Mosaics of Predictability

Lin William Cong[1] Guanhao Feng[2] Jingyu He[2] Yuanzhi Wang[2]

[1]Nanyang Technological University & NBER & CEPR

[2]City University of Hong Kong

Motivation

Motivation: Return Predictability

- Return predictability is well-documented empirically:
  - — Aggregate market (e.g., Campbell and Thompson, 2008, RFS).
  - — Individual stocks (e.g., Fama and French, 2008, JF; Lewellen, 2015, CFR).
- Studies regard predictability as an attribute of predictors or models.
  - — Agg. predictors (e.g., dividend yield) and char. (e.g., size or value).
  - — Models include historical average (e.g., Campbell and Thompson, 2008, RFS) and machine learning (e.g., Gu, Kelly, and Xiu, 2020, RFS).
- We find that predictability is heterogeneous for stocks and varies over time.
  - — Does high predictability imply high return?
  - — It might be a characteristic!
  - — Predictability differentials introduce a model misspecification risk.

Return Predictability Facts for Individual Stocks (1973-2022)

- Simple pooled or homogeneous model (e.g., Gu, Kelly, and Xiu, 2020, RFS).
- Positive OOS R^2 (relative to the zero benchmark).
- Exceptional Long-Short portfolio performance even for OLS!

	In-Sample (1973-2002)			Out-of-Sample (2003-2022)		
	OLS	Lasso	Ridge	OLS	Lasso	Ridge
OOS R^2	1.49	0.52	0.54	0.27	0.40	0.35
Avg	3.04	2.22	3.60	1.02	0.84	0.99
Std	4.45	4.76	5.66	3.74	4.42	4.98
SR	2.37	1.60	2.20	0.94	0.65	0.69
Alpha	3.07***	2.28***	3.71***	1.19***	1.23***	1.31***

Motivation: Heterogeneous Predictability

- Empirical evidence suggests that predictability is NOT homogeneous:

  — Certain stocks (e.g., small-cap, distressed) are more predictable than others (e.g., Avramov, Cheng, and Metzker, 2023, MS).

  — Predictability might be time-varying (e.g., Henkel et al., 2011, JFE) or with structural breaks (e.g., Smith and Timmermann, 2021, RFS).

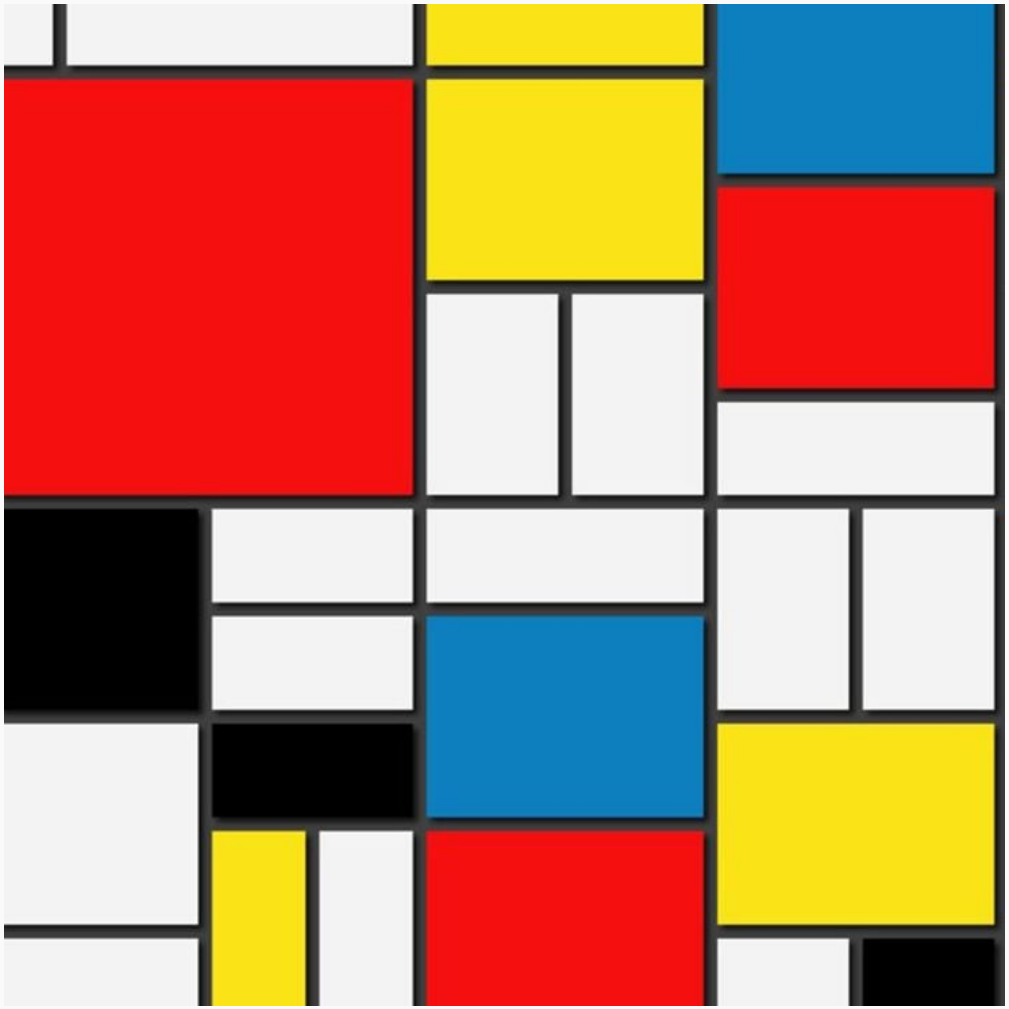
- However, predictability is:

  — An unobservable characteristic.

  — Even not well-defined (e.g., anomaly average return, predictor significance, out-of-sample R^2, forecast-implied portfolios).

- Before exploring heterogeneous predictability, we first need to measure it.

Mosaics of Predictability: Mondrian



We partition the panel of returns by their heterogeneous predictability!

Our Clustering Solution

- Tree-based clustering (goal-oriented) to separate and group asset returns

  $\Longrightarrow$ **Mosaics of Predictability**.

- Objective: Maximize differences in predictability across groups.

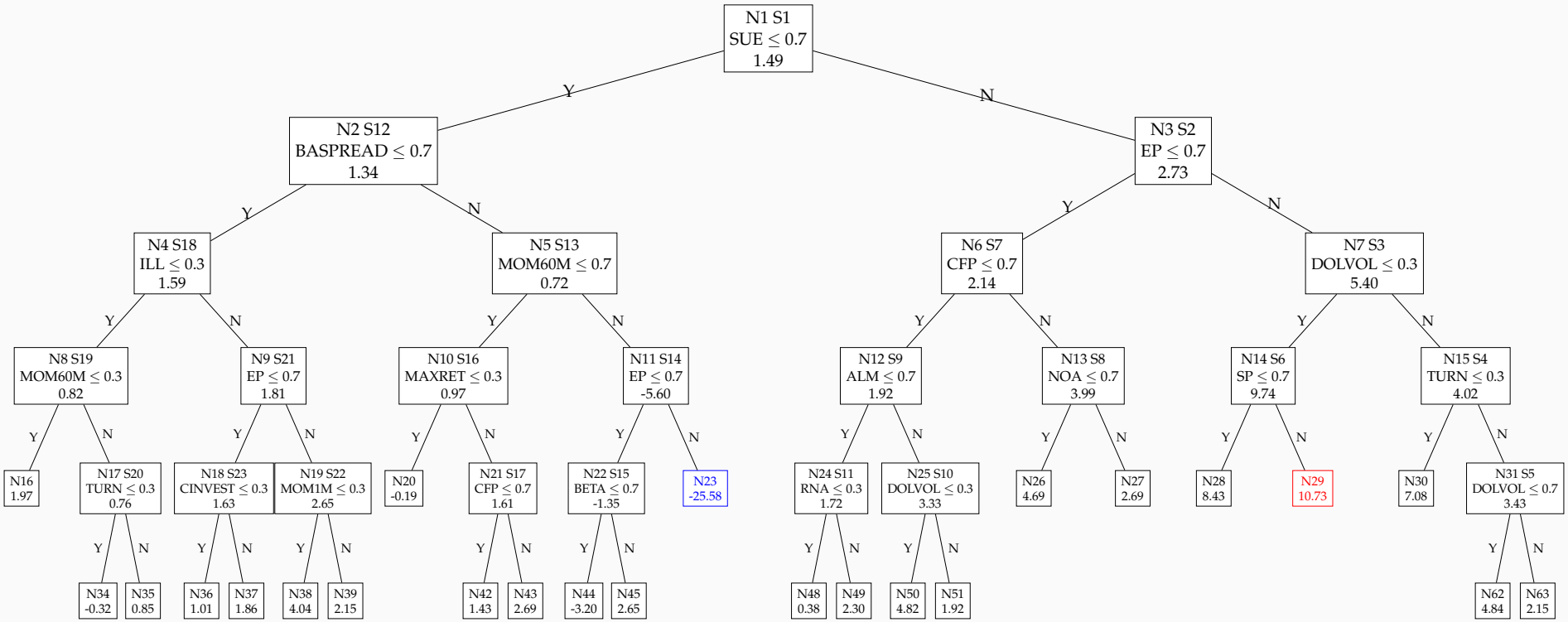
- Interpretable: A decision tree based on firm char. and/or agg. predictors.

- NOT a horse race of return prediction accuracy!

  — New Angle: Our focus is on the heterogeneity of predictability.

  — Which stocks (CS) and when (TS) exhibit higher predictability?

Empirical Highlights: Cross-Sectional Tree-based Clusters



Highly Predictable: N29 (10.73%):

$$\mathbb{1}\{\text{SUE} > 0.7\}\mathbb{1}\{\text{EP} > 0.7\}\mathbb{1}\{\text{DOLVOL} \leq 0.3\}\mathbb{1}\{\text{SP} > 0.7\}$$

Less Predictable: N23 (-25.58%):

$$\mathbb{1}\{\text{SUE} \leq 0.7\}\mathbb{1}\{\text{BASPREAD} > 0.7\}\mathbb{1}\{\text{MOM60M} > 0.7\}\mathbb{1}\{\text{EP} > 0.7\}$$

Methodology

Clustering

- Clustering is inherently challenging since the true labels are unknown.
- "Unsupervised" clustering:
  - Ahn et al. (2009, RFS), Patton and Weller (2022, RFS), Evgeniou et al. (2023, JFQA) apply K-means for clustering individual stocks or portfolios.
- Sub-sample analysis based on size or industry classifications.
- "Goal-oriented" clustering by decision trees:
  - Decision tree provides an indirect outcome for clustering observations.
    — Interpretable (graphable) for variable-based clusters.
  - Panel Tree of Cong et al. (2025, JFE) splits the individual stock returns panel to estimate the mean-variance efficient frontier.

We construct a decision tree to cluster stock return observations based on heterogeneous predictability — maximizing the between-cluster differences.

Measurement of Return Predictability

- A predictive model is typically

$$r_{i,t} = E_{t-1}(r_{i,t}) + \epsilon_{i,t}$$

- We define predictability as signal-to-noise ratio

$$R^2_{i,t} = 1 - \frac{\text{Var}(\epsilon_{i,t})}{\text{Var}(r_{i,t})}$$

  — Higher R^2 indicates the signal is stronger and easier to predict.

- Hard to estimate $R^2_{i,t}$ directly for each i and t. But running a panel model will induce a single R^2 and ignore heterogeneity.

- Balance: Cluster-wise homogeneous predictive model
  — Assets in different clusters show different R^2.
  — Assets in the same cluster share the same R^2.

Split Criterion: Cluster-Wise Predictability

- In-sample model fitness R^2:

  ▸ Why not OOS?

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

- $\widehat{r}_{i,t}$ is the return forecasts from the cluster-specific model.

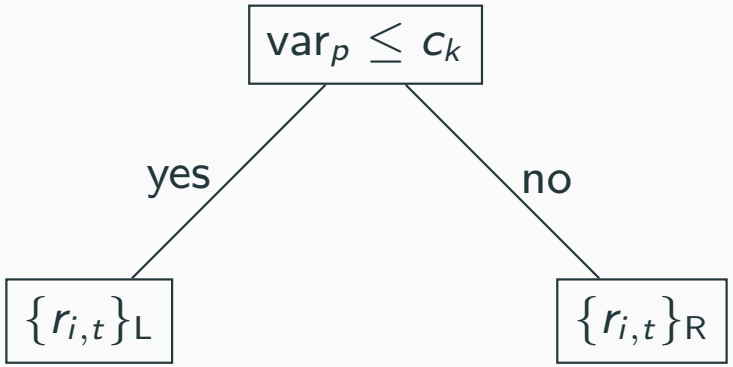
  — Volatility-weighted (avoid dominance of microcaps, e.g., Fama and French, 2008, JF; Hou et al., 2020, RFS) Ridge regression (Shen and Xiu, 2024, WP).

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

- To separate a sample (parent leaf) into two (child leaves), we maximize the R^2 difference between them.

  — Calculate values for the maximal difference for each split candidate.

Split Criterion and Tree Growth



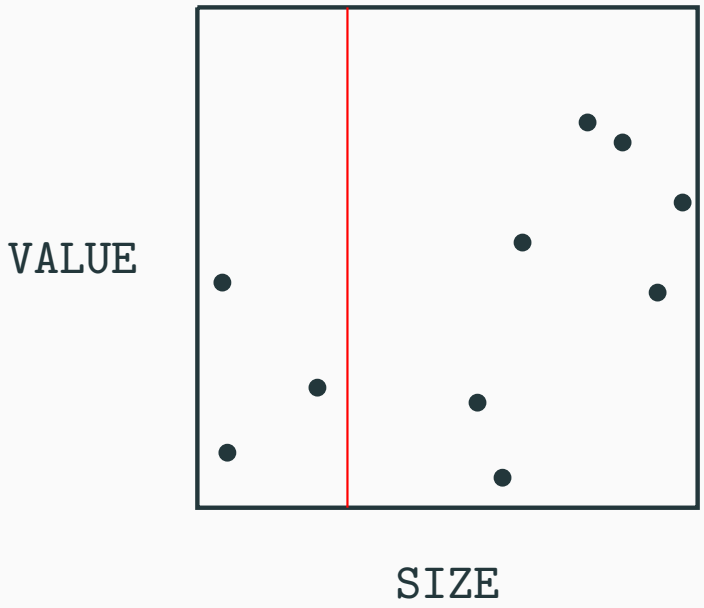
- Candidate cutpoints: $\{0.3, 0.7\}$ among standardized range $[0, 1]$.
  - — Monthly cross-sectional ranked char.
  - — For example, small-cap on the left and non-small-cap on the right.
- Rolling 10-year demeaning macro predictors compared with 0.
- Calculate the model objective for a splitting candidate (var_p, c_k) on the R^2 difference, which differentiates predictability:

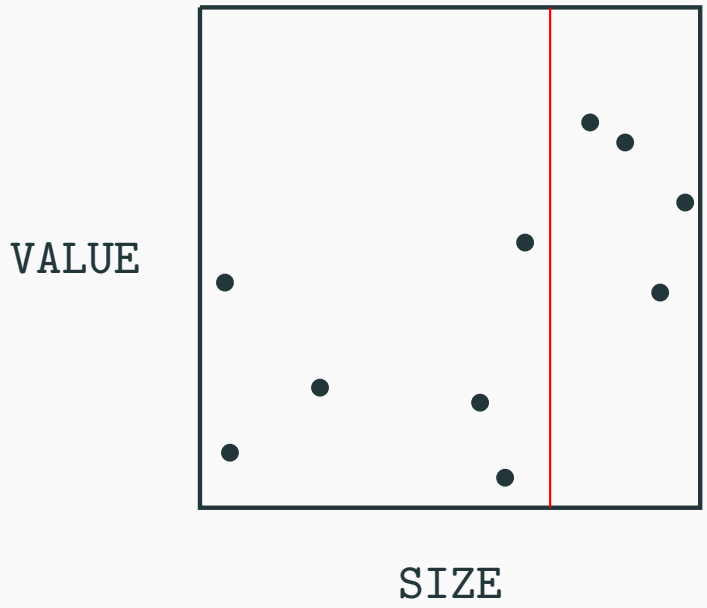
$$S_{\{\text{leaf}_L, \text{leaf}_R\}}(\text{var}_p, c_k) = |R^2_{\text{left}} - R^2_{\text{right}}|$$

- Greedy algorithm: Pick the split variable and cut point that maximize the split criterion to partition the subsample into two.

Split Criterion: Cross-Sectional Demo

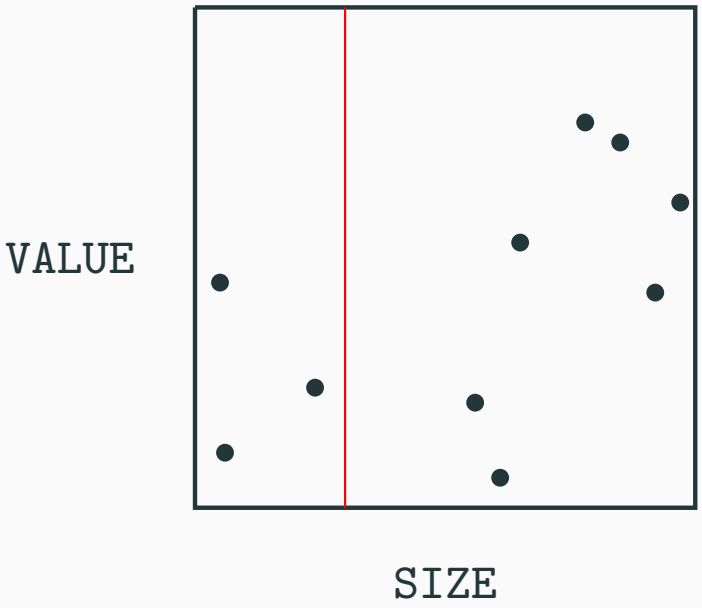
There are many split candidates. Shall we split at "SIZE ≤ 0.3" or "≤ 0.7"?

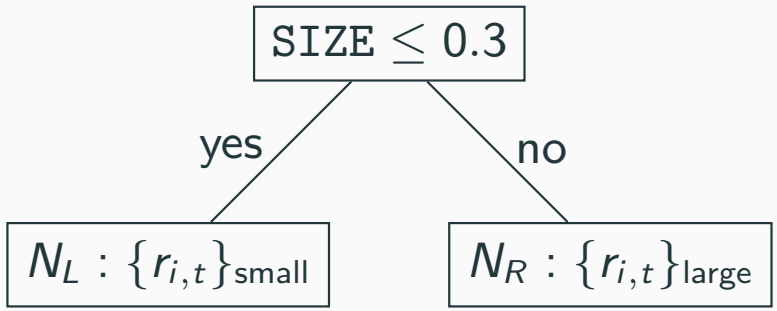




Split Criterion: Cross-Sectional Demo

Consider one split candidate. It partitions the space to the left and right child:



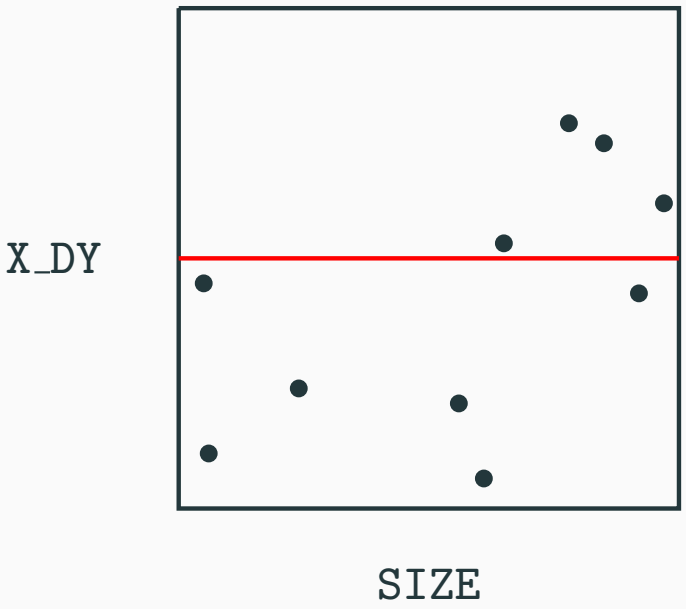


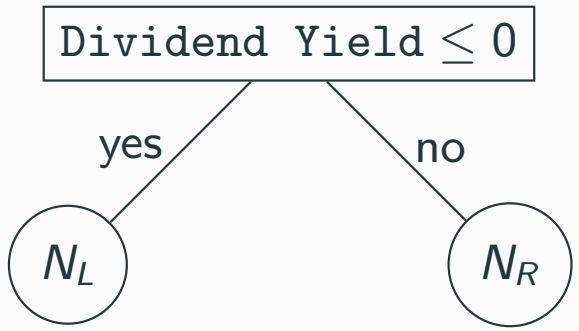
We calculate R^2 for N_L and N_R, and get the split criterion value as:

$$S_{\{N_L,N_R\}}(\texttt{SIZE}, 0.3) = |R^2_{N_L} - R^2_{N_R}|$$

Split Criterion: Time-Series Demo

Time-series dimension $\Longrightarrow$ heterogeneous regime-based predictability.

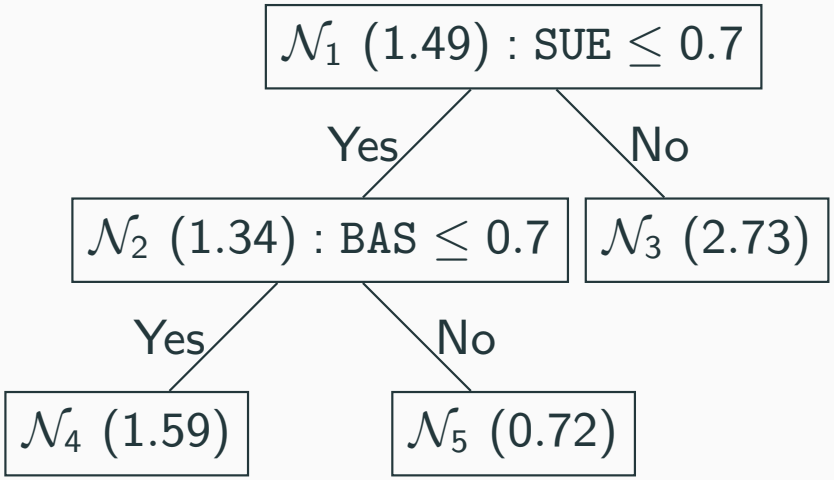




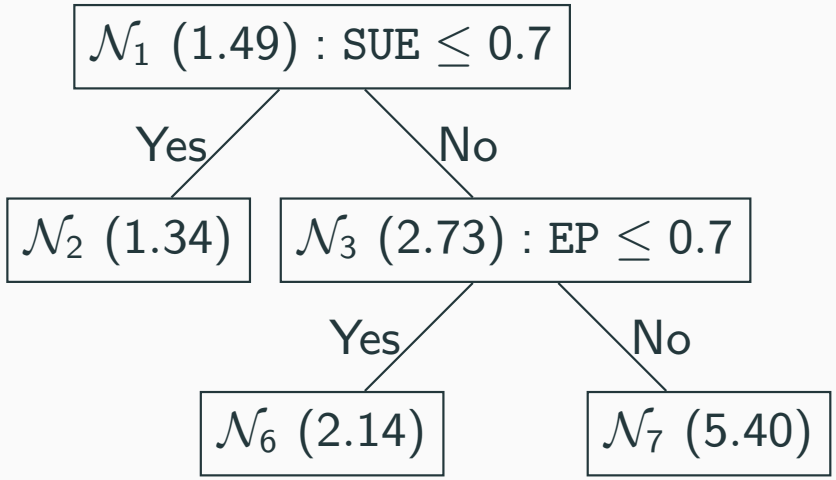
We calculate R^2 for N_L and N_R, and get the split criterion value similarly:

$$S_{\{N_L,N_R\}}(\texttt{X\_DY}, 0) = |R^2_{N_L} - R^2_{N_R}|$$

Growing Tree-based Clusters Demo



(a) If splitting node $\mathcal{N}_2$ at BASPREAD

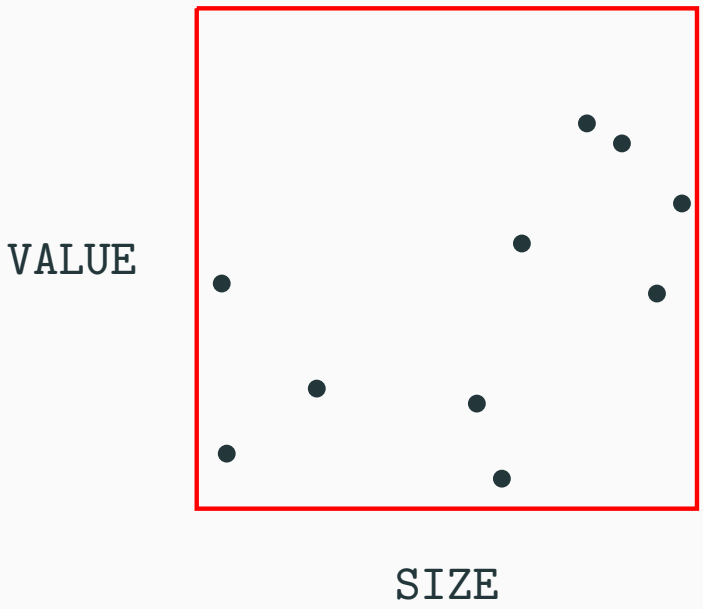


(b) If splitting node $\mathcal{N}_3$ at EP

- 1st split: SUE ≤ 0.7, Max $S_{\{\mathcal{N}_2,\mathcal{N}_3\}}(\text{SUE}, 0.7) = |1.34 - 2.73| = 1.39$.
- 2nd split: $\mathcal{N}_3$, EP ≤ 0.7, Max $S_{\{\mathcal{N}_6,\mathcal{N}_7\}}(\text{EP}, 0.7) = |2.14 - 5.40| = 3.26$.
- Suppose P variables, K cut points, N-th split (Total $\# = P \times K \times N$)
  — Iteratively test all possible leaf nodes and split candidates.

When to Stop the Tree Growth?

Stop splitting is equivalent to leave all data in one node.



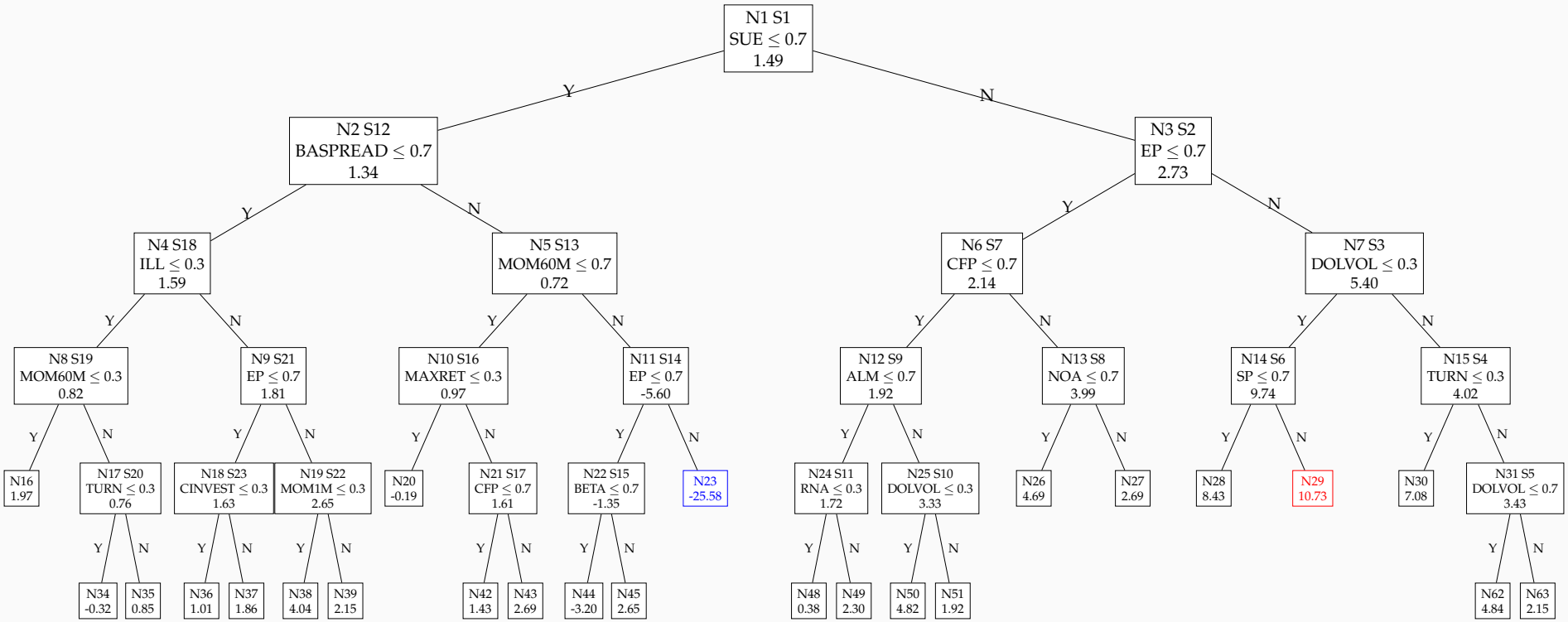
Stop Criteria: max depth / min leaf size / no more improvement (two child nodes not better than parent node)

Empirical Results

U.S. Equity Data

- 1973 - 2022 individual stock monthly returns.
- 51 char. ($z_{i,t}$) for both cross-sectional splits and predictors.
- 8 macro predictors ($\mathbf{x}_t$) for both time series splits and predictors.
- Cross-section splits (1973-2002 for training and 2003-2022 for test)
- Time-series regimes: entire 50-year sample.
- Performance evaluation metrics:
  - Predictability: R^2
  - Investment: Avg Return, Sharpe ratio, CAPM Alpha, Maximum Drawdown.

Cross-Sectional Tree-based Clusters



Highly Predictable: N29 (10.73%):

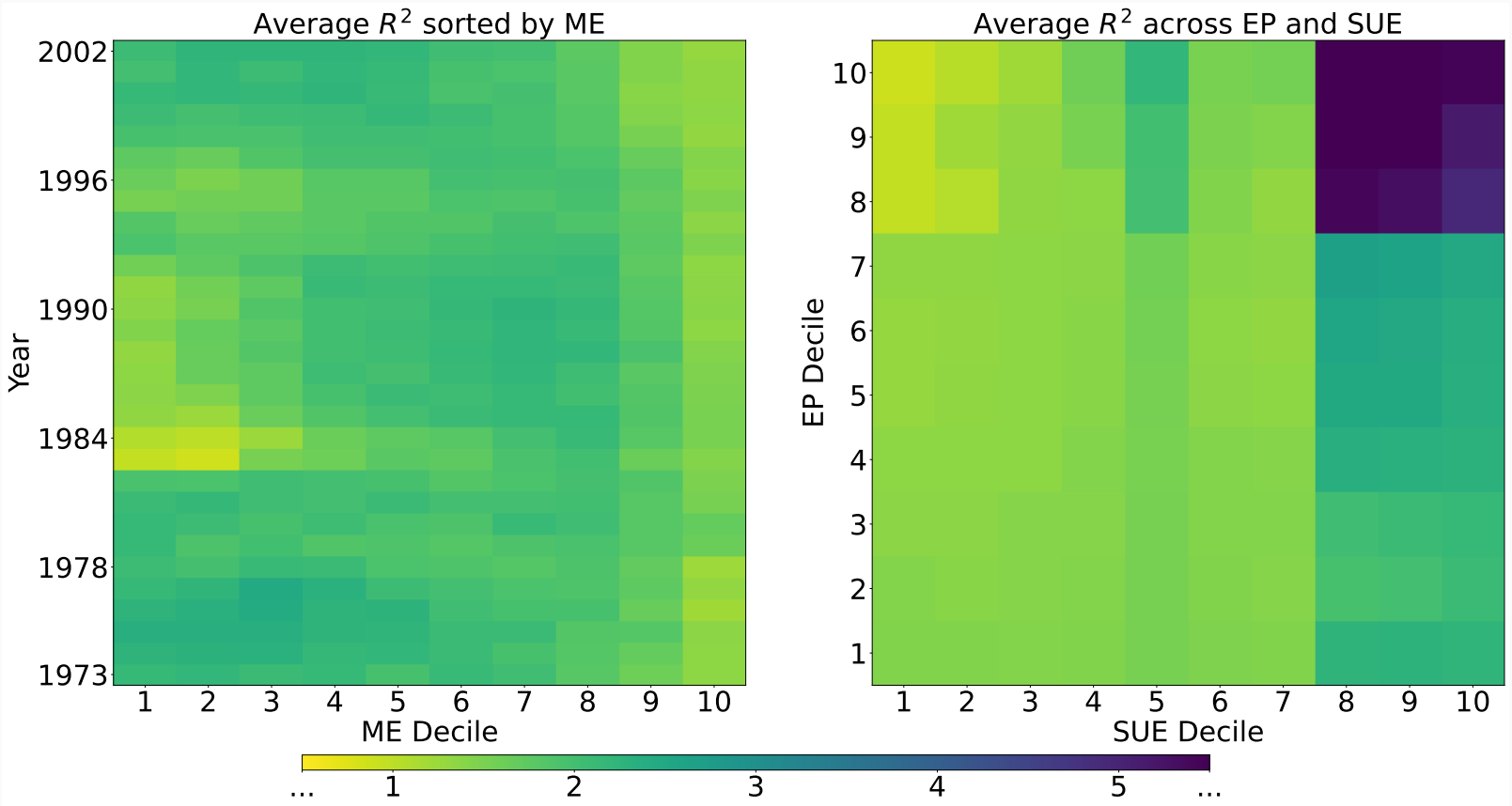
$$\mathbb{1}\{\text{SUE} > 0.7\}\mathbb{1}\{\text{EP} > 0.7\}\mathbb{1}\{\text{DOLVOL} \leq 0.3\}\mathbb{1}\{\text{SP} > 0.7\}$$

Less Predictable: N23 (-25.58%):

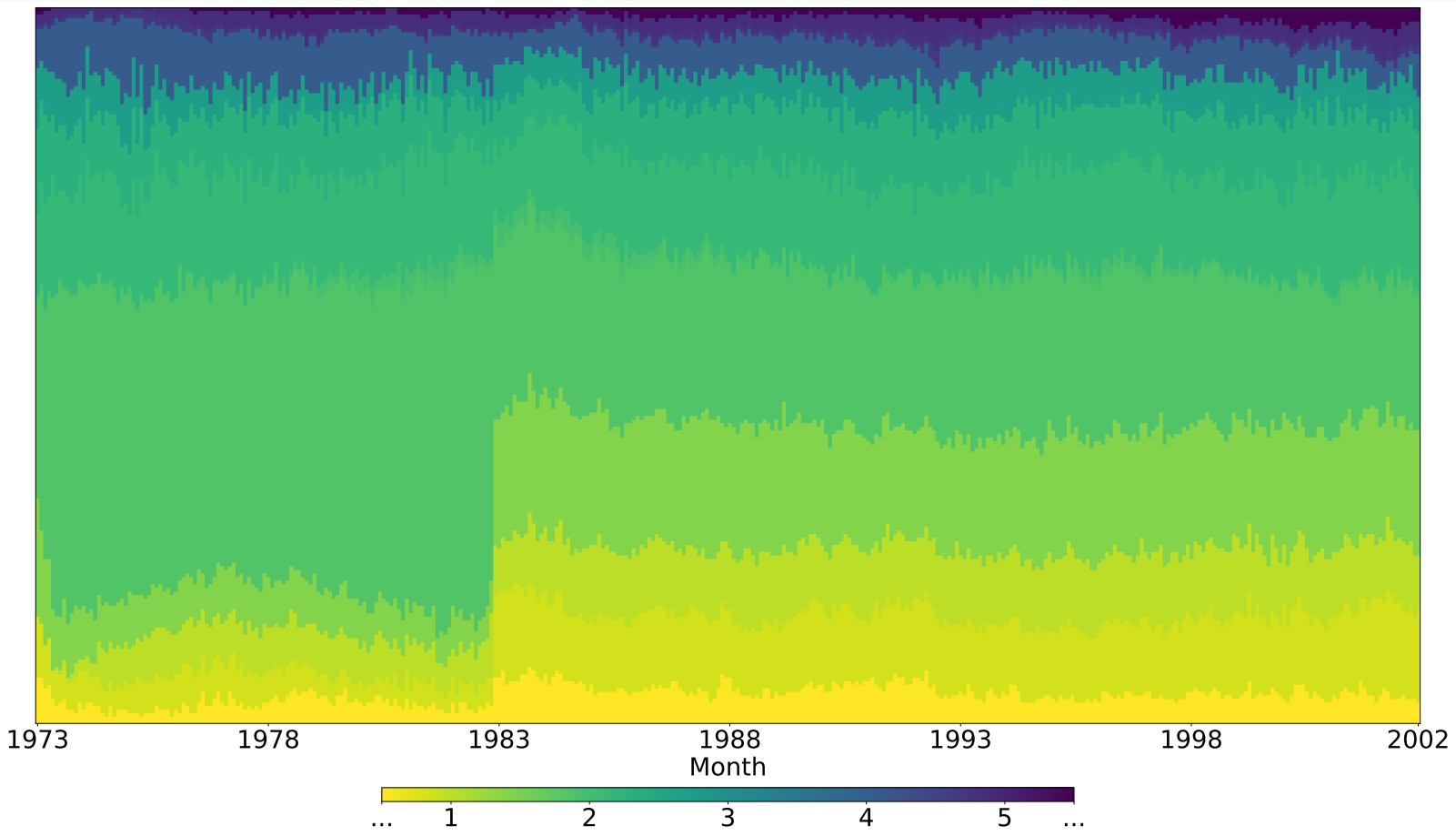
$$\mathbb{1}\{\text{SUE} \leq 0.7\}\mathbb{1}\{\text{BASPREAD} > 0.7\}\mathbb{1}\{\text{MOM60M} > 0.7\}\mathbb{1}\{\text{EP} > 0.7\}$$

Mosaics of Predictability by Deciles

Ascending order of chars-sorted decile R^2s.



Mosaics of Predictability by Months



- Descending order of cluster-based R^2s (y-axis).
- Length of each color is the proportion of each cluster every month.
- Only the top dark parts are highly predictable.

Out-of-Sample Robustness

- **Performance: Cluster-wise $>$ Homogeneous.**
- **Highly predictable clusters show persistently high R^2s out of sample.**

	1973 - 2002 (in-sample)				2003 - 2022 (out-of-sample)			
Forecasts	All	High	Medium	Low	All	High	Medium	Low
Global	0.54	2.28	0.47	0.31	0.35	2.06	0.27	0.17
CW	0.74	4.69	0.59	0.06	0.53	**3.84**	0.39	**0.11**

— Global: Homogeneous predictive model (similar to GKX2020).

— CW (Cluster-Wise): Heterogeneous predictive models.

Predictability Differential Strategy

- We highlight a strategy associated with model misspecification risk.

  — Homogeneous models underperform regime-dependent heterogeneous counterparts (e.g., Smith and Timmermann, 2022, JFE).

- If investors ignore heterogeneity and fit a global model to predict returns, it produces an R^2_{G}, which is the same for all assets.

- Cluster-wise predictive models provide R^2_{C}, same value within each cluster.

- Differentials: R^2 difference between heterogeneous and global models.

  $\Longrightarrow$ Model misspecification risk:

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

▸ Summary Table

Predictability Differential Strategy

- Long (short) similar # of top (bottom) clusters (around 2.5% for 3 leaves).
- Long-leg dominates the large significant OOS alphas (above 1.4%).

	1973 - 2002 (in-sample)			2003 - 2022 (out-of-sample)		
	T3	B3	T3 - B3	**T3**	B3	**T3 - B3**
			Panel A: Performance			
Avg (%)	2.76	-0.26	3.03	2.26	0.52	1.74
Ann. SR	1.69	-0.16	2.28	1.35	0.32	1.48
			Panel B: Unexplained monthly alphas (%)			
CAPM	2.40***	-0.66***	3.07***	1.42***	-0.39**	1.81***
FF3	1.97***	-0.86***	2.83***	1.45***	-0.36**	1.81***
FF5	1.82***	-0.90***	2.72***	1.62***	-0.38**	2.00***
FF5+MOM+IVOL	1.96***	-0.58***	2.53***	1.68***	-0.37**	2.05***
Q5	1.88***	-0.53**	2.41***	1.75***	-0.34**	2.09***
BS6	1.74***	-0.76***	2.49***	1.64***	-0.38**	2.03***
DHS3	2.58***	-0.12	2.70***	1.67***	-0.32*	1.98***
SY4	1.88***	-0.58***	2.46***	2.38***	-0.40**	2.78***

Investment Gains on Cluster-wise Models

- Notably, our approach generates cluster-wise predictive models.

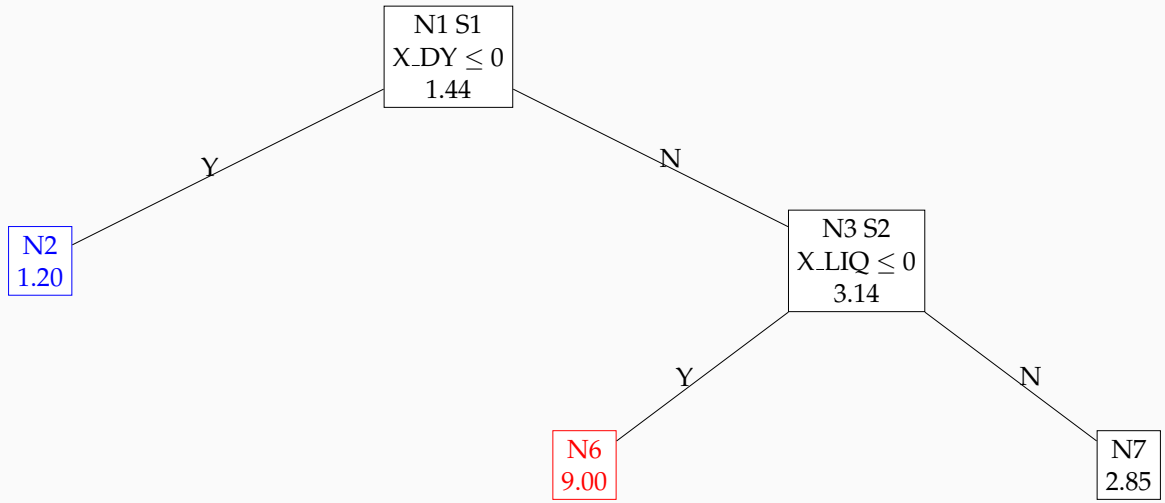
  — Global models (e.g., Gu et al., 2020, RFS) do not account for modeling heterogeneity (e.g., Feng and He, 2022, JoE; Evgeniou et al., 2023, JFQA).

- Forecast-Weighted portfolio (based on the normalized predictions) ▸ Equation

- Same strategy achieves highest gains in the highly predictable clusters!

	In-Sample (1973 - 2002)					Out-of-Sample (2003 - 2022)				
	Avg	Std	SR	Alpha	MDD	Avg	Std	SR	Alpha	MDD
Global	1.58	4.69	1.17	1.26^{***}	20.52	1.12	5.32	0.73	0.25^{*}	23.04
CW	2.01	4.26	1.64	1.74^{***}	16.52	1.56	4.77	1.13	0.79^{***}	21.31
High	3.42	6.08	1.95	3.01^{***}	26.23	3.22	5.99	1.86	2.30^{***}	24.82
Medium	1.74	4.05	1.48	1.50^{***}	16.10	1.25	4.57	0.95	0.53^{***}	21.96
Low	0.88	4.93	0.62	0.53^{***}	17.95	0.88	5.60	0.54	-0.03	19.98

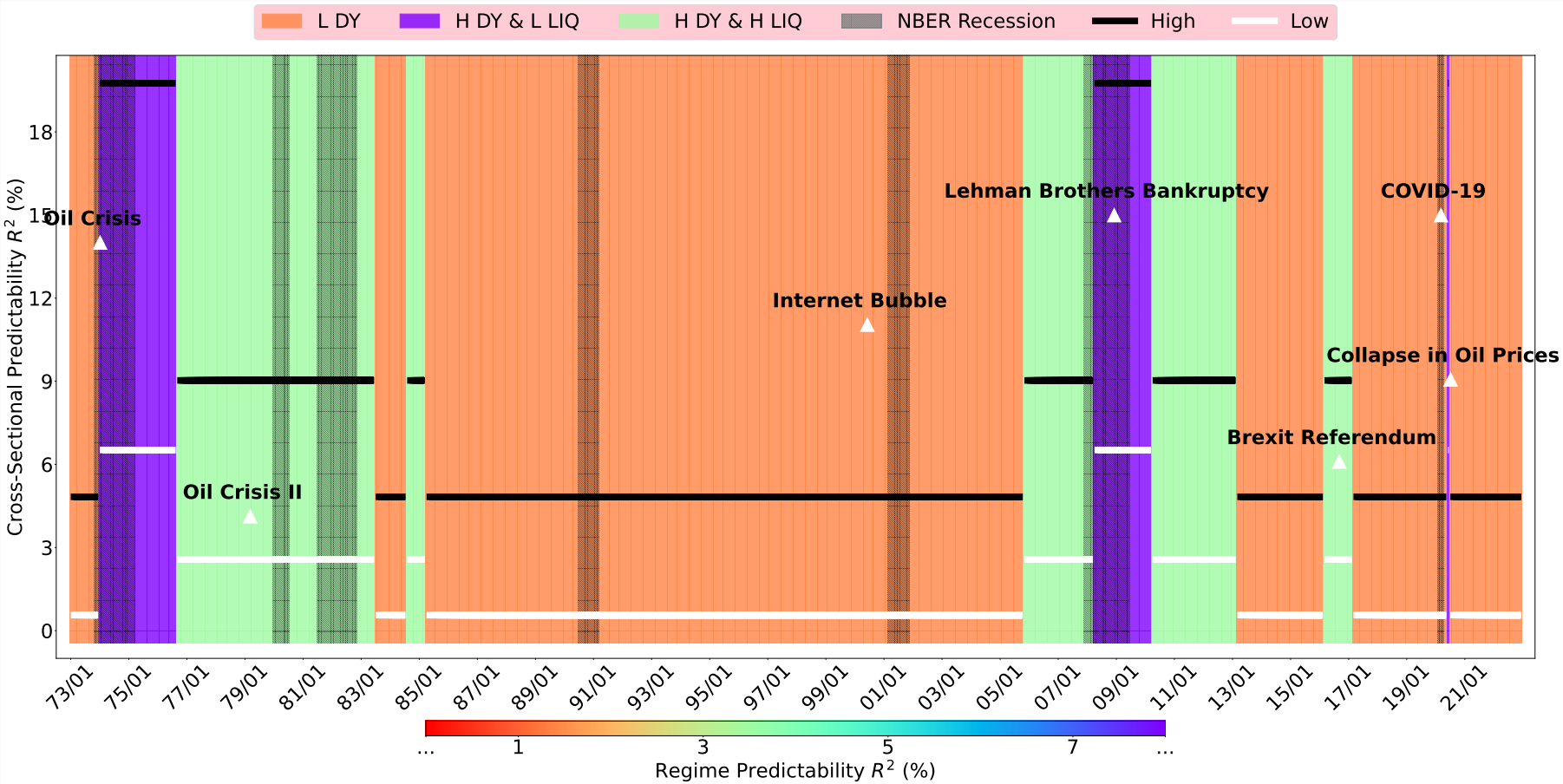
Time-Series Tree-based Clusters



Highly Predictable Period (9.00%, 57 months):
$\mathbb{1}\{\text{X\_DY} > 0\}\mathbb{1}\{\text{X\_LIQ} \leq 0\}$, high Dividend Yield & low Agg. Liquidity.

Less Predictable Period (1.20%, 377 months):
$\mathbb{1}\{\text{X\_DY} \leq 0\}$, low Dividend Yield.

Time-Varying Predictability under Market Regimes



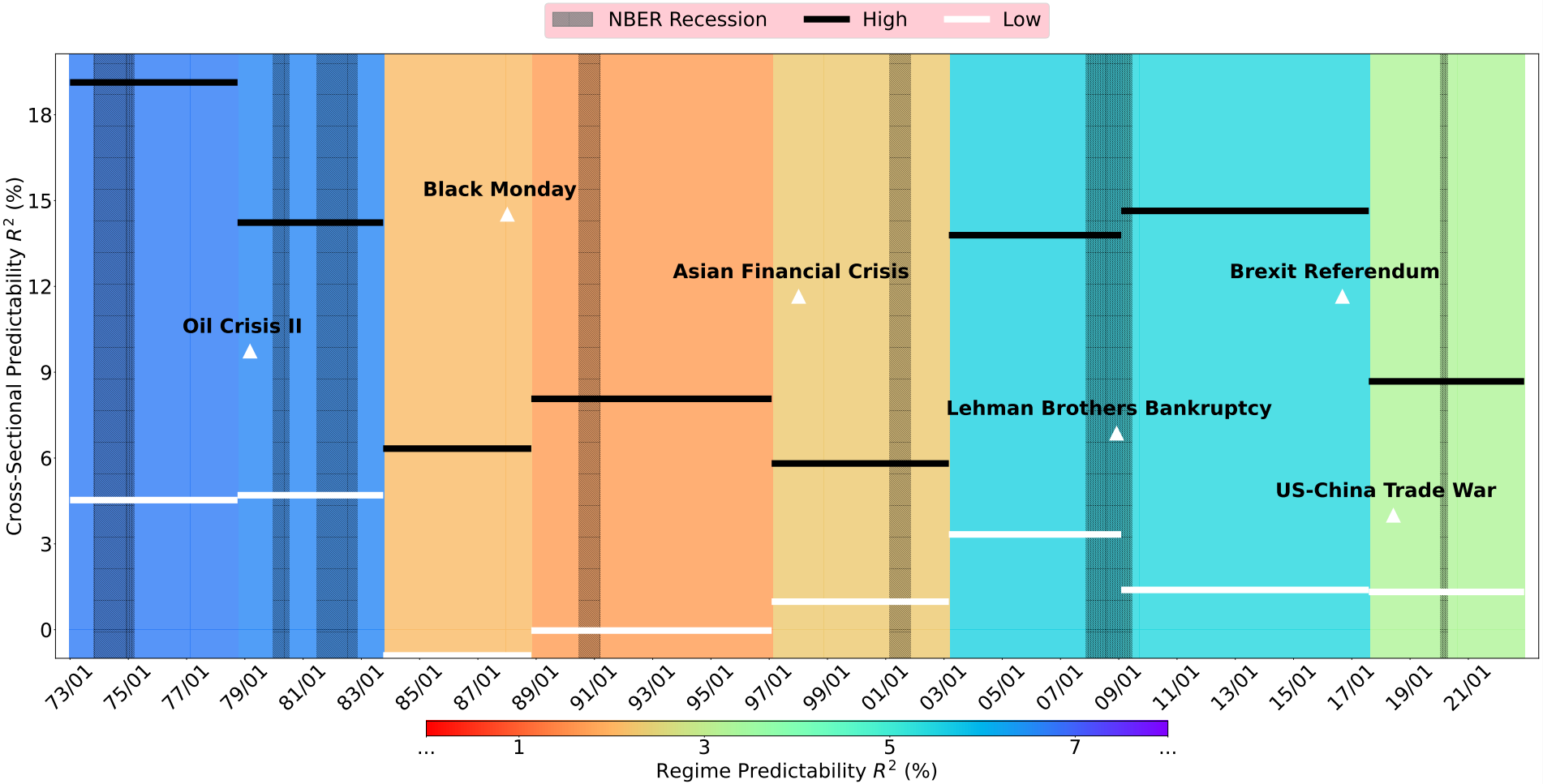
- Variables: S&P 500 dividend yield (DY) and market liquidity (LIQ).
- Time-series partitions display larger predictability heterogeneity (color bar).
- Further cross-sections enlarge the gaps (y-axis).
- Numerous events trigger regime changes (e.g., Oil Crisis, COVID-19).

Comparing Global and Cluster-wise Models

- Predictability levels and cross-sectional differences vary in regimes!

	Sample A: All Stocks				Sample B: Large-Cap			
Forecasts	All	High	Medium	Low	All	High	Medium	Low
				Regime I: $\mathbb{1}\{\texttt{X\_DY} \leq 0\}$				
Global	0.84	3.00	0.96	0.51	0.87	2.15	0.85	0.56
CW	0.98	4.46	1.45	0.16	0.94	2.54	1.17	-0.10
				Regime II: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} \leq 0\}$				
Global	8.68	10.58	9.26	6.29	12.66	16.07	12.41	-
CW	9.28	17.35	9.24	4.95	12.91	20.44	12.36	-
				Regime III: $\mathbb{1}\{\texttt{X\_DY} > 0\}\mathbb{1}\{\texttt{X\_LIQ} > 0\}$				
Global	2.39	6.53	3.18	1.65	2.88	4.16	3.22	2.01
CW	3.00	8.43	3.92	2.11	3.35	5.95	3.74	2.13

Time-varying Predictability over Structural Breaks



- Structural break model shows similar heterogeneity.
- Numerous events trigger regime changes (e.g., Black Monday, Brexit Referendum, US-China Trade War)

Summary

Summary

- Mosaics of predictability — Heterogeneity in return predictability.
- Tree-based clustering approach — based on firm char. and agg. predictors.
- Investment strategy: R^2 differentials $\Longrightarrow$ model misspecification risk.
- **All comments are welcome! Thank you!**

References i

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

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

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.

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

Feng, G. and J. He (2022). Factor Investing: A Bayesian Hierarchical Approach. *Journal of Econometrics 230*(1), 183–200.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies 33*(5), 2223–2273.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-Varying Short-Horizon Predictability. *Journal of Financial Economics 99*(3), 560–580.

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

References ii

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

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.

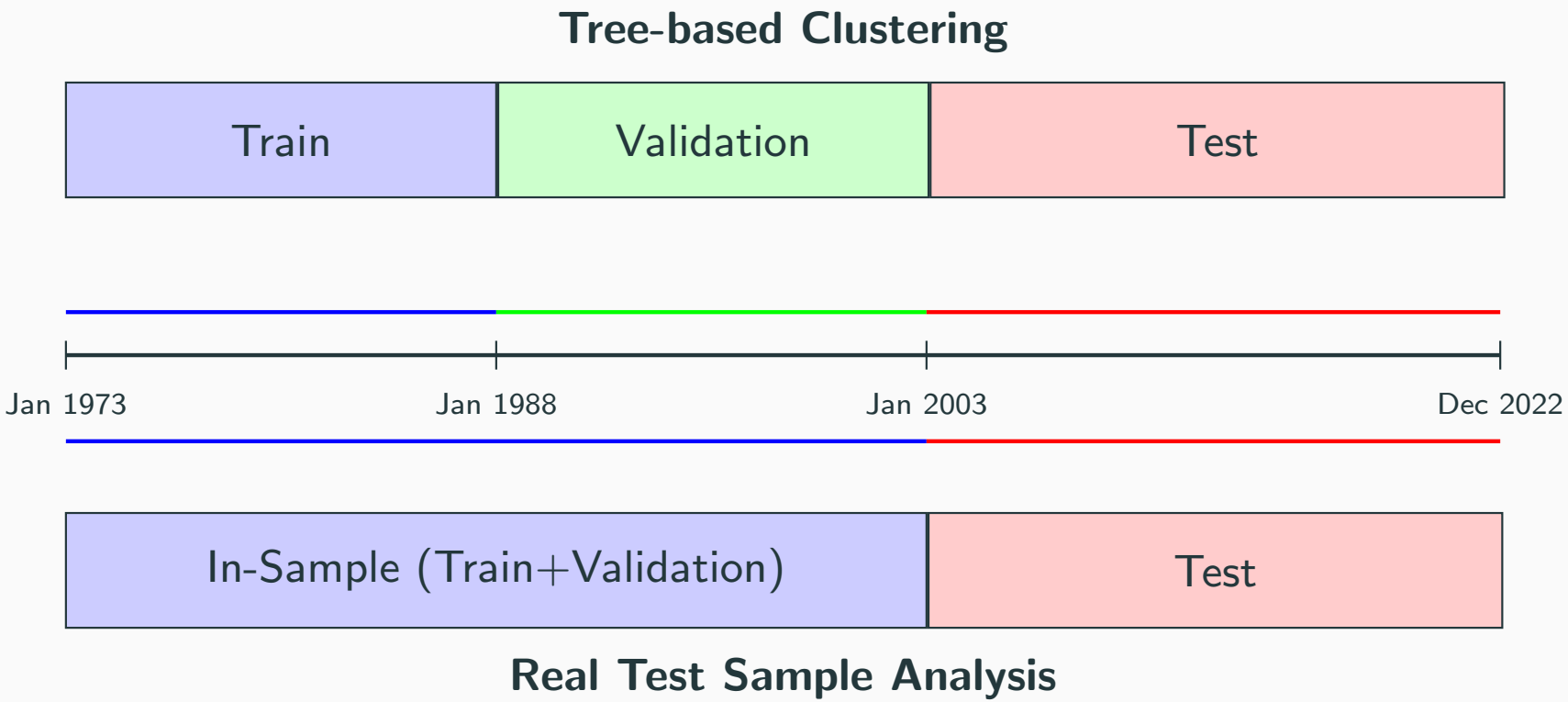
Shen, Z. and D. Xiu (2024). Can Machines Learn Weak Signals? Technical report, University of Chicago, Becker Friedman Institute for Economics Working Paper.

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

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

Appendix

Why not OOS?



Back

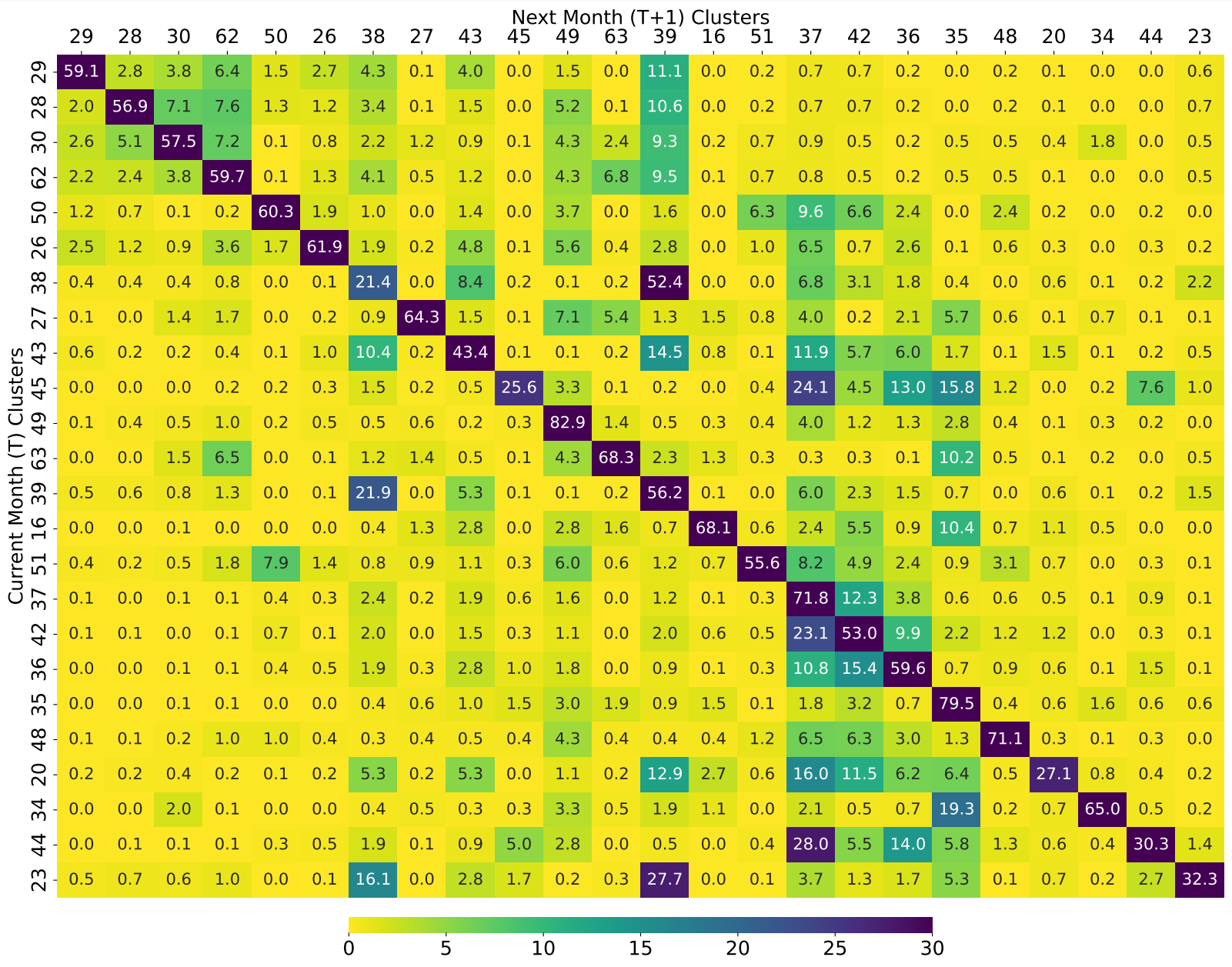
Performance of Asset Clusters

Obvious gaps for predictability and profitability from top to bottom.

Back

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

Persistence of Clusters



Forecast-Implied Portfolio

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

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

- $\widehat{r}_{i,t}$ are return forecasts from the cluster-based model (by Ridge, etc.).
- The corresponding j-th leaf node forecast-implied portfolio return is:

$$R_{j,t} = \sum_{\{i,t\} \in \text{leaf}_j} \widehat{w}_{i,t-1} r_{i,t}. \tag{2}$$

- Include ALL observations rather than only deciles (long-short).
- Test the order between realized ($r_{i,t}$) and predicted ($\widehat{r}_{i,t}$) returns.

Back