Growing the Efficient Frontier on Panel Trees

Will Cong[2] Guanhao Feng [1] **Jingyu He[1]** Xin He[3]

City University of Hong Kong[1]

Cornell University[2]

Hunan University[3]

July. 12th, 2024

2nd Quarterly Journal of Finance Conference

Motivation

- Asset pricing relies on factor models to estimate the mean-variance efficient frontier.

  — reduces the dimension of individual assets ($\{r_t\}$) to left-hand side (LHS) test assets ($\{R_t\}$) and right-hand side (RHS) factors (f_t).

  — $\text{EF}(f_t) \Longrightarrow \text{EF}(\{R_t\}) \Longrightarrow \text{EF}(\{r_t\})$

$$R_{i,t} = \alpha_i + \beta_i^{\intercal} f_t + \epsilon_{i,t}$$

- Fama and French (1993, 1996) propose risk factors (f_t, long-short portfolios of char-sorted portfolios) to explain basis portfolios ($\{R_t\}$).

  — these basis portfolios ($\{R_t\}$, i.e., 5×5 ME-B/M), are criticized for not capturing the EF of $\{r_t\}$ (e.g., Lewellen et al., 2010; Ang et al., 2020).

Motivation

- On the RHS: researchers seek to identify a few common factors to achieve multi-factor portfolio efficiency (Fama, 1996), expected to be statistically insignificant from the EF spanned by the test assets (Gibbons et al., 1989).

  — popular factor models (f_t) hardly span the common test assets ($\{R_t\}$) (e.g., Kozak et al., 2018; Daniel et al., 2020), let alone the ultimate efficient frontier of $\{r_t\}$.

- On the LHS: Rather than generating test assets,

  (i) some evaluate different sets of test assets separately for robustness checks (e.g., Fama and French, 1996, 2016),

  (ii) select test assets (e.g., Daniel et al., 2020; Bryzgalova et al., 2023),

  or (iii) combine test assets (e.g., Feng et al., 2020; Giglio and Xiu, 2021).

Motivation

- The key to bridging the gap between the true efficient frontier and tangency portfolios of assets or factor portfolios lies in systematically utilizing the high-dimensional firm char that contain rich information on the joint distribution of asset returns dynamics.
- Cochrane (2011) asserts that expected returns, variances, and covariances are stable functions of char (also see, e.g., Kelly, Pruitt, and Su, 2019; Kozak, Nagel, and Santosh, 2020).
- Conditional SDF on stocks & Unconditional SDF on char-managed portfolios.
- We introduce Panel Tree (P-Tree) to generate test assets that span the Efficient Frontier and the SDF.

Panel Tree

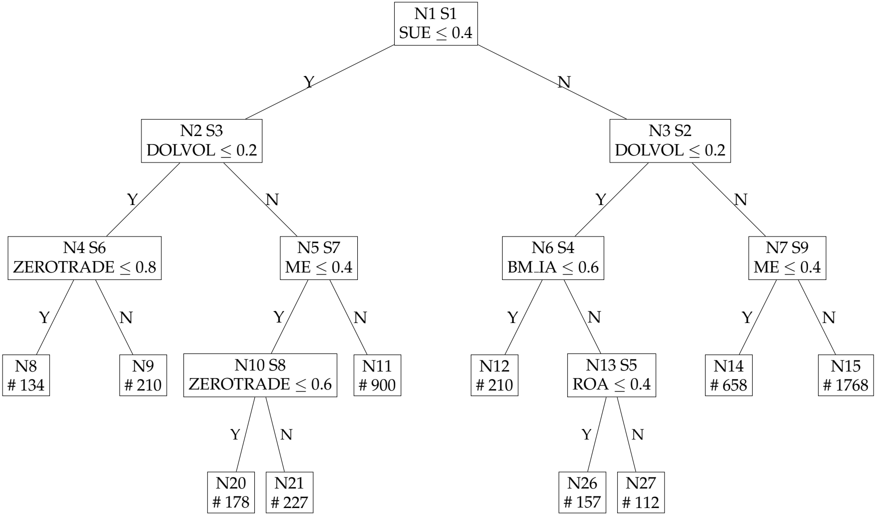
(i) It is adept at analyzing (unbalanced) panel data and inherits the advantages of tree-based models while allowing interpretability under economics-guided global split criteria for goal-oriented search.

(ii) P-Trees utilize high-dimensional firm char under MVE framework to jointly generate diversified test assets and latent factors for recovering SDF.

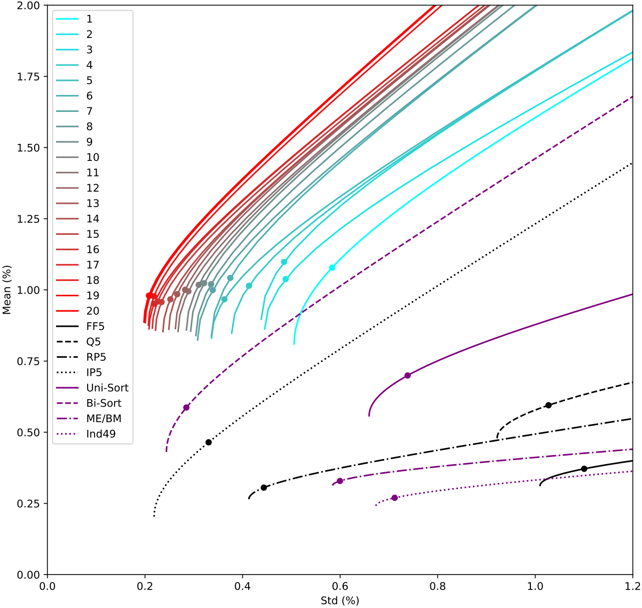
— According to Hansen and Jagannathan (1991), the resulting SDF that maximizes the Sharpe ratio also minimizes the HJ-distance.

(iii) P-Trees employ a top-down approach, splitting the cross section of thousands of individual assets and grouping them into a small number of clusters with similar char values to form value-weighted portfolios.

Panel Tree


N1 S1
SUE ≤ 0.4
Y
N
N2 S3
DOLVOL ≤ 0.2
N3 S2
DOLVOL ≤ 0.2
Y
N
Y
N
N4 S6
ZEROTRADE ≤ 0.8
N5 S7
ME ≤ 0.4
N6 S4
BM_IA ≤ 0.6
N7 S9
ME ≤ 0.4
Y
N
Y
N
Y
N
Y
N
N8
134
N9
210
N10 S8
ZEROTRADE ≤ 0.6
N11
900
N12
210
N13 S5
ROA ≤ 0.4
N14
658
N15
1768
Y
N
Y
N
N20
178
N21
227
N26
157
N27
112

Empirical Highlights

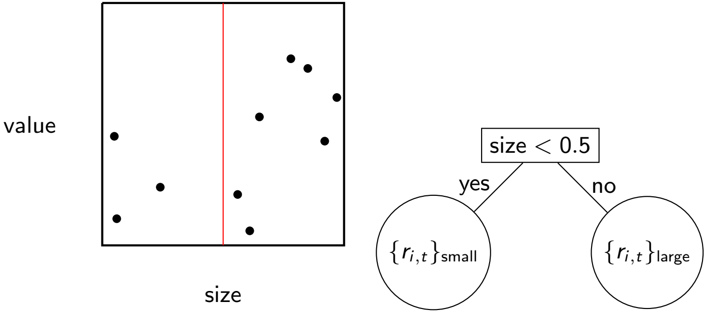


Empirical Highlights

- P-Trees tremendously advance the efficient frontier for exceptionally high annualized SR, ranging from 6.37 for a single P-Tree to 15.63 for 20 boosted P-Trees.
- We identify many unexplained test assets, indicated by most monthly alphas larger than 1%, and an extremely high GRS test statistic of 141.27 for the first P-Tree against FF5.
- We extend P-Tree to Random P-Forest: an over-parameterized ($P > T$) SDF has excellent out-of-sample performance.
- We confirm the same small set of char (e.g., `SUE`, `DOLVOL`, and `BM_IA`) is more likely to proxy for the true fundamental risk inherent in the SDF, which may be overlooked in a linear factor model.

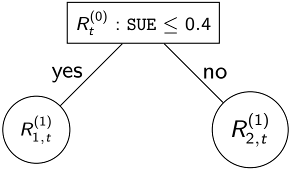
Model: Panel Tree

Splitting the Cross Section



- Pick the split variable and cut point that optimize the split criterion.
- The split criterion is the goal, which can be customized to various studies.
- P-Tree is designed to maximize the collective SR of leaf basis portfolios when splitting or partitioning the cross section.

Grow A P-Tree: First Split — Split Point Candidate

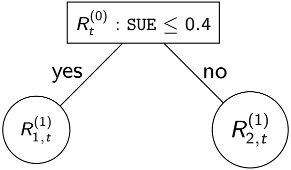


- Before splitting, $R_t^{(0)}$ denote the vector of value weighted portfolio returns of all assets (market) at the root node.

- For example, we use "SUE ≤ 0.4" as a split point candidate.

- $R_{n,t}^{(j)}$ is the leaf-basis portfolio of the n-th leaf node after the j-th split, which is (value weighted) portfolio of the corresponding leaf node.

Grow A P-Tree: First Split — Generating Latent Factor



- After every split, we generate $f_t^{(1)}$ based on leaf basis portfolios, which is a tangency portfolio for $R_t^{(1)} = [R_{1,t}^{(1)}, R_{2,t}^{(1)}]$.

$$f_t^{(1)} \propto \widehat{\Sigma}_1^{-1}\widehat{\mu}_1 R_t^{(1)}.$$

- After j-th split, $f_t^{(j)}$ is the updated latent factor generated jointly with (J+1) test assets $R_t^{(j)}$.

- Each split partitions the cross section, sequentially updating $f_t^{(j)}$ and $R_t^{(j)}$ when maximizing the collective SR.

Grow A P-Tree: First Split — Maximizing Sharpe Ratio

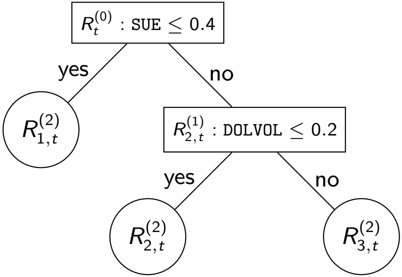
- The goal of test asset construction is to estimate the EF, such that their tangency or MVE portfolio has the maximal SR.

- Therefore, we customize the split criterion by jointly generating latent factors and test assets under the MVE framework.

- CART-type Greedy search, which loops over all char and cut points, is used to avoid NP-hard enumeration.

- Maximizing the collective SR of the basis portfolios is a Global criterion.

Grow A P-Tree: Second Split



- The second split gives us three leaf basis portfolios and an updated $f_t^{(2)}$:

$$f_t^{(2)} \propto \widehat{\Sigma}_2^{-1} \widehat{\mu}_2 R_t^{(2)},$$

- For the second split, the algorithm again searches over all leaf nodes, char, and cut points.
- [Tuning Parameter]: We stop the tree growth by # leaves as test assets should be diversified and representative.

Boosted P-Trees: Multiple Factors

Boosting is a popular ML method for generating additional models to complement previous ones in model combination.

- Generate multiple P-Trees to increase the number of test assets and build a multi-factor model.
- Given benchmark factors $\{f_{1,t}, f_{2,t}, \cdots, f_{P-1,t}\}$.
- To generate the p-th factor $f_{p,t}$, we train a boosted P-Tree model:

$$\max \boldsymbol{\mu}_{\mathbf{F}}' \boldsymbol{\Sigma}_{\mathbf{F}}^{-1} \boldsymbol{\mu}_{\mathbf{F}}$$

  where $\mathbf{F} = [f_{1,t}, f_{2,t}, \cdots, f_{P-1,t}, f_{P,t}]$.
- The p-th factor is generated to complement benchmark factors, which can include pre-specified ones, such as Fama-French factors.

Boosted P-Trees: Multiple Sets of Test Assets

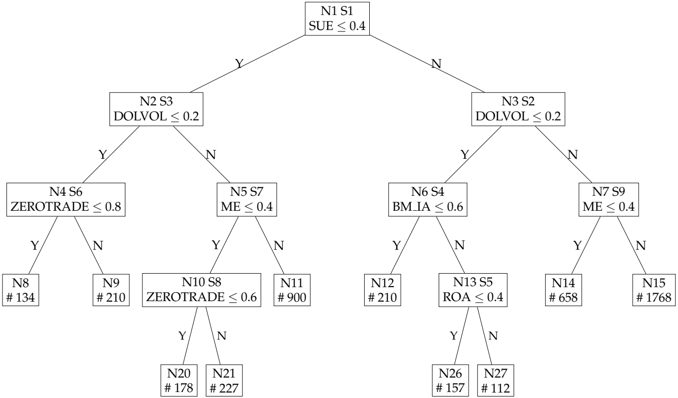
- The additional set of factors (test assets) are generated to complement the benchmark factors (test assets) — achieving the multi-factor efficiency.
- However, combining the increasing sets of leaf basis portfolios may create the high-dimensional problem again.
- The block-diagonal structure of Boosted P-Trees mitigates the issue of high dimensionality.
  - Every tree generates one factor.
  - The multi-factor EF is estimated based on factors.

Empirical Findings

Data: U.S. Equity

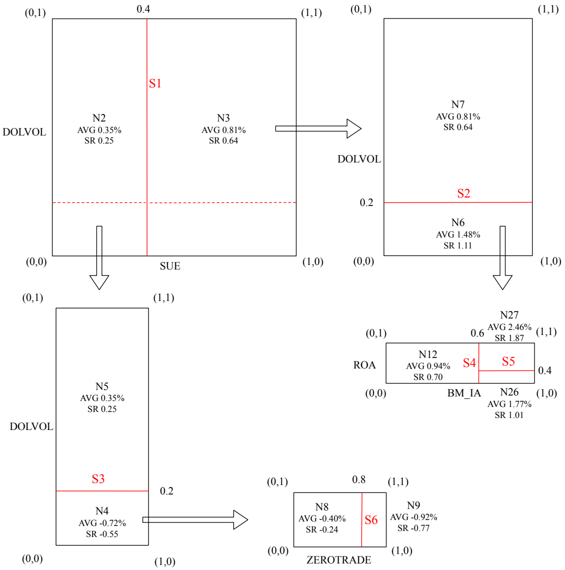
- Returns and lag-one-month char
- Standardize the char in the cross-section into Uniform $[0, 1]$
- 61 char in 6 categories: momentum, value-versus-growth, investment, profitability, intangibles, and frictions
- 1981-2020 monthly observation for US equities
  - 40 years for full sample analysis
  - 20 years (1981-2000) and 20 years (2001-2020) for sub-sample analysis

P-Tree Structure: Graphical ML / Clustering



- Node index, split sequence, char, cut point, portfolio size.
- Split 1: SUE (standard unexpected earnings); Split 2 & 3: DOLVOL (dollar trading volume); Split 4: BM_IA (industry-adjusted book-to-market ratio)

P-Tree Structure: Partitioning the Cross Section



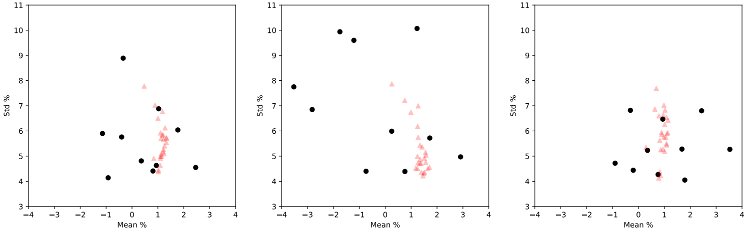
Hard to-price Test Asset

ID	# Median	AVG	α_{CAPM}	β_{CAPM}	R^2_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}	R^2_{FF5}	R^2_{Q5}	R^2_{RP5}	R^2_{IP5}
N8	134	-0.40	-0.97***	0.84	0.42	-0.91***	-0.66***	-1.51***	-0.58*	0.65	0.64	0.73	0.76
N9	210	-0.92***	-1.35***	0.63	0.47	-1.46***	-1.25***	-1.84***	-0.36*	0.68	0.64	0.73	0.78
N20	179	-0.34	-1.34***	1.45	0.53	-1.10***	-0.49*	-1.55***	-1.78***	0.83	0.81	0.89	0.9
N21	227	-1.14***	-1.83***	1.00	0.58	-1.83***	-1.56***	-2.34***	-1.19***	0.81	0.79	0.84	0.88
N11	900	0.36*	-0.35***	1.04	0.93	-0.30***	-0.15**	-0.29***	0.16	0.96	0.96	0.96	0.71
N12	210	0.94***	0.42**	0.76	0.54	0.29**	0.48***	-0.22	1.17***	0.71	0.68	0.76	0.77
N26	157	1.77***	1.16***	0.88	0.43	1.08***	1.37***	0.40**	1.40***	0.62	0.60	0.74	0.79
N27	112	2.46***	1.97***	0.71	0.49	1.80***	2.00***	1.18***	2.18***	0.71	0.67	0.76	0.75
N14	658	1.03***	0.21	1.18	0.59	0.32*	0.64***	-0.33**	0.06	0.84	0.82	0.90	0.95
N15	1768	0.81***	0.14***	0.98	0.98	0.10***	0.03	-0.26***	0.29	0.99	0.99	0.98	0.64

- Smallest (value-weighted) leaves have 100+ for the median portfolio size.
- Two big leaves, N11 (#900) and N15 (#1700+), for high R^2.
- Economically and statistically significant α's against famous factor models.
  - N9 (Low SUE, Low DOLVOL, High ZEROTRADE) — -0.92% average return and -1.35% CAPM α.
  - N27 (High SUE, Low DOLVOL, High BM_IA, High ROA) — 2.46% average return and 1.97% CAPM α.

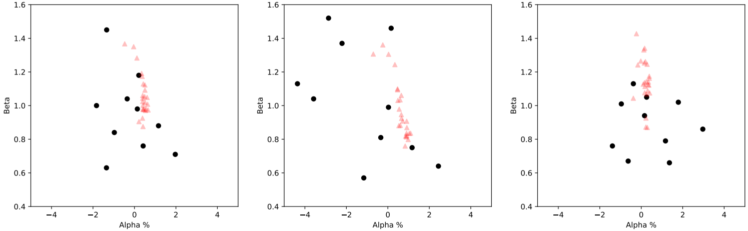
Diversified Test Assets: (P-Tree 10 Assets) v.s. (5 × 5 ME-BM)

Panel A: P-Tree Test Asset: Mean-Standard Deviation



(A1) 1981-2020 (A2) 1981-2000 (A3) 2001-2020

Panel B: P-Tree Test Asset: CAPM Alpha-Beta



(B1) 1981-2020 (B2) 1981-2000 (B3) 2001-2020

Comparing Test Assets against FF5

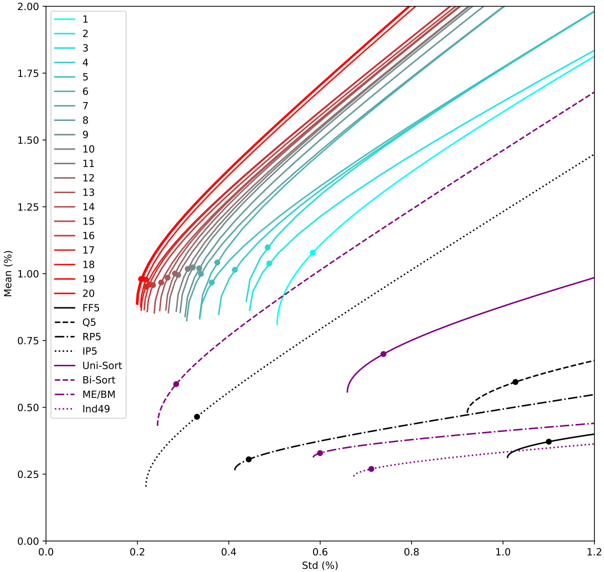
	N	GRS	p-GRS	p-PY	$\overline{\lvert\alpha\rvert}$	$\sqrt{\overline{\alpha^2}}$	$\overline{R^2}$	$\%\alpha_{10\%}$	$\%\alpha_{5\%}$	$\%\alpha_{1\%}$
P-Tree1	10	141.27	0.00	0.00	0.92	1.11	75	100	90	80
P-Tree1-5	50	60.32	0.00	0.00	0.44	0.61	80	70	62	44
P-Tree6-10	50	4.60	0.00	0.00	0.29	0.37	79	56	50	34
P-Tree11-15	50	4.74	0.00	0.00	0.20	0.26	80	38	36	24
P-Tree16-20	50	4.21	0.00	0.00	0.31	0.42	77	52	44	30
P-Tree1-20	200	41.31	0.00	0.00	0.31	0.43	79	54	48	33
Uni-Sort	150	1.62	0.00	0.00	0.10	0.14	88	25	18	7
Bi-Sort	285	2.50	0.00	0.00	0.12	0.17	89	30	23	15
ME/BM	25	5.01	0.00	0.00	0.12	0.16	92	36	28	20
Ind49	49	1.99	0.00	0.00	0.28	0.35	60	39	31	18

- P-Tree test assets have larger and more significant α's than four benchmarks.
- First five P-Trees are hard to price, but alphas are not decreasing too much in the third and fourth sets of five P-Trees.
- GRS and PY (Pesaran and Yamagata, 2023) test for AP models.

$$H_0 : \alpha_i = 0, \quad \forall i = 1, \ldots, N.$$

- Pesaran and Yamagata (2023) accommodate $N > T$ cases.

Characterizing the Efficient Frontier with P-Trees



EF-1 (10 portfolios) achieve higher efficiency than **EF-Bi-Sort** (285 portfolios).

Asset Pricing Model: Cross-sectional R^2

$$\text{Cross-sectional } R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\bar{R}_i - \widehat{\bar{R}}_i\right)^2}{\sum_{i=1}^{N}\left(\bar{R}_i\right)^2}.$$

- Factor models are in the columns.
- Test assets are in the rows.
- PTree factors can price PTree leaf portfolios and classic test assets.

	P-Tree1F	P-Tree5F	P-Tree10F	P-Tree20F	FF5	Q5	RP5	IP5
Panel A: 40 Years (1981-2020)								
P-Tree1	98.3	99.2	–	–	36.8	33.9	51.1	93.4
P-Tree1-5	65.0	82.6	90.2	95.9	54.3	55.9	56.8	66.1
P-Tree1-10	68.5	74.0	87.9	92.1	66.9	70.0	69.1	73.2
P-Tree1-20	72.1	77.9	84.3	88.9	73.8	76.8	77.8	80.3
Uni-Sort	92.7	97.3	98.5	98.7	97.0	97.9	98.0	97.6
Bi-Sort	88.7	97.2	98.0	98.5	96.1	97.5	97.3	97.5
ME/BM	88.4	98.6	98.8	99.4	96.8	97.4	97.0	97.3
Ind49	82.0	92.9	95.3	97.8	96.1	95.9	95.7	92.8

Factor Investing by Boosted P-Trees

	SR	α_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}
Panel A: 40 Years (1981-2020)						
P-Tree1	6.37	1.39***	1.30***	1.36***	1.29***	1.12***
P-Tree1-5	9.18	0.97***	0.91***	0.93***	0.90***	0.82***
P-Tree1-10	11.20	1.01***	0.95***	0.98***	0.96***	0.89***
P-Tree1-15	13.81	0.95***	0.90***	0.93***	0.92***	0.87***
P-Tree1-20	15.63	0.97***	0.93***	0.95***	0.94***	0.90***
Panel B1: 20 Years Train (1981-2000), In-Sample (1981-2000)						
P-Tree1	7.12	1.85***	1.88***	1.71***	1.65***	1.24***
P-Tree1-5	12.71	1.54***	1.59***	1.48***	1.35***	1.27***
P-Tree1-10	19.18	1.49***	1.49***	1.47***	1.39***	1.35***
P-Tree1-15	28.37	1.42***	1.39***	1.39***	1.36***	1.36***
P-Tree1-20	37.85	1.34***	1.34***	1.32***	1.29***	1.29***
Panel B2: 20 Years Train (1981-2000), Out-of-Sample (2001-2020)						
P-Tree1	3.24	1.34***	1.31***	1.23***	1.01***	0.91***
P-Tree1-5	3.41	1.02***	1.02***	0.95***	0.79***	0.63***
P-Tree1-10	3.21	0.94***	0.91***	0.87***	0.76***	0.57***
P-Tree1-15	3.12	0.89***	0.89***	0.83***	0.71***	0.49***
P-Tree1-20	3.13	0.83***	0.82***	0.77***	0.66***	0.49***

- Multi-factor P-Tree strategies provide unexplained α's against multiple models.
- Results are consistently positive for past-to-future out-of-sample analysis.

Virtue of Complexity: Literature

- "SURPRISES IN HIGH-DIMENSIONAL RIDGELESS LEAST SQUARES INTERPOLATION" Hastie et al. (2022) AoS.
  - Randomness.
  - High-dimension.
- Double Decent: prediction errors decrease, increase, and decrease again with growing model complexity (number of parameters).
- Classic statistics/econometrics literature prefers the wisdom of parsimony.
- Neural networks, deep learning, and large language models are examples of high-dimensional models, with excellent prediction accuracy.
- The virtue of complexity is swiftly applied in finance for return prediction (Kelly et al., 2024) and spanning the SDF (Didisheim et al., 2024), i.e., extremely large models have excellent oos performance.

Virtue of Complexity and Random P-Forest

- Tree has a natural extension to large models, the forest models.
- We construct the Random P-Forest as follows:
  1. randomly draw L characteristics for one P-Tree, grow the P-Tree, and store the leaf basis portfolio returns.
  2. repeat the procedure B times to get a forest of B P-Trees, and this operation allows parallel computing.
  3. estimate the SDF weight on all the leaf basis portfolios in the Random P-Forest with the above equation based on in-sample data.

$$\hat{w} = (\gamma \mathbf{I} + \hat{E}[\mathbf{R}\mathbf{R}'])^{-1} \hat{E}[\mathbf{R}]$$

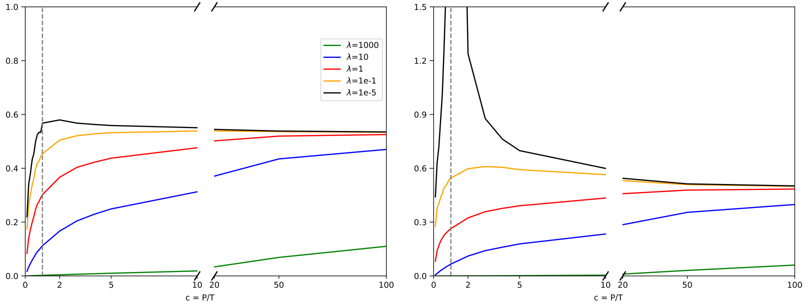
  4. report the out-of-sample performance metrics of the empirical SDF, as complexity grows.

$$\text{Complexity } c = P/T$$

  5. Discuss the role of split criteria.
     - Random P-Forest SDF.
     - Random Split SDF.

Random P-Forest SDF

Out-of-sample Performance of Random P-Forest SDF, #Char = 5

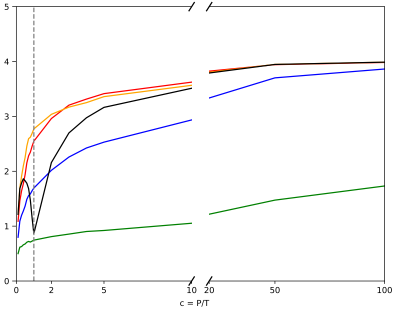


(a) Expected Return (b) Second Moment

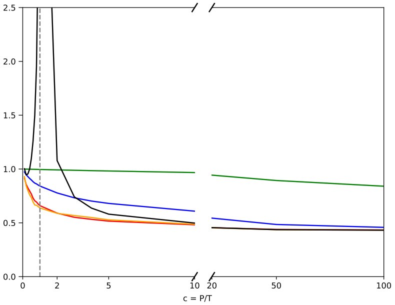
- λ is the ridge shrinkage parameter.
- the average return $E[F_t]$ rises and then be stable.
- the second moment $E[F_t^2]$ spikes as c approaches 1 and decreases afterward for low shrinkage cases, and the second moment increases monotonically for high shrinkage cases.

Random P-Forest SDF

Out-of-sample Performance of Random P-Forest SDF, #Char = 5



(c) Sharpe Ratio

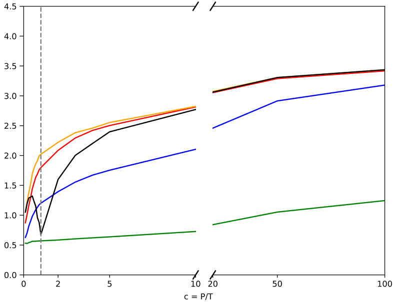


(d) OOS HJ Distance

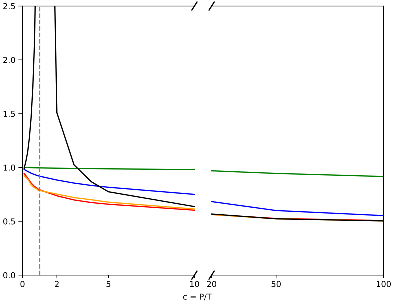
- the Sharpe ratio $E[F_t]/\text{sd}[F_t]$ exhibits Double Ascent for low shrinkage cases and permanent ascent for high shrinkage cases.
- the OOS HJ Distance decrease for high shrinkage cases, with a spike around $c = 1$ for low shrinkage cases.

Random Split SDF: A Comparison

Out-of-sample Performance of SDF via Random Split P-Tree, #Leaf = 10



(c) Sharpe Ratio



(d) OOS HJ Distance

- Random Split SDF
  - Pick split candidate randomly, similar to Didisheim et al. (2024).
- Random P-Forest SDF
  - Pick split candidate to Maximize Sharpe ratio.
- Random Split SDF is less efficient than Random P-Forest SDF, i.e., lower Sharpe ratio and larger OOS HJ Distance with the same c.

Random P-Forest SDF: the Role of #Char = 5,10,20,30,40

	λ=1000	λ=10	λ=1	λ=1e-1	λ=1e-5
			Panel A: Sharpe Ratio		
SR+5	1.05	2.94	3.62	3.57	3.51
SR+10	1.25	3.55	4.22	4.14	4.07
SR+20	1.46	4.08	4.75	4.52	4.38
SR+30	1.57	4.13	4.72	4.34	4.10
SR+40	1.70	4.12	4.53	3.93	3.48
			Panel B: Pricing Error		
SR+5	0.97	0.61	0.48	0.49	0.50
SR+10	0.95	0.51	0.41	0.41	0.42
SR+20	0.92	0.44	0.35	0.37	0.39
SR+30	0.91	0.43	0.35	0.39	0.43
SR+40	0.89	0.42	0.37	0.46	0.57

- #Char is recommend $\sqrt{K}$ or $K/3$ for classification and regression Trees Hastie et al. (2009).
- In this study, #Char = 20 is good enough.

Summary

Summary

Estimating the mean-variance efficient frontier and generating diversified test assets using individual asset returns remain long-term challenges in empirical asset pricing (e.g., Markowitz, 1952; Lewellen et al., 2010; Daniel et al., 2020).

- generate test assets with the maximal collective SR
- generate risk factors that explain the cross section
- KEY: utilize high-dimensional firm char, which contain rich information on the joint distribution of asset returns dynamics

P-Tree offers the first coherent general approach for creating diversified basis portfolios under the MVE framework.

(i) Clustering based on high-dimensional, interactive, and nonlinear char space

—— Graphically display important char by generalizing dependent sorting.

(ii) Diversified test assets and factors that span the efficient frontier

—— Outperform traditional ones with exceptional collective SR.

(iii) Random P-Forest SDF accommodates the virtue of complexity.
—— Sharpe ratio criteria makes SDF Sharpe ratio increase fast.

Ang, A., J. Liu, and K. Schwarz (2020). Using Stocks or Portfolios in Tests of Factor Models. *Journal of Financial and Quantitative Analysis 55*, 709–750.

Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. *Journal of Finance, Forthcoming*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance 66*(4), 1047–1108.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020). The cross-section of risk and returns. *Review of Financial Studies 33*(5), 1927–1979.

Didisheim, A., S. Ke, B. T. Kelly, and S. Malamud (2024). Complexity in factor pricing models. Technical report, Yale University.

Fama, E. F. (1996). Multifactor portfolio efficiency and multifactor asset pricing. *Journal of Financial and Quantitative Analysis 31*(4), 441–465.

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 (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance 75*(3), 1327–1370.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica*, 1121–1152.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*(2), 225–262.

Hastie, T., A. Montanari, S. Rosset, and R. J. Tibshirani (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Annals of statistics 50*(2), 949.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.

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

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *Journal of Finance 73*(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*, 77–99.

Pesaran, M. H. and T. Yamagata (2023). Testing for alpha in linear factor pricing models with a large number of securities. *Journal of Financial Econometrics*.