng and He (2022) and Evgeniou et al. (2023) emphasize that such homogeneous modeling overlooks the heterogeneity in predictive power across different assets. Moreover, this approach implicitly assumes homogeneous return predictability — i.e., the same $R^2$ for every asset — which is inconsistent with empirical findings (Hou et al., 2020; Avramov et al., 2023). Additionally, while rolling-window estimation provides robustness, it fails to capture macroeconomic-driven regime shifts in the stock market.

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

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

Figure 1: **Clustering Illustration via Partitions**

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




Rather than fitting individual predictive models, $g_{i,t}(\cdot)$, for each asset $i$ in period $t$, or a single pooled model, we estimate a small number of cluster-wise homogeneous models that vary across clusters. As a result, the cluster-wise predictive model is:

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

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

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

### 2.3 Clustering: First Split

The objective of our clustering approach is to group asset returns with similar predictability into the same cluster. As discussed in the previous subsection, predictabil-

ity is measured using the in-sample $R^2$ of a cluster-wise predictive model, such as Ridge regression. Traditional clustering methods in the machine learning literature, such as $K$-means, are not well-suited for this task. These methods typically minimize within-cluster distances and maximize cross-cluster distances based on characteristics, but they do not account for return predictability. Moreover, their resulting clusters often lack clear interpretability.

We propose a novel model-based clustering approach that partitions the panel of stock returns based on unobservable predictability, distinguishing between highly and less predictable observations. This method optimizes the separation of samples into two subgroups by maximizing the difference in their S2N ratios or $R^2$ values[5].

Our approach is implemented iteratively to partition the entire panel of stock-return observations, adding one cluster at a time. The results are visualized using a decision tree structure. Unlike the Random Forest algorithm, which constructs multiple trees for prediction only, our method uses lagged (observable) information to iteratively partition the sample, resulting in a deterministic clustering output.

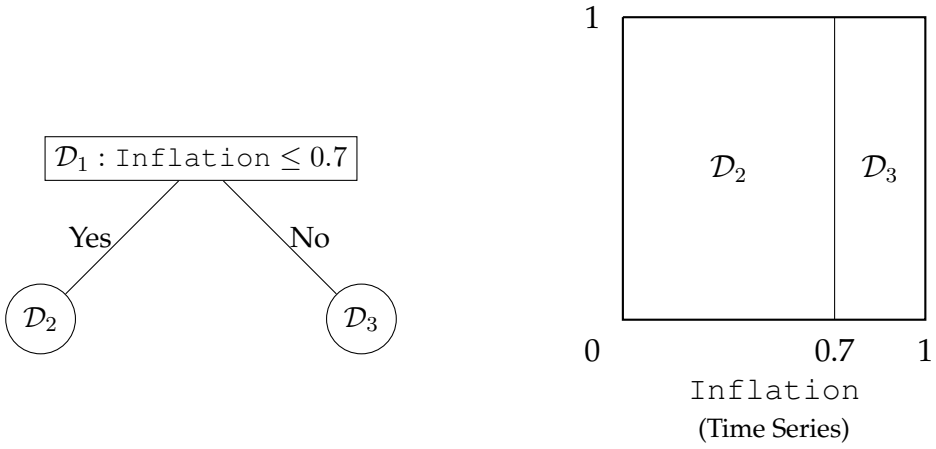
The initial step tries to find the first split point (split predictor and cut-point value) to divide the data into two sets. This process includes fitting cluster-wise predictive models and calculating the absolute $R^2$ difference for each split candidate, ultimately selecting the one that maximizes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

Figure 2 illustrates a *candidate* for the first split in a decision tree, where the root node, $\mathcal{D}_1$, containing the entire dataset, is divided into two clusters, or *leaf nodes*—$\mathcal{D}_2$ and $\mathcal{D}_3$. Following terminologies in Cong et al. (2025), leaf nodes refer to the nodes at the bottom of decision trees without subsequent branches. The split is based on the rule "$\text{var}_p(\texttt{Inflation}) \leq c_k(0.7)$," where the $p$-th variable (firm characteristic, aggregate predictor, or calendar month) is inflation, and the $k$-th cut-point value is 0.7. To evaluate the quality of this split candidate, we define a goal-oriented split criterion that measures its effectiveness in distinguishing returns with high predictability from those with low predictability.

[5]Alternative clustering objectives could achieve similar goals, but maximizing the $R^2$ difference is both interpretable and practical.

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

This figure illustrates one of the first split candidates, such as `Inflation`$\leq 0.7$. The left figure shows a decision tree that divides the sample, and the right figure shows the corresponding partition plot that only partitions over time.




The first split candidate divides the entire return sample $\mathcal{D}_1$ into two clusters, $\mathcal{D}_2$ and $\mathcal{D}_3$. We fit two cluster-wise predictive models, denoted as $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$, respectively. Generally, for the $j$-th leaf node, we fit a specific predictive model $g_j(\cdot)$ in Equation (5), and the return predictions will be 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 stock characteristics and aggregate macro predictors, respectively.

Notably, Fama and French (2008) have criticized that small-cap stocks with high return variance largely dominate the panel regressions, and Hou et al. (2020) show that many anomalies are replicable due to the dominance of micro-caps (about 60% of all firms) in the cross-sectional regressions. Therefore, we consider using the volatility-weighted Ridge regression for the cluster-wise predictive regressions and further split criterion calculations.

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

where $\mathbf{s}_{i,t-1} = \{\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}\}$ include lagged firm characteristics and aggregate predictors, and $w_{i,t-1}$ is the inverse of idiosyncratic return variance. The volatility, $\sigma^2_{i,t-1}$, is

estimated on a rolling-window basis, which helps to incorporate both cross-sectional and time-series variation for observation weights within the leaf cluster. The hyper-parameter $\lambda$ is determined by cross-validation.

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

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

Since our goal is to separate returns with high predictability from those with lower predictability, it is natural to use the absolute value of the $R^2$ difference between the left and right clusters as the split criterion:

$$S_{\{\text{leaf}_l, \text{leaf}_r\}} (\text{var}_p, c_k) = \left| R^2_{\text{leaf}_l} - R^2_{\text{leaf}_r} \right|. \tag{7}$$

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

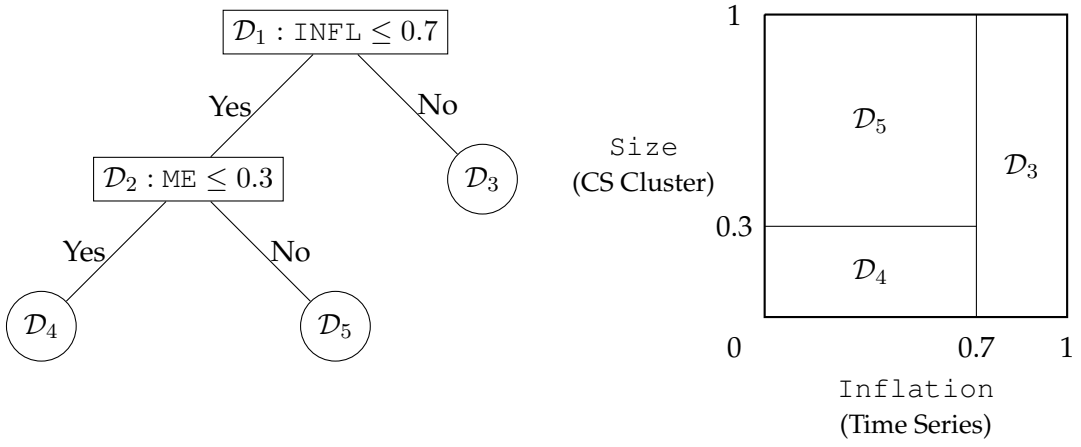
We then evaluate the criterion in Equation (7) across all potential split candidates, considering the full combination of split variables and cut-point values. Different pairs of split candidates, $\text{var}_p, c_k$, result in various partitions of the data, producing non-overlapping sub-samples as leaf nodes, $\mathcal{D}_2$ and $\mathcal{D}_3$, which correspond to cluster-wise predictive models, $\widehat{g}_2(\cdot)$ and $\widehat{g}_3(\cdot)$. These partitions ultimately yield different values for the split criterion in Equation (7), with a successful split candidate maximizing this criterion. Consequently, $P$ candidate variables and $K$ potential cut-point values provide a total of $P \times K$ possible split combinations for the first split, and we evaluate each of them to pick the optimal one as the first split. Then the root node will be divided into two child nodes (clusters).

### 2.4 Clustering: Subsequent Splits and Stopping Criteria

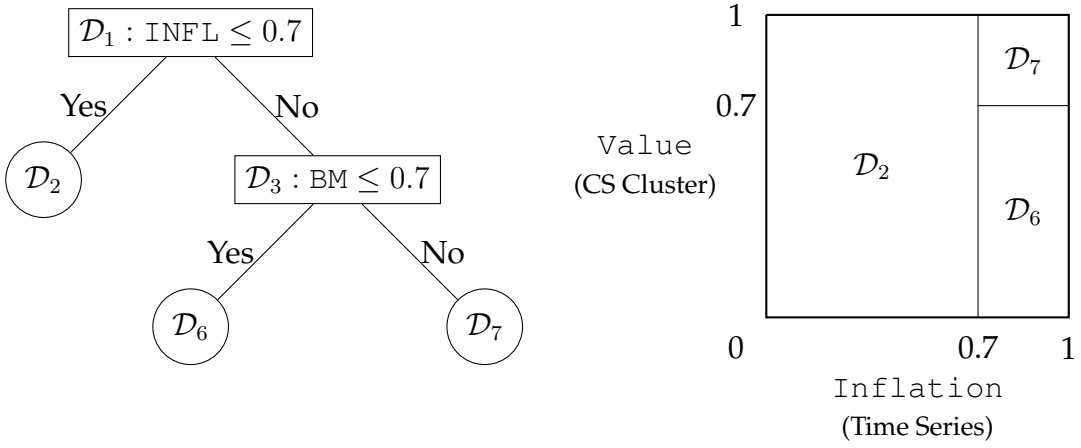
We essentially take a sequential approach to partition the panel into multiple clusters. Once the first split is determined, it creates two leaf nodes, and subsequent splits can occur at either of these nodes to further separate the clusters. Figure 3 illustrates two potential candidates for the second split, which may occur on the left (non-high inflation) or right (high inflation) leaf, depending on different characteristics, highlighting the asymmetric interaction of split predictors.

Figure 3: **Illustration for the Second Split**

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




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




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

Specifically, the second split can occur within the left leaf node $\mathcal{D}_2$, partitioning it

into $\mathcal{D}_4$ and $\mathcal{D}_5$, or within the right leaf node $\mathcal{D}_3$, dividing it into $\mathcal{D}_6$ and $\mathcal{D}_7$. Each side still evaluates $P \times K$ split candidates, resulting in a total of $2 \times P \times K$ 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 adopts a local-global approach: it evaluates the benefits of splitting each leaf node based on local $R^2$ differences between sub-samples and globally selects the split that best differentiates return predictability.

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

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

**Cluster-wise Predictions.** Once tree growth terminates, the entire panel of observations is partitioned into multiple non-overlapping clusters based on variable interactions, such as firm characteristics, aggregate predictors, or calendar months. Within

each cluster, we refit cluster-specific predictive models using the corresponding observations. As outlined in Section 2.2, we employ Ridge regression for return predictions. The final output of our approach is a decision tree structure that delineates the clustering pattern, with each cluster linked to its own Ridge prediction model. We refer to this as a heterogeneous prediction model, contrasting to a global prediction model that fits a single Ridge regression across all stock-return observations.

Notably, while we demonstrate our clustering approach using Ridge regression in Equation (5) for simplicity, the framework is not restricted to any specific predictive model. Instead, it provides a flexible and generalizable framework for heterogeneous predictions that can incorporate a wide range of ML models.

## 3 Empirical Data and Evaluations

We apply our approach to U.S. individual stock returns to study their heterogeneous predictability.[6] The monthly sample spans from 1973[7] to 2022, with the first 30 years used for model training and the most recent 20 years for the OOS analyses. The average and median number of stocks in the training sample are 4,840 and 4,772, respectively, while in the test sample, these numbers are 3,909 and 3,694.[8]

**Aggregate Predictors.** Following Welch and Goyal (2008), we analyze eight aggregate predictors to define and select macroeconomic regimes characterized by time-varying return predictability. As detailed in Table A.6, these predictors include the 3-month Treasury bill rate, inflation, term spread, default yield, and market-level characteristics such as dividend yield, volatility, net equity issues, and liquidity.

To ensure comparability of these predictors over time, we standardize them by subtracting their respective 10-year rolling window average. We also apply a 12-month weighted moving average, assigning greater weights to more recent observa-

[6] We use standard filters (see, e.g., Fama and French, 1992), including: (1) restricting the sample to stocks listed on NYSE, AMEX, or NASDAQ for more than one year; (2) selecting observations for firms with CRSP share codes of 10 and 11; and (3) excluding stocks with negative book equity or negative lagged market equity.

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

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

tions, to further smooth the data and reduce volatility. A positive value of a macro indicator indicates that its current level exceeds its 10-year average, signaling a high regime in the economy. This standardized approach enhances the effectiveness of the tree clustering method in detecting macroeconomic regimes based on recent data.

**Characteristics.** As detailed in Table A.7, our dataset comprises 51 firm-level characteristics categorized into eight major groups: size, value, investment, momentum, profitability, liquidity, volatility, and intangibles. Each characteristic is standardized cross-sectionally and scaled uniformly to the range [0, 1] for every month. To mimic the "top-middle-bottom" sorting approach, we define two cut-point values, 0.3 and 0.7, as split-value candidates for each characteristic. These characteristics are utilized for forecasts to construct tree-based clustering and cluster-wise predictive models.

**Model Fitting Design.** The baseline analyses for cross-sectional partitions use the first 30 years of data for both tree-based clustering and model estimation, with the most recent 20 years serving as testing samples. To enhance the predictive performance, we update the tree-based clustering and cluster-specific predictive models every five years, employing a 30-year rolling window of in-sample data to retrain the decision tree structure. This process is repeated four times over the 20-year OOS data.

For extended analyses involving time-series splits, we conduct a full-sample analysis to account for the long and overlapping nature of business cycles. Additionally, we use cross-validation to optimize the hyperparameters during the post-cluster predictive model training phase.

**Performance Evaluation.** As mentioned in Section 2.1, we use in-sample $R^2$ to measure predictability. In addition to the in-sample $R^2$, we evaluate the OOS $R^2$ 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 OOS $R^2$ ($R^2_{OOS}$) using zero forecasts as the benchmark,

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

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

# 4 Cross-Sectional Heterogeneous Return Predictability

The baseline model examines cross-sectional heterogeneity in return predictability, addressing two key questions: Does heterogeneity exist, and which stock returns are more predictable? This section shows that our goal-oriented clustering approach effectively captures this heterogeneity. Unlike methods assigning stocks to fixed clusters across all periods (e.g., Patton and Weller, 2022; Evgeniou et al., 2023), we allow stocks to transition between clusters over time. Since clusters are defined by characteristic values, these transitions naturally reflect evolving characteristics-return predictability. We refer to the clustering model in this section, which focuses solely on cross-sectional splits, as "CS clusters" to distinguish it from the "TS+CS clusters" model discussed in Section 5, which integrates time-series regime shifts with cross-sectional heterogeneity.

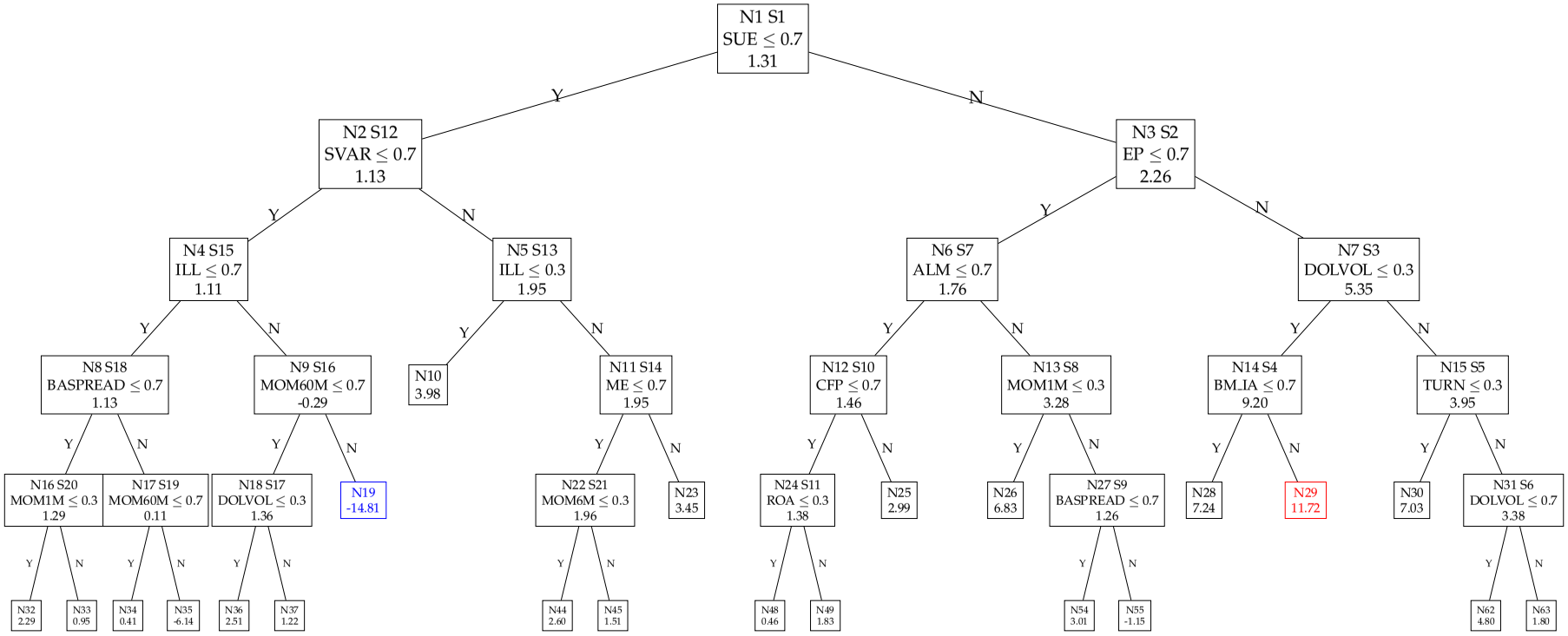
## 4.1 Cluster and Heterogeneous Predictability

**Clustering Pattern.** We present our self-supervised cross-sectional clustering results in Figure 4, derived from training the model on data from the first 30-year period (1973–2002). Following our predetermined stopping criteria, this tree stops growing after reaching 22 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, representing the initial split, is labeled as leaf node $N_1$ and split $S_1$. The second row specifies the optimal split rule, directing observations satisfying the condition to the left child node, while others proceed to the right. Terminal leaves, which undergo no further splits, are denoted only by their index in the first row. The bottom row reports the $R^2$ value, capturing cluster-wise return predictability at each node. This

[9] A moderately deep decision tree with large leaves (sample size) is both robust and interpretable. To achieve this, we set the maximum depth to 6 (up to 32 leaves) and impose a minimum leaf size of 30 monthly stock-return observations to ensure robust training. Under these constraints, the algorithm terminates after 21 splits.

Figure 4: **Tree-Based Cluster (CS Cluster)**

This figure illustrates the cross-sectional tree-based clustering derived from monthly data (1973–2002). The tree partitions stock returns using monthly cross-sectional standardized characteristic ranks within the [0,1] range. Terminal leaves represent clusters defined by specific characteristic intervals. Each node, including intermediate and terminal leaves, is labeled with an ID ($N\#$) and split order ($S\#$). Cluster-level $R^2$ values, indicating return predictability, are provided for each node.



structure offers a clear visualization of clustering processes and interactions among characteristics.

Before any split, the aggregate return predictive ability ($R^2$) of the homogeneous model is 1.31% (root node). After the first split, the $R^2$ improves significantly for stock returns with high unexpected earnings ($\texttt{SUE} > 0.7$), increasing to 2.26% at N3, while slightly declining for the complement set, reducing to 1.13% at N2. The $R^2$ difference (2.26 - 1.13)% represents the maximum value identified among all split candidates ($51 \times 2$) by the algorithm. Stocks with higher unexpected earnings reflect increasing predictability, related to the post-earnings-announcement drift (e.g., Bernard and Thomas, 1989).

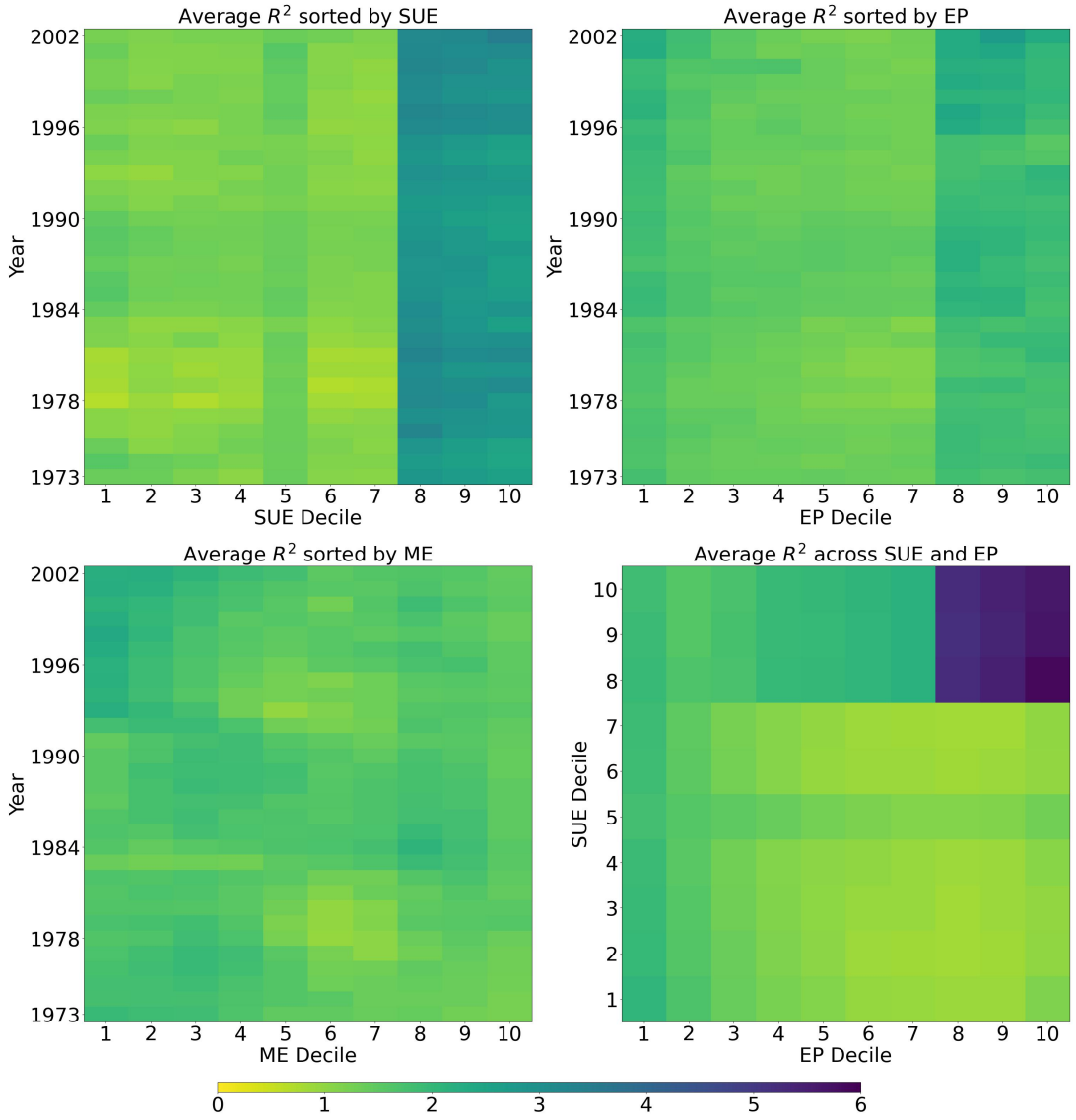
Additionally, Figure 5 highlights marginal information extracted from the decision tree shown in Figure 4. It illustrates the average decile $R^2$ values for each year and/or decile cluster sorted by various characteristics. A clear pattern emerges, indicating an advance in return predictability for stocks with higher earning surprises, which has persisted over time. The bottom-left sub-figure further corroborates the relatively superior forecast accuracy for small-cap stocks.

The second split selects high earnings-to-price stocks ($\texttt{EP} > 0.7$), resulting in a higher $R^2$ value of 5.35% at N7, under the top earning surprises condition. Valuable stocks exhibit greater return predictability, while non-value stocks present lower forecasting accuracy (1.76% at N6). The maximum difference in $R^2$ values (5.35 - 1.76)% is determined through a comprehensive search involving the left cluster N2 (stocks with non-high unexpected earnings, $\texttt{SUE} \leq 0.7$) and the right leaf N3 (stocks with higher earning surprises, $\texttt{SUE} > 0.7$).

In addition, as shown in Figure 5, the EP-sorted decile clusters in the middle exhibit the lowest predictability compared to those at both extremes, with higher value stocks consistently achieving greater accuracy. The bottom-right sub-figure, sorted by SUE and EP, highlights an interaction pattern where stocks with both high SUE and EP exhibit the largest forecast accuracies, indicated by the darkest shading. These mosaic-like patterns emphasize the differences in return predictability arising from

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

These heat maps summarize return predictability ($R^2$ values, % in the color bar) for stock returns based on the tree-based structure in Figure 4. The first three heat maps show average $R^2$ values for groups sorted by years and deciles of unexpected earnings, earnings-to-price, and market equity. The fourth heat map displays average $R^2$ values for $10 \times 10$ groups formed by bivariate decile sorting of the top two characteristics. Colors range from light to dark, indicating increasing levels of return predictability.

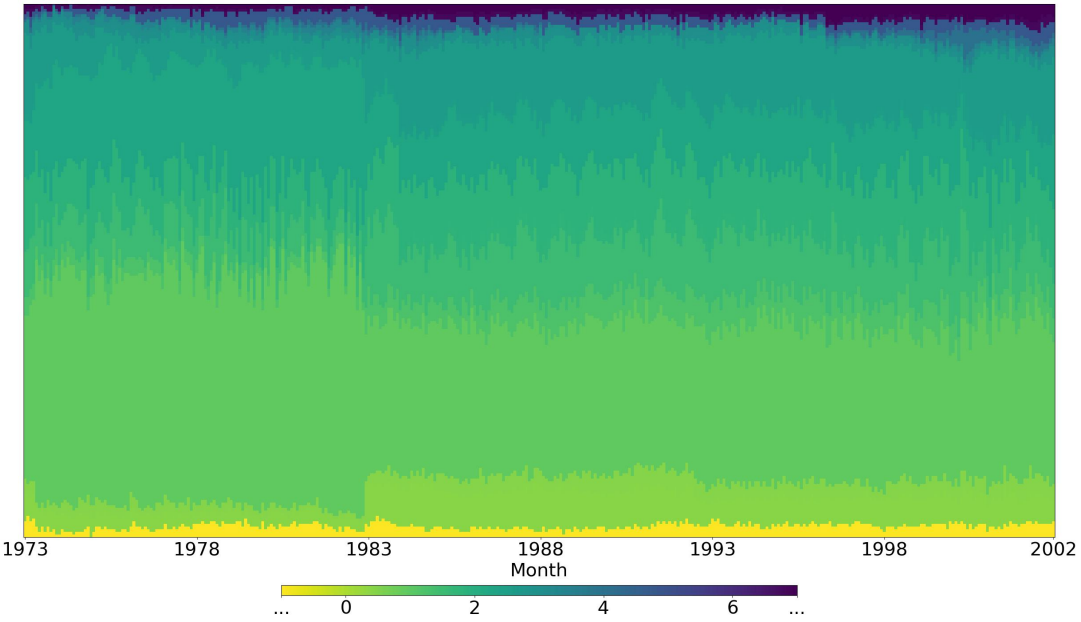



the interactions among various predictors.

Figure 6 presents the return predictability derived from tree-based clustering in another visually distinct mosaic-like format. The sample period spans 360 months, corresponding to the tree structure shown in Figure 4, with approximately 4,000 to 5,500 observations per month. Notably, since clusters are defined based on characteristic values, the number of stocks within each cluster is not fixed over time. For each month, clusters are arranged in ascending order of predictability, with the height of

Figure 6: **Mosaics of Predictability by Clusters**

This heat map depicts monthly return predictability ($R^2$ values, % in the color bar) for stock returns based on the tree structure in Figure 4. The vertical axis indicates the proportion of observations per cluster, while the horizontal axis represents months. Colors transition from light to dark, denoting increasing predictability.




each segment reflecting the proportion of stocks within that cluster for that month.[10] As $R^2$ values increase (darker colors), the proportion of observations decreases, indicating that the majority of stock returns are hard to predict. By setting the lower bound of the color bar to 0, it becomes apparent that more than half of the observations shift toward lighter shades each month. In contrast, only a small fraction of observations exhibit high predictability, represented by the deepest colors, corresponding to $R^2$ values exceeding 10%. This horizontal, mountain-like cascading pattern, with its variations across cross-sectional, time-series, and color dimensions, provides valuable insights into the heterogeneous nature of stock return predictability.

**Cluster-Wise Performance.** To provide a more detailed description, Panel A of Table 1 provides summary statistics for each cluster. Ordered by descending predictability of cluster-wise model forecasts, the table reveals substantial heterogeneity across clusters, with $R^2_C$ values ranging from 11.72% to -14.81%, reflecting a difference exceeding

[10]The denominators for the monthly proportions are based on the total number of observations in each month. The gap around 1982 may be attributed to a surge in IPOs during that period, potentially linked to the launch of the NASDAQ National Market in April 1982, which included larger and more actively traded NASDAQ securities.

Table 1: **Performance of Asset Clusters**

This table summarizes key metrics for each cluster from the cross-sectional tree structure in Figure 4. Panel A reports the number of observations (# obs), return predictability ($R^2$ in %) based on cluster-specific ($R^2_C$) 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 | $R^2_C$ | $R^2_G$ | $R^2_{CMG}$ | $\text{Avg}_{EW}$ | $\text{SR}_{EW}$ | $\text{Avg}_{VW}$ | $\text{SR}_{VW}$ |
| N29 | 11,622 | 11.72 | 7.56 | 4.16 | 4.51 | 2.21 | 3.65 | 1.92 |
| N28 | 16,327 | 7.24 | 5.23 | 2.02 | 3.24 | 1.76 | 2.54 | 1.60 |
| N30 | 19,357 | 7.03 | 6.25 | 0.78 | 2.43 | 1.68 | 1.78 | 1.25 |
| N26 | 17,658 | 6.83 | 2.74 | 4.09 | 5.00 | 1.86 | 2.21 | 0.85 |
| N62 | 34,818 | 4.80 | 4.38 | 0.43 | 2.39 | 1.36 | 1.82 | 1.13 |
| N10 | 14,191 | 3.98 | 1.41 | 2.57 | -0.81 | -0.26 | -1.05 | -0.34 |
| N23 | 11,842 | 3.45 | 1.55 | 1.89 | -1.18 | -0.41 | -1.41 | -0.49 |
| N54 | 21,116 | 3.01 | 2.36 | 0.65 | 2.10 | 1.16 | 1.22 | 0.63 |
| N25 | 19,908 | 2.99 | 2.41 | 0.58 | 2.00 | 1.14 | 0.95 | 0.59 |
| N44 | 180,186 | 2.60 | 1.14 | 1.46 | -0.42 | -0.15 | -1.44 | -0.52 |
| N36 | 17,843 | 2.51 | 1.18 | 1.33 | 0.76 | 0.37 | 0.52 | 0.26 |
| N32 | 218,020 | 2.29 | 1.74 | 0.56 | 1.07 | 0.57 | 0.63 | 0.37 |
| N49 | 173,910 | 1.83 | 1.51 | 0.33 | 1.39 | 0.77 | 0.68 | 0.44 |
| N63 | 30,990 | 1.80 | 0.92 | 0.88 | 1.30 | 0.75 | 1.08 | 0.67 |
| N45 | 161,143 | 1.51 | 1.05 | 0.46 | -0.31 | -0.13 | -0.58 | -0.23 |
| N37 | 75,041 | 1.22 | 0.96 | 0.26 | 0.41 | 0.22 | 0.44 | 0.25 |
| N33 | 537,210 | 0.95 | 0.86 | 0.08 | 0.37 | 0.27 | 0.25 | 0.20 |
| N48 | 47,114 | 0.46 | 0.61 | -0.16 | 0.52 | 0.22 | 0.47 | 0.20 |
| N34 | 87,370 | 0.41 | 0.95 | -0.54 | 0.16 | 0.10 | -0.09 | -0.05 |
| N55 | 13,548 | -1.15 | 1.07 | -2.22 | 2.09 | 0.90 | 1.12 | 0.45 |
| N35 | 16,638 | -6.14 | 1.02 | -7.16 | -0.11 | -0.07 | -0.57 | -0.26 |
| N19 | 16,635 | -14.81 | 1.22 | -16.04 | 0.36 | 0.20 | 0.44 | 0.27 |

20%. The predictive abilities based on homogeneous models follow a similar downward trend ($R^2_G$ values between 0.61% and 7.56%). These variations in $R^2$ demonstrate that each cluster exhibits a distinct level of predictability, underscoring the diversity in the cross-sectional dimension. Furthermore, the bottom six less predictable clusters (N33 to N19, with $R^2$ lower than 1%) account for a large portion of stocks (over 40%), whereas the top four most predictable clusters (N29 to N26, with $R^2$ above 5%) includes only a small subset (around 4%). This disparity highlights the limited availability of highly predictable stocks.

We further examine the relationship between return predictability and investment profits by constructing equal- and value-weighted portfolios of stocks within each cluster over time. Panel B of Table 1 reports the average returns and Sharpe ratios for these portfolios. Notably, the most predictable cluster (N29, with $R^2_C = 11.72\%$)

delivers an average monthly return of 3.65% (4.51%) and an annualized Sharpe ratio of 1.92 (2.21) for value- (equal-) weighted strategies, reflecting both strong profitability and an attractive risk-return trade-off. In contrast, the least predictable cluster (N19, with $R^2_C = -14.81\%$) exhibits significantly weaker investment performance.

These results highlight a clear relationship between average returns and return predictability, suggesting the potential for identifying anomalies associated with predictability. A more detailed discussion of the anomaly is postponed to Section 4.2.

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

Using the tree structure and cluster-wise performance detailed in Table 1, we categorize all leaves into three groups based on return predictability rankings: high, medium, and low. Observations from clusters such as N29, N28, N30, and N26 are aggregated to form the highly predictable group, while the last five leaves represent data with the worst forecast accuracies. The remaining leaves are combined to create the medium predictability set.[12] We use OOS $R^2$ values, as defined in Equation (8), to validate the cross-sectional patterns of stock return predictability.

Table 2 reports the $R^2$ statistics for global models (Panel A) and cluster-wise predictive models (Panel B). The "Global Forecasts" fits all data using a single predictive model without any clustering and is also evaluated on sub-samples of high, medium,

[11]The in-sample data is divided into two equally continuous periods. The model is trained on one period and validated on the other. The optimal hyperparameters are selected based on the average MSE, and the model is then retrained using all in-sample data. Finally, the resulting coefficients are applied to predict the next five years of OOS values.

[12]Predictability levels are determined by balancing $R^2$ values with the proportion of observations, prioritizing clusters with larger $R^2$ gaps while acknowledging that most returns show low predictability.

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

This table reports return predictability ($R^2$ in %) across predictive methods, with both in-sample and out-of-sample results from the cross-sectional tree structure in Figure 4. 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 predictions), and High, Medium, and Low, based on sub-sample predictability rankings across forecasts.

| | In-Sample (1973 - 2002) | | | Out-of-Sample (2003 - 2022) | | |
|---|---|---|---|---|---|---|
| Sample | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| | | | Panel A: Global Forecasts | | | |
| Global | 1.31 | 0.57 | 0.79 | 0.44 | 0.36 | 0.42 |
| High | 4.32 | 2.96 | 2.98 | 3.50 | 2.38 | 2.59 |
| Medium | 1.25 | 0.51 | 0.74 | 0.33 | 0.28 | 0.34 |
| Low | 0.84 | 0.27 | 0.46 | 0.33 | 0.34 | 0.34 |
| | | | Panel B: Cluster-Wise Forecasts | | | |
| Aggregate | 1.84 | 1.39 | 1.29 | 0.12 | 0.54 | 0.62 |
| High | 7.69 | 6.70 | 7.03 | 4.24 | 4.29 | 4.58 |
| Medium | 1.99 | 1.44 | 1.19 | -0.02 | 0.45 | 0.48 |
| Low | -0.85 | -0.46 | 0.26 | -0.06 | 0.25 | 0.51 |

and low predictability clusters identified by our approach. The "Aggregate" model in Panel B represents the aggregation results of predictions from all cluster-wise models. For instance, the global Ridge model in Panel A only achieves 2.59% $R^2$ out of the sample for highly predictable clusters, while the cluster-wise model achieves 4.58% in Panel B. The aggregate model demonstrates improvements in $R^2$ values, highlighting the benefits of the cluster-wise predictive model.

When comparing clusters with different levels of predictability, highly predictable clusters consistently outperform their less predictable counterparts, regardless of the model used. For instance, when comparing the values in each row for the highly predictable cluster against the low predictability cluster, every metric — whether in-sample or out-of-sample, OLS or Ridge regression, global or cluster-wise model — shows higher $R^2$ for the more predictable cluster. These results confirm that predictability is a persistent characteristic of stocks, remaining consistent out of the sample. Furthermore, they validate the effectiveness of our tree-based clustering approach in capturing and identifying this heterogeneity.

### 4.2 Anomaly of Predictability

Thus far, we have introduced a novel clustering approach that groups stocks based on return predictability and demonstrated its variation among the cross-section. It provides a new measure of the unobservable return predictability of individual stocks. Our findings highlight that cluster-specific predictability can be improved (reduced) for highly (less) predictable stock returns using heterogeneous forecasts instead of global predictions. Recent studies, such as Smith and Timmermann (2022) and Feng et al. (2024), explore the regime-switching risk for individual stocks and currencies, which naturally motivates us to raise the question: Do those cross-sectional predictability differences between heterogeneous and homogeneous measures capture the predictive model misspecification risks that are orthogonal to established risk factor models? For instance, if an investor estimates a model without incorporating heterogeneous predictability and constructs portfolios based on global forecasts, an unforeseen cross-sectional shift could materially alter the S2N ratios, potentially exposing the investor to unexpected losses.

To investigate this issue, we first evaluate the cluster-specific aggregated differences in return predictability between cluster-wise and global predictive models:

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

This metric quantifies the gaps in S2N ratios between the heterogeneous forecasts ($R^2_{\text{C}}$, derived from cluster-wise predictions) and the unconditional models ($R^2_{\text{G}}$, by global predictions without clustering) for each cluster $j$, as reported in Table 1. It captures the extent to which a predictive model estimated on the full sample either underestimates or overestimates stock returns across different characteristic-based groupings. A higher value of $R^2_{\text{CMG}}$ (R2CMG) signals a greater divergence between the two model predictions, reflecting stronger exposure to cluster-specific model misalignment risk. Stocks exhibiting such characteristics require higher risk-adjusted compensation, thereby uncovering a novel dimension of risk in the risk-return trade-off.

We construct both value-weighted and equal-weighted long-only portfolios for each cluster and examine whether their average returns exhibit significant spreads. Panel B of Table 1 highlights the return differentials, showing a positive relationship: clusters with larger return predictability differences tend to yield higher average returns and Sharpe ratios.

These results motivate the design of novel long-short strategies to exploit anomalies in return predictability. Our decision tree-based clustering reveals that each cluster exhibits distinct predictability levels and systematic divergences between predictive models. This approach generalizes sorting by leveraging latent discrete characteristics of cluster-specific heterogeneous predictability.

Specifically, we focus on the value-weighted portfolios depicted in Table 1. We begin by projecting the previously estimated return predictability difference ($R^2_{\text{CMG}}$), which is calculated at the cluster level within all in-sample data, onto all individual stock observations in both the in-sample and out-of-sample periods.[13] We rank the entire cross-section based on this discrete predictability-implied score at the end of each month and implement a long-short strategy. This mechanism simultaneously operates on equal numbers of cluster-wise portfolios at the top and bottom level of model misspecification risks (from "T1-B1" to "T3-B3"). To assess the abnormal returns associated with these anomalies, we conduct the spanning regressions of the portfolio returns on various widely used risk factor models.

Panel A of Table 3 summarizes the performance of the anomalies, demonstrating that portfolios constructed based on return predictability differences exhibit distinct return spreads. Long-short strategies with larger cluster-specific risk differentials (e.g., "T1-B1") consistently outperform those with smaller gaps (e.g., "T3-B3"), achieving higher average returns, Sharpe ratios, and profitability across both sample periods. Even the strategy incorporating the largest number of clusters ("T3-B3") delivers robust performance, with an average monthly return of 0.59% and an annualized Sharpe

[13]This implies that all observations sharing the same cluster label, whether in-sample or out-of-sample, are assigned an identical value. Moreover, we reiterate that the value for an individual stock $i$ may vary across different months $t$.

Table 3: **A New Anomaly on Predictability Difference**

This table presents summary statistics (Panel A) and abnormal returns (Panel B) for the anomaly based on cross-sectional return predictability differences ($R^2_{\text{CMG}}$ in Table 1) across clusters. The numbers following "T" and "B" indicate the number of top and bottom clusters, ranked by $R^2_{\text{CMG}}$, 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-values (in parentheses) from various factor models.

| | In-Sample (1973 - 2002) | | | Out-of-Sample (2003 - 2022) | | |
|---|---|---|---|---|---|---|
| | T1-B1 | T2-B2 | T3-B3 | T1-B1 | T2-B2 | T3-B3 |
| | | | Panel A: Summary Statistics | | | |
| Avg | 3.20 | 2.99 | 1.27 | 2.91 | 1.41 | 0.59 |
| Std | 4.63 | 4.63 | 4.03 | 5.49 | 4.83 | 3.98 |
| SR | 2.39 | 2.24 | 1.09 | 1.84 | 1.01 | 0.52 |
| | | | Panel B: Abnormal Returns | | | |
| CAPM | 3.22*** | 2.99*** | 1.18*** | 2.95*** | 1.33*** | 0.43* |
| | (13.14) | (12.17) | (5.71) | (8.19) | (4.20) | (1.67) |
| FF3 | 3.07*** | 2.75*** | 1.36*** | 2.98*** | 1.36*** | 0.45* |
| | (12.25) | (11.25) | (6.61) | (8.36) | (4.42) | (1.89) |
| FF5 | 2.93*** | 2.66*** | 1.48*** | 3.09*** | 1.39*** | 0.57** |
| | (11.47) | (10.57) | (7.00) | (8.42) | (4.39) | (2.34) |
| FF5+MOM+IVOL | 2.93*** | 2.73*** | 1.54*** | 3.24*** | 1.47*** | 0.66*** |
| | (11.12) | (10.53) | (7.12) | (9.50) | (4.70) | (2.87) |
| Q5 | 2.83*** | 2.71*** | 1.60*** | 3.46*** | 1.61*** | 0.77*** |
| | (9.43) | (9.16) | (6.43) | (9.75) | (5.09) | (3.27) |
| BS6 | 2.76*** | 2.55*** | 1.54*** | 3.23*** | 1.40*** | 0.69*** |
| | (9.91) | (9.35) | (6.64) | (9.41) | (4.51) | (3.04) |
| DHS3 | 3.24*** | 3.03*** | 1.58*** | 3.20*** | 1.47*** | 0.64** |
| | (11.45) | (10.60) | (6.79) | (8.83) | (4.56) | (2.49) |
| SY4 | 2.87*** | 2.73*** | 1.69*** | 3.77*** | 1.89*** | 0.87*** |
| | (10.40) | (10.03) | (7.52) | (8.85) | (5.26) | (3.04) |

ratio of 0.52 for the OOS data.

Panel B of Table 3 reports the evaluation of these anomalies using various factor models, including CAPM, FF3, FF5, FF5+MOM+IVOL, Q5 (Hou, Mo, Xue, and Zhang, 2021), BS6 (Barillas and Shanken, 2018), DHS3 (Daniel, Hirshleifer, and Sun, 2020), and SY4 (Stambaugh and Yuan, 2017).[14] All models exhibit statistically significant unexplained alphas, both in-sample and out-of-sample. The OOS monthly alphas for the strategy with larger predictability gaps ("T1 - B1") are particularly strong, ranging from 2.95% to 3.77%, with high $t$-statistics. While strategies with smaller cluster-specific risk gaps ("T3 - B3") show slightly weaker performance, their alphas remain

[14]The data on Stambaugh's website is updated only through 2016. Thus, the OOS results for SY4 are based on the 2003–2016 period.

significant at the 10% level across all models. These findings underscore the robustness of the strategy's unexplained alphas, even after adjusting for risk factors in standard asset pricing models.

We further investigate whether the abnormal returns primarily originate from the long-leg, the short-leg, or both components of the portfolio. Table A.1 reveals that long-leg portfolios with larger predictability differences consistently contribute the most to the abnormal returns, whereas the short-leg portfolios generally exhibit insignificant, near zero or marginally negative alphas. Therefore, those top clusters — stocks with the largest model misspecification risks (long-leg) — play a particularly important role in generating abnormal returns.

In conclusion, we identify a novel anomaly arising from the cross-sectional heterogeneity in return predictability. Strategies constructed based on aggregate predictability gaps between cluster-wise and homogeneous predictive models exhibit significant spreads in average returns, with the corresponding anomaly generating substantial alpha even after accounting for popular risk factor models in the literature. These findings underscore predictability differences as a latent model misalignment risk that provides information orthogonal to established factor models, which offer new insights into the role of predictability in stock market dynamics. Furthermore, predictability-driven strategies deliver superior risk-adjusted returns, presenting an empirical contrast to the findings of Rapach et al. (2010) and Kelly et al. (2024).

## 5 Heterogeneous Predictability and Regime Switches

The prior section explored the clustering of predictability in the cross-section. A natural extension is to investigate whether such heterogeneity evolves over time. This section incorporates the time-series dimension into the analysis. Notably, the split variables need not be limited to characteristics (CS Clusters, Section 4); they may also include macroeconomic variables (TS+CS Clusters). By employing aggregate predictors, we analyze clustering in stock return predictability across varying horizons (Section 5.1). For robustness, we also show that calendar months can serve as valid split

variables, aiding in identifying structural breaks in return predictability (Section 5.2). Unlike cross-sectional clusters, time-series heterogeneity is analyzed by splitting the full sample period (1973–2022) to uncover regime shifts across the entire sample.

### 5.1 Macro-Driven Regime Change

We begin by using aggregate macro predictors to split the time horizons and identify macroeconomic regimes of predictability. This approach divides the entire sample based on macroeconomic variables, providing clear and intuitive interpretations. Notably, because the regimes are defined by macroeconomic values, a single regime may not correspond to a continuous time horizon.

In particular, we restrict the use of aggregate predictors to time-series data for the first two splits, while utilizing firm characteristics for subsequent cross-sectional partitions.[15] This methodology establishes a systematic framework that learns the clustering of predictability across both time-series and cross-sectional dimensions, enabling the implementation of timing and stock selection strategies.

**Clustering Pattern.** For the macro-driven regime case, we fit the model using the full sample of data, resulting in a single tree-based clustering structure for the 50-year sample. Figure 7 illustrates the initial decision tree structure based on macroeconomic variables, while Figure A.1 continues the further partitions by firm characteristics under each time horizon. This structure allows the clustering pattern to jointly account for both time-series and cross-sectional heterogeneity. It ultimately produces 46 leaves corresponding to different levels of predictive abilities.[16]

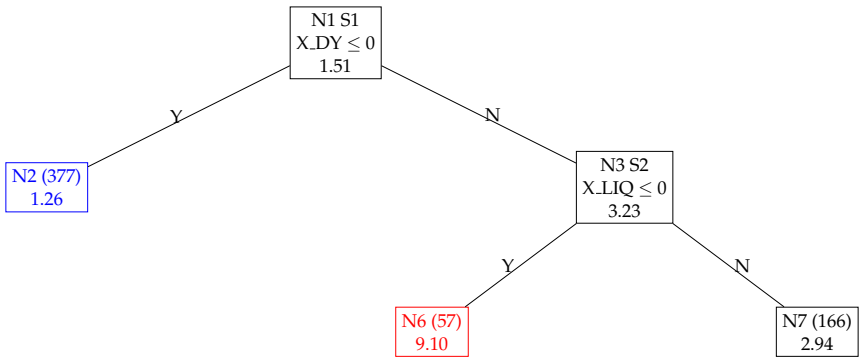
Specifically, the overall in-sample $R^2$ for the global model across the entire sample is 1.51% before any partitions. Our approach identifies aggregated dividend yields and liquidity as the two most important aggregate predictors for detecting regimes of

[15] We have experimented with relaxing these constraints, allowing both aggregate predictors and firm characteristics to be considered as candidates for all splits. However, the trees consistently select aggregate predictors for the first two layers of partitions in all cases. This suggests a clear preference for initial separation based on time series data. To mitigate the potential for excessive regime changes, we add those restrictions.

[16] We limit the maximum tree depth within each regime to 5 (resulting in at most 48 leaves) and set the minimum leaf size to 30 stock-return observations per month.

Figure 7: **Tree-Based Cluster under Macroeconomic Regimes**

This figure presents the tree-based clustering structure under macroeconomic regimes using monthly data from 1973 to 2022. The tree segments stock return panels based on aggregate predictors, applying a subtraction of the previous 10-year rolling window average and smooth, and splits the entire sample into three regimes (aggregated dividend yield and liquidity). Terminal leaves represent distinct regimes with nonconsecutive month indices in brackets. Each node, including both intermediate and terminal ones, is assigned a unique ID ($N\#$), while split order is indicated by ($S\#$). Cluster-specific $R^2$ values are reported for all nodes.




predictability, and creates the following three macroeconomic regimes:

1. Regime I (377 months): $\texttt{X_DY} \leq 0$, when the dividend yield is low.

2. Regime II (57 months): $\texttt{X_DY} > 0$ and $\texttt{X_LIQ} \leq 0$, when the dividend yield is high, and the liquidity is low.

3. Regime III (166 months): $\texttt{X_DY} > 0$ and $\texttt{X_LIQ} > 0$, when the dividend yield and the liquidity are both high.

Notably, these two predictors are highly related to business cycles (Fama and French, 1988; Pástor and Stambaugh, 2003). Thus, we find that the business cycle affects the return predictability of stock returns.

After the first two splits based on macroeconomic variables, the $R^2$ value increases significantly during periods with higher dividend yields and lower liquidity compared with the previous rolling 10-year average (Regime II), reaching 9.10% at node N6. Examining this subset more closely, the regime characterized by low aggregated liquidity ($\texttt{X_LIQ} \leq 0$) reflects low market returns or increased market uncertainty, leading to greater predictability compared to the opposite regime (Regime III), which has an $R^2$ of 2.94% at node N7.

Notably, these two periods are typically associated with recessions (consistent with Henkel et al., 2011 and Dangl and Halling, 2012) but correspond to different phases of the recession. Regime II, with high dividend yields and low liquidity, often indicates the end of economic expansion and the early stages of a recession. During this period, stock returns exhibit higher predictability. In contrast, Regime I, where the dividend yield is lower than the previous average, shows reduced predictability, with $R^2$ decreasing slightly to 1.26% at node N2. Overall, Regime II consistently demonstrates superior return predictability compared to the other macroeconomic regimes.

Furthermore, subsequent partitions based on cross-sectional characteristics further uncover the heterogeneity in stock return predictability within each regime. For example, leaf node N96 (red labeled in Regime II) exhibits the highest $R^2$ of 35.59%, while the poorest cluster, node N39 (blue labeled in Regime I), shows the lowest value at -13.37%. This range is wider than the cross-sectional clustering results presented in Section 4, suggesting that incorporating time-series heterogeneity allows us to identify larger gaps of return predictability.

Figure 8: **Time-varying Predictability between Macroeconomic Regimes**

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

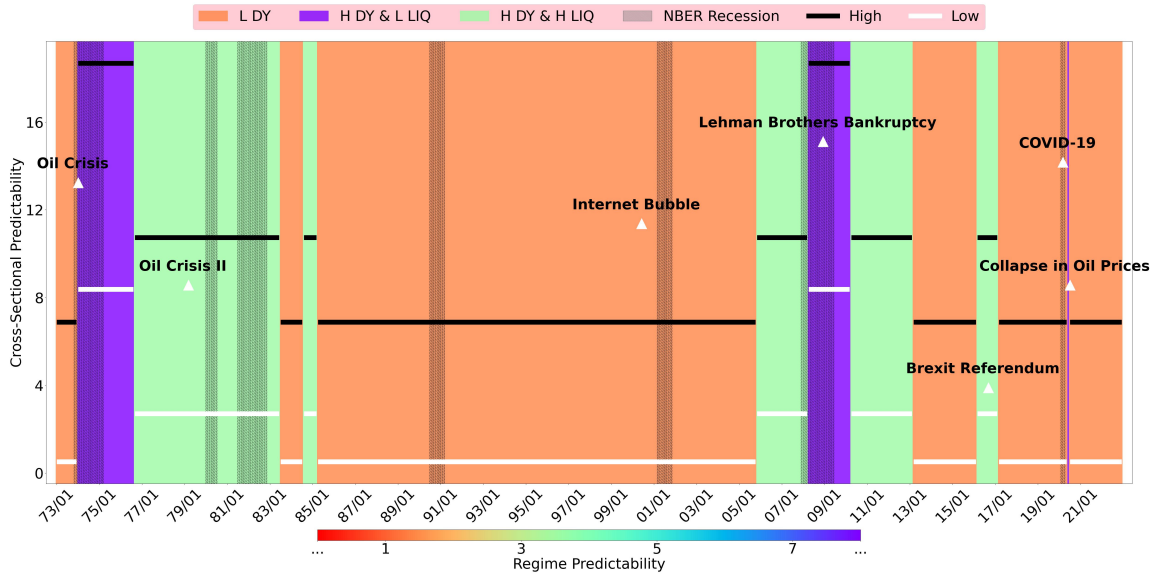
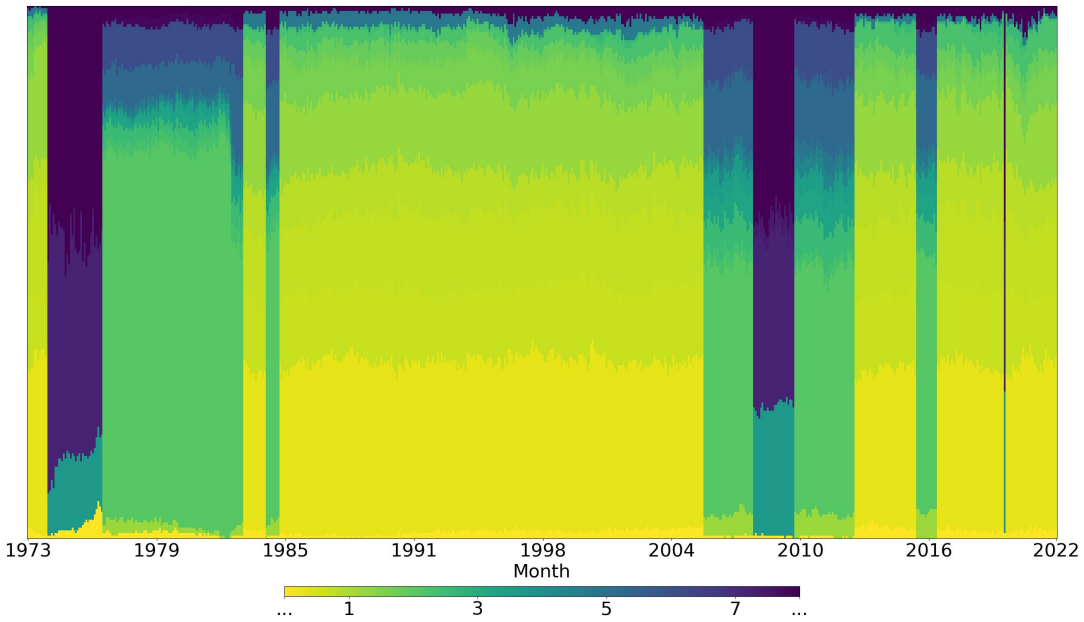


Figure 8 illustrates macroeconomic regime switches from another perspective. The color changes reflect regime shifts detected by the decision tree in Figure 7, with the colors indicating different levels of heterogeneity. The range is broader than in the cross-sectional case (from around 1% in orange to more than 7% in purple). Cross-sectional splits by firm characteristics further widen the intervals across sub-samples. Highly predictable clusters (black lines) consistently outperform less predictable ones (white lines) and are more robust before additional clustering. Regime transitions are linked to recessions (e.g., 1974, 2008, 2020) and global events like the Oil Crisis (1973), the Lehman Brothers Bankruptcy (2008), the Brexit Referendum (2016), and the 2020 Oil Price Collapse. Other events, such as the 1979 Oil Crisis and the 2000 Internet Bubble, likely triggered NBER recessions.

Figure 9: **Mosaics of Predictability by Clusters (TS+CS Cluster)**

This heat map depicts the predictability ($R^2$ values, % as shown in the color bar) of stock returns for the clusters identified via tree-based clustering in Figure 7. The horizontal axis represents months, with colors transitioning from light to dark, indicating increasing return predictability within each cluster. Vertically, the height of each color bin reflects the proportion of observations in each cluster.




Compared to the CS clusters, the mosaics of stock return predictability for time series in Figure 9 appear more diverse. The horizontal axis represents all 600 months. 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 show lower predictability in light green, while specific regimes (e.g., around 1975 and 2009) show extremely higher predictability for almost all clusters compared with other regimes. The frequent color shifts across clusters highlight noncontinuous macroeconomic regime variations.

**Cluster-Wise Performance.** Next, we analyze single-leaf statistics to evaluate predictive abilities. Incorporating time-series partitions, Table A.2 translates the tree structure in Figure 7 into descending rankings of predictability within each macroeconomic regime, highlighting heterogeneity.

The gaps between the highest and lowest $R^2$ values are more pronounced than in the cross-sectional case (Table 1). Examining the table longitudinally reveals substantial variability across regimes, each exhibiting a wide range of return predictability. Sorting $R^2$ values in descending order within specific time periods further illustrates the mosaic patterns of predictability across clusters. Unlike in the cross-sectional analysis, where slight positive correlations are observed between predictability and investment gains in certain time horizons, such as Regimes I and III, investment performance does not uniformly decline with lower predictability. Regime II, spanning only 57 months, represents the most predictable period but shows almost zero correlations between predictability and returns. This suggests significant opportunities to anticipate return patterns during this challenging financial landscape, warranting alternative strategies like long-short approaches.

Overall, these findings emphasize the importance of analyzing the diversity of return prediction accuracy at the time-series level.

**Aggregate Evaluations.** Following a similar approach to the cross-sectional analysis, we perform aggregate evaluations to assess improvements in heterogeneous return predictability after incorporating time-series information. We focus on a large-cap sub-sample analysis. Clusters are aggregated into several sub-samples, with Table 4 presenting the resulting predictability metrics.

When comparing performance without cross-sectional splits, most aggregated clustering results have higher prediction accuracy. Filtering out large-cap stocks fur-

Table 4: **Further Comparing Global and Cluster-Wise Predictive Models**

Similar to Table 2, this table reports return predictability ($R^2$, %) across predictive methods under macroeconomic regimes. For each regime, two panels are presented: global and cluster-wise predictions. Results include five samples: Global (homogeneous forecasts), Aggregate (cluster-aggregated predictions), and High, Medium, and Low, based on predictive rankings within tree clusters.

| | 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 | 1.05 | 0.29 | 0.90 | 0.99 | 0.42 | 0.90 |
| High | 4.49 | 2.76 | 4.30 | 3.54 | 2.65 | 3.58 |
| Medium | 1.17 | 0.33 | 1.01 | 1.02 | 0.45 | 0.93 |
| Low | 0.69 | 0.09 | 0.55 | 0.68 | 0.14 | 0.54 |
| *Panel B: Cluster-Wise Forecasts* | | | | | | |
| Aggregate | 1.08 | 0.56 | 1.01 | 0.96 | 0.48 | 0.93 |
| High | 6.75 | 5.82 | 6.66 | 3.72 | 3.57 | 3.87 |
| Medium | 1.56 | 0.90 | 1.45 | 1.27 | 0.67 | 1.20 |
| Low | 0.13 | -0.19 | 0.12 | -0.27 | -0.41 | -0.15 |
| | 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 | 9.24 | 4.22 | 8.77 | 13.01 | 6.57 | 12.67 |
| High | 12.17 | 5.53 | 11.59 | 14.67 | 7.03 | 14.31 |
| Medium | 9.72 | 4.41 | 9.25 | 11.26 | 6.11 | 10.95 |
| Low | 7.93 | 3.64 | 7.51 | 10.93 | 5.89 | 10.53 |
| *Panel B: Cluster-Wise Forecasts* | | | | | | |
| Aggregate | 11.40 | 9.10 | 9.39 | 15.32 | 11.31 | 12.92 |
| High | 18.36 | 13.93 | 15.88 | 18.11 | 12.31 | 15.06 |
| Medium | 12.43 | 11.07 | 10.33 | 12.62 | 10.27 | 11.02 |
| Low | 8.34 | 6.59 | 6.53 | 11.05 | 9.98 | 8.99 |
| | 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 | 2.65 | 1.37 | 2.47 | 3.06 | 1.85 | 2.88 |
| High | 9.36 | 6.11 | 9.10 | 6.88 | 6.09 | 7.14 |
| Medium | 4.49 | 2.85 | 4.23 | 3.50 | 2.38 | 3.34 |
| Low | 1.91 | 0.80 | 1.76 | 2.34 | 1.01 | 2.11 |
| *Panel B: Cluster-Wise Forecasts* | | | | | | |
| Aggregate | 3.19 | 2.16 | 3.02 | 3.50 | 2.40 | 3.33 |
| High | 10.44 | 9.60 | 10.46 | 7.87 | 7.10 | 8.10 |
| Medium | 5.25 | 3.75 | 4.98 | 4.37 | 2.75 | 4.05 |
| Low | 2.38 | 1.46 | 2.22 | 2.27 | 1.74 | 2.24 |

ther reveals specific predictability trends under varying market conditions.

Among the three regimes, Regime II ($\texttt{X_DY} > 0$ and $\texttt{X_LIQ} \leq 0$) consistently outperforms the others. This regime corresponds to periods characterized by high dividend yields and low liquidity, typically signaling the end of a business expan-

sion phase. Declining trends are evident in the last three rows within each period. The highly predictable sub-sample achieves the largest $R^2$ values, while the least predictable sub-sample underperforms all others, including the "Global" sample, without cross-sectional clustering.

Despite differences in magnitude, the patterns of improvement and decline are similar across methods, particularly for return predictions based on cluster-wise models. These results provide insights into the mosaics of return predictability and the heterogeneity of stocks under different characteristics and market conditions.

### 5.2 Structural Breaks by Calendar Months

Splitting by aggregate predictors generates non-continuous regimes on the time horizon. For robustness check, we also examine results from splitting by calendar months. One advantage of this approach is that calendar months increase over time, leading to single, continuous regimes that resemble structural breaks (e.g., Smith and Timmermann, 2021). This method continues to assess whether dividing continuous periods effectively captures the heterogeneity in stock return predictability.

Figure 10: **Time-varying Predictability over Structural Breaks**

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

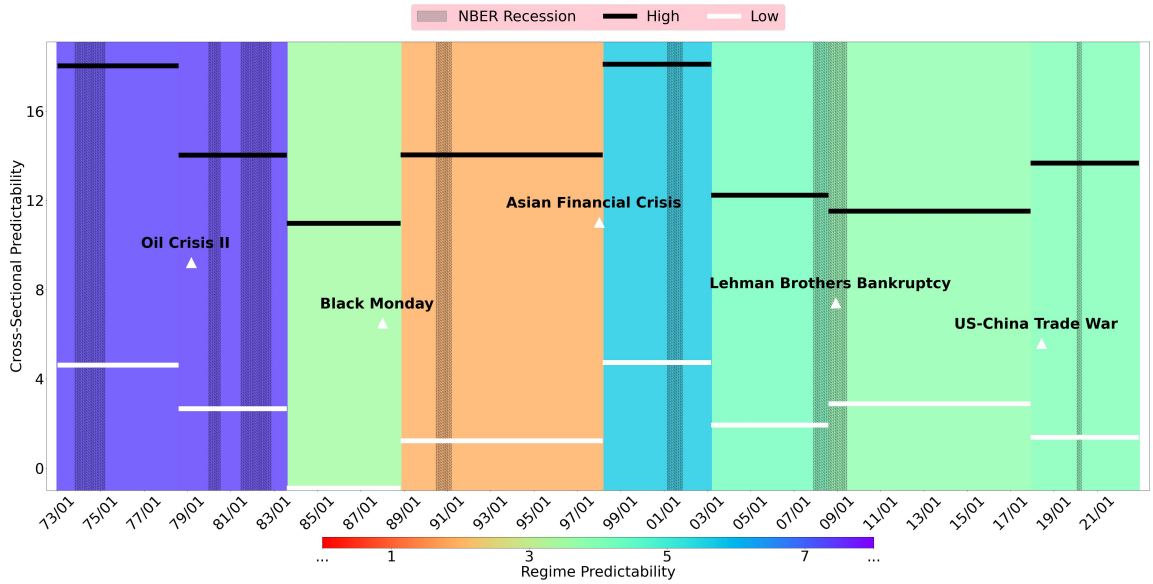


Figure 10 illustrates that stock return predictability varies across time horizons, although the variability is less than the version using aggregate predictors. For instance, predictability peaks at 7.39% from September 1978 to August 1983 (second purple) but falls to just 1.93% from December 1988 to March 1998 (orange). Moreover, certain events, such as Black Monday (1987) and the Asian Financial Crisis (1997–1998), likely triggered structural breaks to different levels of predictability. Conversely, structural breaks may also have contributed to other incidents, such as the Oil Crisis (1979), Lehman Brothers' bankruptcy (2008), and the US-China Trade War (2018).

Splitting further in the cross-section within each time-series period would likely increase these heterogeneity gaps. The predictive abilities and investment performance across regimes remain between relatively high (black) and low (white) predictable sub-samples, with trends showing some consistency across periods. These findings confirm the heterogeneous predictability of stock returns in the time-series dimension, highlighting their mosaic-like nature.

## 6 Investment Gains on Cluster-wise Models

The previous sections primarily focus on illustrating the heterogeneous nature of return predictability, both across the cross-section and over time, using our novel clustering approach. In doing so, we uncover an anomaly associated with return predictability. A key byproduct of our clustering analysis is a cluster-wise predictive model. Specifically, we partition assets into multiple clusters, each linked to its cluster-specific predictive model. Together, these cluster-wise models form a heterogeneous predictive model, unlike a global predictive model that employs a single, uniform approach to all asset returns.

In this section, we explore the investment gains derived from the cluster-wise predictive model by constructing a new type of forecast-implied portfolio, which evaluates predictive accuracy through portfolio returns. We construct three types of portfolio weights based on model predictions. For equal-weighted and value-weighted portfolios, the stock weights are determined by the absolute value of the respective

weighting schemes, but the sign is determined by the model's prediction. Additionally, we construct forecast-weighted portfolios based on the normalized predictions, controlling for leverage (e.g., Guijarro-Ordonez et al., 2021). In this framework, stocks that are predicted to have high returns will receive a larger long weight, while those with lower predicted negative returns will receive a larger short weight.

Specifically, let weights $w_{i,t}$ be absolute value of original equal-weighted ($1/N_t$) or value-weighted ($\texttt{size}_{i,t-1}/\sum_i \texttt{size}_{i,t-1}$) portfolios for each cross-sectional dimension at time $t$. The three portfolio weights are as follows:

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

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

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

It is important to note that these forecast-implied portfolios differ from cluster portfolios discussed in Section 4.2, where only value-weighted portfolios were created for each cluster to investigate the relationship between predictability and average returns. In contrast, the forecast-implied portfolios in this section are long-short, with the long and short positions determined by the model's predictions, and the absolute weights being either equal- or value-weighted. Intuitively, if the predictive model is more accurate, the resulting portfolio should generate greater investment gains.

Table 5 presents the results of direction-adjusted value- and equal-weighted portfolios, as well as forecast-weighted portfolios, constructed from all observations. These results are derived from the cross-sectional clustering model described in Section 4.

First, our results confirm that highly predictable stocks remain easier to predict out of the sample, and as a result, the corresponding trading strategies generate sig-

Table 5: **Forecast-Implied Investment Performance (CS Cluster)**

This table summarizes the in-sample and out-of-sample performance of forecast-implied portfolios (Equation 10). Panels A and B present sign-adjusted value- and equal-weighted portfolios, while Panel C reports forecast-weighted portfolios constructed using all observations (Equation 9). Results are shown for five samples: Global (no clustering), Aggregate (cluster aggregation), and High, Medium, and Low (predictive rankings within tree clusters). Each panel includes five metrics: monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), market 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 | 0.80 | 3.10 | 0.89 | $0.58^{***}$ | 19.15 | 0.68 | 4.01 | 0.58 | $-0.04$ | 15.99 |
| Aggregate | 0.69 | 3.45 | 0.70 | $0.41^{***}$ | 20.43 | 0.75 | 4.06 | 0.64 | 0.02 | 16.69 |
| High | 1.99 | 4.86 | 1.42 | $1.65^{***}$ | 19.60 | 1.36 | 5.62 | 0.84 | $0.56^{**}$ | 24.35 |
| Medium | 0.68 | 3.50 | 0.67 | $0.39^{***}$ | 20.47 | 0.83 | 4.00 | 0.72 | $0.12^{**}$ | 15.75 |
| Low | 0.55 | 3.65 | 0.52 | $0.39^{**}$ | 18.96 | 0.58 | 4.24 | 0.47 | $-0.16^{*}$ | 17.77 |
| | Panel B: Sign-adjusted Equal-Weighted | | | | | | | | | |
| | Avg | Std | SR | Alpha | MDD | Avg | Std | SR | Alpha | MDD |
| Global | 1.39 | 2.99 | 1.62 | $1.23^{***}$ | 11.70 | 0.98 | 4.55 | 0.75 | $0.25^{*}$ | 17.69 |
| Aggregate | 1.41 | 2.45 | 1.99 | $1.25^{***}$ | 8.86 | 1.05 | 3.92 | 0.93 | $0.43^{***}$ | 18.37 |
| High | 3.84 | 6.36 | 2.09 | $3.44^{***}$ | 25.12 | 2.89 | 5.61 | 1.79 | $2.08^{***}$ | 23.06 |
| Medium | 1.38 | 2.45 | 1.95 | $1.24^{***}$ | 8.61 | 1.06 | 3.64 | 1.01 | $0.56^{***}$ | 17.70 |
| Low | 0.82 | 2.66 | 1.07 | $0.67^{***}$ | 9.49 | 0.87 | 5.05 | 0.59 | 0.02 | 20.00 |
| | Panel C: Forecast-Weighted | | | | | | | | | |
| | Avg | Std | SR | Alpha | MDD | Avg | Std | SR | Alpha | MDD |
| Global | 2.06 | 3.94 | 1.82 | $1.86^{***}$ | 16.02 | 1.27 | 5.12 | 0.86 | $0.43^{***}$ | 21.82 |
| Aggregate | 2.85 | 3.71 | 2.66 | $2.68^{***}$ | 11.99 | 1.84 | 4.51 | 1.42 | $1.19^{***}$ | 19.64 |
| High | 4.87 | 7.19 | 2.35 | $4.43^{***}$ | 25.91 | 3.60 | 6.34 | 1.97 | $2.71^{***}$ | 24.14 |
| Medium | 2.66 | 3.72 | 2.48 | $2.53^{***}$ | 19.56 | 1.69 | 4.36 | 1.35 | $1.14^{***}$ | 19.90 |
| Low | 1.42 | 3.57 | 1.38 | $1.24^{***}$ | 14.05 | 1.32 | 5.57 | 0.82 | $0.41^{**}$ | 21.46 |

nificant investment gains. The sign-adjusted value-weighted portfolio also performs well, indicating that the results are not dominated by small stocks. For example, in Panel A of Table 5, the highly predictable sub-sample shows an OOS Sharpe ratio of 0.84 and an average return of 1.36%. In contrast, clusters with the lowest predictability achieve only about half of these values, with a Sharpe ratio of 0.47 and an average return of 0.58%. Similarly, in Panels B and C, highly predictable stocks generate Sharpe ratios of 1.79 and 1.97, and averages of 2.89% and 3.60%, respectively. These figures significantly outperform the portfolio of medium- and low-predictability stocks. Sim-

ilar patterns are observed for market alphas and maximum drawdowns.

Second, the cluster-wise predictive models ("Aggregate") have better performance compared to the homogeneous predictive model ("Global"), which trains on all stocks without accounting for clustering patterns. For example, in Panel C, the OOS average return for the Aggregate model is 1.84%, higher than the 1.27% achieved by the Global model, with Sharpe ratios of 1.42 and 0.86, respectively. The gap is smaller in Panels A and B, primarily because the absolute weights of the two portfolios are identical, with only the sign adjusted by the predictive model.

Overall, while both Aggregate and Global models perform similarly in predicting whether a stock price will go up or down in the next month (Panel A and B), the Aggregate model demonstrates superior accuracy in forecasting *how much* the price will go up or down (Panel C). Furthermore, the time-series and cross-sectional clustering models exhibit similar economic gains for the heterogeneous predictive models. For more details, see Appendix II..4.

All evaluations confirm that highly predictable stocks, identified through our cross-sectional tree clusters, significantly dominate the economic profitability of the entire sample. Excluding these highly predictable stocks from the sample markedly reduces potential gains for investors. Extending the findings from the stock return predictability mosaics to the contributions of investment strategies, it is clear that investors can achieve higher economic benefits by allocating funds to stocks with high earnings surprises, high valuations, or low trading volumes. Furthermore, the aggregate performance of the cluster-wise predictive model consistently delivers superior forecasts compared to the Global homogeneous model.

## 7 Conclusion

Existing studies often treat return predictability as a feature of the predictors or models employed, evaluating it at an aggregate level and applying it uniformly to assets. This paper offers a novel perspective, casting return predictability as an inherent yet unobserved characteristic of assets at particular points in time, and potentially

linked to the cross-section of expected returns. Specifically, we demonstrate that return predictability can vary significantly across assets and over time. More importantly, we show that asset characteristics and lagged macroeconomic variables can generate asset groups through a data-driven yet goal-oriented clustering algorithm to maximally separate assets that are highly predictable in returns from the ones that are less predictable. The resulting "mosaics of predictability" essentially tell us what assets are more predictable under what macroeconomic regimes.

To this end, we introduce a novel tree-based clustering method that groups asset-return observations based on similar levels of predictability that $R^2$ intuitively captures. Our empirical analysis reveals substantial heterogeneity in return predictability among individual U.S. stocks, advancing our understanding of asset return predictability in several key dimensions. First, we find that asset clusters characterized by high unexpected earnings, high earnings-to-price ratios, and low trading volumes are the most predictable. Second, we observe that return predictability declines sharply when the dividend yield is low, but peaks during periods of high dividend yield and low liquidity. Finally, we identify a new predictability-implied anomaly: long-short portfolios based on predictability differences generate significantly unexplained alphas of over 1% across various risk factor models over the OOS period of two decades. Moreover, high cluster-specific risk long-only portfolios achieve even greater abnormal returns.

## References

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

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

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

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*(2), 715–754.

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

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *Review of Financial Studies 34*(2), 1046–1089.

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

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

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.

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. *Working Paper*.

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.

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

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies 33*(4), 1673–1736.

DeMiguel, V., J. Gil-Bazo, F. J. Nogales, and A. A. Santos (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics 150*(3), 103737.

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

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

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

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

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

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

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

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

Feng, G., X. He, Y. Wang, and C. Wu (2025). Predicting individual corporate bond returns. *Journal of Banking and Finance 171*, 107372.

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 (2021). Deep learning statistical arbitrage. Technical report, Stanford University.

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.

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

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

Hou, K., 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.

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

Jensen, M. C., F. Black, and M. S. Scholes (1972). The capital asset pricing model: Some empirical tests.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023). Machine-learning the skill of mutual fund managers. *Journal of Financial Economics 150*(1), 94–138.

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

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

Kelly, B. and S. Pruitt (2013). Market expectations in the cross-section of present values. *Journal of Finance 68*(5), 1721–1756.

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

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

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political economy 111*(3), 642–685.

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

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

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

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: What is the role of the united states? *Journal of Finance 68*(4), 1633–1662.

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

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

Smith, S. C. and A. Timmermann (2022). Have risk premia vanished? *Journal of Financial Economics 145*(2), 553–576.

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

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

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

# Appendix

## I. Panel Regression Tree Algorithm

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

**Algorithm Panel Regression Tree**

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

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

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

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

# II. Additional Empirical Results

## II..1 More results for CS cluster

This section presents additional results related to cross-sectional clustering. Table A.1 reports the performance of three specific long- and short-leg portfolios associated with the "long-short cluster" anomalies detailed in Table 3. Notably, the long-only portfolios comprising highly predictable stocks significantly drive the abnormal returns observed in the long-short strategies.

Table A.1: **Predictability-Implied Anomaly Testing**

This table reports summary statistics (Panel A) and abnormal returns (Panel B) for the long- and short-leg portfolios corresponding to the long-short strategies in Table 3. The numbers following "T" and "B" denote the number of top and bottom clusters, ranked by $R^2_{\text{CMG}}$, 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-values (in parentheses) from various factor models.

| | In-Sample (1973 - 2002) | | | | | | Out-of-Sample (2003 - 2022) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | B1 | B2 | B3 | T1 | T2 | T3 | B1 | B2 | B3 |
| | | | | | | Panel A: Summary Statistics | | | | | | |
| Avg | 3.65 | 2.93 | 1.60 | 0.44 | -0.06 | 0.33 | 3.54 | 2.65 | 1.89 | 0.63 | 1.23 | 1.29 |
| Std | 6.57 | 6.96 | 7.42 | 5.75 | 6.15 | 6.38 | 6.92 | 7.10 | 7.54 | 5.34 | 6.29 | 5.97 |
| SR | 1.92 | 1.46 | 0.75 | 0.27 | -0.03 | 0.18 | 1.77 | 1.29 | 0.87 | 0.41 | 0.68 | 0.75 |
| | | | | | | Panel B: Abnormal Returns | | | | | | |
| CAPM | 3.29*** | 2.47*** | 1.05*** | 0.06 | -0.51*** | -0.12 | 2.74*** | 1.62*** | 0.76*** | -0.21 | 0.29 | 0.34* |
| | (12.07) | (10.06) | (4.95) | (0.32) | (-2.87) | (-0.62) | (7.83) | (5.73) | (2.71) | (-1.25) | (1.25) | (1.83) |
| FF3 | 2.77*** | 2.09*** | 0.95*** | -0.31** | -0.66*** | -0.42*** | 2.80*** | 1.67*** | 0.81*** | -0.18 | 0.31 | 0.36** |
| | (12.50) | (10.65) | (5.72) | (-2.10) | (-4.27) | (-2.68) | (8.74) | (7.06) | (3.81) | (-1.27) | (1.49) | (2.29) |
| FF5 | 2.61*** | 2.10*** | 1.15*** | -0.31** | -0.56*** | -0.33** | 2.93*** | 1.78*** | 0.99*** | -0.16 | 0.39* | 0.42*** |
| | (11.77) | (10.33) | (7.01) | (-2.10) | (-3.58) | (-2.08) | (8.91) | (7.45) | (4.66 ) | (-1.08) | (1.82) | (2.67) |
| FF5+MOM+IVOL | 2.81*** | 2.36*** | 1.46*** | -0.11 | -0.37** | -0.08 | 3.07*** | 1.89*** | 1.11*** | -0.17 | 0.43** | 0.45*** |
| | (12.53) | (11.80) | (9.54) | (-0.78) | (-2.36) | (-0.54) | (10.23 ) | (8.80) | (6.28) | (-1.14) | (2.03) | (2.94) |
| Q5 | 2.79*** | 2.48*** | 1.53*** | -0.04 | -0.22 | -0.07 | 3.15*** | 1.93*** | 1.11*** | -0.31** | 0.32 | 0.33** |
| | (10.08) | (10.32) | (8.09) | (-0.22) | (-1.24) | (-0.41) | (9.78) | (8.33) | (5.63) | (-2.03) | (1.49) | (2.05) |
| BS6 | 2.44*** | 2.14*** | 1.35*** | -0.32** | -0.40** | -0.18 | 3.00*** | 1.85*** | 1.11*** | -0.22 | 0.45** | 0.43*** |
| | (10.35) | (10.28) | (7.85) | (-2.10) | (-2.46) | (-1.14) | (9.99) | (8.39) | (6.23) | (-1.51) | (2.15) | (2.73) |
| DHS3 | 3.56*** | 3.00*** | 1.85*** | 0.32 | -0.03 | 0.26 | 3.01*** | 1.95*** | 1.15*** | -0.19 | 0.49** | 0.51*** |
| | (11.36) | (10.63) | (8.17) | (1.39) | (-0.14) | (1.17) | (8.61) | (7.08) | (4.31) | (-1.07) | (2.09) | (2.82) |
| SY4 | 2.72*** | 2.35*** | 1.47*** | -0.15 | -0.38** | -0.22 | 3.60*** | 2.28*** | 1.32*** | -0.17 | 0.38 | 0.46** |
| | (10.79) | (10.72) | (8.71) | (-0.91) | (-2.31) | (-1.27) | (9.10) | (8.04) | (5.62) | (-1.09) | (1.62) | (2.50) |

## II..2 More results for TS+CS cluster

Figure A.1 separately exhibits the further cross-sectional splits under each regime determined by tree structure in Figure 7. Table A.2 summarizes the cluster-wise information under each regime.

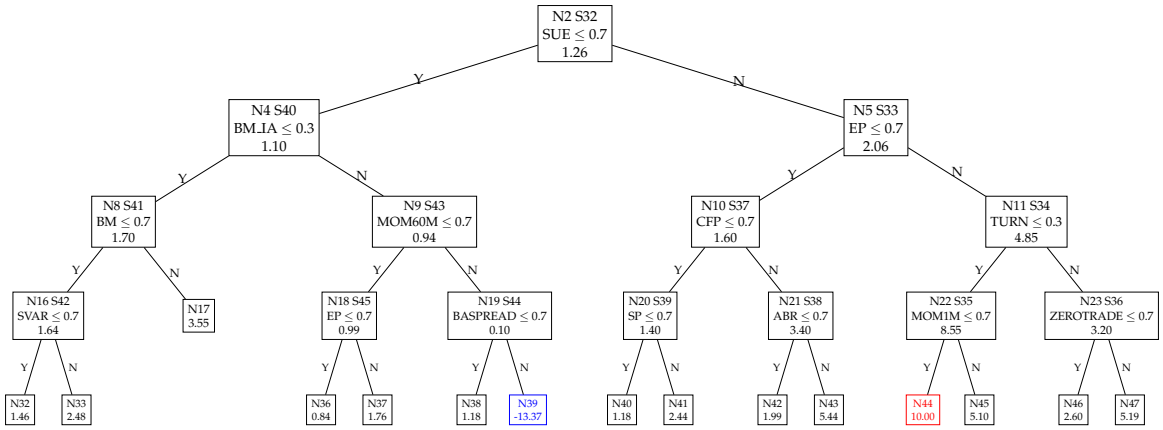
Table A.2: **Cluster-Wise Performance (TS+CS Cluster)**

This table summarizes the cluster-based information under each regime within three panels corresponding to the cross-sectional and time-series tree structure in Figure 7 and A.1. 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.

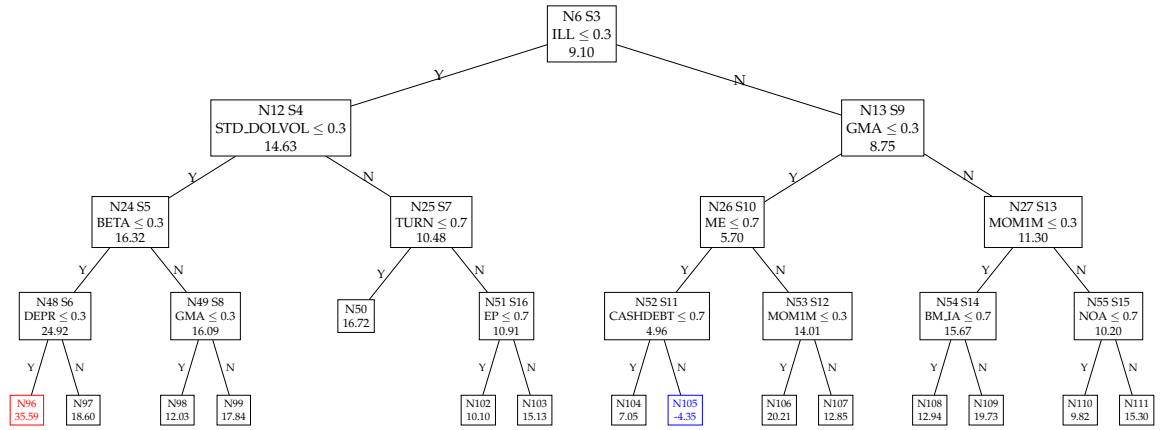
| | | | Regime I: $\mathbb{1}\{\texttt{X_DY} \leq 0\}$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Leaf | N44 | N43 | N47 | N45 | N17 | N46 | N33 | N41 | N42 | N37 | N32 | N38 | N40 | N36 | N39 |
| # obs | 36,565 | 12,440 | 27,349 | 12,709 | 13,766 | 56,720 | 105,029 | 49,601 | 21,191 | 247,000 | 246,605 | 140,208 | 243,811 | 555,839 | 27,364 |
| $R^2$ | 10.00 | 5.44 | 5.19 | 5.10 | 3.55 | 2.60 | 2.48 | 2.44 | 1.99 | 1.76 | 1.46 | 1.18 | 1.18 | 0.84 | -13.37 |
| $\text{Avg}_{\text{EW}}$ | 2.75 | 2.95 | 1.93 | 1.69 | -0.85 | 1.47 | -1.93 | 2.31 | 2.09 | 0.67 | -0.01 | 0.42 | 0.96 | -0.05 | -0.22 |
| $\text{SR}_{\text{EW}}$ | 2.14 | 1.52 | 1.30 | 1.13 | -0.41 | 0.83 | -0.80 | 1.11 | 1.23 | 0.46 | -0.01 | 0.30 | 0.55 | -0.03 | -0.11 |
| $\text{Avg}_{\text{VW}}$ | 1.59 | 1.17 | 0.91 | 0.78 | -0.24 | 0.94 | -1.47 | 1.06 | 0.89 | 0.79 | 0.37 | 0.63 | 0.70 | 0.41 | -0.46 |
| $\text{SR}_{\text{VW}}$ | 1.22 | 0.64 | 0.62 | 0.49 | -0.11 | 0.59 | -0.52 | 0.64 | 0.57 | 0.61 | 0.28 | 0.46 | 0.49 | 0.29 | -0.20 |
| | | | Regime II: $\mathbb{1}\{\texttt{X_DY} > 0\}\mathbb{1}\{\texttt{X_LIQ} \leq 0\}$ | | | | | | | | | | | | |
| Leaf | N96 | N106 | N109 | N97 | N99 | N50 | N111 | N103 | N108 | N107 | N98 | N102 | N110 | N104 | N105 |
| # 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 |
| $R^2$ | 35.59 | 20.21 | 19.73 | 18.60 | 17.84 | 16.72 | 15.30 | 15.13 | 12.94 | 12.85 | 12.03 | 10.10 | 9.82 | 7.05 | -4.35 |
| $\text{Avg}_{\text{EW}}$ | 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 |
| $\text{SR}_{\text{EW}}$ | 0.00 | 0.17 | 1.36 | -0.24 | 0.28 | 0.13 | 0.52 | 0.32 | 0.71 | 0.03 | 0.20 | 0.06 | 0.34 | 0.43 | 0.76 |
| $\text{Avg}_{\text{VW}}$ | 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 |
| $\text{SR}_{\text{VW}}$ | 0.13 | 0.11 | 0.97 | -0.14 | 0.13 | 0.09 | 0.39 | 0.33 | 0.50 | -0.05 | -0.07 | -0.11 | 0.18 | 0.30 | 0.40 |
| | | | Regime III: $\mathbb{1}\{\texttt{X_DY} > 0\}\mathbb{1}\{\texttt{X_LIQ} > 0\}$ | | | | | | | | | | | | |
| Leaf | N127 | N126 | N124 | N117 | N123 | N116 | N115 | N121 | N125 | N118 | N122 | N114 | N112 | N119 | N120 | N113 |
| # obs | 5,791 | 12,867 | 11,300 | 4,982 | 54,211 | 5,727 | 65,179 | 8,479 | 12,100 | 16,401 | 26,028 | 29,114 | 14,666 | 370,807 | 19,795 | 5,431 |
| $R^2$ | 15.08 | 9.38 | 8.60 | 7.11 | 6.97 | 6.07 | 5.81 | 5.76 | 5.27 | 4.33 | 4.00 | 3.18 | 3.07 | 2.82 | 1.36 | -0.33 |
| $\text{Avg}_{\text{EW}}$ | 4.43 | 2.54 | 2.68 | 1.31 | 2.64 | 0.83 | 1.15 | 3.25 | 1.57 | 1.43 | 1.33 | 0.48 | 0.53 | 0.92 | 1.40 | 0.33 |
| $\text{SR}_{\text{EW}}$ | 2.66 | 1.92 | 1.66 | 0.73 | 1.55 | 0.49 | 1.11 | 1.56 | 1.04 | 0.92 | 0.85 | 0.47 | 0.49 | 0.54 | 0.77 | 0.27 |
| $\text{Avg}_{\text{VW}}$ | 2.41 | 1.73 | 2.02 | 0.91 | 1.94 | 0.61 | 0.65 | 1.85 | 1.43 | 0.96 | 0.94 | 0.41 | 0.55 | 0.71 | 0.88 | 0.46 |
| $\text{SR}_{\text{VW}}$ | 1.55 | 1.35 | 1.36 | 0.52 | 1.23 | 0.36 | 0.65 | 0.88 | 1.06 | 0.56 | 0.71 | 0.39 | 0.44 | 0.43 | 0.50 | 0.36 |

Figure A.1: **Tree-Based Cluster (CS Clusters under Regimes)**

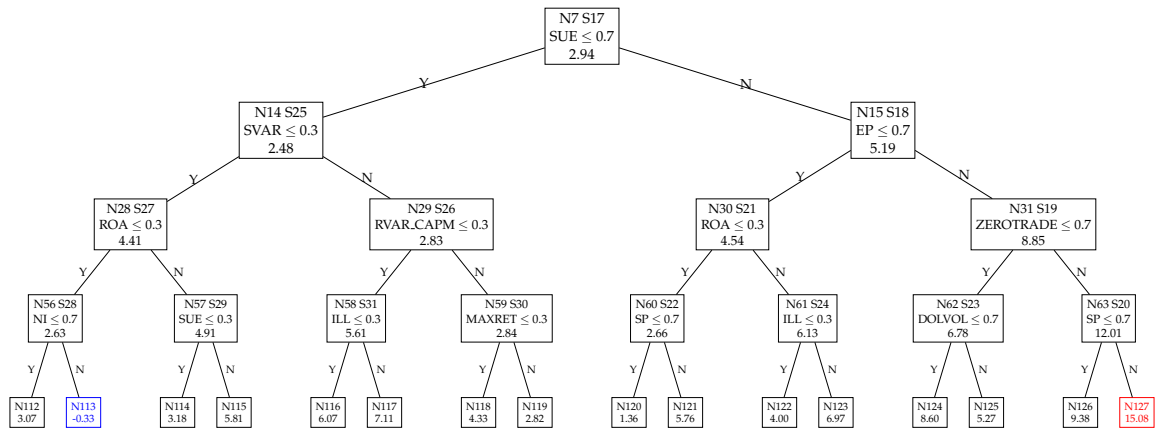
This figure shows the further cross-sectional split tree-based clustering structure under each regime determined by Figure 7 using monthly data from 1973 to 2022. Each tree splits into monthly cross-sectional standardization of firm characteristic ranks in [0,1]. The terminal leaves correspond to clusters identified by the interaction of aggregate predictors and firm characteristics ranges. Each node, including bottom leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. All nodes are labeled with cluster-wise model $R^2$.




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




(b) Regime II: $\mathbb{1}\{\texttt{X_DY} > 0\}\mathbb{1}\{\texttt{X_LIQ} \leq 0\}$




(c) Regime III: $\mathbb{1}\{\texttt{X_DY} > 0\}\mathbb{1}\{\texttt{X_LIQ} > 0\}$

## II..3 Structural Break

Table A.3: **Cluster-Wise Performance (+ Structural Break)**

This table shows the performance of the tree-based cluster model using both cross-sectional and time-series splits. In-sample return predictability ($R^2$, %) is calculated using Equation (6). "# obs" indicates the stock returns count for each cluster. "Avg" and "SR" denote the monthly average return (in %) and annualized Sharpe ratio for cluster-wise equal-weighted and value-weighted forecast-implied portfolios. Regimes are ordered by descending $R^2$ from left to right.

| 197301-197808 | N21 | N6 | N30 | N19 | N23 | N20 | N31 | N29 | N18 | N16 | N22 | N28 | N17 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # obs | 2,247 | 2,739 | 2,095 | 6,996 | 6,179 | 7,484 | 3,873 | 11,998 | 37,929 | 8,646 | 24,518 | 5,056 | 126,345 | | | |
| $R^2$ | 23.69 | 21.18 | 19.60 | 17.25 | 12.79 | 12.56 | 10.84 | 10.26 | 7.96 | 7.67 | 7.15 | 6.36 | 4.81 | | | |
| $Avg_{EW}$ | 6.64 | 2.54 | 1.81 | 0.98 | 2.39 | 2.66 | 1.00 | 0.83 | 1.17 | -0.17 | 1.02 | 1.17 | 0.57 | | | |
| $SR_{EW}$ | 2.02 | 0.63 | 0.64 | 0.31 | 1.15 | 1.11 | 0.38 | 0.31 | 0.56 | -0.08 | 0.55 | 0.43 | 0.29 | | | |
| $Avg_{VW}$ | 3.39 | 2.27 | 1.43 | 0.07 | 1.31 | 0.89 | 0.89 | 0.50 | -0.09 | -0.46 | 0.07 | 0.92 | -0.38 | | | |
| $SR_{VW}$ | 1.21 | 0.57 | 0.53 | 0.02 | 0.67 | 0.44 | 0.35 | 0.21 | -0.05 | -0.22 | 0.04 | 0.39 | -0.26 | | | |
| 197809-198308 | N31 | N29 | N30 | N18 | N21 | N22 | N13 | N19 | N28 | N16 | N23 | N20 | N17 | N12 | | |
| # obs | 6,111 | 2,923 | 24,520 | 8,042 | 2,032 | 33,859 | 3,026 | 35,040 | 4,240 | 8,998 | 7,012 | 6,700 | 94,653 | 2,642 | | |
| $R^2$ | 18.51 | 16.56 | 13.62 | 13.18 | 10.88 | 10.07 | 8.89 | 8.49 | 8.43 | 8.32 | 7.61 | 6.03 | 5.07 | 1.17 | | |
| $Avg_{EW}$ | 3.44 | 1.43 | 2.23 | 1.32 | -0.10 | 0.88 | 3.00 | 1.25 | 1.97 | 0.42 | 0.99 | 0.60 | 0.95 | 1.12 | | |
| $SR_{EW}$ | 1.69 | 0.85 | 1.09 | 0.89 | -0.04 | 0.44 | 1.45 | 0.68 | 0.92 | 0.18 | 0.56 | 0.30 | 0.54 | 0.62 | | |
| $Avg_{VW}$ | 1.40 | 0.63 | 1.04 | 0.74 | -0.03 | 0.14 | 1.07 | 0.88 | 0.96 | 0.26 | 0.78 | 0.54 | 0.22 | 0.76 | | |
| $SR_{VW}$ | 0.76 | 0.37 | 0.60 | 0.53 | -0.01 | 0.07 | 0.54 | 0.54 | 0.48 | 0.13 | 0.40 | 0.25 | 0.19 | 0.36 | | |
| 198309-198811 | N25 | N27 | N7 | N18 | N24 | N26 | N19 | N8 | N23 | N20 | N22 | N21 | | | | |
| # obs | 3,679 | 2,191 | 2,360 | 3,634 | 11,060 | 43,642 | 54,895 | 2,002 | 154,037 | 14,710 | 1,977 | 5,205 | | | | |
| $R^2$ | 14.17 | 10.84 | 10.40 | 9.46 | 8.85 | 6.77 | 6.65 | 3.68 | 2.81 | 2.45 | -1.47 | -53.36 | | | | |
| $Avg_{EW}$ | 0.49 | -1.34 | -1.37 | 0.95 | 0.73 | 0.49 | -0.94 | -0.16 | -0.20 | -0.21 | 0.29 | 0.56 | | | | |
| $SR_{EW}$ | 0.23 | -0.60 | -0.62 | 0.54 | 0.37 | 0.30 | -0.45 | -0.08 | -0.13 | -0.13 | 0.24 | 0.39 | | | | |
| $Avg_{VW}$ | 1.03 | -0.36 | -0.60 | 1.02 | 0.94 | 0.58 | -0.41 | -0.06 | 0.45 | 0.29 | 0.79 | 1.04 | | | | |
| $SR_{VW}$ | 0.52 | -0.19 | -0.30 | 0.58 | 0.54 | 0.40 | -0.20 | -0.03 | 0.31 | 0.19 | 0.53 | 0.72 | | | | |
| 198812-199803 | N28 | N24 | N31 | N25 | N19 | N29 | N11 | N30 | N21 | N26 | N20 | N17 | N18 | N27 | N16 | |
| # obs | 11,000 | 6,955 | 12,471 | 5,658 | 7,909 | 3,607 | 5,977 | 14,504 | 34,494 | 107,970 | 55,916 | 76,518 | 31,570 | 23,626 | 221,638 | |
| $R^2$ | 14.36 | 13.52 | 8.63 | 6.32 | 5.90 | 5.64 | 5.23 | 5.16 | 4.57 | 3.26 | 2.41 | 2.05 | 1.67 | 0.97 | 0.88 | |
| $Avg_{EW}$ | 3.69 | 1.77 | 2.94 | 1.60 | 3.63 | 2.58 | -0.42 | 2.04 | 1.12 | 1.08 | 0.81 | 2.08 | 0.94 | 0.60 | -0.05 | |
| $SR_{EW}$ | 3.51 | 1.29 | 2.51 | 0.86 | 2.80 | 1.93 | -0.18 | 1.35 | 1.10 | 1.03 | 0.70 | 1.41 | 0.77 | 0.54 | -0.04 | |
| $Avg_{VW}$ | 2.24 | 1.39 | 1.54 | 1.71 | 2.52 | 0.84 | -0.07 | 1.34 | 1.18 | 1.11 | 0.99 | 1.55 | 0.74 | 0.19 | 0.24 | |
| $SR_{VW}$ | 2.15 | 0.83 | 1.33 | 0.78 | 1.78 | 0.66 | -0.03 | 0.97 | 1.09 | 1.15 | 0.95 | 0.93 | 0.60 | 0.12 | 0.17 | |
| 199804-200303 | N29 | N12 | N17 | N30 | N13 | N10 | N28 | N11 | N18 | N31 | N16 | N19 | | | | |
| # obs | 3,062 | 1,969 | 2,947 | 2,855 | 2,411 | 1,860 | 9,163 | 2,520 | 14,130 | 86,207 | 34,315 | 190,795 | | | | |
| $R^2$ | 20.44 | 18.85 | 12.11 | 11.46 | 10.80 | 10.65 | 9.47 | 6.66 | 6.63 | 6.28 | 5.74 | 2.68 | | | | |
| $Avg_{EW}$ | -1.39 | -1.44 | 0.48 | -0.82 | -0.87 | -0.63 | -1.91 | 0.06 | -0.30 | -0.46 | -0.19 | 0.25 | | | | |
| $SR_{EW}$ | -0.26 | -0.30 | 0.14 | -0.17 | -0.20 | -0.19 | -0.40 | 0.02 | -0.19 | -0.13 | -0.08 | 0.18 | | | | |
| $Avg_{VW}$ | -4.61 | -3.25 | -0.97 | -0.67 | -1.41 | -0.91 | -1.78 | 1.67 | -0.76 | -1.41 | -0.87 | -0.45 | | | | |
| $SR_{VW}$ | -0.91 | -0.66 | -0.29 | -0.14 | -0.28 | -0.23 | -0.36 | 0.42 | -0.33 | -0.35 | -0.32 | -0.32 | | | | |
| 200304-200808 | N29 | N26 | N19 | N28 | N31 | N25 | N27 | N22 | N30 | N18 | N17 | N21 | N23 | N24 | N20 | N16 |
| # obs | 4,404 | 2,278 | 2,047 | 6,838 | 2,942 | 7,912 | 6,789 | 7,249 | 5,242 | 2,461 | 44,109 | 16,630 | 48,209 | 44,048 | 85,711 | 8,801 |
| $R^2$ | 15.13 | 14.78 | 13.35 | 10.57 | 10.55 | 7.75 | 7.69 | 7.64 | 6.00 | 5.99 | 5.47 | 4.28 | 3.92 | 3.46 | 2.19 | 1.97 |
| $Avg_{EW}$ | 4.19 | 1.84 | 0.28 | 3.42 | 2.46 | 2.87 | 1.19 | 1.34 | 1.41 | 1.38 | 0.75 | 0.35 | 1.02 | 1.63 | 0.44 | 0.52 |
| $SR_{EW}$ | 3.11 | 1.81 | 0.14 | 2.85 | 1.74 | 1.69 | 1.18 | 1.05 | 1.23 | 0.90 | 0.41 | 0.37 | 0.93 | 1.19 | 0.36 | 0.53 |
| $Avg_{VW}$ | 2.92 | 1.02 | 1.16 | 2.82 | 2.08 | 1.53 | 0.95 | 0.79 | 0.96 | 0.64 | 0.63 | 0.32 | 1.16 | 0.54 | 0.46 | 0.73 |
| $SR_{VW}$ | 2.07 | 0.90 | 0.48 | 2.45 | 1.55 | 1.04 | 1.02 | 0.67 | 0.91 | 0.32 | 0.44 | 0.32 | 1.20 | 0.57 | 0.56 | 0.80 |
| 200809-201712 | N11 | N8 | N29 | N10 | N13 | N25 | N19 | N28 | N30 | N24 | N31 | N18 | | | | |
| # obs | 4,342 | 4,449 | 4,146 | 4,276 | 3,841 | 5,312 | 107,760 | 15,850 | 69,905 | 6,492 | 184,309 | 3,438 | | | | |
| $R^2$ | 12.93 | 11.21 | 9.76 | 9.69 | 9.55 | 8.53 | 5.73 | 5.28 | 4.90 | 4.62 | 2.89 | 0.94 | | | | |
| $Avg_{EW}$ | 1.14 | 1.27 | 1.05 | 1.50 | 1.51 | 1.36 | 1.12 | 1.17 | 0.89 | 1.14 | 0.99 | 1.00 | | | | |
| $SR_{EW}$ | 0.85 | 1.48 | 0.60 | 1.25 | 0.93 | 1.27 | 0.84 | 0.65 | 0.56 | 0.88 | 0.47 | 0.79 | | | | |
| $Avg_{VW}$ | 1.01 | 1.29 | 0.59 | 1.36 | 0.71 | 1.35 | 0.96 | 1.00 | 0.86 | 0.96 | 1.05 | 0.89 | | | | |
| $SR_{VW}$ | 0.75 | 1.56 | 0.27 | 1.07 | 0.35 | 0.96 | 0.83 | 0.55 | 0.46 | 0.71 | 0.53 | 0.62 | | | | |
| 201801-202212 | N24 | N20 | N27 | N25 | N14 | N22 | N21 | N30 | N26 | N19 | N23 | N17 | N31 | N18 | N16 | |
| # obs | 2,104 | 1,967 | 4,899 | 7,088 | 2,691 | 1,946 | 5,132 | 15,587 | 16,204 | 27,542 | 23,496 | 20,680 | 12,655 | 66,260 | 5,323 | |
| $R^2$ | 27.47 | 14.20 | 13.40 | 12.58 | 10.74 | 9.42 | 8.62 | 7.93 | 7.70 | 5.93 | 5.58 | 4.25 | 4.21 | 2.56 | -19.71 | |
| $Avg_{EW}$ | 0.94 | 0.89 | 1.34 | 0.78 | 0.47 | 0.87 | 0.56 | 0.69 | 0.88 | 0.75 | 0.71 | -0.02 | 0.70 | -0.32 | -0.38 | |
| $SR_{EW}$ | 0.46 | 0.38 | 0.51 | 0.38 | 0.26 | 0.47 | 0.24 | 0.34 | 0.51 | 0.29 | 0.34 | -0.01 | 0.32 | -0.15 | -0.19 | |
| $Avg_{VW}$ | 0.89 | 0.73 | 0.71 | 0.67 | 0.85 | 1.81 | -0.15 | 0.65 | 0.98 | 0.50 | 0.86 | 0.83 | 0.63 | 0.08 | -0.35 | |
| $SR_{VW}$ | 0.42 | 0.35 | 0.34 | 0.33 | 0.38 | 0.81 | -0.07 | 0.41 | 0.67 | 0.20 | 0.44 | 0.37 | 0.36 | 0.04 | -0.21 | |

Table A.4: **Evaluating Return Predictability (+ Structural Break)**

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

| | Sample A: All Stocks | | | Sample B: Large-Cap | | |
|---|---|---|---|---|---|---|
| 197301-197808 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 7.82 | 6.63 | 7.36 | 6.62 | 5.58 | 6.28 |
| High | 19.32 | 16.21 | 18.04 | 11.33 | 11.35 | 11.62 |
| Medium | 8.64 | 7.35 | 8.02 | 7.28 | 6.18 | 6.90 |
| Low | 4.80 | 4.07 | 4.62 | 5.19 | 4.16 | 4.86 |
| 197809-198308 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 7.65 | 6.46 | 7.39 | 10.79 | 9.40 | 10.55 |
| High | 14.58 | 12.72 | 14.03 | 13.43 | 11.57 | 13.01 |
| Medium | 6.79 | 5.66 | 6.56 | 9.84 | 8.57 | 9.64 |
| Low | 1.18 | 3.04 | 2.66 | -0.38 | 2.04 | 1.18 |
| 198309-198811 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 3.83 | 1.96 | 3.48 | 6.55 | 3.47 | 5.86 |
| High | 11.89 | 9.03 | 10.97 | 12.85 | 9.65 | 11.81 |
| Medium | 4.08 | 2.16 | 3.68 | 6.90 | 3.71 | 6.15 |
| Low | -28.78 | -25.22 | -24.04 | -41.15 | -39.27 | -37.52 |
| 198812-199803 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 1.97 | 1.19 | 1.93 | 2.99 | 1.79 | 2.90 |
| High | 14.19 | 13.36 | 14.04 | 13.11 | 11.38 | 12.94 |
| Medium | 4.56 | 3.03 | 4.40 | 4.66 | 2.73 | 4.44 |
| Low | 1.25 | 0.63 | 1.23 | 1.73 | 0.95 | 1.72 |
| 199804-200303 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 6.16 | 4.70 | 5.68 | 6.18 | 4.93 | 5.79 |
| High | 20.01 | 17.69 | 18.11 | 21.86 | 20.84 | 21.23 |
| Medium | 10.28 | 8.80 | 9.73 | 10.45 | 9.25 | 10.15 |
| Low | 5.15 | 3.72 | 4.73 | 3.60 | 2.30 | 3.18 |
| 200304-200808 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 4.54 | 3.07 | 4.24 | 4.50 | 3.25 | 4.21 |
| High | 12.58 | 11.21 | 12.24 | 10.93 | 10.11 | 10.57 |
| Medium | 4.95 | 3.35 | 4.63 | 4.62 | 3.33 | 4.34 |
| Low | 2.18 | 0.99 | 1.93 | 3.05 | 1.77 | 2.74 |
| 200809-201712 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 3.80 | 3.40 | 3.79 | 5.41 | 4.10 | 5.40 |
| High | 12.22 | 7.23 | 11.52 | - | - | - |
| Medium | 5.39 | 4.38 | 5.39 | 6.16 | 3.53 | 6.15 |
| Low | 2.87 | 2.85 | 2.88 | 4.67 | 4.66 | 4.67 |
| 201801-202212 | OLS | Lasso | Ridge | OLS | Lasso | Ridge |
| Aggregate | 3.99 | 3.51 | 4.05 | 6.86 | 6.46 | 6.87 |
| High | 14.16 | 11.85 | 13.67 | 13.93 | 11.49 | 13.39 |
| Medium | 5.68 | 5.15 | 5.54 | 6.44 | 6.12 | 6.33 |
| Low | 1.01 | 0.76 | 1.38 | -1.97 | 0.43 | -0.29 |

### II..4 Investment Gains: Forecast-implied Portfolio

This appendix explores the investment gains of forecast-implied portfolios constructed from TS+CS clusters. The observed patterns are consistent with those of the CS clusters: highly predictable clusters outperform medium- and low-predictability clusters, and the aggregate forecast model delivers superior performance compared to the global model.

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

This table reports full-sample and large-cap sub-sample baseline long-short investment performance for forecast-implied portfolios (Equation 10). The first two panels show sign-adjusted value- and equal-weighted portfolios, while Panel C reports forecast-weighted portfolios, all constructed from observations in specific samples (Equation 9). We show five samples: Global (no clustering), Aggregate (aggregation of clustering results), and High, Medium, and Low, based on predictive rankings within the tree clusters. We show results of monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), market 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 | 0.82 | 2.81 | 1.01 | 0.58*** | 13.38 | 0.79 | 2.81 | 0.97 | 0.56*** | 13.27 |
| Aggregate | 1.15 | 2.82 | 1.41 | 0.98*** | 6.68 | 1.11 | 2.83 | 1.36 | 0.95*** | 6.83 |
| High | 1.70 | 3.99 | 1.48 | 1.49*** | 17.00 | 1.50 | 4.29 | 1.21 | 1.30*** | 16.32 |
| Medium | 1.22 | 3.39 | 1.24 | 1.04*** | 10.56 | 0.92 | 2.84 | 1.12 | 0.72*** | 10.84 |
| Low | 0.76 | 2.98 | 0.89 | 0.59*** | 9.40 | 0.80 | 3.00 | 0.92 | 0.58*** | 10.37 |
| | Panel B: Sign-adjusted Equal-Weighted | | | | | | | | | |
| | Avg | Std | SR | Alpha | MDD | Avg | Std | SR | Alpha | MDD |
| Global | 1.49 | 3.28 | 1.58 | 1.30*** | 14.79 | 0.98 | 3.15 | 1.08 | 0.74*** | 16.91 |
| Aggregate | 1.86 | 3.34 | 1.92 | 1.77*** | 11.95 | 1.35 | 3.16 | 1.48 | 1.18*** | 8.00 |
| High | 3.00 | 4.38 | 2.37 | 2.67*** | 19.91 | 1.79 | 4.14 | 1.50 | 1.53*** | 19.74 |
| Medium | 2.10 | 3.44 | 2.12 | 1.97*** | 12.36 | 1.21 | 2.98 | 1.41 | 1.00*** | 10.16 |
| Low | 1.14 | 3.12 | 1.27 | 1.10*** | 14.56 | 0.91 | 2.95 | 1.07 | 0.74*** | 11.25 |
| | Panel C: Forecast-Weighted | | | | | | | | | |
| | Avg | Std | SR | Alpha | MDD | Avg | Std | SR | Alpha | MDD |
| Global | 2.13 | 3.97 | 1.86 | 1.91*** | 18.17 | 1.33 | 3.79 | 1.22 | 1.06*** | 20.08 |
| Aggregate | 2.82 | 4.23 | 2.31 | 2.74*** | 16.47 | 1.84 | 3.85 | 1.66 | 1.68*** | 14.85 |
| High | 3.95 | 5.23 | 2.62 | 3.56*** | 22.13 | 2.13 | 4.76 | 1.55 | 1.82*** | 21.52 |
| Medium | 3.23 | 4.48 | 2.50 | 3.15*** | 16.89 | 1.83 | 4.14 | 1.53 | 1.71*** | 17.49 |
| Low | 1.60 | 3.92 | 1.41 | 1.55*** | 20.39 | 1.18 | 3.58 | 1.14 | 0.99*** | 13.91 |

# III. Predictor Descriptions

Table A.6: **Aggregate 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 aggregated liquidity |

Table A.7: **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 | chpm | Industry-adjusted change in profit margin | Profitability |
| 14 | chtx | Change in tax expense | Momentum |
| 15 | cinvest | Corporate investment | Investment |
| 16 | depr | Depreciation / PPandE | 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 |
| 22 | ill | Illiquidity rolling 3m | Liquidity |
| 23 | lev | Leverage | Value |

Table A.7: Equity Characteristics (Continued)

| No. | Acronym | Description | Category |
|---|---|---|---|
| 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, Quarterly Asset Turnover | 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 |