

NBER WORKING PAPER SERIES

GROWING THE EFFICIENT FRONTIER ON PANEL TREES

Lin William Cong
Guanhao Feng
Jingyu He
Xin He

Working Paper 30805
<http://www.nber.org/papers/w30805>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

December 2022, revised December 2023

The authors acknowledge financial support from INQUIRE Europe and the Hong Kong Research Grants Council (Project No. CityU 11507022). Will Cong is also an advisor to BlackRock, Inc. We are grateful to Fahiz Baba-Yara (discussant), Svetlana Bryzgalova (discussant), John Campbell, Andrei Goncalves, Bing Han, Kewei Hou, Bob Jarrow, Ron Kaniel, Chris Malloy, Alberto Rossi (discussant), Gideon Saar, Shrihari Santosh (discussant), Peixuan Yuan (discussant), and Guofu Zhou for detailed comments and suggestions. We also thank Doron Avramov, Jie Cao, Zhanhui Chen, Eric Ghysels, P. Richard Hahn, Dashan Huang, Sophia Zhengzi Li, Asaf Manela, Stefan Nagel, Adam Reed, Artem Streltsov, Yinan Su, Fred Sun, Yaki Tsaig, Mao Ye, Dacheng Xiu, Bobby Yu, and seminar and conference participants at ArrowStreet Capital, Baylor University, BUAA, Ca' Foscari University of Venice, Cambridge University Algorithmic Trading Society Quant Conference, CFTRC 2022, CityUHK, CKGSB, Conference on FinTech, Innovation and Development at Hangzhou (2nd), Cornell University, Goethe University Frankfurt, GSU-RFS FinTech Conference 2022, HKAIIFT-Columbia joint seminar, Hunan University, INFORMS Annual Meeting 2021, 4th International FinTech Research Forum (RUC), 2023 Mid-South DATA Conference (Memphis), KAIST Digital Finance Conference, NFA 2022, University of Oxford SML-Fin Seminar, OSU, PKU Guanghua, PKU-NSD, PKU-NUS Annual International Conference on Quantitative Finance and Economics, Qube Research and Technology, Reichman University (IDC Herzliya), Schrodgers Investments, Shanghai Financial Forefront Symposium (3rd), SHUFE, SMU, SUSTech, TAMU, Tsinghua SEM, UNC, University of Bath, University of Hawai'i, USC, USTC, University of Macau, World Online Seminars on Machine Learning in Finance, WFA 2022, and 2023 XJTLU AI and Big Data in Accounting and Finance Conference for constructive discussions and feedback. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Lin William Cong, Guanhao Feng, Jingyu He, and Xin He. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Growing the Efficient Frontier on Panel Trees
Lin William Cong, Guanhao Feng, Jingyu He, and Xin He
NBER Working Paper No. 30805
December 2022, revised December 2023
JEL No. C1,G11,G12

ABSTRACT

We develop a new class of tree-based models (P-Trees) for analyzing (unbalanced) panel data using economically guided, global (instead of local) split criteria that guard against overfitting while preserving sparsity and interpretability. To generalize security sorting and better estimate the efficient frontier, we grow a P-Tree top-down to split the cross section of asset returns to construct test assets and re-cover the stochastic discount factor under the mean-variance efficient framework, visualizing (asymmetric) nonlinear interactions among firm characteristics. When applied to U.S. equities, boosted (multi-factor) P-Trees significantly advance the efficient frontier relative to those constructed with established factors and common test assets. P-Tree test assets are diversified and exhibit significant unexplained alphas against benchmark models. The unified P-Tree factors outperform most known observable and latent factor models in pricing cross-sectional returns, delivering transparent and effective trading strategies. Beyond asset pricing, our framework offers a more interpretable and computationally efficient alternative to recent machine learning and AI models for analyzing panel data through goal-oriented, high-dimensional clustering.

Lin William Cong
SC Johnson College of Business
Cornell University
Sage Hall
Ithaca, NY 14853
and NBER
will.cong@cornell.edu

Guanhao Feng
City University of Hong Kong
83 Tat Chee Avenue
Hong Kong, China
gavin.feng@cityu.edu.hk

Jingyu He
City University of Hong Kong
83 Tat Chee Avenue
Kowloon Tong
Hong Kong, China
jingyuhe@cityu.edu.hk

Xin He
College of Finance and Statistics
Hunan University
Changsha, China
xinh@hnu.edu.cn

A P-Tree Github Package is available at <https://github.com/Quantactix/TreeFactor>

1 Introduction

Estimating the mean-variance efficient (MVE) frontier remains crucial in asset pricing and investment management. However, attempting to reach the efficient frontier by estimating the tangency portfolio from the high-dimensional, unbalanced panel of individual asset returns proves impracticable (Cochrane, 2009). Empirical asset pricing studies address the challenge by considering a “diversified” set of assets as basis portfolios and identifying a few common factors to achieve multi-factor portfolio efficiency (Fama, 1996). These approaches for using left-hand-side test assets and right-hand-side risk factors are statistically equivalent in terms of spanning the efficient frontier (Gibbons, Ross, and Shanken, 1989), because under the MVE framework with zero alphas, the collective Sharpe ratios of test assets equal those of risk factors.

However, popular factor models hardly span the common test asset universe (e.g., Kozak et al., 2018; Daniel et al., 2020; Lopez-Lira and Roussanov, 2020), let alone the ultimate efficient frontier constructed by individual asset returns.¹ The empirical literature mostly uses ad hoc characteristics-sorted portfolios as test assets (e.g., Fama and French, 1993, 1996), which are typically limited to univariate- and bivariate-sorted approaches, due to the difficulties of high-dimensional sorting (Cochrane, 2011). Even with dependent sorting (e.g., Daniel et al., 1997), researchers leave out nonlinearity and asymmetric interactions, instead focus on interactions uniformly applicable to all assets. The inability to capture the efficient frontier hampers the effectiveness of model estimations and evaluations (Lewellen et al., 2010; Ang et al., 2020). 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 characteristics that contain rich information on the joint distribution of asset returns dynamics.²

To this end, we introduce the Panel Tree (P-Tree) framework, a new and versatile model family that clusters panel observations to achieve given economic objec-

¹To improve the estimation, numerous characteristics are considered in constructing characteristics-managed factors (Kozak and Nagel, 2022) or machine-learning-based latent factors (Kelly et al., 2019; Lettau and Pelger, 2020b; Chen et al., 2023), despite interpretation challenges.

²Cochrane (2011) asserts that expected returns, variances, and covariances are stable functions of characteristics (also see, e.g., Kelly et al., 2019; Kozak et al., 2020). Based on high-dimensional characteristics, Feng et al. (2023) estimate the tangency portfolio via deep learning.

tives. 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. When applied to the panel of individual asset returns, P-Trees utilize high-dimensional firm characteristics under the MVE framework to jointly generate diversified test assets and latent factors for recovering the stochastic discount factor (SDF). Therefore, they extend the scope of tree models beyond pure prediction tasks, particularly expanding the empirical efficient frontier and constructing diversified basis portfolios, with rich asset pricing and investment implications.³

Specifically, 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 characteristic values to form value-weighted portfolios. Guided by asset pricing considerations, we grow P-Trees following global split criteria to construct diversified test assets and latent factor models, estimating the efficient frontier. According to [Hansen and Jagannathan \(1991\)](#), the resulting SDF that maximizes the Sharpe ratio by spanning the efficient frontier also minimizes the HJ-distance for explaining the cross section. Generated under the unified framework, the resulting leaf basis portfolios provide empirical researchers with a diversified set of test assets for model evaluation. Furthermore, the decision trees have an intuitive layout, allowing visualized economic interpretability and identifying a sparse set of useful firm characteristics that interact to jointly generate test assets and latent factors.

Fundamentally, P-Tree is a greedy algorithm guided by economic goals for optimal clustering based on firm characteristic values within a large clustering space. In asset pricing, basis portfolios are created by grouping assets based on classifications such as country or industry, or through unsupervised clustering based on the correlation of returns or firm characteristics. By contrast, P-Trees generate characteristics-managed leaf basis portfolios to maximize the collective Sharpe ratio of the resulting

³Tree-based models excel in predicting complex data with high dimensionality, nonlinearity, and variable interactions, even in low signal-to-noise environments and small sample sizes (e.g., [Sorensen et al., 2000](#); [Rossi and Timmermann, 2015](#); [Gu et al., 2020](#); [Bianchi et al., 2021](#); [Bryzgalova et al., 2023](#)). Decision trees are easy to visualize and interpret, but no algorithm currently customizes tree generation for security sorting and constructs test assets. Most off-the-shelf machine learning (ML) methods, including tree-based models, are designed for predicting tasks on independent and identically distributed (i.i.d.) observations and are not guided by economic principles or tailored for financial panel data.

MVE portfolios, without being either supervised or unsupervised. The objective can be flexibly specified, making the P-Tree framework applicable for a wide range of panel data analyses.⁴

Methodological innovations. Our emphasis on the asset pricing application effectively demonstrates the key methodological innovations. First, off-the-shelf ML methods, including the famous Classification and Regression Trees (CART, Breiman et al., 1984), typically assume that observations are independent and identically distributed (i.i.d.), and not designed for extracting heterogeneous information from panel data. Though naively fitting CART, boosted trees, or random forests on characteristics-return data shows positive performance (e.g., Gu et al., 2020), they ignore the panel data structure. By contrast, P-Tree fully embraces it and incorporates vectorized leaf parameters tailored for multi-period observations, enabling a time-invariant tree structure for all periods. The time-invariant P-Tree allows for interpretability when building ML models on panel data. As P-Tree grows, its nodes evolve into vectorized values of basis portfolio returns, with the P-Tree factor being their MVE portfolio.

Second, the standard tree-based models in ML, including CART, focus purely on prediction. Furthermore, CART-based models grow recursively, optimizing the quadratic loss of prediction. These models optimize locally at each node without considering sibling nodes, mainly for computational efficiency. However, this “myopic” strategy often leads to overfitting, because it operates on fewer observations in each node as the tree grows. In contrast, P-Trees broaden the applications beyond prediction, encompassing the generation of test assets and latent factors and explaining cross-sectional returns. With an economics-guided optimization objective, the iterative growth of P-Trees is designed to utilize data from the entire cross section to guard against overfitting. P-Tree uses a nonlinear ML approach to generate test assets in asset pricing, maximizing their collective Sharpe ratio with global split criteria.

⁴We have developed and shared the P-Tree statistical package, `TreeFactor`, in a public repository in R for other researchers to explore (see <https://github.com/Quantactix/TreeFactor>). For instance, we can adopt different investment objectives, such as constructing the minimum variance portfolio using individual asset returns, which can be achieved by modifying the split criteria and latent factor generation. We can also consider the global split criteria to minimize aggregated pricing errors for pricing individual assets, with further details provided in Appendix III.

The approach combines economic principles and statistical properties while ensuring transparency and interpretability of individual trees. P-Tree offers the first coherent general approach for creating diversified basis portfolios under the MVE framework.

Third, P-Tree seamlessly integrates with ML's boosting strategy, enabling multiple factors to demonstrate exceptional asset pricing and investment performance. As detailed in section 2.3, boosted and multi-factor P-Trees expand the set of assets and factors by exploiting the unspanned efficient frontier under the MVE framework. The multi-factor efficiency of our boosted P-Trees enjoys a natural block-diagonal covariance structure in portfolio construction, achieved by grouping individual assets into P-Tree leaf test assets, without estimating a large covariance matrix of thousands of test assets. Furthermore, P-Trees can assess incremental information on benchmark factor models, such as the Fama-French factors, for interactive and nonlinear modeling of additional characteristics. Overall, when boosted to generate multiple factors, P-Trees offer an alternative to PCA methods, with greater interpretability and sparsity, while capturing (asymmetric) interactions.⁵

Empirical findings. First, we apply P-Tree to monthly U.S. individual stock returns from 1981 to 2020 using 61 firm characteristics. P-Trees significantly advance the efficient frontier with high annualized Sharpe ratios, ranging from 6.37 for a single P-Tree (with 10 test assets) to 15.63 for 20 boosted P-Trees (with 200 test assets). These numbers are substantially higher than those constructed by conventional basis portfolios (hundreds of univariate- or bivariate-sorted portfolios) or commonly used factor models. These findings provide strong evidence of a significant gap between the current empirical literature and the potential limits of the efficient frontier. Moreover, generated under the unified MVE framework, boosted P-Trees produce multi-factor models that effectively price the cross-sectional returns and offer annualized Sharpe ratios over 3 out-of-sample in the recent 20 years.

Second, these diversified P-Tree test assets span the efficient frontier and pose a significant challenge for alternative factor models. We identify many economically

⁵For instance, P-Trees might choose ME to split low-volatility stocks and BM to split high-volatility stocks, providing a more nuanced analysis than conventional symmetric sorting, which applies the same characteristic interactions to all.

and statistically unexplained test assets, indicated by most monthly alphas larger than 1%, and an extremely high GRS test statistic of 141.27 (aggregated weighted pricing error) for the first P-Tree against the Fama-French five-factor (FF5) model. These asymmetric, interactive, and nonlinear characteristics-managed leaf basis portfolios are difficult to explain using other well-known observable factor models (e.g., [Hou et al., 2021](#)) and ML-based latent factor models (e.g., [Kelly et al., 2019](#); [Lettau and Pelger, 2020b](#)). Given the insufficient MVE spanning and low hurdles of univariate- or bivariate-sorted portfolios for model testing, we recommend including these multi-characteristic, systematically sorted P-Tree test assets in any future model evaluation.

Finally, we design the random P-Forest with bagging to evaluate characteristic importance, account for variable selection risk, and validate transparent and interpretable P-Tree models. We confirm the same small set of characteristics (e.g., `SUE`, `DOLVOL`, and `BM_LIA`) selected by the P-Tree algorithm is more likely to proxy for the true fundamental risk inherent in the SDF, which may be overlooked in a linear factor model with ad hoc selected factors. In addition, P-Tree offers a novel approach to assess the incremental information of newly added characteristics by controlling for baseline characteristics of size and value, uncovering some characteristics associated with investment, profitability, frictions, and intangibles that contain incremental signals. Last but not least, P-Trees display all split rules, aiding researchers in understanding the interactions among firm characteristics graphically.⁶

Literature. Foremost, our paper relates to the selection and construction of test assets used for asset pricing model evaluations. The empirical literature, including regularized linear models (e.g., [Ao et al., 2019](#); [Kozak et al., 2020](#); [Bryzgalova et al., 2023](#)) and the PCA approaches (e.g., [Kozak et al., 2018](#); [Lettau and Pelger, 2020b](#); [Kim et al., 2021](#)), largely takes pre-specified or characteristics-managed portfolios as given. Some evaluate different sets of test assets separately for robustness checks (e.g., [Fama and French, 1996, 2016](#)), select test assets (e.g., [Daniel et al., 2020](#); [Giglio et al., 2022](#); [Bryz-](#)

⁶P-Trees allow the long and short legs of a long-short portfolio to interact with different characteristics, thereby loading the portfolio on different leaf basis portfolios. This asymmetric interactive sorting contrasts with the traditional treatment of a long-short portfolio as a single asset and complements the pioneering work of [Jarrow et al. \(2021\)](#) in modeling the two legs of anomaly portfolios separately.

galova et al., 2023), or combine test assets (e.g., [Feng et al., 2020](#); [Giglio and Xiu, 2021](#)).

P-Trees constitute the first goal-oriented, systematic clustering of individual securities and generation of leaf basis portfolios (test assets) under the MVE framework. [Hoberg and Welch \(2009\)](#) propose constructing factors and test portfolios by optimizing objective functions instead of sorting. For unsupervised clustering, [Ahn et al. \(2009\)](#) generates basis portfolios by modeling within- and between-group return correlations of individual stocks, [Chen et al. \(2017\)](#) groups hedge fund alpha values for performance evaluation, and [Patton and Weller \(2022\)](#) groups assets in the cross section based on their heterogeneous risk premia on risk factors. We join [Cong et al. \(2022\)](#) as the earliest studies that iteratively cluster panel data by maximizing specific economic objectives such as spanning the efficient frontier.

Including a large set of test assets presents empirical challenges, such as large portfolio estimation or weak factors ([Giglio et al., 2022](#)), leading to the development of new methods for selecting test assets in multiple studies (e.g., [Daniel et al., 2020](#); [Zhu et al., 2020](#)). Notably, [Bryzgalova, Pelger, and Zhu \(2023\)](#) shrinks useless assets for estimating the optimal portfolio on a large number of basis portfolios and links security sorting to a collection of decision trees. The authors focus on pruning or regularizing the tree bottom-up without specifying a split criterion for tree growth and manually specify a small set of split candidates and a shallow depth for initial trees (since enumerating all possible tree configurations to prune from is computationally infeasible). P-Trees differ by growing top-down to provide a goal-oriented clustering approach and efficiently scanning the large space of generalized sequential sorting on high-dimensional characteristics, allowing for asymmetric interactions.

Moreover, our study contributes to the growing literature on latent factor models in asset pricing. Early works apply principal component analysis (PCA) to extract latent factors with a purely statistical criterion, such as [Chamberlain and Rothschild \(1983\)](#) and [Connor and Korajczyk \(1986, 1988\)](#). For recent developments, the projected PCA ([Kim et al., 2021](#)) and the instrumental PCA ([Kelly et al., 2019, 2023](#)) use firm characteristics as instruments for modeling the time-varying factor loadings and estimating principal components. [Lettau and Pelger \(2020a,b\)](#) develop the risk premia

PCA and provide a regularized estimation for risk premia. Recent studies, such as the auto-encoder (Gu et al., 2021), generative adversarial network (Chen et al., 2023), characteristics-sorted factor approximation (Feng et al., 2023), and structural neural network (Fan et al., 2022), have also developed customized deep learning models for nonlinear latent factor modeling. Following these developments, P-Trees provide a graphical representation for variable nonlinearity and asymmetric interactions, which PCA or deep learning methods do not offer.

Finally, ML and tree-based methods are widely adopted in finance due to their powerful nonlinear modeling capabilities using high-dimensional characteristics (e.g., Gu et al., 2020; Bianchi et al., 2021; van Binsbergen et al., 2023; Leippold et al., 2022; Bali et al., 2023). However, most studies either apply off-the-shelf ML models for prediction without considering the panel data structure or incorporating economic guidance, or focus on supervised or un-supervised learning. Recent exceptions include Cong et al. (2022), who combine Bayesian spike-and-slab priors and tree models for cross-sectional asset pricing, Cong et al. (2021), who develop the first “large” model in finance for portfolio management using reinforcement learning, and Creal and Kim (2023), who adapt Bayesian trees to currency returns. We also contribute to recent studies advocating imposing economic restrictions when estimating and evaluating ML or factor models (e.g., Gagliardini et al., 2016; Avramov et al., 2023).

The remainder of the paper follows: Section 2 introduces the P-Tree framework for the economic-guided split criteria. Section 3 discusses the implementation of P-Tree, estimation performance on the efficient frontier, and generation of diversified test assets. Section 4 introduces the boosted and multi-factor P-Trees and compares performance. Section 5 interprets P-Trees for character importance, demonstrates another application, and evaluates robust performance across macro states. Section 6 concludes. Simulation exercises demonstrating P-Trees’ superiority can be found in Internet Appendix I. Internet Appendices II and III illustrate P-Tree’s flexibility by allowing adjustments to benchmark factors and using alternative economic objectives.

2 Panel Tree for Asset Pricing

P-Tree is designed for economic panel data and expands tree models beyond predictions. Section 2.1 reviews CART and motivates the P-Tree framework. Section 2.2 delves into the growth of P-Trees, before section 2.3 introduces boosted P-Trees for a multi-factor model. Section 2.4 briefly presents the simulation settings and findings.

2.1 CART and P-Tree Innovations

The standard CART (Breiman et al., 1984) model and its variants use “divide-and-conquer” strategies for making predictions.⁷ CART partitions the predictor space into distinct, non-overlapping regions of leaf nodes and assigns a constant leaf parameter to each region (Hastie et al., 2009). It searches and records the partition using a decision tree structure with binary split rules. Suppose the data constitute K predictors, the i -th observation is denoted by $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,K})$. The j -th split rule of the decision tree is denoted by $\tilde{\mathbf{c}}^{(j)} = (z_{\cdot,k}, c_j)$, which partitions data by checking whether the value of the k -th predictor $z_{\cdot,k}$ is greater or smaller than cutoff c_j . CART considers only binary splits, since any multiway-split tree can be represented as one with multiple binary splits. The optimal split rule is chosen to minimize the prediction error.

After J splits, CART partitions the predictor space into $J + 1$ leaf nodes denoted by $\{\mathcal{R}_n\}_{n=1}^{J+1}$ and assigns a constant leaf parameter μ_n to each node. The regression tree $\mathcal{T}$, with parameters $\Theta_J = \{\{\tilde{\mathbf{c}}^{(j)}\}_{j=1}^J, \{\mu_n\}_{n=1}^{J+1}\}$, can be expressed as high-dimensional step functions:

$$\mathcal{T}(\mathbf{z}_i | \Theta_J) = \sum_{n=1}^{J+1} \mu_n \mathbb{I}\{\mathbf{z}_i \in \mathcal{R}_n\}. \quad (1)$$

The indicator $\mathbb{I}\{\mathbf{z}_i \in \mathcal{R}_n\}$ takes value 1 for one leaf node and 0 for all others. The leaf parameters, which represent the height of the step function, are estimated by averaging the training data within each leaf node. This nonparametric approach adapts CART to low signal-to-noise environments and small sample sizes. After training, the tree model predicts new observations by locating the corresponding leaf and using its

⁷CART is a binary decision tree model and serves as the foundation for ensemble methods such as random forest (Breiman, 2001) and boosting trees (Freund and Schapire, 1997). Other notable Bayesian tree models include BART (Chipman et al., 2010) and XBART (He et al., 2019; He and Hahn, 2023).

parameter as the prediction.

Figure 1: **Example of a Decision Tree**

Left: A decision tree has two splits, three leaf nodes, and three leaf parameters. Right: The corresponding partition plot is generated for the sample of predictor space on value and size.



For a general example, Figure 1 illustrates how CART uses firm characteristics to predict stock returns.⁸ The characteristics-based prediction model is achieved by recursively partitioning the characteristic space, providing the clustering outcome (small-cap, large-cap value, and large-cap growth). The bottom nodes, often called leaf nodes, are associated with constant leaf parameters, meaning assets within the same node share the same expected return estimate.

Conventional tree-based models such as CART have multiple limitations, rendering them unsuitable for economic or financial panel data and for generating test assets or pricing kernels. First, they treat asset returns as i.i.d in the cross section or the time series. Second, they focus on predictions using statistical local split criteria, such as minimizing the sum of squared error objectives in the child nodes of a particular split. Therefore, they do not incorporate panel structure or economic restrictions. As the tree grows, the number of observations in each node decreases, and idiosyncratic noise causes overfitting. Although ensemble methods such as random forests reduce overfitting, they are less interpretable than single trees.

P-Tree solves both problems by (i) using a vectorized time series of leaf basis portfolio and (ii) using economic objectives that consider all observations instead of just those in a parent node to guide the tree growth. (i) allows us to extract panel data to estimate the efficient frontier thoroughly. (ii) iteratively builds P-Trees using global split criteria to include economic principles and prevent overfitting, extending the scope of trees from pure prediction to generating test assets.

⁸Gu et al. (2020) has a similar Figure 1 for demonstrating tree-based predictive models.

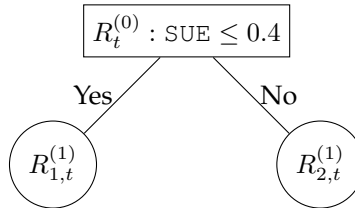
2.2 Growing a P-Tree

P-Tree partitions the universe of individual assets into non-overlapping leaf nodes based on the values of ranked characteristics with a time-invariant structure over T periods, with I_t assets in each period. The leaf basis portfolios are (value-weighted) portfolios of stocks within each leaf node, and their time series of returns are saved in a vectorized leaf parameter. P-Tree splits sequentially from the root node and generates one extra leaf basis portfolio after each additional split. Specifically, it produces $j + 1$ leaf basis portfolios after the j -th split, reducing the dimension from thousands of individual stocks to $j + 1$ portfolios. This concept of characteristics-managed leaf basis portfolios generalizes the popular characteristics-sorted portfolios to allow for asymmetric interactions of multiple characteristics.

Let $\mathbf{R}_t^{(j)}$ denote all return vectors of leaf basis portfolios after the j -th split of the tree, and let $f_t^{(j)}$ denote the factor spanned by $\mathbf{R}_t^{(j)}$. The tree has $j + 1$ basis portfolios in all leaf nodes, denoted by $\mathbf{R}_t^{(j)} = [R_{1,t}^{(j)}, \dots, R_{n,t}^{(j)}, \dots, R_{j+1,t}^{(j)}]$, where $R_{n,t}^{(j)}$ represents a length T vector of returns over T periods for the n -th leaf node. To find the first split, we begin with a portfolio of all assets, denoted as $\mathbf{R}_t^{(0)} = [R_t^{(0)}]$, which serves as the root node for P-Tree. The tree is expanded through iterative updates of $\{\mathbf{R}_t^{(j)}, f_t^{(j)}\}$. First, the leaf basis portfolios $\mathbf{R}_t^{(j)}$ are expanded as the tree grows and splits into additional leaves. Second, using the expanded leaf basis portfolios, we re-estimate the P-Tree factor $f_t^{(j)}$. We use a greedy algorithm to grow the tree sequentially to expedite the process instead of enumerating all possible configurations.

Figure 2: Illustration of the First Split

To determine the optimal characteristic and cutpoint value, we evaluate a list of candidate splits (i.e., idiosyncratic volatility (SUE) ≤ 0.4) for calculating the split criterion value.



First split. The root node contains all assets corresponding to the value-weighted market factor. Firm characteristics are uniformly normalized cross-sectionally to the range

of $[0, 1]$ within each period. We assess different split threshold options c_m for characteristics $z_{\cdot,k}$, such as quintile splits $c_m \in [0.2, 0.4, 0.6, 0.8]$.⁹ Consider a candidate split $\tilde{\mathbf{c}}_{k,m} = (z_{\cdot,k}, c_m)$ (Figure 2). The candidate partitions the root node into left and right child nodes based on whether $z_{\cdot,k}$ is less than or equal to c_m .

The observations of stock returns within each potential child leaf form a leaf basis portfolio, with vectors of return $R_{1,t}^{(1)}$ and $R_{2,t}^{(1)}$, respectively. Note they are vectors of returns, thus maintaining the panel structure of asset returns. The P-Tree factor is estimated as the MVE portfolio of all leaf basis portfolios,

$$f_t^{(1)} = \mathbf{w}^{(1)\top} \mathbf{R}_t^{(1)}, \quad \mathbf{w}^{(1)} \propto \Sigma_1^{-1} \boldsymbol{\mu}_1, \quad (2)$$

where Σ_1 and $\boldsymbol{\mu}_1$ are the covariance matrix and average excess returns, respectively, for leaf basis portfolios $\mathbf{R}_t^{(1)} = [R_{1,t}^{(1)}, R_{2,t}^{(1)}]$ after the first split.¹⁰ Note that generating latent factors is consistent with the P-Tree split criteria, because the P-Tree factor represents the MVE portfolio of all leaf basis portfolios.

Our split criterion aims to maximize the squared Sharpe ratio of the MVE portfolio as follows:

$$\mathcal{L}(\tilde{\mathbf{c}}_{k,m}) = \boldsymbol{\mu}'_{\mathbf{F}} \Sigma_{\mathbf{F}}^{-1} \boldsymbol{\mu}_{\mathbf{F}}, \quad (4)$$

where $\mathbf{F} = f_t^{(1)}$ is the generated latent factor, with mean and covariance $\boldsymbol{\mu}_{\mathbf{F}}$ and $\Sigma_{\mathbf{F}}$, respectively. The criterion allows P-Tree to construct and estimate basis portfolios, latent factors, and efficient portfolio weights simultaneously, for the global objective of estimating the efficient frontier.

Each candidate split $\tilde{\mathbf{c}}_{k,m}$ generates a different partition of the data, resulting in unique leaf basis portfolios, corresponding P-Tree factors, and ultimately, varying valuations of the split criteria in (4). Consequently, we loop over all candidate splits and

⁹Quintile sorting is a popular and efficient method that reduces search space. Our results are robust to different specifications, including denser breakpoints.

¹⁰Following the regularization approaches in Kozak et al. (2020) and Bryzgalova et al. (2023), we utilize small shrinkage parameters, $\gamma_{\Sigma} = 10^{-4}$, for robustly estimating efficient portfolio weights:

$$\mathbf{w}^{(j)} = \left(\text{Cov}(\mathbf{R}_t^{(j)}) + \gamma_{\Sigma} \mathbf{I}_{k+1} \right)^{-1} \mathbf{E}(\mathbf{R}_t^{(j)}), \quad (3)$$

where $\mathbf{I}_{k+1}$ is the identity matrix. Shrinkage parameters stabilize portfolio weight estimation and prevent over-leveraging, which can be adjusted to control the degree of regularization. Also, portfolio weights are normalized such that the sum of absolute weights equals one.

select the one that maximizes the split criteria as our initial split rule.

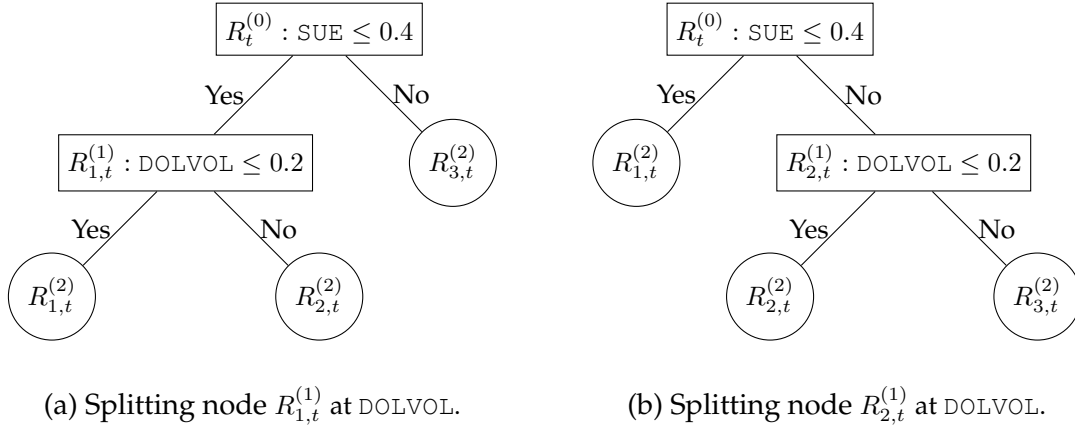
Second split. The second split can occur at the root's left or right child node. We assess the split criteria for all candidate splits for *both* leaf nodes and choose the split that minimizes the criteria as in (4). Figure 3 depicts the tree of the candidates for the second split. In either scenario, one leaf node splits, becoming an internal node and generating two new leaf nodes. The P-Tree factor is then constructed based on *all three* leaf basis portfolios:

$$f_t^{(2)} = \mathbf{w}^{(2)\top} \mathbf{R}_t^{(2)}, \quad \mathbf{w}^{(2)} \propto \Sigma_2^{-1} \boldsymbol{\mu}_2, \quad (5)$$

where Σ_2 and $\boldsymbol{\mu}_2$ are the covariance matrix and average excess returns for leaf basis portfolios $\mathbf{R}_t^{(2)} = [R_{1,t}^{(2)}, R_{2,t}^{(2)}, R_{3,t}^{(2)}]$ after the second split. The construction of the three basis portfolios depends on which node the candidate splits, as shown in Figure 3. The updated P-Tree factor is plugged in to maximize the squared Sharpe ratio, $\mathcal{L}(\tilde{\mathbf{c}}_{k,m}) = \boldsymbol{\mu}'_{\mathbf{F}} \Sigma_{\mathbf{F}}^{-1} \boldsymbol{\mu}_{\mathbf{F}}$, where $\mathbf{F} = f_t^{(2)}$ is the updated P-Tree factor.

Figure 3: Illustration of the Second Split

These two figures illustrate two potential candidates for the second split, where one of the nodes splits. During the second iteration of the tree model, three leaf basis portfolios remain available to produce the latent factor.



Notably, our proposed objective is a *global* split criterion, because it considers all leaf basis portfolios when constructing $\mathbf{F}$ (the global MVE portfolio). Unlike CART, which focuses on a specific leaf node, our model explores all candidate splits in all leaf nodes to find the one with the largest investment improvement.

Growth termination. All subsequent splits are determined in the same manner. The tree-growing procedure is outlined in Algorithm 1. Determining the number of leaves

is a natural turning point in tree growth, which is also the only tuning parameter needed for P-Tree. We consider P-Trees with $J + 1 = 10$ leaf nodes in the baseline specification. Furthermore, we mandate a minimum leaf size of 20 for growing the tree because the leaf needs to serve as a basis portfolio, and any leaves that do not meet this criterion are not subjected to further splitting. Once the tree growing process terminates, it outputs the P-Tree split sequence, leaf basis portfolios $\mathbf{R}_t^{(J)} = [R_{1,t}^{(J)}, \dots, R_{J+1,t}^{(J)}]$, and the P-Tree factor $f_t^{(J)}$. Note each leaf portfolio can be interpreted through economic fundamentals revealed by the sequential splits.

Other economically guided criteria The P-Tree framework is flexible to accommodate objectives alternative to the global split criteria in (4) for estimating the MVE portfolio using individual asset returns. For instance, one might seek to create the minimum-variance portfolio or construct latent factors that best explain test assets, such as individual assets or basis portfolio returns. The key lies in specifying a clear economic goal and using the greedy growth algorithm embedded in trees to search the large observation clustering space to optimize the objective effectively.

In Internet Appendix III, we illustrate P-Tree’s flexibility using a split criterion based on realized asset pricing errors after estimating a latent factor model with time-varying beta. The asset pricing goal of minimizing pricing errors shares the same spirit in the recent literature for latent factor generation via deep learning (e.g., Gu et al., 2021; Chen et al., 2023; Feng et al., 2023). We run a regression (factor model) for the tree structure and provide a conditional expected value for the leaf parameter. Such P-Trees again produce superior investment and pricing performances.

Regime detection by macro variables The panel data has multiple dimensions, so besides splitting the cross section, one can also split the time series, which is an alternative to structural break modeling in econometrics. For future flexible extensions, P-Tree can be extended to consider different macroeconomic regimes and detect those regimes when recursively optimizing the asset pricing objective, fitting regime-specific P-Trees. The data-driven time-series splits of P-Tree are shown by detecting macro-instrumented regimes for cross-sectional asset pricing in Appendix III.D.

2.3 Boosted P-Trees

Boosting is an ML technique that combines weak learners to form a strong learner (Freund and Schapire, 1997; Rossi and Timmermann, 2015). Boosted P-Trees sequentially grow a list of additive trees that augment previous trees under the MVE framework, which helps further span the efficient frontier with additional basis portfolios and generate multi-factor models. Considering existing factors, the boosting procedure seamlessly integrates into the P-Tree framework by creating additional P-Tree factors to maximize the collective Sharpe ratio of the tangency portfolio. The boosting procedure is outlined below.

1. The initial factor, denoted as $f_{1,t}$, can be either generated by the first single P-Tree or benchmark factors selected by researchers (e.g., the market factor).
2. The second factor $f_{2,t}$ is generated by the second P-Tree to complement the first factor and maximize the squared Sharpe ratio. The second tree growth follows a similar boosting split criterion:

$$\mathcal{L}(\tilde{\mathbf{c}}_{k,m}) = \boldsymbol{\mu}_{\mathbf{F}}' \boldsymbol{\Sigma}_{\mathbf{F}}^{-1} \boldsymbol{\mu}_{\mathbf{F}}, \quad (6)$$

where $\mathbf{F} = [f_{1,t}, f_{2,t}]$. The MVE portfolio is generated from all the factors directly.

3. Repeat Step 2 to sequentially generate P factors $\mathbf{f}_t = [f_{1,t}, \dots, f_{P,t}]$ until the stopping criteria are triggered (e.g., one finds insignificant squared Sharpe ratio improvement Barillas and Shanken, 2017).

The boosted P-Tree utilizes a two-step process to produce either the MVE portfolio or the SDF by incorporating multiple sets of leaf basis portfolios. Initially, the tree-specific MVE portfolio of all leaf basis portfolios within each P-Tree is computed to determine the current P-Tree factor. Subsequently, this factor is merged with all previous P-Tree factors or benchmarks to create an all-tree MVE portfolio for the multi-factor P-Tree model. Boosted P-Trees indirectly impose regularization on the all-tree MVE portfolio weights through the sequential tree block structure, even though the

resulting Sharpe ratio may not surpass the one generated directly from all basis portfolios in all P-Trees. This avoids high-dimensional estimation issues and, in turn, leads to robust estimation and out-of-sample performance.

2.4 Simulation Evidence

Simulation studies in this section demonstrate the ability of P-Tree models to select characteristics and the effectiveness of P-Tree test assets in capturing non-linear interactions among characteristics. In Appendix I, we calibrate the return-generating process on the same empirical data sample, considering three true characteristics, their interactions, and nonlinear terms. Additionally, we include a large set of redundant characteristics that are correlated with the true characteristics but do not drive returns.

First, we compare the Sharpe ratios and alphas of P-Tree test assets with conventional characteristics-sorted portfolios, which cannot model the nonlinear characteristics interactions. The single P-Tree and boosted P-Trees consistently produce leaf basis portfolios with higher Sharpe ratios and larger unexplained pricing errors under various signal-to-noise-ratio levels, both in-sample and out-of-sample. Even though sorting is performed on the true characteristics directly, univariate- and bivariate-sorted portfolios still underperform P-Tree test assets generated on high-dimensional characteristics. This result confirms our empirical findings that P-Tree test assets achieve better mean-variance efficiency than sorted portfolios as in Section 4.2.

Second, we demonstrate that P-Tree helps select the sparse true characteristics and a large number of redundant or useless ones. The interpretability of the single P-Tree relies on its ability to select variables. With repeated simulations, we show high frequencies for P-Tree for selecting the true characteristics in the first few splits. These correct selection rates are higher for cases with a high signal-to-noise ratio. This finding can only be identified from simulations, because we remain agnostic regarding the true sparse characteristics in the real-world data. The consistent selection properties are related to evaluating characteristics' importance in Section 5.1.

Finally, we evaluate the efficiency loss without observing all true characteristics but still observe a large set of redundant or useless characteristics. This exercise evaluates the model's performance with incomplete or useless information, which may

occur for various reasons. Under our calibrated return-generating process, the Sharpe ratios and CAPM alphas decline substantially, and P-Tree selects redundant or useless characteristics more frequently. Therefore, when a large set of characteristics is provided, we should include all of them rather than performing ad hoc variable selection in empirical studies. For more details, please refer to Appendix I.

3 Splitting the Cross Section

The P-Tree framework easily applies to U.S. public equities or corporate bonds, focusing on the former in this paper. In an era of emergent large models, P-Tree offers an interpretable alternative that does not require excessive computation resources.¹¹

3.1 Data on U.S. Public Equities

Equity data and characteristics. The standard filters (e.g., same as in Fama-French factor construction) are applied to the universe of U.S. equities. This universe includes only stocks listed on NYSE, AMEX, or NASDAQ for more than one year and uses those observations for firms with a CRSP share code of 10 or 11. We exclude Stocks with negative book equity or lag market equity. We use 61 firm characteristics with monthly observations for each stock, covering six major categories: momentum, value, investment, profitability, frictions (or size), and intangibles. Characteristics are standardized cross-sectionally to the range $[0, 1]$. Table A.1 lists these input variables.¹²

The monthly data ranges from 1981 to 2020. The average and median monthly stock observations are 5,265 and 4,925, respectively, for the first 20 years, and 4,110 and 3,837, respectively, for the latter 20 years. We apply cross-sectional winsorization at 1% and 99% to mitigate the impact of outliers on individual stock returns. The entire 40-year sample is used for benchmark analysis. The sample is split into two parts—the first 20 years from 1981 to 2000 and the recent 20 years from 2001 to 2020—for a subsample robustness check and as training and test samples.

¹¹The baseline single P-Tree model trained on U.S. data runs about 20 minutes on a server with an Intel Xeon Gold 6230 CPU, for a training dataset with 61 characteristics and 2.2 million observations.

¹²For example, market equity values in Dec. 2020 are uniformly standardized into $[0, 1]$. The firm with the lowest value is 0, and the highest is 1. All others are distributed uniformly in between. Missing values of characteristics are imputed as 0.5, which implies the firm is neutral in the security sorting.

Macroeconomic variables. In addition, we use 10 macro variables to capture potential regime switches. Table A.2 summarizes the macro variables, which include market timing macro predictors, bond market predictors, and aggregate characteristics for S&P 500. We standardize these macro predictor data by the historical percentile numbers for the past 10 years.¹³ This rolling-window data standardization is useful when comparing the predictor level to detect different macroeconomic regimes.

3.2 Visualizing a Single P-Tree

Figure 4 plots the P-Tree diagram. In each leaf node, S# represents the order of sequential splits, N# is the node index, and the # numbers are the monthly median number of stocks within the node. We provide the numbers in the terminal leaves for the portfolio size: the monthly median number of stock observations. The split rules are the selected splitting characteristics and cross-sectional quintile cutpoints [0.2, 0.4, 0.6, 0.8].¹⁴ Before the first split, the tree grows from the root node (N1), whose leaf basis portfolio represents the value-weighted market portfolio.

The data-driven P-Tree first splits along the unexpected quarterly earnings (SUE, Rendleman Jr et al., 1982) at 0.4 (40% quantile). After this split, 40% of the stocks go to the left leaf (labeled N2), and 60% go to the right (N3). Then, the second split is on the dollar trading volume (DOLVOL, Chordia et al., 2001) at 0.2 of the right leaf (N3), and the third split is also on DOLVOL at 0.2 of the left leaf (N2). Furthermore, subsequent splits include the industry-adjusted book-to-market ratio (BM_LIA), return-on-equity (ROA), zero trade (ZEROTRADE), and market equity (ME). After nine splits, we stop the P-Tree growth and obtain 10 leaf basis portfolios.

Clustering patterns and asymmetric interactions. P-Tree helps cluster similar assets under the economics-guided split criterion based on underlying characteristics. Figure 4 illustrates the (asymmetric) characteristics interactions for splitting the cross section. By jointly defining the partition corresponding to the leaf node, P-Tree learns the interaction of characteristics appearing in the same path. For instance, ZEROTRADE

¹³For example, inflation greater than 0.7 implies the current inflation level is higher than 70% of observations during the past decade.

¹⁴Our R package offers more flexibility and finer grid searches than conventional univariate security sorting, which uses quintiles to ensure adequate observations in each portfolio.

of liquidity is a valuable indicator for further splitting low-SUE low-DOLVOL stocks. However, for low-SUE non-low-DOLVOL stocks, ME of size might be a better indicator for a further split under the MVE framework.

The partition plot in Figure 5 is an alternative way to visualize the clustering and asymmetric patterns. We report each leaf's monthly average excess returns and annualized Sharpe ratios. First, the P-tree splits on SUE at 0.4, which yields a low-SUE portfolio including 40% of the stocks and a high-SUE portfolio constituted by 60% of the stocks in the cross section. The spread of monthly expected returns between N2 and N3 is 0.46%, and the Sharpe ratios range from 0.25 to 0.64.

Second, we split on DOLVOL at 0.2 on N3 to harvest N6 and N7, where the low-DOLVOL portfolio N6 has *higher* expected returns and Sharpe ratio than N7. Third, we split on DOLVOL at 0.2 on N2 to harvest N4 and N5, where the low-DOLVOL portfolio N4 has *lower* expected returns and Sharpe ratio than N5. We find DOLVOL has dif-

Figure 4: Panel Tree from 1981 to 2020

We provide splitting characteristics, cutpoint values for each parent node, and their respective node and splitting indexes. For example, node N1 is split by SUE at 0.4 as the first split S1, and the second split S2 is on node N3 by DOLVOL at 0.2. The median monthly number of assets in the terminal leaf basis portfolios is also included. For example, node N8 has 134 stocks by monthly median. Table A.1 describes equity characteristics. Figure A.1 reports the P-Tree diagrams for subsamples.

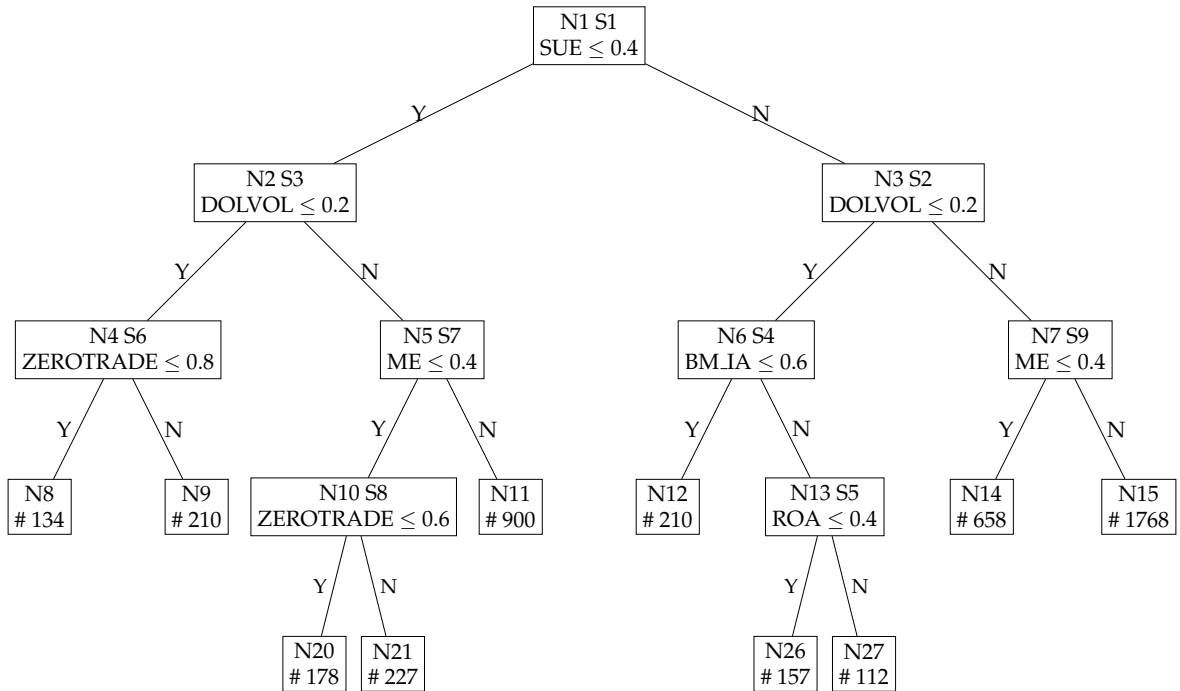
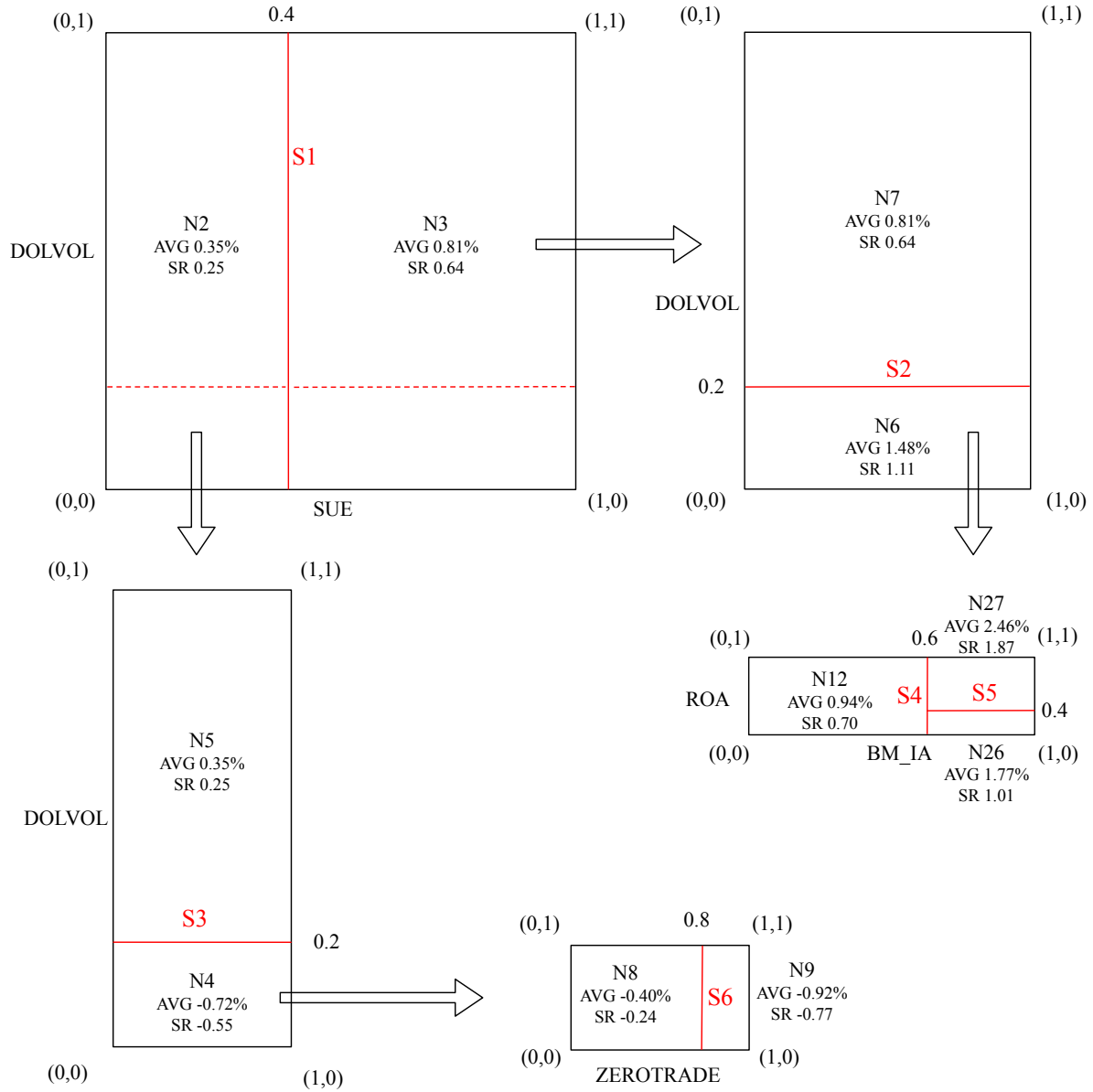


Figure 5: Visualizing Nonlinear Interactions with Partition Plots

This diagram illustrates the partitions for the first few splits of the tree structure in Figure 4. The first split (S1) occurs at 40% of SUE on the entire stock universe, and the second split occurs at 80% of DOLVOL on the high SUE partition. The portfolio results for each partition are provided (monthly average excess returns (AVG) and annualized Sharpe ratios (SR)). The arrows indicate the next split is implemented on the partitioned area from the previous one.



ferent impacts on N2 and N3. On the high-SUE side, DOLVOL positively correlates to asset returns. However, on the low-SUE side, DOLVOL negatively correlates with asset returns. This is an example of asymmetric interaction between SUE and DOLVOL. A simple trading strategy that shorts N4 (Low-SUE -Low-DOLVOL portfolio) and longs N6 (High-SUE - Low-DOLVOL portfolio) makes over 2% monthly expected return. The return gaps among the leaves show the usefulness of splitting the cross section via the

asymmetric interaction of characteristics.

3.3 P-Tree Leaf Basis Portfolios

Asset clustering. P-Tree generates leaves, that is, leaf basis portfolios. Unlike the scalar output in CART for pure prediction, a leaf of P-Tree represents a time series of portfolio returns from a time-invariant P-Tree structure. These leaf basis portfolios are nonlinear, interactive, and high-dimensional characteristics-managed portfolios. Table 1, panel A, summarizes leaf basis portfolios generated by the first P-Tree in Figure 4: the index of nodes, median number of stocks in the leaf basis portfolio, average returns, CAPM α (%), β , time series regression R^2 , and alphas with respect to FF5, Q5, RP-PCA, and IPCA five-factor models (Fama and French, 2015; Hou et al., 2015; Kelly et al., 2019; Lettau and Pelger, 2020b).

We observe two large leaves: N11 containing 900 stocks and N15 containing over 1,700 stocks by monthly median. Their CAPM R^2 are over 90%, and the β are close to one, meaning their return time series are highly correlated with the market returns. No further splitting results in higher Sharpe ratios on these two leaves, yet other leaves offer higher investment improvement under the MVE framework. For instance, N21 represents the low-SUE, high-DOLVOL, low-ME, and high-ZEROTRADE leaf, which has -1.14% average return and -1.83% CAPM α , and N27 is the high-SUE, low-DOLVOL, high-BM_LIA, and high-ROA leaf, which has 2.46% average returns and 1.97% CAPM α . By employing a simple monthly rebalanced strategy of buying N27 and selling short N21, one can expect an average return of 3.5%.

These leaf basis portfolios are generated via interactive and nonlinear splits on 61 characteristics. We expect them to be very hard to price by the prevalent linear factor models. Nine leaves have significant CAPM alphas among the 10 leaf basis portfolios. At a confidence level of 10%, the number of statistically significant non-zero alphas for FF5, Q5, RP-PCA 5, and IPCA 5 factor models are 10, 9, 9, and 7, respectively.

P-Tree is a goal-oriented clustering algorithm that maximizes the Sharpe ratio. Therefore, we anticipate these test assets having economically large and statistically significant alphas against those “inadequate” benchmark factor models. With such distinct basis portfolios, we expect the efficient frontier they span cannot be spanned

by benchmark factor models, possibly achieving higher mean-variance efficiency. Panels B and C of Table 3 provide subsample analysis for the first and the latter 20 years. Although the tree structure and leaf identity change because of refitting, we still find similar patterns. For example, pricing these portfolios based on leaf factors remains difficult using benchmark factor models.

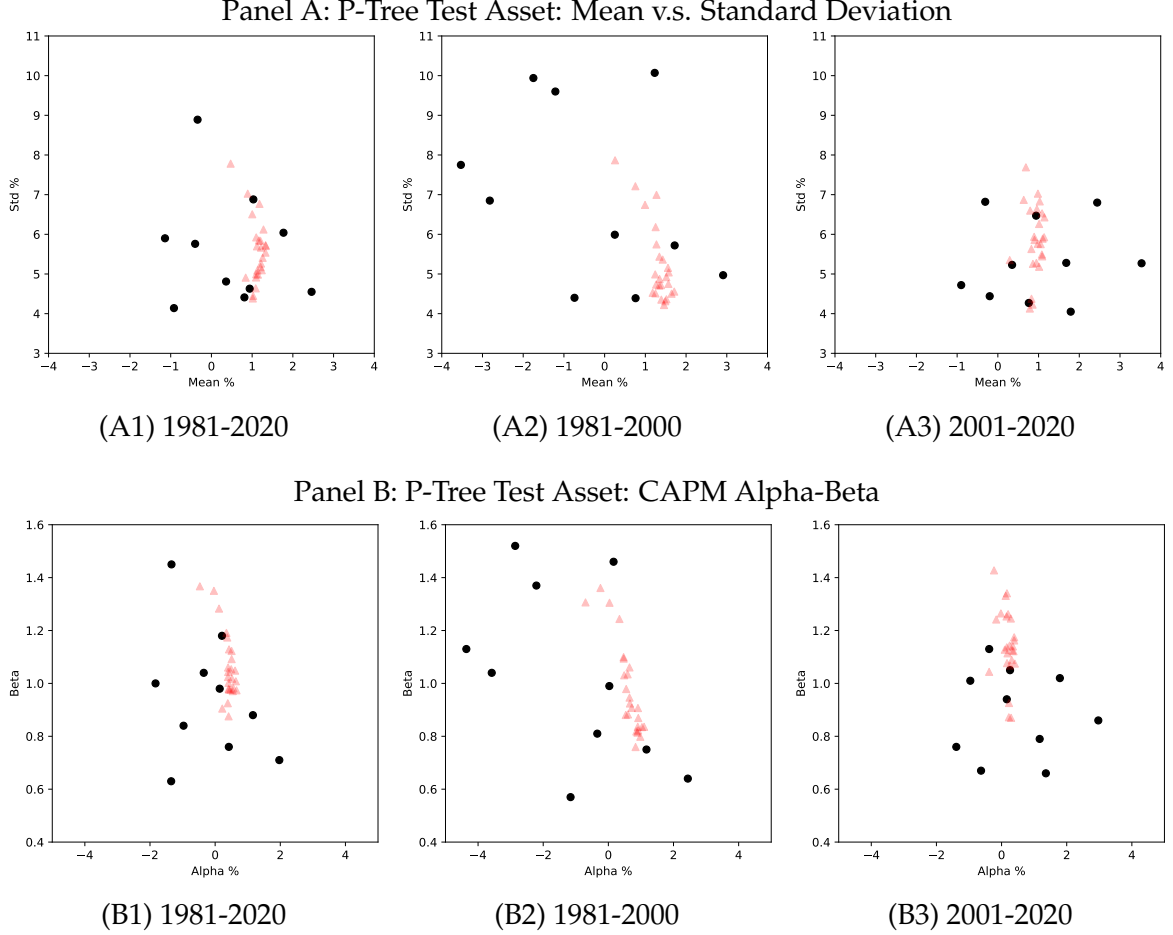
Table 1: **Evaluation for Leaf Basis Portfolios**

This table reports the performance of 10 leaf basis portfolios generated by the first P-Tree. Panel A corresponds to Figure 4, whereas Panels B and C correspond respectively to (a) and (b) in Figure A.1. We report the median number of stocks (leaf size), average returns, CAPM α (%), β , R^2 , and the alphas (%) with respect to FF5, Q5, RP-PCA, and IPCA five-factor models. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

ID	# Median	AVG	STD	α_{CAPM}	β_{CAPM}	R^2_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}
Panel A: 40 Years (1981-2020)										
N8	134	-0.40	5.76	-0.97***	0.84	0.42	-0.69***	-0.66***	-1.51***	-0.58*
N9	210	-0.92***	4.14	-1.35***	0.63	0.47	-1.02***	-1.25***	-1.84***	-0.36*
N20	179	-0.34	8.89	-1.34***	1.45	0.53	-0.36	-0.49*	-1.55***	-1.78***
N21	227	-1.14***	5.90	-1.83***	1.00	0.58	-1.06***	-1.56***	-2.34***	-1.19***
N11	900	0.36*	4.81	-0.35***	1.04	0.93	-0.24***	-0.15**	-0.29***	0.16
N12	210	0.94***	4.63	0.42**	0.76	0.54	0.64***	0.48***	-0.22	1.17***
N26	157	1.77***	6.04	1.16***	0.88	0.43	1.74***	1.37***	0.40**	1.40***
N27	112	2.46***	4.55	1.97***	0.71	0.49	2.07***	2.00***	1.18***	2.18***
N14	658	1.03***	6.88	0.21	1.18	0.59	1.15***	0.64***	-0.33**	0.06
N15	1768	0.81***	4.41	0.14***	0.98	0.98	0.08**	0.03	-0.26***	0.29
Panel B: 20 Years (1981-2000)										
N16	97	0.25	5.99	-0.34	0.81	0.36	0.39	-0.08	-0.27	0.22
N17	256	-0.74***	4.40	-1.16***	0.57	0.33	-0.95*	-1.17***	-1.56***	-0.34
N18	95	1.72***	5.72	1.17***	0.75	0.34	2.07***	1.27***	0.84**	1.56**
N19	48	2.91***	4.97	2.44***	0.64	0.33	3.75***	2.50***	2.36***	2.26***
N5	3443	0.76***	4.39	0.03*	0.99	1.00	-0.03	-0.04***	-0.49***	0.41
N24	160	-3.53***	7.75	-4.36***	1.13	0.42	-1.74***	-3.29***	-3.58***	-2.54**
N25	250	-2.82***	6.85	-3.58***	1.04	0.45	-2.17***	-2.54***	-1.89***	-0.45
N13	286	-1.75***	9.94	-2.86***	1.52	0.46	-0.91	-1.19***	-1.29***	0.74
N14	217	-1.21*	9.60	-2.21***	1.37	0.40	0.00	-0.63	-0.09	1.61
N15	116	1.23*	10.07	0.16	1.46	0.41	2.35***	1.52***	1.39**	2.29
Panel C: 20 Years (2001-2020)										
N8	153	-0.20	4.44	-0.63**	0.67	0.45	-0.54**	-0.46**	-0.93***	-0.10
N18	58	-0.31	6.82	-0.96***	1.01	0.44	-1.17***	-0.53*	-1.36***	-0.88**
N19	112	-0.90***	4.72	-1.39***	0.76	0.53	-1.41***	-1.31***	-1.89***	-0.93***
N5	1206	0.35	5.23	-0.38***	1.13	0.95	-0.27***	-0.25***	-0.21***	0.37*
N24	171	0.94**	6.47	0.26	1.05	0.53	0.49	0.57**	-0.13	0.59**
N25	104	2.44***	6.80	1.79***	1.02	0.45	2.17***	2.17***	1.16***	1.43***
N52	47	1.68***	5.28	1.17***	0.79	0.46	0.90***	1.45***	0.52**	1.56***
N53	76	3.53***	5.27	2.97***	0.86	0.54	2.68***	3.10***	2.24***	3.19***
N27	76	1.79***	4.05	1.36***	0.66	0.53	1.50***	1.37***	0.92***	1.6***
N7	1896	0.76***	4.27	0.16***	0.94	0.98	0.09**	0.09**	-0.05	0.63***

Figure 6: **Diversifying P-Tree Test Assets**

This figure reports the performance of P-Tree test assets (presented in Table 1) and the 5×5 ME-BM portfolios. The 10 black circles represent P-Tree test assets, and the 25 light-red triangles represent ME-BM portfolios. Panel A shows the scatter plots of the mean and standard deviation for portfolio returns in percentage. Panel B shows the scatter plots of CAPM alpha and beta, with alphas in percentage.



Diversified test assets. Figure 6, Panel A, depicts the performance statistics of P-Tree test assets. Subfigure A1 shows the mean-standard-deviation scatter plot for leaf basis portfolios of the first P-Tree on the 40-year sample from 1981 to 2020 in black circles. The expected returns are in the range of -2% to 3%, and the standard deviations are in the range of 4% to 9%. We observe large variations in these portfolios' expected returns and risks, which means they are diversified under the MVE framework. Significant variations exist among the portfolios in Panels A2 and A3 for subsample analysis, with the earlier sample showing more diversification and the latter sample being more concentrated. By contrast, the 5×5 ME-BM portfolios in light-red triangles cluster around the same level of average returns, which are much smaller than those of P-Tree

test assets. This finding indicates that P-Tree test assets are more diversified under the MVE framework than the ME-BM portfolios.

Panel B of Figure 6 shows the CAPM model-implied alpha-beta plots. In subfigure B1's 40-year sample, we find CAPM β s of P-Tree test assets are scattered around one in the range of 0.5 to 1.4. CAPM alphas are around zero, ranging from -2% to 2%. These basis portfolios have a lot of variations in CAPM alphas and β s. In the first 20 years in B1, alpha and beta ranges were larger, whereas the variation was smaller in the second 20 years in B2. The 5×5 ME-BM portfolios are more easily explained for the CAPM model than P-Tree test assets because they align closely with the vertical line of zero alpha with more negligible diversification in all figures. In summary, this is highly positive evidence for P-Tree test assets' better-diversified patterns than conventional ones, even when priced by CAPM.

4 Boosted P-Trees for Multiple Factors and Test Assets

4.1 Growing the Efficient Frontier

Our investigation focuses on the efficient frontiers spanned by test assets generated using (boosted) P-Tree. Many P-Tree test assets have significant alphas over multiple well-known benchmark factor models, indicating those factor models cannot effectively span the efficient frontier. The boosting technique enhances our ability to generate more effective test assets that expand the efficient frontier better, that is, "growing the efficient frontiers on P-Trees."

Unspanned efficient frontier. Figure 7 shows the MVE frontiers of P-Trees and boosted P-Trees in gradient color from blue to red. Benchmark factor models are printed in black, and benchmark test assets in purple. Notably, the P-Tree1 frontier (generated by the first P-Tree test assets) is already more efficient than any frontier spanned by the benchmark factor models and test assets. Among the benchmark test assets, 285 bivariate-sorted portfolios span a frontier very close to our P-Tree1 frontier. However, P-Tree1 contains only 10 portfolios, much smaller than 285. Among the factor models, the IPCA five-factor model is a strong candidate but still less efficient than the P-Tree1 frontier. Overall, these 10 leaf basis portfolios consist of promising test assets

to span the efficient frontier, which is more efficient than the benchmarks under the MVE framework. In other words, the listed benchmark factor models and test assets are insufficient to span the real efficient frontier.

Boosting efficient frontier. Beyond the first P-Tree, boosted P-Trees improve portfolio efficiency from one factor to 20 factors. In Figure 7, the spanned frontiers move toward the top left corner of the mean-variance diagram. Our approach consistently advances the efficient frontier throughout the 20 P-Trees. Combining 200 assets in one MVE

Figure 7: **Characterizing the Efficient Frontier with P-Trees**

This figure shows the mean-variance efficient frontiers of the first P-Tree, the sequence of boosted P-Tree factor models, benchmark factor models, and benchmark test assets. The dots on the frontiers represent the tangency portfolios. The mean and standard deviation are on a monthly scale.

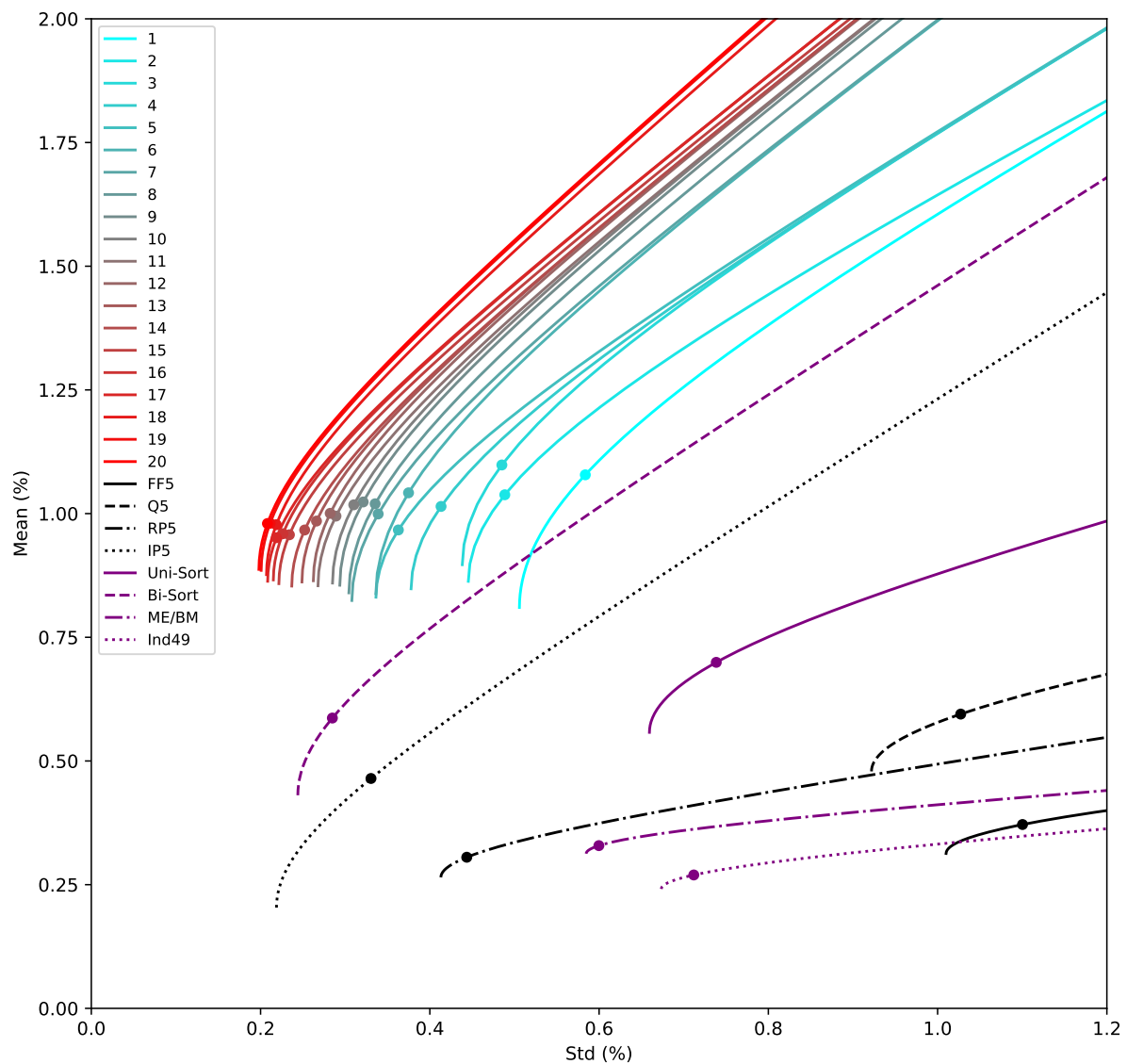


Table 2: Testing the Boosted P-Tree Growth

This table presents the performance for each boosted P-Tree factor generated sequentially, including the Sharpe ratio, CAPM test, FF5 test, expanding factors test, and the Barillas and Shanken (2017) test. The “Single” column shows the Sharpe ratio of the single factor, and the “Cumulative” (cumulative) column displays the Sharpe ratio of the MVE spanned by multiple factors from the first to the current P-Tree factors. The columns reporting on CAPM and FF5 spanning regression tests show the α (%) and t -statistic. For the expanding factor tests, we regress each P-Tree factor on all the boosted P-Tree factors generated and report the α (%), t -statistic, and R^2 . Table A.3 shows the results for subsample analysis.

	Sharpe Ratio		CAPM Test		FF5 Test		Expanding Factors Test			BS Test
	Single	Cumu.	α (%)	t -stat	α (%)	t -stat	α (%)	t -stat	R^2	p -value
1	6.37	6.37	1.39	39.63	1.37	37.48	-	-	-	-
2	3.20	7.35	0.52	19.74	0.48	18.01	0.62	9.55	0.01	0.00
3	1.18	7.80	0.34	6.59	0.18	3.82	-0.86	-5.43	0.26	0.00
4	2.06	8.46	0.44	12.46	0.38	10.77	0.69	7.62	0.16	0.00
5	1.99	9.18	0.48	11.96	0.41	10.41	0.84	5.61	0.21	0.00
6	1.01	9.57	0.18	5.38	0.08	2.88	-0.50	-5.47	0.45	0.00
7	1.42	10.11	0.28	8.27	0.22	6.62	0.63	7.03	0.36	0.00
8	1.32	10.40	0.28	7.56	0.20	5.47	-0.50	-5.33	0.41	0.00
9	1.83	10.88	0.53	10.91	0.43	9.05	0.85	6.88	0.34	0.00
10	1.48	11.20	0.44	8.64	0.30	6.45	-0.68	-5.49	0.46	0.00
11	1.78	11.72	0.38	10.65	0.31	8.85	0.72	6.70	0.30	0.00
12	1.02	12.06	0.20	5.44	0.10	3.15	-0.55	-5.76	0.55	0.00
13	1.37	12.57	0.29	7.85	0.22	6.21	0.76	6.01	0.33	0.00
14	1.37	13.01	0.48	7.90	0.32	5.90	-0.91	-5.68	0.60	0.00
15	1.37	13.81	0.31	7.89	0.21	5.59	0.97	7.11	0.48	0.00
16	1.24	14.28	0.28	7.02	0.17	4.49	-0.74	-6.24	0.54	0.00
17	1.54	14.60	0.46	9.04	0.40	7.70	-0.89	-4.80	0.40	0.00
18	1.64	14.92	0.32	9.69	0.27	8.12	-0.65	-5.42	0.34	0.00
19	1.48	15.43	0.43	8.63	0.36	7.46	1.18	5.86	0.34	0.00
20	1.35	15.63	0.33	7.72	0.23	6.47	-0.59	-4.19	0.44	0.00

portfolio is challenging because of the curse of dimensionality. Unlike Ait-Sahalia and Xiu (2017), who use exogenous industry classifications to construct block-diagonal patterns, our boosted P-Trees entertain an endogenous block-diagonal structure in the return covariance. We combine the 10 leaf basis portfolios of each P-Tree into a single P-Tree factor and then estimate the tangency portfolio of the boosted P-Tree factors. As such, we can generate multiple factors and combine multiple sets of test assets in one frontier under the MVE framework.

Moving from a single P-Tree to 20 P-Trees results in more efficient frontiers, allowing investors to achieve higher expected returns while bearing the same level of risk. Although a significant difference exists between the frontiers of P-Tree1 and P-Tree10, the frontier lines increase after P-Tree10, indicating a more negligible improvement in mean-variance efficiency with each subsequent boosting.

We employ asset pricing tests to demonstrate the incremental efficiency of boosted P-Trees. First, we show the Sharpe ratios of each boosted P-Tree factor and the multi-factor model. Second, we evaluate each boosted factor with CAPM and FF5 by checking for a significant alpha. As we know from Table 1, P-Tree test assets are hard to price, and we expect the P-Tree factors to pass the CAPM test and FF5 test. Third, we regress each P-Tree factor on all the boosted P-Tree factors generated for the expanding factor tests. The efficiency of the frontier increases if the previous P-Tree factors do not span the additional P-Tree factor.

The test results are presented in Table 2. The Sharpe ratios of each boosted factor are above 1, even for the 20-th factor. Meanwhile, the cumulative multi-factor Sharpe ratio increases monotonically from 6 to 15, which means that the test assets of the boosted P-Tree add incremental pricing information to the existing test assets. Furthermore, all CAPM and FF5 alphas are positive and highly significant, indicating that all the boosted P-Tree factors cannot be explained by CAPM or FF5. In the expanding factor test, we find the significant α and t -statistic and the medium values for the time series R^2 . After applying the nested asset pricing test (Barillas and Shanken, 2017), we do not find any evidence that supports a sufficient factor model to explain the assets in the boosted P-Tree test.¹⁵ In summary, Table 2 shows that boosted P-Trees generate unexplained alphas, and their leaf portfolios thus better span the efficient frontier. By contrast, conventional univariate- or bivariate-sorted test assets used for testing asset pricing models do not fully cover the efficient frontier.

4.2 Generating Test Assets via Boosted P-Trees

First, we show the pricing performance of leaf basis portfolios, implying they serve as diversified test assets and are challenging against the FF5 model. Then, we utilize boosted P-Trees that generate multiple sets of basis portfolios. Empirically, we grow up to 20 P-Trees, yielding 200 leaf basis portfolios.

¹⁵The first row contains "-" signs, which indicate that the expanding factor and BS tests are not applicable for the single factor case.

P-Tree test assets. Table 3, Panel A, lists the number of portfolios, the GRS test statistic and its p -value, the p -value of PY test,¹⁶ the average absolute α , the root mean squared α , the average time series regression R^2 , and the percentage of significant alphas with respect to 10%, 5%, and 1% confidence levels. In the rows, P-Tree1 represents a set of test assets comprising 10 leaf basis portfolios in the first P-Tree. P-Tree1-5 combines all the leaf basis portfolios generated by the first five P-Trees, and P-Tree6-10 contains the 6-th to the 10-th P-Trees. Similarly, we define P-Tree11-15, P-Tree16-20, and P-Tree1-20.

We find the GRS test and PY test always reject the null hypothesis that the alphas of test assets are jointly zero for all specifications of P-Tree test assets, which means the expected returns of P-Tree test assets can hardly be explained by the FF5 model. Furthermore, the last three columns show many assets have statistically significant alphas. For comparison, we include four sets of test assets that are commonly used in literature: 150 10×1 univariate-sorted portfolios, 285 bivariate-sorted portfolios, 5×5 ME-BM portfolios, and 49 industry portfolios, all of which can be downloaded from Ken French's website. P-Tree test assets have larger GRS test statistics and alphas than these benchmark test assets.

Furthermore, we observe the decline of alphas from the first P-Tree to the subsequently boosted P-Trees as anticipated, indicating diminishing marginal convergence to the limit of the efficient frontier. The test assets of the first P-Tree have larger GRS statistics, higher average alphas, and higher proportions of significant alphas than the test assets generated by the follow-up P-Trees. In addition, we see an evident decline in GRS statistics and alphas from P-Tree1-5 to P-Tree6-10, P-Tree11-15, and P-Tree16-20. P-Tree generates the most informative test assets in the first P-Tree, whereas the incremental information in the boosted P-Trees is complementary and declining. Although the alphas decline along boosting, the test assets generated by P-Tree16-20 cannot be explained by FF5, with a 4.21 GRS test statistic, 0.31% mean absolute alpha, and 52% test assets having significant alphas at the 10% level. The test assets of P-Tree16-20 are still harder to price than univariate- and bivariate-sorted portfolios.

Overall, these 200 test assets created by 20 P-Trees on high-dimensional charac-

¹⁶Pesaran and Yamagata (2023) adapts to cases where the number of assets being tested is larger than the number of periods, which Gibbons et al. (1989) cannot address.

Table 3: Comparing Test Assets

The table displays performance statistics of P-Tree test assets and others for comparison, including P-Tree1, which consists of 10 portfolios in the first P-Tree. P-Tree1-5, P-Tree6-10, P-Tree11-15, P-Tree16-20, and P-Tree1-20 are sets of test assets that combine leaf basis portfolios generated by the first 5, 6-10, 11-15, 16-20, and 1-20 P-Trees, respectively. “Uni-Sort” has 150 univariate-sorted portfolios, “Bi-Sort” has 285 bivariate-sorted portfolios, ME/BE has 25 portfolios, and “Ind49” includes 49 industry portfolios. The reported statistics include the number of test assets, GRS test statistics and p-values (Gibbons et al., 1989), p-values of PY test (Pesaran and Yamagata, 2023), average absolute α (%), root mean squared α (%), average R^2 (%) of regressing the portfolios on Fama-French five factors, and the proportion of test assets with unexplained significant alphas under 10%, 5%, and 1%. The three panels report on the re-trained models for different sample periods.

	N	GRS	p -GRS	p -PY	$ \bar{\alpha} $	$\sqrt{\alpha^2}$	$\bar{R}^2$	% $\alpha_{10\%}$	% $\alpha_{5\%}$	% $\alpha_{1\%}$
Panel A: 40 Years (1981-2020)										
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
Panel B: 20 Years (1981-2000)										
P-Tree1	10	84.36	0.00	0.00	1.58	1.95	70	90	80	80
P-Tree1-5	50	50.84	0.00	0.00	0.79	1.26	76	62	60	52
P-Tree6-10	50	7.27	0.00	0.00	0.58	0.87	75	56	44	38
P-Tree11-15	50	6.39	0.00	0.00	0.55	0.82	76	66	60	42
P-Tree16-20	50	8.42	0.00	0.00	0.52	0.74	76	62	54	50
P-Tree1-20	200	112.90	0.00	0.00	0.61	0.95	76	62	55	46
Uni-Sort	150	1.94	0.00	0.00	0.17	0.23	88	35	27	19
Bi-Sort	285	-	-	0.00	0.21	0.30	89	42	33	21
ME-BM	25	4.75	0.00	0.00	0.21	0.25	91	56	48	32
Ind49	49	2.44	0.00	0.00	0.52	0.61	61	59	49	27
Panel C: 20 Years (2001-2020)										
P-Tree1	10	56.76	0.00	0.00	1.09	1.35	68	90	90	90
P-Tree1-5	50	30.35	0.00	0.00	0.43	0.68	76	52	38	24
P-Tree6-10	50	5.17	0.00	0.00	0.29	0.37	75	34	28	14
P-Tree11-15	50	2.20	0.00	0.00	0.27	0.35	75	30	22	10
P-Tree16-20	50	2.52	0.00	0.00	0.31	0.40	76	42	28	10
P-Tree1-20	200	83.91	0.00	0.00	0.33	0.47	76	40	29	14
Uni-Sort	150	1.46	0.03	0.94	0.09	0.12	89	12	7	0
Bi-Sort	285	-	-	0.01	0.11	0.15	91	21	15	6
ME-BM	25	2.58	0.00	0.10	0.11	0.14	93	24	8	8
Ind49	49	1.29	0.11	0.36	0.25	0.32	62	18	8	2

teristics challenge the FF5 model. They are more difficult to price than benchmark test assets, setting a high standard for testing factor models, which respond to the concerns in empirical studies (Lewellen et al., 2010; Ang et al., 2020).

For robustness, Table 3, Panels B and C, report the first and the latter 20-year subsamples. In the recent 20 years, FF5 has performed well in pricing the benchmark test asset; that is, the GRS and PY tests cannot be rejected at the 10% level, and less significant alphas testing on 49 industry portfolios and univariate-sorted portfolios. However, the P-Tree test assets always reject the null hypothesis for GRS and PY tests and are consistently challenging to price in both subsamples.¹⁷

Advantages of P-Tree clustering. P-Tree has two advantages that make these test assets efficient. First, P-Tree is a goal-oriented clustering algorithm tailored for spanning the efficient frontier iteratively. Each leaf clusters individual asset returns based on similar characteristic values, aggregating as a time series of portfolio returns. The goal is to maximize the collective Sharpe ratio of the MVE portfolio spanned by the leaf basis portfolios. The economic objective of P-Tree is vastly different from the CART and other off-the-shelf ML models, which focus on statistical return prediction without economic utilities, and cannot ever tackle the task of growing efficient frontier.

Second, P-Tree exploits the complex asymmetric interactions and nonlinearities on high-dimensional characteristics. The commonly used test assets are characteristics-sorted portfolios, such as univariate- and bivariate-sorted portfolios, which only consider decile or quintile sorting, along with up to three characteristics chosen ad hoc or based on researchers' domain expertise. In contrast, P-Tree is more flexible, allowing for general sequential and simultaneous sorting with more characteristics and asymmetric interactions in a unified framework. P-Trees enable the inclusion of large leaves, like N15, as shown in Figure 4, where further splitting does not result in a higher Sharpe ratio, as well as small leaves, such as N8. The asymmetric splitting allows P-Trees to focus on the finer clusters of informative assets while paying less attention to non-informative assets under the MVE framework.

¹⁷The GRS test is not applicable in the presence of more test assets than observations, as is the case for Bi-Sort in Panels B and C, marked with “-”.

4.3 P-Tree Factors and Asset Pricing Performance

Using the findings from the previous analysis, we have established that P-Tree test assets are distinct from conventional test assets. Moving forward, we proceed with cross-examining P-Tree factors and various benchmark factor models on different sets of test assets, including P-Tree leaf basis portfolios. We adopt the cross-sectional R^2 to evaluate the asset pricing performance:

$$\text{Cross-sectional } R^2 = 1 - \frac{\sum_{i=1}^N (\bar{R}_i - \widehat{\bar{R}}_i)^2}{\sum_{i=1}^N (\bar{R}_i)^2}, \quad (7)$$

where $\widehat{\bar{R}}_i = \widehat{\beta}^T \tilde{\lambda}$. The risk premium estimation adopts the cross-sectional regression estimates of factors, and factor loadings are estimated from time series regressions. The cross-sectional R^2 represents the fraction of assets' average returns explained by the model-implied expected returns. Various test assets, including P-Tree test assets, conventional sorted portfolios, and industry portfolios, are used to test the factor models. In Table 4, we display the factor models in the columns to explain the expected return of test assets in the rows. Panel A provides the asset pricing performances for the 40-year sample, and Panels B and C report for the 20-year subsamples.¹⁸

First, the P-Tree1 factor (from the first P-Tree) explains 98.3% of the cross-sectional return variation of the P-Tree1 leaf basis portfolios (10 assets), whereas the benchmark factor models provide much lower performance. This finding is not surprising, since the first P-Tree factor is constructed jointly with these test assets. However, the FF5, Q5, and RP-PCA five-factor models have much smaller R^2 s than the P-Tree1 factor, indicating large pricing errors in explaining the expected returns of the highly diversified P-Tree1 test assets. One exception is that the IPCA five-factor model performs well in pricing the P-Tree1 test assets, with a 93.4% R^2 .

Second, the P-Tree five-factor model (P-Tree1-5, read as model containing from the first to the fifth factors), which includes five P-Tree factors from the first five P-Trees, demonstrates tremendous asset pricing performance. It outperforms the benchmark

¹⁸The first row contains two “-”, which indicate the number of assets $N = 10$ is equal to or smaller than the number of factors (10 or 20). Thus, the cross-sectional regression is not applicable.

five-factor models in pricing the first 10 P-Trees' (P-Tree1-10) test assets and 5×5 ME-BM portfolios. For other specifications of test assets, P-Tree1-5 shows comparable levels of cross-sectional R^2 among all the five-factor models for a fair comparison. The superior pricing performance remains consistent across Panels A to C for the entire sample and subsample analysis.

Table 4: Asset Pricing Performance: Cross-Sectional R^2

This table displays the cross-sectional R^2 (%) of pricing different sets of test assets by the P-Tree factor models and benchmark factor models. The factor models are listed in the columns, and the different specifications of test assets are listed in the rows. Specifically, the table includes 1, 5, 10, and 20 P-Tree factor models and FF5, Q5, RP-PCA, and IPCA five-factor models in the columns. In the rows, the test assets are from the top 1, 5, 10, and 20 P-Trees, univariate-sorted portfolios, bivariate-sorted portfolios, ME-BM 5×5 portfolios, and 49 industry portfolios.

	P-Tree1F	P-Tree5F	P-Tree10F	P-Tree20F	FF5	Q5	RP5	IP5
<u>Panel A: 40 Years (1981-2020)</u>								
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
<u>Panel B: 20 Years (1981-2000)</u>								
P-Tree1	99.3	99.3	–	–	54.9	59.7	37.6	76.0
P-Tree1-5	49.2	88.5	92.6	96.0	25.8	31.8	41.1	32.2
P-Tree1-10	47.7	59.7	77.8	82.3	37.7	40.5	47.8	28.9
P-Tree1-20	42.8	54.0	70.2	74.5	42.2	45.6	52.5	30.4
Uni-Sort	80.4	96.5	96.9	97.6	94.6	95.9	97.0	87.6
Bi-Sort	66.4	92.1	93.9	94.9	89.1	91.5	92.7	78.8
ME-BM	66.6	94.0	96.2	99.0	93.4	95.3	96.4	88.3
Ind49	60.9	86.5	87.9	95.0	91.0	89.7	92.1	73.5
<u>Panel C: 20 Year (2001-2020)</u>								
P-Tree1	99.2	99.7	–	–	52.2	43.6	69.7	53.9
P-Tree1-5	71.4	87.6	92.3	94.3	60.0	60.7	64.9	61.3
P-Tree1-10	70.4	78.1	86.4	88.4	66.4	65.7	69.8	65.4
P-Tree1-20	70.5	76.4	79.8	82.8	71.7	69.9	75.4	65.2
Uni-Sort	87.5	96.6	97.6	97.9	97.3	97.2	97.4	95.7
Bi-Sort	93.0	97.2	97.9	98.4	97.9	97.5	98.0	96.9
ME-BM	92.9	95.7	98.6	99.3	97.5	97.5	97.4	96.1
Ind49	75.3	92.6	94.8	97.7	90.2	90.8	92.8	86.0

Third, these endogenously generated P-Tree test assets better expand the MVE frontier and raise the bar for testing asset pricing models. The commonly used test assets set a low bar for testing factor models, so benchmark factor models seem sufficient to price them. As we see from Panel A of Table 4, the listed benchmark five-factor models have over 90% cross-sectional R^2 s on the benchmark test assets, and these numbers are consistently large in Panels B and C for subsample analyses. However, their R^2 s decline dramatically when tested against P-Tree test assets. P-Tree test assets use nonlinear and asymmetric interactions among a large set of characteristics, whereas benchmark test assets involve only up to two characteristics.¹⁹

In summary, P-Tree factor models accurately price P-Tree test assets, while benchmark factors do not, and perform similarly or better than benchmark factor models in pricing benchmark test assets. P-Tree test assets set a high standard for testing asset pricing models, surpassing common sorted and industry portfolios.

4.4 Investing P-Tree Factors

The P-Tree framework clusters thousands of assets into a number of leaf basis portfolios, reducing portfolio optimization complexity. These P-Tree test assets can be used as factors directly or as building blocks toward MVE portfolios for investment. Additionally, the frequency of rebalancing P-Tree investment strategies can be reduced to quarterly or annually to decrease transaction costs.

Full-sample evaluation. We evaluate the investment performance of P-Tree factors by combining them into one — the tangency portfolio of P-Tree factors. Table 5, Panel A, reports the P-Tree investment strategy’s annualized Sharpe ratios and monthly alphas in the 40-year sample. The Sharpe ratio increases from 6.37 to 9.20, 11.19, 13.80, and 15.59 for the one-factor strategy to 5, 10, 15, and 20 factors, respectively. To assess their additional investment information, we further evaluate these P-Tree investment strategies over benchmark factor models. These model-adjusted alphas are all greater than 0.80% and highly statistically significant.

The existing research finds that P-Tree investment strategies have highly com-

¹⁹Industry portfolios are an exception because they are not sorted based on characteristics.

Table 5: Factor Investing by Boosted P-Trees

This table presents the investment performance of the initial factors produced by P-Tree and the mean-variance efficient portfolio of multi-factors generated by boosted P-Trees. The reported statistics include the annualized Sharpe ratio and alphas with respect to CAPM, FF5, Q5, RP-PCA, and IPCA five-factor models. Panel A shows the 40-year sample from 1981 to 2020. Panels B1 and B2 present in-sample and out-of-sample results from 1981 to 2000 and 2001 to 2020, respectively. Panels C1 and C2 display in-sample and out-of-sample results from 2001 to 2020 and 1981 to 2000, respectively. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

	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***	1.30***	1.12***
P-Tree1-5	9.20	0.96***	0.90***	0.93***	0.89***	0.81***
P-Tree1-10	11.19	1.00***	0.95***	0.98***	0.95***	0.88***
P-Tree1-15	13.80	0.94***	0.89***	0.91***	0.90***	0.86***
P-Tree1-20	15.59	0.96***	0.92***	0.94***	0.92***	0.89***
Panel B1: 20 Years 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 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***
Panel C1: 20 Years In-Sample (2001-2020)						
P-Tree1	5.82	1.51***	1.46***	1.50***	1.25***	1.36***
P-Tree1-5	9.32	1.32***	1.34***	1.30***	1.20***	1.24***
P-Tree1-10	14.33	1.13***	1.12***	1.11***	1.05***	1.08***
P-Tree1-15	20.45	1.04***	1.05***	1.03***	0.98***	1.00***
P-Tree1-20	26.11	1.05***	1.05***	1.04***	1.01***	1.02***
Panel C2: 20 Years Out-of-Sample (1981-2000)						
P-Tree1	4.35	1.50***	1.59***	1.35***	1.43***	1.10***
P-Tree1-5	3.88	1.20***	1.24***	0.98***	1.11***	0.75***
P-Tree1-10	4.29	1.03***	1.10***	0.85***	0.90***	0.76***
P-Tree1-15	4.04	0.91***	0.91***	0.75***	0.78***	0.65***
P-Tree1-20	3.91	0.92***	0.91***	0.76***	0.78***	0.61***

petitive Sharpe ratios. The correlation-based clustering in [Ahn et al. \(2009\)](#) generates basis portfolios that underperform bivariate-sorted portfolios on ME and BM. Further-

more, [Daniel et al. \(2020\)](#) nearly doubles the Sharpe ratio of the characteristics-sorted portfolio from 1.17 to 2.13 by constructing the corresponding characteristic-efficient portfolios. However, a single P-Tree delivers a 6.37 Sharpe ratio over 40 years from 1981 to 2020, representing a significant improvement over [Ahn et al. \(2009\)](#) and [Daniel et al. \(2020\)](#), although the empirical samples differ.

Out-of-sample evaluation. The results in Panel A of Table 5 are based on a full-sample analysis, which may raise concerns about in-sample overfitting. Therefore, we provide two exercises on out-of-sample investment strategies. We follow [Kozak et al. \(2020\)](#) and perform a half-half split to construct the training and test samples. It suffices to demonstrate the out-of-sample performance when the past and future samples predict each other. This approach further mitigates concerns regarding the look-ahead bias when models trained by the future sample predict the past.

On the one hand, the past-predicting-future results are reported in Panels B1 and B2. The first twenty-year in-sample Sharpe ratios are even higher than the forty-year ones in Panel A, and the out-of-sample Sharpe ratios are over 3. Despite decreasing performance, many out-of-sample alphas, adjusted by benchmark models, are still close to 1% and highly significant. On the other hand, for the future-predicting-past result, the out-of-sample Sharpe ratios are larger than 3.88, and the model-adjusted alphas are over 0.60%. These findings demonstrate strong out-of-sample evidence for P-Tree test assets and factors for spanning the MVE frontier.

These out-of-sample investment performance results are comparable to those reported in recent studies. The deep reinforcement learning strategy in [Cong et al. \(2021\)](#) achieves a Sharpe ratio of 2.21 after excluding small stocks from 1990 to 2016. The regularized portfolio, based on sorted portfolios on size, operating profitability, and investment in [Bryzgalova et al. \(2023\)](#), achieves an out-of-sample monthly Sharpe ratio of 0.69 from 2004 to 2016, equivalent to a 2.39 annualized Sharpe ratio. The deep neural network approach in [Feng et al. \(2023\)](#) reports out-of-sample Sharpe ratios of 2.95-3.00 for 2002 to 2021. P-Tree strategies stand out for its transparency, ease of implementation, and flexibility as an all-round framework for test asset generation, asset pricing, and investment management.

5 Interpretability, Characteristics, and Macro Regimes

5.1 Characteristic Importance via Random P-Forest

First, our simulation evidence in Appendix I demonstrates the efficiency loss resulting from missing true characteristics and illustrates how the nonlinear and interactive P-Tree can recover the true set of characteristics, despite a large number of redundant or useless characteristics. Second, the conventional sequential or independent asset sorting scheme primarily focuses on no more than three characteristics in practice, due to the high-dimensional challenge. Because enumerating all possible sequential sorting cases with multiple characteristics is NP-hard, whether using tree models or not, P-Tree’s iterative growth is an efficient algorithm for overcoming this problem and make computation feasible. Therefore, the characteristics selected by the P-Tree algorithm are better proxies for the true fundamental risk inherent in the SDF, which may be overlooked in a linear factor model with ad hoc selected factors.

Random P-Forest. We adopt another famous ensembling scheme in ML to evaluate characteristic importance and account for variable selection risk. Random Forest fits multiple trees to avoid overfitting and quantifying model uncertainty.²⁰ We utilize this approach to assess the significance of variables within the P-Tree framework. In the subsequent text, we refer to our approach as Random P-Forest.

The ensemble steps for constructing the Random P-Forest are as follows: (i) bootstrap the data on the time-series horizon with replacement, preserving the complete cross section of the panel data for the selected time periods to exploit the low serial correlations among asset returns. (ii) randomly draw 20 characteristics, and P-Trees grow on each bootstrap sample independently. We repeat the procedure 1,000 times to form a forest of 1,000 P-Trees. (iii) evaluate the importance of each characteristic by counting its appearance in the P-Forest.

Our empirical exercises focus on the frequency of selecting one variable as a split

²⁰Standard Random Forest (Breiman, 2001) grows many trees on bootstrap training samples, which come from random samples of the observations with replacement and sample a subset of variables from the complete training set. The idea behind bagging or Random Forest is that sampling data from the empirical distribution approximates sampling from the true underlying distribution, thus enabling quantification of the uncertainty of model estimation.

in a decision tree. Intuitively, if a characteristic is chosen more frequently as a split rule, it becomes more important for splitting the cross section to maximize the mean-variance diversification. For each bootstrap sample, a random subset of characteristics is drawn to grow the tree, and only a fraction of them are selected as split rules. We count the number of times a specific l -th characteristic z_l is used in the first J splits and the total number of appearances in bootstrap subsamples. We define the measure of characteristic importance as follows:

$$\text{Selection Probability } (z_l) = \frac{\#(z_l \text{ is selected at first } J \text{ splits})}{\#(z_l \text{ appears in the bootstrap subsamples})}. \quad (8)$$

Table 6 summarizes the selection probabilities (8) of characteristics. Three rows report the selection probabilities for $J = 1, 2, 3$, respectively. The selection probability increases as J grows from 1 to 3, so we must compare within the row. For each row, we list the top five characteristics in descending order of selection probability.

Table 6: Characteristics Importance by Selection Probability

This table reports the most frequently selected characteristics from the random P-Forest of 1,000 trees. The “Top 1” rows only count the first split for 1,000 trees. The “Top 2” or “Top 3” rows only count the first two or three splits. The numbers reported are the selection frequency for these top characteristics selected out of the 1,000 ensembles. The descriptions of characteristics are listed in Table A.1.

	1	2	3	4	5
Top1	SUE 0.51	SVAR 0.34	CHPM 0.25	RVAR_CAPM 0.23	BASPREAD 0.21
Top2	SUE 0.66	SVAR 0.38	DOLVOL 0.34	CHPM 0.31	BM_IA 0.31
Top3	SUE 0.71	DOLVOL 0.53	BM_IA 0.48	SVAR 0.41	ME_IA 0.41

First, we find the earnings surprise (SUE), one of the most important fundamentals, has a 50% chance of being the first splitting characteristic and a 66% chance of being one of the top two splitting characteristics in the bootstrap sample. Though not considered in linear factor models like FF5 and Q5, SUE seems most useful in non-linear interactive modeling to display the cross-sectional variation of asset returns. Other frequently selected characteristics include return volatility (SVAR), dollar trading volume (DOLVOL), idiosyncratic volatility (RVAR_CAPM), change in profit margin

(CHPM), and industry-adjusted book-to-market ratio (BM_{IA}). The characteristics in Table 6 cover almost all major categories—momentum, size, value, frictions/volatility, and profitability—in the top splits. Notice that characteristics (e.g., investment) not being selected does not imply they are useless, because this table only shows the top three important characteristics.

Interpretability v.s. overfitting. Using tree ensembles such as random forests or boosted trees reduces the risk of overfitting but also makes understanding a single tree more difficult. P-Trees are easily interpretable because they graphically display characteristics for non-linearities and asymmetric interactions, but the model’s effectiveness might be constrained if we cannot show it does not overfit.

In addition to the out-of-sample empirical evaluation presented in Table 5, we compare important characteristics between a random P-Forest and a single P-Tree, using importance metrics to detect overfitting. The similarity of characteristics implies that the global split criteria of P-Tree avoid overfitting, just like ML ensembling methods, but achieve greater interpretability and provide transparent insights into non-linearities and asymmetric interactions. The significant characteristics obtained from the random P-Forest, such as earning surprise (SUE), dollar trading volume (DOLVOL), and industry-adjusted book-to-market ratio (BM_{IA}), are also chosen in the top splits of P-Tree for the 40 years, as illustrated in Figure 4. In addition, the important characteristics listed in Table 6 are also selected in the top splits displayed in the P-Tree diagram for the subsample analysis (refer to Figure A.1).

5.2 Evaluating a Characteristic with P-Tree.

As an alternative pricing kernel, P-Tree can directly evaluate a new characteristic’s incremental (and nonlinear) contribution over a benchmark to complement the literature. Researchers create a new factor by sorting based on a specific characteristic in empirical studies. A crucial research task is to test the independent information of new factors by controlling benchmark models through spanning regression (Fama and French, 1996). However, due to the three-dimensional data structure of firm characteristics, empirical studies rarely examine whether the associated characteristics provide

new information and may suffer information loss (Lettau, 2023). To address this issue, P-Tree provides a new perspective for evaluating the usefulness of characteristics in generating test assets and latent factor models, given a set of benchmark characteristics, such as ME and BM.

In Figure 8, we demonstrate two baseline P-Trees: a three-layer ME baseline P-Tree where all splits are based on ME and a 3-layer ME-BM baseline P-Tree where the first layer is split on ME and the second layer is split on BM. To evaluate the incremental utility for estimating the efficient frontier of a characteristic against ME, we expand the ME baseline P-Tree with one more split on the characteristic of interest or ME. One might expect a split to increase the Sharpe ratio of an MVE portfolio for a shallow P-Tree. Therefore, we only determine if a characteristic provides new information, by comparing the Sharpe ratio improvement of adding this characteristic to the baseline P-Tree over a split on ME. This approach allows us to measure the marginal contribution of the characteristic of interest.

Figure 8: **Demonstration for ME baseline and ME-BM baseline P-Trees**

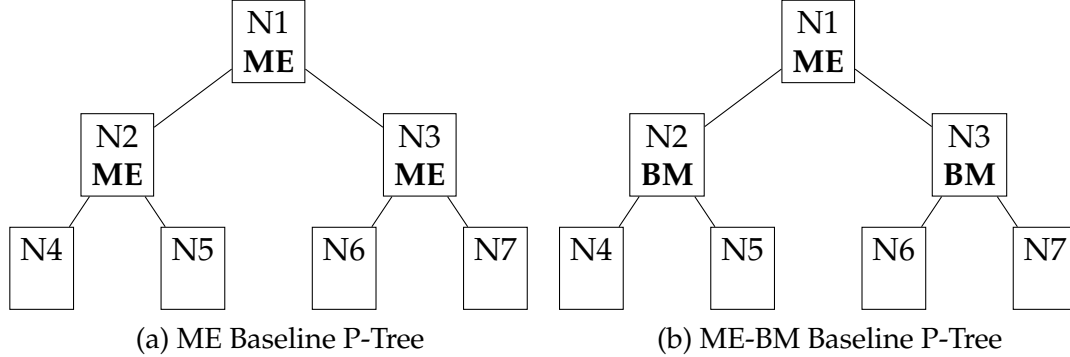
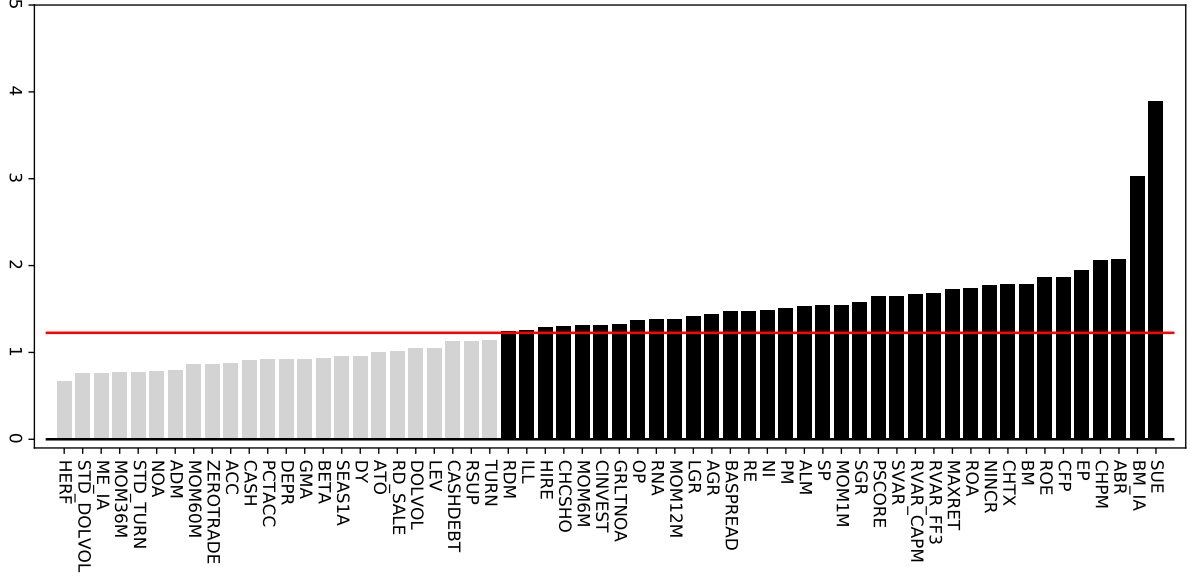


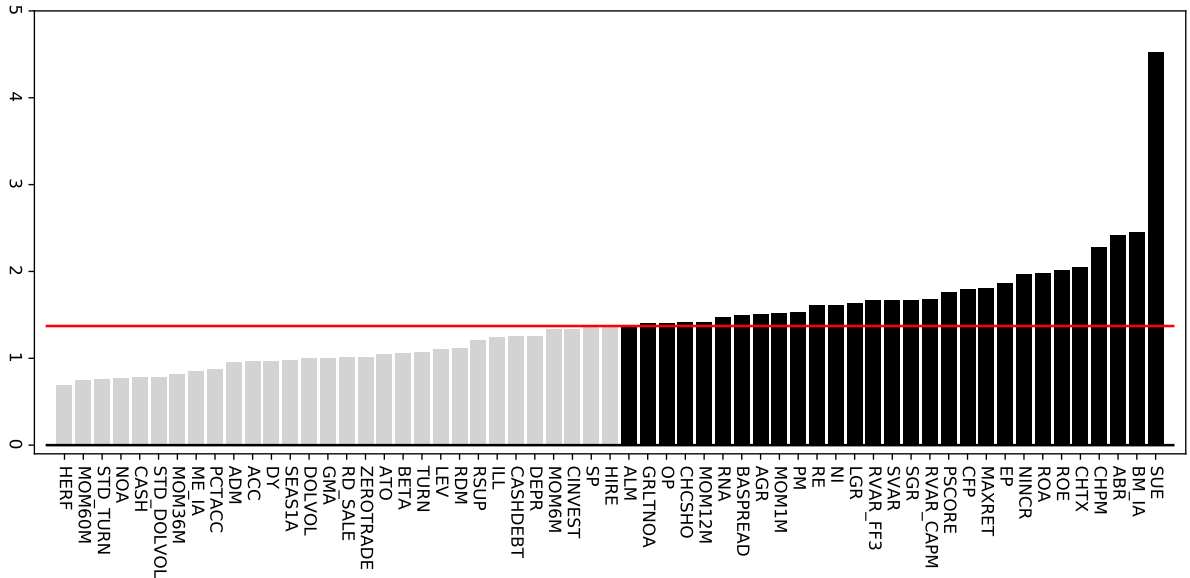
Figure 9, subfigure (a), summarizes the incremental Sharpe ratio of each characteristic on the ME baseline P-Tree. We color those characteristics outperforming ME in black, and those underperforming ME in grey. Also, we sort the characteristics by the Sharpe ratios in ascending order. Even for a shallow P-Tree, we find that 24 characteristics do not provide incremental information against the ME baseline P-Tree. Further, we identify categories other than size, including momentum (SUE, ABR, NINCR), value (BM_LIA, BM, CFP, EP), profitability (CHPM, ROA, ROE), and frictions/volatility (MAXRET, SVAR), that vastly improve investment performance by a simple further split on the ME baseline P-Tree.

Figure 9: Evaluating a Characteristic with P-Tree

This figure shows the Sharpe ratio of P-Tree basis portfolios after splitting on a characteristic, based on either ME or ME-BM baseline P- Trees. Sub-figures (a) and (b) display the results for ME and ME-BM baseline P-Trees, respectively. We sort the characteristics by their Sharpe ratios in ascending order. The red lines show the Sharpe ratios of P-Tree basis portfolios after splitting one step further on the controlling characteristics. A bar above the red line indicates incremental information against the benchmark characteristics and is colored black. Otherwise, the grey bar indicates the characteristic does not provide incremental information. The characteristics are listed in Table A.1.



(a) Controlling ME



(b) Controlling ME-BM

In summary, we assess each characteristic on a benchmark P-Tree by controlling ME and/or BM for its marginal contribution to the investment objective. We can expect these marginal contributions to be smaller when considering a deeper tree as the

benchmark. The relative importance of a characteristic may change when the benchmark P-Tree or controlling characteristics are altered, highlighting the adaptable nature of P-Tree evaluation. This evaluation can complement the conventional factor-spanning regression in evaluating firm characteristics.

Similarly, we assess each characteristic using the ME-BM baseline P-Tree shown in subfigure (b) of Figure 8. This P-Tree has three layers and is based on ME and BM. With a stronger benchmark, some characteristics, such as `RDM`, `ILL`, `HIRE`, `MOM6M`, `CINVEST`, `SP`, do not provide additional information on this upgraded ME-BM baseline P-Tree. Against the ME-BM baseline P-Tree, we find that 30 characteristics do not offer incremental information for estimating the efficient frontier. Consistent with the ME baseline P-Tree results, we find momentum, value, and profitability characteristics are useful in the nonlinear interactive model.

5.3 P-Tree with Macroeconomic Regimes

Increasing empirical evidence indicates factor risk premia/price and factor loading betas can vary significantly under various macroeconomic conditions. Extensive literature studies regime changes and market timing in capital markets, including bull versus bear markets (Fabozzi and Francis, 1977), regime changes (Ang and Timmermann, 2012), time-varying inflation risk premia (Boons et al., 2020), time-varying factor models (Cui et al., 2023), and so on. Given that the cross-sectional driving forces may change over time, examining how P-Tree performs in different states and which characteristics are selected for the mean-variance diversification is valuable. This subsection presents the subsample analysis for implementing P-Trees under various macro regimes and demonstrates their robust performance.

Macroeconomic regimes and characteristics evaluation. For robustness check, We analyze ten macro variables from Table A.2 for subsample analysis in various regimes. The empirical outcomes of P-Tree models under different regimes are demonstrated in Table 7. P-Tree can identify macro regime shifts through time-series splits, similar to the approach of Cong et al. (2022). For demonstration, we detect regime switches

under alternative split criteria and data-driven time-series splits in Appendix III.D.²¹ However, for our baseline P-Tree, we split the sample into "Top" and "Btm" (Bottom) regimes using the 50% 10-year rolling-window quantiles and fit a separate model for each regime. Therefore, we evaluate the usefulness of P-Tree in all positive and negative macroeconomic environments.

Table 7: **P-Tree Performance under Regime Switches**

This table shows the performance of P-Tree models across various macroeconomic states. The time series of each macro variable is split into "Top" and "Btm" regimes using the 50% 10-year rolling-window quantiles. A simple P-Tree is then built for each regime, and similar performance statistics are reported in Table 3, along with the top three splitting characteristics. The rows labeled "Com" report the average return and Sharpe ratio of the combined P-Tree factor of the "Top" and "Btm" regimes.

	Regime	AVG	SR	GRS	p -value	$ \bar{\alpha} $	$\sqrt{\alpha^2}$	$\bar{R}^2$	Top Three Characteristics		
DFY	Top	1.43	8.97	73.21	0.00	1.12	1.43	56.22	SUE	DOLVOL	CFP
	Btm	1.3	8.16	63.08	0.00	1.58	1.88	43.70	BASPREAD	SUE	ZEROTRADE
	Com	1.37	8.52								
DY	Top	1.35	9.43	55.21	0.00	1.10	1.35	65.88	CHPM	BM.IA	ME
	Btm	1.31	7.7	72.39	0.00	1.09	1.33	47.08	ABR	ZEROTRADE	DOLVOL
	Com	1.32	8.21								
EP	Top	1.29	11.39	68.73	0.00	1.15	1.41	68.00	RVAR.CAPM	ME.IA	ME
	Btm	1.4	7.87	82.93	0.00	1.23	1.52	47.03	SVAR	SUE	ME
	Com	1.37	8.43								
ILL	Top	1.24	8.48	89.34	0.00	1.17	1.45	48.88	RVAR.CAPM	SUE	ZEROTRADE
	Btm	1.26	11.01	74.48	0.00	0.91	1.18	56.27	SUE	DOLVOL	SP
	Com	1.24	9.18								
INFL	Top	1.15	7.3	31.89	0.00	1.07	1.33	66.79	CFP	ATO	ME
	Btm	1.44	7.73	70.41	0.00	1.27	1.57	44.41	RVAR.CAPM	DOLVOL	SUE
	Com	1.34	7.41								
LEV	Top	1.24	7.88	44.34	0.00	0.97	1.32	57.73	CFP	ALM	STD.DOLVOL
	Btm	1.45	7.03	55.18	0.00	1.23	1.52	46.24	NI	MOM12M	BM.IA
	Com	1.37	7.18								
NI	Top	1.69	7.91	54.29	0.00	1.73	2.23	45.54	RVAR.CAPM	ME.IA	RD.SALE
	Btm	1.58	8.25	66.19	0.00	1.17	1.45	52.90	ROA	DOLVOL	SUE
	Com	1.63	8.05								
SVAR	Top	1.44	9.45	79.11	0.00	1.09	1.49	55.79	SUE	DOLVOL	CFP
	Btm	1.41	9.54	85.86	0.00	1.16	1.40	49.93	ROA	ZEROTRADE	DOLVOL
	Com	1.42	9.51								
TBL	Top	1.08	10.5	61.48	0.00	0.97	1.25	54.85	CHCSHO	SUE	DOLVOL
	Btm	1.43	8.98	108.59	0.00	1.19	1.49	50.87	SUE	DOLVOL	CFP
	Com	1.32	8.74								
TMS	Top	1.27	9.33	75.74	0.00	1.12	1.28	62.06	OP	STD.DOLVOL	SUE
	Btm	1.21	8.12	62.59	0.00	0.82	1.13	52.94	SUE	DOLVOL	BM.IA
	Com	1.24	8.65								

First, the set of useful characteristics might change across different macroeconomic conditions. P-Tree has robust performance, adapts to macro regimes, and selects different sets of characteristics in the top three splits. For example, when the term spread is low, or market volatility is high, the roles of SUE and DOVOL for firm fundamentals and trading volume are more important for cross-sectional mean-variance

²¹ Appendix III.D defines the time-series split criteria on the macro variables. For example, Figure I.4 shows a time-series split P-Tree on Market Volatility (SMVAR), which identifies a high-volatility regime and a low-volatility regime. Then, two separate cross-sectional P-Trees are developed on each regime. This asset-pricing regime determination remains data-driven and instrumented by macro predictors.

diversification. However, characteristics such as `BASPREAD` and `RVAR_CAPM`, representing frictions for liquidity and volatility, are often chosen for scenarios involving high net equity issuance, low default yield, low market liquidity, or low inflation.

Second, P-Tree test assets consistently provide informative results for asset pricing considerations: the GRS tests against Fama-French five factors are rejected for each regime, and the Sharpe ratios of the MVE portfolios spanned by P-Tree test assets are all greater than 7. Interestingly, the GRS statistics (weighted pricing errors) are significantly larger for scenarios that involve low inflation or a low t-bill rate, which implies the weakest moments of commonly used factors capturing the cross-sectional signals. The average alphas are larger than those of “P-Tree1” in Table 5, Panel A, implying the test assets under extreme macro regimes are even more challenging to price by common factors. Finally, we combine the P-Tree factors in two regimes and find the “Com” (Combined) factor has larger Sharpe ratios than the first P-Tree factor in Table 5, Panel A, which implements an unconditional model.

6 Conclusion

Estimating the mean-variance efficient (MVE) 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](#)). Our paper presents a novel interpretable AI model, Panel Tree (P-Tree), that effectively addresses the traditional empirical challenge of generalized security sorting and delivers superior performance for analyzing the (unbalanced) panel of individual asset returns. P-Tree utilizes high-dimensional characteristics containing rich information on the joint distribution of asset returns, to generate basis portfolios as diversified test assets and building blocks of the pricing kernel. More generally, P-Trees broaden tree-based models beyond pure prediction and analyze panel data effectively with economics-guided tree growth, while retaining interpretability and handling asymmetric and nonlinear interactions in noisy environments.

Our empirical study of U.S. equities shows that the efficient frontiers spanned by P-Tree test assets vastly outperform those from commonly used observable and ma-

chine learning latent factor models, as well as conventional test assets. Boosted P-Trees generate sequences of test assets to challenge all existing factor models, providing a diversified test asset alternative to characteristic-sorted portfolios. In addition to leaf basis portfolios, our unified approach under the MVE framework provides tradable P-Tree factors and out-of-sample annualized Sharpe ratios over 3. The factor models also outperform benchmark models in explaining the cross-sectional return variation. Finally, P-Tree offers a new nonlinear and interactive framework to assess the conditional information of high-dimensional characteristics, complementing traditional factor-spanning regression in evaluating firm characteristics.

References

- Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009). Basis assets. *Review of Financial Studies* 22(12), 5133–5174.
- Ait-Sahalia, Y. and D. Xiu (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* 201(2), 384–399.
- 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.
- Ang, A. and A. Timmermann (2012). Regime changes and financial markets. *Annu. Rev. Financ. Econ.* 4(1), 313–337.
- Ao, M., L. Yingying, and X. Zheng (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies* 32(7), 2890–2919.
- Avramov, D., S. Cheng, and L. Metzker (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science* 69(5), 2587–2619.
- Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert (2023). Option return predictability with machine learning and big data. *Review of Financial Studies* 36(9), 3548–3602.
- Barillas, F. and J. Shanken (2017). Which alpha? *Review of Financial Studies* 30(4), 1316–1338.
- Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *Review of Financial Studies* 34(2), 1046–1089.
- Boons, M., F. Duarte, F. De Roon, and M. Szymanowska (2020). Time-varying inflation risk and stock returns. *Journal of Financial Economics* 136(2), 444–470.
- Breiman, L. (2001). Random forests. *Machine learning* 45(1), 5–32.

- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984). *Classification and regression trees*. Routledge.
- Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance* 78(1), 487–557.
- Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. *Journal of Finance, Forthcoming*.
- Buraschi, A. and A. Jiltsov (2005). Inflation risk premia and the expectations hypothesis. *Journal of Financial Economics* 75(2), 429–490.
- Chamberlain, G. and M. Rothschild (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica* 51, 1281–1304.
- Chen, L., M. Pelger, and J. Zhu (2023). Deep learning in asset pricing. *Management Science, Forthcoming*.
- Chen, Y., M. Cliff, and H. Zhao (2017). Hedge funds: The good, the bad, and the lucky. *Journal of Financial and Quantitative Analysis* 52(3), 1081–1109.
- Chipman, H. A., E. I. George, and R. E. McCulloch (2010). BART: Bayesian additive regression trees. *Annals of Applied Statistics* 4, 266–298.
- Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001). Trading activity and expected stock returns. *Journal of financial Economics* 59(1), 3–32.
- Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton University Press.
- Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance* 66(4), 1047–1108.
- Cong, L. W., G. Feng, J. He, and J. Li (2022). Sparse Modeling Under Grouped Heterogeneity with an Application to Asset Pricing. Technical report, Cornell University.
- Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021). AlphaPortfolio: Direct Construction Trough Deep Reinforcement Learning and Interpretable AI. Technical report, Cornell University.
- Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15(3), 373–394.
- Connor, G. and R. A. Korajczyk (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21(2), 255–289.
- Creal, D. and J. Kim (2023). Empirical Asset Pricing with Bayesian Regression Trees. Technical report, University of Notre Dame.
- Cui, L., G. Feng, Y. Hong, and J. Yang (2023). Time-Varying Factor Selection: A Sparse Fused GMM Approach. Technical report, City University of Hong Kong.

- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52(3), 1035–1058.
- 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.
- Fabozzi, F. J. and J. C. Francis (1977). Stability tests for alphas and betas over bull and bear market conditions. *Journal of Finance* 32(4), 1093–1099.
- 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 (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1–22.
- Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29(1), 69–103.
- Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2022). Structural deep learning in conditional asset pricing. Technical report, Princeton University.
- Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance* 75(3), 1327–1370.
- Feng, G., J. He, N. Polson, and J. Xu (2023). Deep Learning of Characteristics-Sorted Factor Models. *Journal of Financial and Quantitative Analysis, Forthcoming*.
- Feng, G., L. Jiang, J. Li, and Y. Song (2023). Deep tangency portfolio. Technical report, City University of Hong Kong.
- Freund, Y. and R. E. Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1), 119–139.
- Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica* 84(3), 985–1046.
- 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.
- Giglio, S., D. Xiu, and D. Zhang (2022). Test Assets and Weak Factors. Technical report, University of Chicago.

- 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.
- Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics* 222(1), 429–450.
- Hamilton, J. D. and G. Lin (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics* 11(5), 573–593.
- 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., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.
- He, J. and P. R. Hahn (2023). Stochastic tree ensembles for regularized nonlinear regression. *Journal of the American Statistical Association* 118(541), 551–570.
- He, J., S. Yalov, and P. R. Hahn (2019). XBART: Accelerated Bayesian additive regression trees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1130–1138.
- Hoberg, G. and I. Welch (2009). Optimized vs. sort-based portfolios. In *AFA 2010 Atlanta Meetings Paper*.
- 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 (2015). Digesting anomalies: An investment approach. *Review of Financial Studies* 28(3), 650–705.
- Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies* 33(5), 2019–2133.
- Jarrow, R. A., R. Murataj, M. T. Wells, and L. Zhu (2021). The low-volatility anomaly and the adaptive multi-factor model. Technical report, Cornell University.
- Kelly, B., D. Palhares, and S. Pruitt (2023). Modeling corporate bond returns. *Journal of Finance* 78(4), 1967–2008.
- 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.
- Kim, S., R. A. Korajczyk, and A. Neuhierl (2021). Arbitrage portfolios. *Review of Financial Studies* 34(6), 2813–2856.

- Kozak, S. and S. Nagel (2022). When do cross-sectional asset pricing factors span the stochastic discount factor? Technical report, University of Michigan.
- 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.
- Leippold, M., Q. Wang, and W. Zhou (2022). Machine learning in the chinese stock market. *Journal of Financial Economics* 145(2), 64–82.
- Lettau, M. (2023). High-dimensional factor models and the factor zoo. Technical report, National Bureau of Economic Research.
- Lettau, M. and M. Pelger (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics* 218(1), 1–31.
- Lettau, M. and M. Pelger (2020b). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies* 33(5), 2274–2325.
- Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96(2), 175–194.
- Lopez-Lira, A. and N. L. Roussanov (2020). Do Common Factors Really Explain the Cross-Section of Stock Returns? Technical report, University of Pennsylvania.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance* 7, 77–99.
- 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 T. Yamagata (2023). Testing for alpha in linear factor pricing models with a large number of securities. *Journal of Financial Econometrics*.
- Rendleman Jr, R. J., C. P. Jones, and H. A. Latane (1982). Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics* 10(3), 269–287.
- Rossi, A. G. and A. Timmermann (2015). Modeling covariance risk in Merton’s ICAPM. *Review of Financial Studies* 28(5), 1428–1461.
- Sorensen, E. H., K. L. Miller, and C. K. Ooi (2000). The decision tree approach to stock selection. *Journal of Portfolio Management* 27(1), 42–52.
- van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man versus machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies* 36(6), 2361–2396.
- Zhu, L., S. Basu, R. A. Jarrow, and M. T. Wells (2020). High-dimensional estimation, basis assets, and the adaptive multi-factor model. *Quarterly Journal of Finance* 10(04), 2050017.

Appendices

Algorithm Growing a Single P-Tree

```

1: procedure GROWTREE(root)
2: Input Asset returns  $\mathbf{r}_{i,t}$  and ranked characteristics  $\mathbf{z}_{i,t}$ 
3: outcome Grow P-Tree from the root node, form leaf basis portfolios
4:   for  $j$  from 1 to  $J$  do                                     ▷ Loop over number of iterations
5:     if current depth  $\geq d_{\max}$  then
6:       return.
7:     else
8:       Search the tree, find all leaf nodes  $\mathcal{N}$ 
9:       for each leaf node  $N$  in  $\mathcal{N}$  do                             ▷ Loop over all current leaf nodes
10:        for each candidate split  $\tilde{\mathbf{c}}_{k,m,N}$  in  $\mathcal{C}_N$  do
11:          Partition data temporally in  $N$  according to  $\tilde{\mathbf{c}}_{k,m,N}$ .
12:          if Either left or right child of  $N$  does not satisfy minimal leaf size then
13:             $\mathcal{L}(\tilde{\mathbf{c}}_{k,m,N}) = -\infty$ .
14:          else
15:            Calculate leaf basis portfolios.
16:            Estimate SDF using all leaf basis portfolios as in (2).
17:            Calculate the split criteria  $\mathcal{L}(\tilde{\mathbf{c}}_{k,m,N})$  in (4).
18:          end if
19:        end for
20:      end for
21:      Find the best leaf node and split rule that maximize split criteria

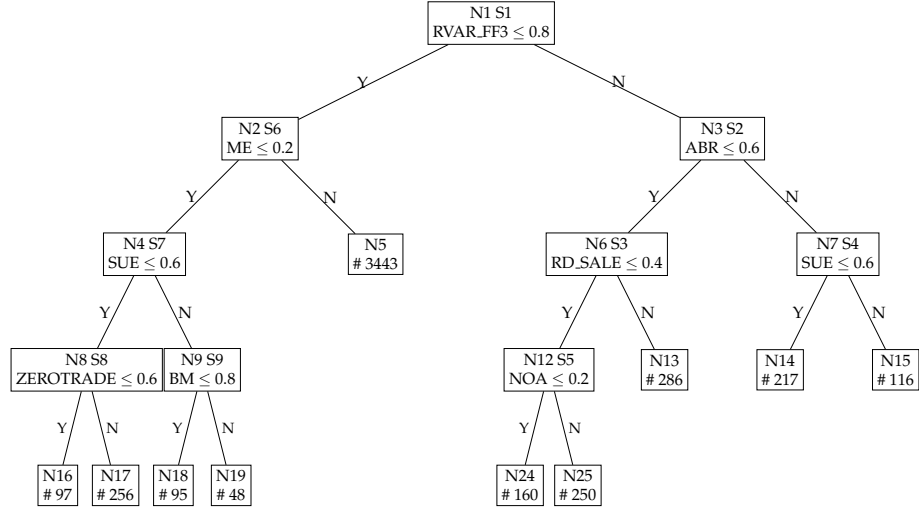

$$\tilde{\mathbf{c}}_j = \arg \max_{N \in \mathcal{N}, \tilde{\mathbf{c}}_{k,m,N} \in \mathcal{C}_N} \{\mathcal{L}(\tilde{\mathbf{c}}_{k,m,N})\}$$


22:      Split the node selected with the  $j$ -th split rule of the tree  $\tilde{\mathbf{c}}_j$ .
23:    end if
24:  end for
25:  return  $\{\mathbf{R}_t^{(J)}, f_t^{(J)}\}$ 
26: end procedure

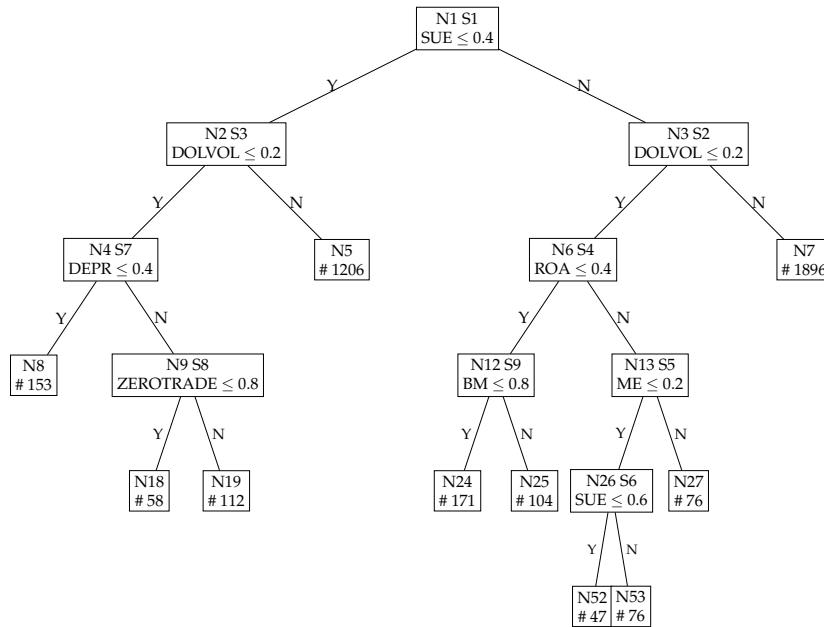
```

Figure A.1: Tree Diagram, 20 Years, 1981-2000

Format follows Figure 4. Subfigure (a) shows P-Tree in the sample from 1981-2000. Subfigure (b) shows P-Tree in the sample from 2001-2020.



(a) 20-year Sample (1981-2000)



(b) 20-year Sample (2001-2020)

Table A.1: Equity Characteristics

This table lists the description of 61 characteristics used in the empirical study.

No.	Characteristics	Description	Category
1	ABR	Abnormal returns around earnings announcement	Momentum
2	ACC	Operating Accruals	Investment
3	ADM	Advertising Expense-to-market	Intangibles
4	AGR	Asset growth	Investment
5	ALM	Quarterly Asset Liquidity	Intangibles
6	ATO	Asset Turnover	Profitability
7	BASPREAD	Bid-ask spread (3 months)	Frictions
8	BETA	Beta (3 months)	Frictions
9	BM	Book-to-market equity	Value-versus-growth
10	BM_IA	Industry-adjusted book to market	Value-versus-growth
11	CASH	Cash holdings	Value-versus-growth
12	CASHDEBT	Cash to debt	Value-versus-growth
13	CFP	Cashflow-to-price	Value-versus-growth
14	CHCSHO	Change in shares outstanding	Investment
15	CHPM	Change in Profit margin	Profitability
16	CHTX	Change in tax expense	Momentum
17	CINVEST	Corporate investment	Investment
18	DEPR	Depreciation / PP&E	Momentum
19	DOLVOL	Dollar trading volume	Frictions
20	DY	Dividend yield	Value-versus-growth
21	EP	Earnings-to-price	Value-versus-growth
22	GMA	Gross profitability	Investment
23	GRLTNOA	Growth in long-term net operating assets	Investment
24	HERF	Industry sales concentration	Intangibles
25	HIRE	Employee growth rate	Intangibles
26	ILL	Illiquidity rolling (3 months)	Frictions
27	LEV	Leverage	Value-versus-growth
28	LGR	Growth in long-term debt	Investment
29	MAXRET	Maximum daily returns (3 months)	Frictions
30	ME	Market equity	Frictions
31	ME_IA	Industry-adjusted size	Frictions
32	MOM12M	Cumulative Returns in the past (2-12) months	Momentum
33	MOM1M	Previous month return	Momentum
34	MOM36M	Cumulative Returns in the past (13-35) months	Momentum
35	MOM60M	Cumulative Returns in the past (13-60) months	Momentum
36	MOM6M	Cumulative Returns in the past (2-6) months	Momentum
37	NI	Net Equity Issue	Investment
38	NINCR	Number of earnings increases	Momentum
39	NOA	Net Operating Assets	Investment
40	OP	Operating profitability	Profitability

Continue: Equity Characteristics

No.	Characteristics	Description	Category
41	PCTACC	Percent operating accruals	Investment
42	PM	Profit margin	Profitability
43	PS	Performance Score	Profitability
44	RD.SALE	R&D to sales	Intangibles
45	RDM	R&D to market	Intangibles
46	RE	Revisions in analysts' earnings forecasts	Intangibles
47	RNA	Return on Net Operating Assets	Profitability
48	ROA	Return on Assets	Profitability
49	ROE	Return on Equity	Profitability
50	RSUP	Revenue surprise	Momentum
51	RVAR_CAPM	Idiosyncratic Volatility - CAPM (3 months)	Frictions
52	RVAR_FF3	Res. var. - Fama-French 3 factors (3 months)	Frictions
53	SVAR	Return variance (3 months)	Frictions
54	SEAS1A	1-Year Seasonality	Intangibles
55	SGR	Sales growth	Value-versus-growth
56	SP	Sales-to-price	Value-versus-growth
57	STD.DOLVOL	Std of dollar trading volume (3 months)	Frictions
58	STD.TURN	Std. of Share turnover (3 months)	Frictions
59	SUE	Unexpected quarterly earnings	Momentum
60	TURN	Shares turnover	Frictions
61	ZEROTRADE	Number of zero-trading days (3 months)	Frictions

Table A.2: Macroeconomic Variables

This table lists the description of macro variables used in the empirical study.

No.	Variable Name	Description
1	DFY	Default yield
2	DY	Dividend yield of S&P 500
3	EP	Earnings-to-price of S&P 500
4	ILL	Pastor-Stambaugh illiquidity
5	INFL	Inflation
6	LEV	Leverage of S&P 500
7	NI	Net equity issuance of S&P 500
8	SMVAR	Stock Market (S&P 500) Volatility
9	TBL	Three-month treasury bill rate
10	TMS	Term spread

Table A.3: Subsample Analysis for Testing the Boosted P-Tree Growth

This table reports the subsample Analysis for results in Table 2. Panel B is for the first 20-year sample from 1981–2000, and panel C is for the recent 20-year sample from 2001–2020.

	Sharpe Ratio		CAPM Test		FF5 Test		Expanding Factors Test			BS Test
	Single	Multiple	α (%)	t -stat	α (%)	t -stat	α (%)	t -stat	R^2	p -value
Panel B: 20 Years (1981-2000)										
1	7.12	7.12	1.86	31.23	1.78	28.53	-	-	-	-
2	1.09	7.75	0.25	4.09	0.11	1.91	-0.83	-4.80	0.28	0.00
3	2.26	9.45	0.39	9.66	0.34	8.30	0.92	8.19	0.17	0.00
4	6.48	11.36	1.80	28.41	1.76	26.07	1.75	10.16	0.03	0.00
5	1.59	12.71	0.33	6.54	0.28	5.37	1.11	6.85	0.36	0.00
6	1.65	13.67	0.37	6.83	0.26	4.74	-1.03	-5.34	0.40	0.00
7	1.72	14.94	0.33	7.20	0.32	6.44	1.25	5.67	0.20	0.00
8	1.52	15.78	0.39	6.19	0.31	4.75	-1.32	-5.12	0.35	0.00
9	3.02	17.89	1.03	13.04	0.93	11.56	2.68	8.42	0.26	0.00
10	1.93	19.18	0.83	8.14	0.67	7.18	-2.41	-6.27	0.52	0.00
11	1.34	20.72	0.38	5.36	0.19	2.85	-2.18	-7.33	0.45	0.00
12	2.22	22.89	0.36	9.49	0.31	8.24	1.46	7.36	0.39	0.00
13	1.51	24.14	0.35	6.15	0.23	4.07	1.82	6.06	0.43	0.00
14	2.11	25.71	0.58	8.97	0.48	7.27	-2.43	-5.93	0.40	0.00
15	2.39	28.38	0.87	10.21	0.76	9.10	3.69	7.64	0.52	0.00
16	2.53	30.37	0.53	10.84	0.44	8.85	2.40	7.12	0.31	0.00
17	1.50	32.42	0.30	6.10	0.31	5.78	2.65	7.17	0.35	0.00
18	1.63	33.89	0.37	6.76	0.28	4.99	-2.79	-5.64	0.37	0.00
19	1.84	35.78	0.47	7.71	0.37	5.95	3.54	7.34	0.30	0.00
20	2.70	37.93	0.89	11.62	0.79	10.21	-4.66	-6.90	0.44	0.00
Panel C: 20 Years (2001-2020)										
1	5.82	5.82	1.51	25.60	1.47	24.01	-	-	-	-
2	2.14	6.64	0.58	9.23	0.52	8.04	0.91	6.44	0.03	0.00
3	0.91	7.53	0.18	3.37	0.09	1.85	-0.86	-5.82	0.36	0.00
4	1.35	8.60	0.29	5.58	0.31	5.72	1.00	8.65	0.17	0.00
5	1.08	9.30	0.31	4.24	0.26	3.46	-1.02	-6.38	0.42	0.00
6	2.63	10.36	0.60	11.41	0.56	10.31	1.04	7.61	0.12	0.00
7	1.39	11.31	0.29	5.75	0.26	4.90	1.00	7.23	0.28	0.00
8	1.60	11.93	0.40	6.71	0.30	5.26	-0.85	-5.62	0.44	0.00
9	1.04	12.73	0.21	4.08	0.18	3.64	-1.01	-7.45	0.36	0.00
10	1.64	14.32	0.30	6.92	0.25	5.67	1.11	7.80	0.44	0.00
11	1.69	15.73	0.36	7.15	0.30	5.94	1.43	6.79	0.33	0.00
12	1.58	16.82	0.60	6.65	0.51	5.64	2.49	5.72	0.24	0.00
13	1.48	18.07	0.41	6.16	0.29	4.61	-1.77	-5.52	0.47	0.00
14	1.39	19.46	0.32	5.73	0.26	4.61	1.68	6.05	0.40	0.00
15	1.48	20.59	0.35	6.15	0.21	4.19	-1.38	-6.25	0.58	0.00
16	1.40	21.91	0.34	5.78	0.27	4.61	1.84	6.29	0.46	0.00
17	1.27	23.40	0.39	5.18	0.25	3.39	-2.55	-6.99	0.51	0.00
18	1.44	24.36	0.34	5.99	0.28	4.99	-1.61	-4.97	0.48	0.00
19	1.58	25.45	0.33	6.62	0.28	5.56	-1.75	-5.21	0.42	0.00
20	1.88	26.52	0.58	8.02	0.50	6.90	-2.76	-4.96	0.39	0.00

Internet Appendices for “Growing the Efficient Frontier on Panel Trees”

Lin William Cong

Guanhao Feng

Jingyu He

Xin He

The online appendix is organized as follows. Section I uses simulation to demonstrate P-Tree’s advantage. Section II illustrates the flexibility of P-Tree in adjusting for benchmark factors such as the market factor. Section III further reveals the framework’s flexibility with an alternative split criterion that minimizes the realized pricing error of individual stock returns, with a subsection extending the criteria to split the time series.

I Simulation

In this section, simulation studies are conducted to demonstrate the practical capability of P-Tree in choosing characteristics and the efficiency of P-Tree test assets in capturing non-linear interactions among characteristics. The positive simulation evidence provides further support for the modeling ability of P-Tree. We calibrate the return-generating process using the same empirical data sample and consider three genuine characteristics and their interactions and nonlinear terms without loss of generality. Moreover, to simulate the true investor data environment, we consider a large set of useless or noisy characteristics that correlate with the true characteristics but do not impact generating returns.

Data-generating process. Specifically, the ranked asset characteristics $\mathbf{z}_{i,t}$ (a vector of size 61×1) are simulated from a VAR(1) model. This design ensures that even redundant characteristics still correlate with genuine ones. Redundant characteristics may be selected in scenarios of low signal-to-noise ratios or small sample sizes. The true return-generating process is assumed only to involve three characteristics among all 61 (ME, BM, and MOM12M). The model incorporates simple nonlinearity and interaction,

as described below:

$$r_{i,t} = \text{mkt}_t + \kappa \left[c_1 \text{ME}_{i,t-1} + c_2 \text{BM}_{i,t-1} + c_3 \text{MEBM}_{i,t-1} + c_4 \text{MOM12M}_{i,t-1} + c_5 \text{MOM12M}_{i,t-1}^2 \right] + \epsilon_{i,t}, \quad (9)$$

where $r_{i,t}$ represents the stock excess return, mkt_t denotes the market excess return, MEBM is the interaction of ME and BM, MOM12M^2 is the quadratic term, and the error terms are simulated from $\epsilon_{i,t} \sim i.i.d.N(0, \sigma^2)$. The parameter σ is calibrated from real data, and so is $[c_1, \dots, c_5]$. The parameter $\kappa = [0.5, 1, 2]$ allows us to adjust the strength of the characteristic signal, with the value of $\kappa = 1$ representing our actual data in the empirical study.

Each simulation sample consists of 1,000 assets, 500 periods for in-sample analysis, and another 500 periods for out-of-sample analysis. Table I.1 presents the average results of 10 repetitions, including the maximal Sharpe ratios and market alphas of P-Tree test assets. The former displays the tangency portfolio on the efficient frontier, whereas the latter demonstrates the pricing difficulty of the market factor.

Conventional characteristics-sorted portfolios. In Table I.1, “DECILE AVG(30)” presents the average performance of three MVE portfolios and test assets for decile portfolios sorted on three true characteristics. This configuration yields a Sharpe ratio of 3.12, an average absolute α of 0.39%, and a root mean squared α of 0.47% when $\kappa = 1$. Meanwhile, the performance of “DECILE AVG(30)” is quite consistent in out-of-sample results, because of the pre-specified true characteristics. Without knowing the true characteristics, “DECILE AVG(610)” on 61 characteristics reduces the Sharpe ratio to 0.97 and the alphas to 0.10% and 0.12%, whereas the out-of-sample results reveal lower Sharpe ratios. Across the different κ 's, we find higher Sharpe ratios and alphas for large κ than for small κ cases.

The return-generating process described in (9) involves interaction and quadratic terms. We also examine bivariate-sorted portfolios (5×5) using the true characteristics. These portfolios consistently yield higher Sharpe ratios and alphas than decile portfolios, demonstrating their superior performance. Our findings indicate that accounting for potential characteristic interactions can lead to pricing information gains.

Table I.1: **Simulation Performance Comparison**

The table presents the maximal Sharpe ratios, average absolute market α (%), and root mean squared α (%) for different sets of test assets. The average results are reported based on 10 simulations. We test up to 20 P-Trees, each containing 10 leaf basis portfolios. The numbers in parentheses represent the average number of leaf basis portfolios. Benchmark results include the following specifications: (1) market factor; (2) average results for 3 sets of decile portfolios of true char.; (3) average results for 61 sets of decile portfolios; and (4 - 6) bivariate-sort (5×5) portfolios on either two true char., including ME and BM, ME and MOM12M, and BM and MOM12M.

	$\kappa = 0.5$			$\kappa = 1.0$			$\kappa = 2.0$		
	SR	$ \alpha $	$\sqrt{\alpha^2}$	SR	$ \alpha $	$\sqrt{\alpha^2}$	SR	$ \alpha $	$\sqrt{\alpha^2}$
Panel A: In-Sample Results									
P-Tree1	3.31	0.43	0.54	6.14	0.73	0.94	12.00	1.54	1.89
P-Tree1-5	4.47	0.24	0.33	7.58	0.34	0.52	14.23	0.63	1.00
P-Tree1-10	6.15	0.19	0.27	9.75	0.26	0.42	17.45	0.49	0.79
P-Tree1-15	7.86	0.17	0.24	12.04	0.23	0.37	21.11	0.41	0.68
P-Tree1-20	9.76	0.16	0.22	14.49	0.22	0.34	25.45	0.37	0.61
MKT(1)	0.55	0.00	0.00	0.55	0.00	0.00	0.55	0.00	0.00
DECILE AVG(30)	1.73	0.21	0.25	3.12	0.39	0.47	5.97	0.76	0.92
DECILE AVG(610)	0.81	0.08	0.09	0.97	0.10	0.12	1.36	0.15	0.19
ME-BM(25)	2.63	0.31	0.40	4.92	0.57	0.78	9.60	1.12	1.54
ME-MOM12M(25)	2.42	0.29	0.37	4.49	0.55	0.70	8.74	1.09	1.37
BM-MOM12M(25)	2.95	0.35	0.46	5.60	0.68	0.89	11.00	1.35	1.75
Panel B: Out-of-Sample Results									
P-Tree1	2.66	0.61	0.72	5.61	0.88	1.06	11.38	1.56	1.90
P-Tree1-5	2.35	0.58	0.63	5.27	0.68	0.77	10.80	0.89	1.14
P-Tree1-10	2.09	0.59	0.62	4.90	0.65	0.72	10.21	0.77	0.98
P-Tree1-15	1.95	0.59	0.62	4.69	0.63	0.70	9.90	0.72	0.91
P-Tree1-20	1.90	0.59	0.62	4.55	0.63	0.68	9.58	0.70	0.87
P-Tree16-20	0.80	0.59	0.61	1.60	0.61	0.65	3.12	0.64	0.73
MKT(1)	0.45	0.00	0.00	0.45	0.00	0.00	0.45	0.00	0.00
DECILE AVG(30)	1.55	0.20	0.25	3.03	0.39	0.48	6.00	0.77	0.94
DECILE AVG(610)	0.49	0.08	0.09	0.70	0.10	0.13	1.17	0.16	0.20
ME-BM(25)	2.23	0.29	0.41	4.58	0.56	0.78	9.20	1.12	1.54
ME-MOM12M(25)	2.20	0.31	0.39	4.40	0.58	0.72	8.70	1.12	1.41
BM-MOM12M(25)	2.72	0.36	0.46	5.41	0.69	0.89	10.73	1.35	1.76

P-Tree test assets. P-Trees are an endogenous solution for cross-sectional splitting, test asset creation, and efficient frontier spanning. In Table I.1, the P-Tree test assets (“P-Tree1”) exhibit an impressive in-sample Sharpe ratio (6.14), absolute alpha (0.73%), squared alpha (0.94%), out-of-sample Sharpe ratio (5.61), absolute alpha (0.88%), and squared alpha (1.06%) when $\kappa = 1$. When the κ increases to 2, we observe larger Sharpe ratios and alphas for both in-sample and out-of-sample results. Meanwhile, the first P-Tree is parsimonious, requiring fewer assets than alternative approaches to

achieve comparable Sharpe ratios; for example, the first P-Tree only needs 10 assets to achieve an out-of-sample Sharpe ratio of 5.61, whereas the alternative bivariate-sorted portfolios on true characteristics require 25 assets to get comparable Sharpe ratios. Further, the test assets of the first P-Tree have larger out-of-sample alphas than other benchmark test assets.

The boosted P-Trees are denoted as “P-Tree1-5” to “P-Tree1-20.” Although the in-sample Sharpe ratio increases when adding boosted factors, the alphas are lower from the first to the 20-th P-Tree. Therefore, P-Trees capture the important information in early boosting steps, whereas the subsequent P-Trees provide declining information. We observe lower Sharpe ratios and higher alphas for out-of-sample results for all P-Tree specifications. Meanwhile, “P-Tree1” provides the largest out-of-sample Sharpe ratios and alphas. This finding confirms the early P-Trees capture more information, given such a simple return-generating process involving only three true characteristics as in (9).

In conclusion, the P-Tree test assets outperform most pre-specified sorted portfolio specifications in risk-adjusted or model-based investment criteria, such as Sharpe ratios and alphas. The outperformance is more pronounced when the predictive signal κ is larger. In this interactive and nonlinear return-generating scenario (9), P-Tree performances are competitive with all the conventional sorted portfolios.

P-Tree with incomplete information. The literature documents hundreds of characteristics, factors, and anomalies (Cochrane, 2011; Green et al., 2017; Hou et al., 2020; Feng et al., 2020; Bryzgalova et al., 2023). We must include all the valuable characteristics to span the efficient frontier fully. However, we still do not know whether all the characteristics are helpful. Also, we are agnostic if more valuable characteristics are still not found.

To this end, we investigate the performance of P-Tree without observing the full set of true characteristics, given the return-generating process (9). Specifically, we ignore MOM1 2M and re-train P-Trees. Table I.2 reports the simulation results. Once we ignore one of the true characteristics, we find a considerable decline in Sharpe ratios and alphas for both in-sample and out-of-sample analysis. Also, the alphas of P-Tree

Table I.2: **Simulation Performance for P-Tree Omitting a Characteristic MOM12M**

This table shows the Sharpe ratios and alphas of P-Tree test assets omitting a true characteristic MOM12M. Table format follows Table I.1.

	$\kappa = 0.5$			$\kappa = 1.0$			$\kappa = 2.0$		
	SR	$ \alpha $	$\sqrt{\alpha^2}$	SR	$ \alpha $	$\sqrt{\alpha^2}$	SR	$ \alpha $	$\sqrt{\alpha^2}$
Panel A: In-Sample Results									
P-Tree1	2.73	0.32	0.41	4.86	0.60	0.77	9.42	1.30	1.66
P-Tree1-5	4.00	0.18	0.26	6.20	0.26	0.40	11.52	0.47	0.81
P-Tree6-10	1.66	0.12	0.15	2.35	0.16	0.21	3.14	0.23	0.31
P-Tree11-15	1.71	0.12	0.16	2.19	0.16	0.21	3.45	0.26	0.35
P-Tree16-20	1.52	0.13	0.16	1.71	0.15	0.19	3.16	0.24	0.31
Panel B: Out-of-Sample Results									
P-Tree1	1.91	0.59	0.68	4.23	0.76	0.94	8.91	1.36	1.72
P-Tree1-5	1.54	0.59	0.63	3.83	0.63	0.70	8.28	0.77	0.99
P-Tree6-10	0.48	0.60	0.62	1.38	0.61	0.64	2.73	0.60	0.67
P-Tree11-15	0.82	0.60	0.62	1.45	0.61	0.64	2.95	0.63	0.71
P-Tree16-20	0.69	0.60	0.62	1.28	0.59	0.63	2.80	0.58	0.66

test assets are smaller, ignoring MOM12M. Therefore, P-Tree would be inefficient if it were missing some key characteristics. In empirical applications, we are agnostic regarding the return-generating process of the real economy. Therefore, we cover many characteristics to relieve the concern about missing key ones.

P-Tree with true characteristics. Even without pre-specifying the true characteristics, P-Tree can produce informative test assets to achieve high Sharpe ratios and large pricing errors against the market factor. Thus, we conjecture that P-Tree can correctly select the true characteristics in a high-dimensional space. Table I.3 tabulates the probability of selecting each characteristic in P-Tree splits. For simplicity, we only report for the first P-Tree over 10 simulation repetitions. We find the true characteristics, ME, BM, and MOM12M, are selected with high probabilities in the top three, top five, and all nine splits.

When κ is large, the probability of selecting the true characteristics is higher. For low signal-to-noise setting $\kappa = 0.5$, the noisy characteristics are selected in top splits; for example, NI and BASPREAD are selected with non-zero probabilities in the top five splits. However, for high signal-to-noise setting $\kappa = 2$, only the true characteristics are selected in the top five splits, over 10 simulation repetitions. Therefore, we confirm

Table I.3: **Characteristics Selection Probability for the First P-Tree**

This table shows the selected characteristics of the first P-Tree with its selection probability over 10 repetitions. Each P-Tree contains nine splits and, thus, ten leaves. The “Top3” columns specify that we only count the first three splits of the P-Tree model, and similarly, we defined “Top5” and “Top9.” The rows only show the five most frequently selected characteristics. For example, in Panel A, the characteristic BM is the most frequently selected, with a 40% probability, in the top three splits. Panels A, B, and C report for $\kappa = 0.5, 1, 2$. The left panel reports for P-Trees with complete characteristics information as in Table I.1, whereas the right panel is for P-Trees omitting MOM12M as in Table I.2.

Complete Characteristics Set						Remove MOM12M				
	1	2	3	4	5	1	2	3	4	5
Panel A: $\kappa = 0.5$										
Top3	BM 0.40	MOM12M 0.33	ME 0.27			ME 0.63	BM 0.37			
Top5	BM 0.40	ME 0.34	MOM12M 0.22	NI 0.02	BASPREAD 0.02	ME 0.46	BM 0.38	BASPREAD 0.02	HIRE 0.02	SEAS1A 0.02
Top9	BM 0.33	ME 0.26	MOM12M 0.12	ALM 0.02	CFP 0.02	ME 0.29	BM 0.26	ILL 0.03	RNA 0.02	CASHDEBT 0.02
Panel B: $\kappa = 1$										
Top3	BM 0.37	MOM12M 0.33	ME 0.30			ME 0.60	BM 0.40			
Top5	BM 0.44	ME 0.32	MOM12M 0.24			BM 0.48	ME 0.44	DEPR 0.02	TURN 0.02	SVAR 0.02
Top9	BM 0.36	ME 0.29	MOM12M 0.16	CHCSHO 0.03	RE 0.02	ME 0.40	BM 0.34	TURN 0.03	DEPR 0.02	ILL 0.02
Panel C: $\kappa = 2$										
Top3	BM 0.37	MOM12M 0.33	ME 0.30			ME 0.67	BM 0.33			
Top5	BM 0.48	ME 0.28	MOM12M 0.24			BM 0.60	ME 0.40			
Top9	BM 0.43	ME 0.27	MOM12M 0.19	MOM36M 0.02	CFP 0.01	ME 0.47	BM 0.42	CFP 0.02	SP 0.01	TURN 0.01

that the superb performance of P-Trees originates from correctly selecting the true characteristics in the return-generating process (9).

Further, we investigate the characteristics selection results when omitting an important characteristic MOM12M, which is the same as the treatment in Table I.2. We find BM and ME are selected with an even higher probability, once we omit MOM12M. More importantly, we find noisy characteristics are selected more frequently. However, the noisy characteristics cannot help P-Tree fully span the efficient frontier, as shown in Table I.2. This phenomenon alerts us that we must feed all true characteristics to P-Tree.

II Benchmark-Adjusted P-Trees

As described in section 2.3, P-Tree boosting is a flexible approach to include any benchmark factor(w), such as CAPM, Fama-French three factors, and PCA factors. Here, we demonstrate an example of fitting boosted P-Trees with the market factor as a benchmark factor model. Specifically, we define $\mathbf{F} = [\text{mkt}_t, f_{1,t}, \dots, f_{K,t}]$ in the boosting split criteria (6).

The initial P-Tree remains unchanged in the empirical findings, even when the market factor is included as a benchmark, suggesting the information usefulness of the cross-sectional characteristics beyond the market factor. Therefore, the P-Tree diagram remains identical to Figure 4. However, the subsequent P-Trees change significantly, and the empirical performances differ from those in section 3. The following tables and figures present the empirical results for the P-Tree test assets and factors with the market factor benchmark.

Figure I.1 depicts the efficient frontiers spanned by P-Tree factors and the market factor. We find similar patterns as our main results without a market benchmark in Figure 7. The P-Tree frontiers move to the top-left corner of the mean-variance diagram, as more boosted P-Trees are generated. However, the market factor only takes a small proportion of the tangency portfolio, and the frontier is flat for the first P-Tree and the market factors.

Table I.4 shows the asset pricing performance of P-Tree test assets. Consistent with our main results in Table 3, we find the test assets of the first P-Tree have larger GRS statistics, higher average alphas, and higher proportions of significant alphas than the test assets generated by the follow-up P-Trees. Additionally, the GRS and PY tests always reject the null hypothesis, except for P-Tree11-15 in subsample 2001-2020. In general, the test assets of P-Trees with a market factor benchmark are challenging to be priced by FF5.

Table I.5 investigates the performance of each P-Tree factor. Consistent with our main results in Table 2, the boosted P-Tree factors cannot be explained by the previous P-Tree factors, as evidenced by the low p -values. In contrast to the results in Table

2, where the first factor could not be tested, we test the first P-Tree factor against the market benchmark. As expected, the first P-Tree factor cannot be explained by the market factor.

Table I.6 reports the investment performances. Adding the market factor slightly improves the risk-adjusted investment performance. The investment Sharpe ratios are over 3 for the out-of-sample periods 2001-2020 and 1981-2020. Again, those listed observable and latent factor models cannot explain our P-Tree factors with large unexplained alphas. In conclusion, this section demonstrates the flexibility of boosting in P-Trees: one can add any given factor model as a control when building P-Tree test assets and factors.

Figure I.1: Market Factor Benchmark: Characterizing the Efficient Frontier

This figure presents complementary results to Figure 7 for considering the market factor benchmark while growing P-Trees.

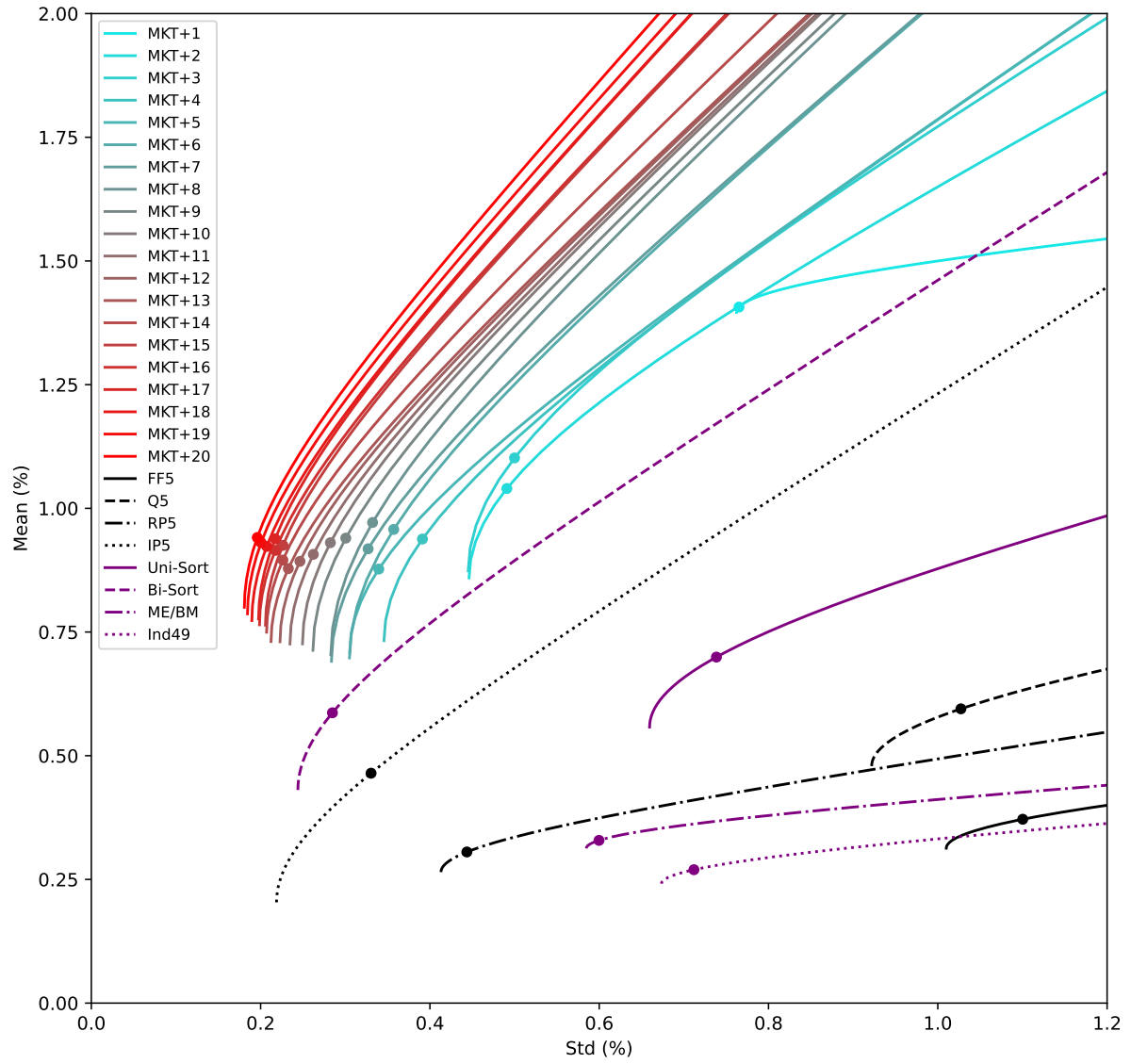


Table I.4: Market Factor Benchmark: Test Assets Generated by P-Tree

This table presents complementary results to Table 3 for considering the market factor benchmark while growing P-Trees.

	N	GRS	p -GRS	p -PY	$ \alpha $	$\sqrt{\alpha^2}$	$\overline{R^2}$	% $\alpha_{10\%}$	% $\alpha_{5\%}$	% $\alpha_{1\%}$
Panel A: 40 Years (1981-2020)										
P-Tree1	10	141.27	0.00	0.00	0.92	1.11	75	100	90	80
P-Tree1-5	50	57.09	0.00	0.00	0.39	0.59	78	64	52	46
P-Tree6-10	50	4.99	0.00	0.00	0.24	0.31	79	48	38	24
P-Tree11-15	50	6.82	0.00	0.00	0.39	0.52	71	50	40	30
P-Tree16-20	50	3.74	0.00	0.00	0.27	0.36	72	48	40	32
P-Tree1-20	200	39.88	0.00	0.00	0.32	0.46	75	53	43	33
Panel B: 20 Years (1981-2000)										
P-Tree1	10	84.36	0.00	0.00	1.58	1.95	70	90	80	80
P-Tree1-5	50	50.61	0.00	0.00	0.65	1.04	77	54	50	42
P-Tree6-10	50	9.05	0.00	0.00	0.46	0.66	75	52	42	26
P-Tree11-15	50	3.55	0.00	0.00	0.29	0.38	76	46	36	28
P-Tree16-20	50	10.01	0.00	0.00	0.46	0.59	78	74	64	48
P-Tree1-20	200	118.72	0.00	0.00	0.47	0.71	77	56	48	36
Panel C: 20 Years (2001-2020)										
P-Tree1	10	56.76	0.00	0.00	1.09	1.35	68	90	90	90
P-Tree1-5	50	31.26	0.00	0.00	0.35	0.66	78	32	30	24
P-Tree6-10	50	3.55	0.00	0.00	0.28	0.43	80	38	28	18
P-Tree11-15	50	1.12	0.28	0.82	0.15	0.21	77	16	6	2
P-Tree16-20	50	1.76	0.00	0.02	0.25	0.36	76	24	12	6
P-Tree1-20	200	51.47	0.00	0.00	0.26	0.45	78	28	19	13

Table I.5: Market Factor Benchmark: Testing the Boosted P-Tree Growth

This table presents complementary results to Table 2 for considering the market factor benchmark while growing P-Trees.

	Sharpe Ratio		CAPM Test		FF5 Test		Expanding Factors Test			BS Test
	Single	Cumu.	α (%)	t -stat	α (%)	t -stat	α (%)	t -stat	R^2	p -value
Panel A: 40 Years (1981-2020)										
1	6.37	6.37	1.39	39.63	1.37	37.48	1.39	39.63	0.01	0.00
2	2.24	7.33	0.40	13.61	0.37	12.65	0.66	10.99	0.13	0.00
3	0.65	7.61	0.12	2.36	-0.03	-0.80	-0.62	-5.65	0.66	0.00
4	1.57	8.29	0.27	9.25	0.26	8.42	0.58	8.51	0.25	0.00
5	1.63	8.92	0.46	9.62	0.42	8.81	0.97	8.06	0.17	0.00
6	0.72	9.24	0.10	3.02	0.05	1.69	-0.47	-5.59	0.70	0.00
7	1.40	9.66	0.36	8.09	0.29	6.58	0.74	6.37	0.29	0.00
8	0.83	10.04	0.17	4.02	0.09	2.20	-0.66	-5.93	0.55	0.00
9	1.39	10.72	0.29	8.03	0.25	6.94	0.77	7.78	0.36	0.00
10	1.37	11.27	0.36	7.88	0.27	6.24	0.93	6.86	0.31	0.00
11	1.22	11.82	0.37	6.87	0.40	7.34	1.14	6.59	0.31	0.00
12	1.24	12.36	0.36	6.99	0.28	5.41	1.09	6.49	0.35	0.00
13	1.27	12.87	0.40	7.22	0.25	4.85	-1.02	-6.07	0.46	0.00
14	1.64	13.47	0.53	9.68	0.48	8.61	1.12	6.58	0.44	0.00
15	0.69	13.88	0.10	2.73	0.05	1.56	-0.68	-5.40	0.73	0.00
16	1.16	14.31	0.41	6.47	0.29	5.18	-1.00	-5.16	0.58	0.00
17	0.80	14.67	0.15	3.71	0.11	2.62	-0.79	-4.84	0.55	0.00
18	1.69	15.09	0.34	10.07	0.28	8.51	0.71	5.24	0.33	0.00
19	1.37	15.59	0.33	7.92	0.29	6.68	-1.00	-6.07	0.40	0.00
20	1.08	16.09	0.23	5.85	0.15	3.72	-0.87	-5.58	0.52	0.00

Table Continued: Market Factor Benchmark: Testing the Boosted P-Tree Growth

Sharpe Ratio			CAPM Test		FF5 Test		Expanding Factors Test			BS Test
Single	Cumu.		α (%)	t -stat	α (%)	t -stat	α (%)	t -stat	R^2	p -value
Panel B: 20 Years (1981-2000)										
1	7.12	7.12	1.86	31.23	1.78	28.53	1.86	31.23	0.01	0.00
2	0.71	8.02	0.12	1.89	-0.01	-0.08	-0.93	-7.26	0.79	0.00
3	3.17	9.43	0.74	13.75	0.74	12.78	1.14	8.72	0.09	0.00
4	0.85	10.49	0.19	2.76	0.13	1.75	-1.20	-7.12	0.58	0.00
5	1.83	12.14	0.39	7.65	0.30	5.71	1.26	8.55	0.29	0.00
6	1.48	13.61	0.33	6.02	0.34	5.89	1.38	7.52	0.32	0.00
7	0.96	14.61	0.20	3.43	0.05	0.87	-1.14	-5.91	0.62	0.00
8	2.14	16.17	0.70	9.08	0.62	7.80	2.24	7.09	0.20	0.00
9	1.30	17.37	0.32	5.12	0.18	3.41	-1.45	-6.00	0.47	0.00
10	3.50	19.16	0.92	15.19	0.82	13.42	2.03	7.10	0.21	0.00
11	1.02	20.95	0.21	3.73	0.21	3.55	1.95	6.76	0.52	0.00
12	1.36	22.02	0.29	5.54	0.21	4.00	-1.40	-5.07	0.52	0.00
13	1.23	24.02	0.22	4.81	0.19	3.76	1.89	6.71	0.39	0.00
14	1.17	25.66	0.26	4.50	0.15	2.60	-2.02	-5.83	0.51	0.00
15	1.00	27.74	0.14	3.65	0.11	2.74	1.73	6.36	0.57	0.00
16	0.91	29.86	0.14	3.12	0.02	0.45	-1.73	-6.24	0.74	0.00
17	4.13	32.95	0.73	18.02	0.65	16.36	2.28	7.48	0.38	0.00
18	1.57	35.18	0.27	6.43	0.24	6.07	2.22	6.00	0.41	0.00
19	1.53	38.25	0.36	6.23	0.29	5.15	-3.28	-6.91	0.56	0.00
20	1.79	41.12	0.39	7.49	0.28	5.36	3.30	7.16	0.57	0.00
Panel C: 20 Years (2001-2020)										
1	5.82	5.82	1.51	25.60	1.47	24.01	1.51	25.60	0.01	0.00
2	1.39	7.13	0.34	5.74	0.34	5.86	0.97	9.42	0.33	0.00
3	0.98	8.23	0.22	3.75	0.20	3.23	0.97	8.00	0.43	0.00
4	0.63	9.03	0.11	1.73	0.03	0.52	-0.95	-6.55	0.76	0.00
5	0.78	9.61	0.12	2.72	0.06	1.57	-0.59	-5.47	0.63	0.00
6	1.51	10.58	0.44	6.30	0.35	5.04	1.18	6.88	0.39	0.00
7	1.07	11.46	0.25	4.22	0.21	3.66	1.08	6.12	0.37	0.00
8	1.02	12.24	0.20	4.00	0.13	2.80	-0.81	-5.62	0.57	0.00
9	2.24	13.22	0.54	9.68	0.50	8.76	1.15	6.24	0.30	0.00
10	0.74	14.24	0.16	2.44	0.10	1.53	-1.41	-6.23	0.68	0.00
11	1.09	15.86	0.25	4.31	0.20	3.28	1.68	7.68	0.51	0.00
12	0.93	16.97	0.23	3.47	0.13	2.03	-1.51	-5.83	0.59	0.00
13	1.46	18.29	0.34	6.09	0.23	4.61	1.49	6.45	0.47	0.00
14	0.91	19.37	0.17	3.40	0.10	2.05	-1.34	-5.66	0.63	0.00
15	0.58	20.52	0.08	1.38	0.05	0.83	-1.77	-5.62	0.83	0.00
16	1.30	21.91	0.27	5.34	0.22	4.36	-1.70	-6.54	0.54	0.00
17	1.05	22.94	0.47	4.09	0.45	3.87	3.93	5.40	0.44	0.00
18	0.65	24.43	0.08	1.92	0.08	1.89	1.66	6.32	0.79	0.00
19	1.33	25.80	0.29	5.48	0.20	3.87	2.10	5.92	0.51	0.00
20	0.84	26.95	0.16	3.01	0.18	3.17	2.19	5.33	0.62	0.00

Table I.6: Market Factor Benchmark: Factor Investing by Boosted P-Trees

This table presents complementary results to Table 5 for considering the market factor benchmark while growing P-Trees.

	SR	α_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}
<u>Panel A: 40 Years (1981-2020)</u>						
P-Tree1	6.37	1.39***	1.30***	1.35***	1.29***	1.12***
P-Tree1-5	8.92	0.75***	0.68***	0.72***	0.70***	0.63***
P-Tree1-10	11.27	0.52***	0.49***	0.51***	0.49***	0.47***
P-Tree1-15	13.88	0.45***	0.43***	0.44***	0.43***	0.42***
P-Tree1-20	16.09	0.39***	0.38***	0.39***	0.38***	0.37***
<u>Panel B1: 20 Years In-Sample (1981-2000)</u>						
P-Tree1	7.12	1.85***	1.88***	1.71***	1.64***	1.24***
P-Tree1-5	12.14	0.80***	0.79***	0.77***	0.75***	0.71***
P-Tree1-10	19.16	0.64***	0.65***	0.62***	0.61***	0.57***
P-Tree1-15	27.74	0.53***	0.53***	0.52***	0.52***	0.49***
P-Tree1-20	41.12	0.47***	0.47***	0.46***	0.46***	0.45***
<u>Panel B2: 20 Years Out-of-Sample (2001-2020)</u>						
P-Tree1	3.24	1.34***	1.31***	1.23***	1.30***	0.91***
P-Tree1-5	2.98	0.52***	0.47***	0.49***	0.48***	0.32***
P-Tree1-10	2.57	0.36***	0.29***	0.33***	0.34***	0.19***
P-Tree1-15	2.48	0.30***	0.23***	0.27***	0.27***	0.16***
P-Tree1-20	2.57	0.26***	0.20***	0.24***	0.25***	0.15***
<u>Panel C1: 20 Years In-Sample (2001-2020)</u>						
P-Tree1	5.82	1.50***	1.46***	1.50***	1.37***	1.36***
P-Tree1-5	9.61	0.66***	0.65***	0.66***	0.64***	0.62***
P-Tree1-10	14.24	0.46***	0.46***	0.46***	0.45***	0.45***
P-Tree1-15	20.52	0.38***	0.38***	0.38***	0.37***	0.37***
P-Tree1-20	26.95	0.34***	0.34***	0.34***	0.33***	0.33***
<u>Panel C2: 20 Years Out-of-Sample (1981-2000)</u>						
P-Tree1	4.35	1.50***	1.59***	1.35***	1.43***	1.10***
P-Tree1-5	4.11	0.57***	0.71***	0.51***	0.60***	0.30***
P-Tree1-10	4.07	0.40***	0.49***	0.35***	0.44***	0.23***
P-Tree1-15	3.98	0.36***	0.45***	0.31***	0.40***	0.18***
P-Tree1-20	3.82	0.31***	0.40***	0.26***	0.35***	0.14***

III P-Tree with Alternative Economically Guided Split Criteria

III.A An Alternative Global Split Criteria: Minimizing Realized Pricing Errors

Our baseline model generates test assets covering the efficient frontier and constructs the corresponding latent factor, recovering the pricing kernel, maximizing the Sharpe ratio of the MVE portfolio, and minimizing the HJ-distance (Hansen and Jagannathan, 1991). We now demonstrate the flexibility of the P-Tree framework using a criterion that directly minimizes the realized pricing errors of individual stocks. We run a regression model and provide a conditional expected value for the leaf parameter:

$$\mathcal{L}_A(\tilde{\mathbf{c}}_{k,m}) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left[\left(r_{i,t} - \beta_{i,t-1}(\mathbf{z}_{i,t-1}) f_t \right) \right]^2, \quad (10)$$

where $r_{i,t}$ is the excess return for stock i at period t , f_t is the $P \times 1$ traded factor return, and $\beta(z_{i,t-1}) = b_0 + b^\top z_{i,t-1}$ are conditional betas on the past firm characteristics $z_{i,t-1}$.

The asset pricing goal of minimizing objectives on pricing errors is similar in spirit to the recent literature about latent factor generation for constructing the SDF via deep learning (e.g., Gu et al., 2021; Chen et al., 2023; Feng et al., 2023). With this criterion, P-Tree solves a joint problem of constructing the latent factors and estimating betas. The realized pricing errors are the return residuals from the generated P-Tree factor model. P-Tree generates latent traded factors to reduce the aggregate quadratic realized pricing errors.

Figures I.2 and I.3 display P-Trees with and without a market benchmark, fitted on the alternative split criteria. The tree structure differs from our main results based on the maximum squared Sharpe ratio shown in Figure 4 for selecting different firm characteristics.

III.B Boosting

We propose the boosting procedure for the alternative criteria as follows:

1. The first factor $f_{1,t}$ can be pre-specified or generated by P-Tree as discussed in section 2.3. Then, we keep $f_{1,t}$ and its corresponding factor loadings $\beta_1(\cdot)$.

2. The second factor is generated to augment the first one. The tree-growing steps are the same as those in section 2.3, except we use the boosting split criteria:

$$\mathcal{L}(\tilde{\mathbf{c}}_{k,m}) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left(r_{i,t} - \beta_1(\mathbf{z}_{i,t-1})f_{1,t} - \beta_2(\mathbf{z}_{i,t-1})f_{2,t} \right)^2. \quad (11)$$

Note $\beta_1(\cdot)$ is fixed from step 1, and then, $r_{i,t} - \beta_1(\mathbf{z}_{i,t-1})f_{1,t}$ represents the unexplained part of returns or the realized pricing error from the previous nested factor model. The second factor is generated to augment the first one. Then, we keep $f_{2,t}$ and its corresponding factor loadings $\beta_2(\cdot)$.

3. The third factor $\mathbf{f}_{3,t}$ is generated similarly to augment the first two factors, whereas factor loadings for the first two factors are fixed at previous estimations:

$$\mathcal{L}(\tilde{\mathbf{c}}_{k,m}) = \sum_{t=1}^T \sum_{i=1}^{N_t} \left(r_{i,t} - \beta_1(\mathbf{z}_{i,t-1})f_{1,t} - \beta_2(\mathbf{z}_{i,t-1})f_{2,t} - \beta_3(\mathbf{z}_{i,t-1})f_{3,t} \right)^2. \quad (12)$$

4. Repeat the above process K times to generate K factors $\mathbf{f}_t = [f_{1,t}, \dots, f_{K,t}]$ and obtain corresponding factor loadings $\beta = [\beta_1, \dots, \beta_K]$.

Figure I.2: P-Tree with Market

This figure displays P-Tree with alternative split criteria trained from 1981 to 2000, with 20 leaves. The first factor represents the market factor, and then P-Tree is applied to generate a boosted tree factor under the asset pricing criterion. The figure format follows Figure 4.

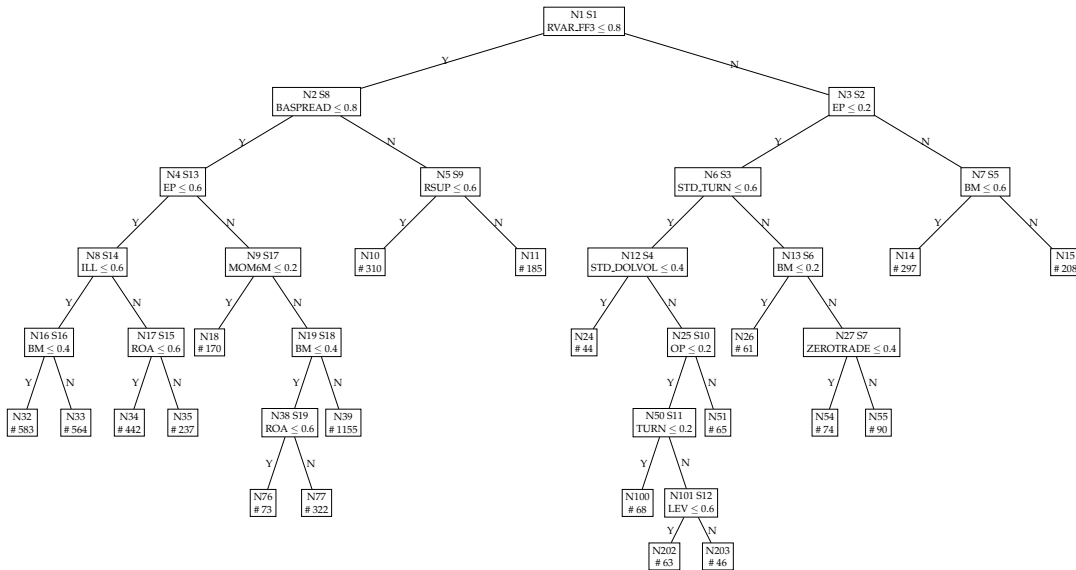
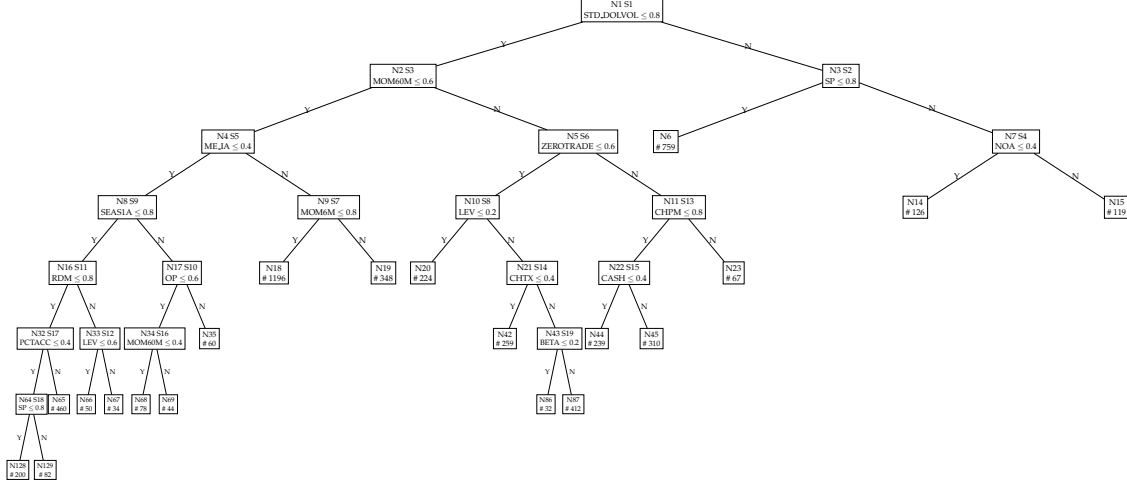


Figure I.3: **P-Tree without Market**

This figure displays a P-Tree on alternative split criteria without controlling for market factors. The training sample spans from 1981 to 2000. The figure format follows Figure I.2.



The boosted P-Tree factors enhance an existing factor model. Assuming we fit the P-Tree model with $K = 3$ factors, we can arrange the three factors in descending order of significance as $[f_{1,t}, f_{2,t}, f_{3,t}]$. In the empirical results, we provide two specifications: one with the market factor as a benchmark and one with a vanilla P-Tree. Generating additional P-Tree factors is similar to generating additional principal components, and each subsequent factor helps reduce the remaining pricing errors. For example, a three-factor boosted P-Tree model follows:

$$\mathcal{T}(\mathbf{z}_{i,t-1}|\Theta) = \beta_1(\mathbf{z}_{i,t-1})f_{1,t} + \beta_2(\mathbf{z}_{i,t-1})f_{2,t} + \beta_3(\mathbf{z}_{i,t-1})f_{3,t}. \quad (13)$$

III.C Empirical Results

Performance metrics. We consider multiple performance metrics. Total R^2 and cross-sectional R^2 metrics evaluate asset pricing performance for time-series and cross-sectional variations, respectively. Specifically, we evaluate the asset pricing performance using these metrics:

$$\text{Total } R^2 = 1 - \frac{\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (r_{i,t} - \hat{r}_{i,t})^2}{\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N r_{i,t}^2}, \quad (14)$$

where $\hat{r}_{i,t} = \hat{\beta}(\mathbf{z}_{i,t-1})f_t$. Total R^2 represents the fraction of realized return variation explained by the model-implied contemporaneous returns.

$$\text{Stock CS } R^2 = 1 - \frac{\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T (r_{i,t} - \hat{r}_{i,t}) \right)^2}{\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T r_{i,t} \right)^2}, \quad (15)$$

where $\hat{r}_{i,t} = \hat{\beta}(\mathbf{z}_{i,t-1})f_t$. Cross-sectional R^2 for individual stocks represents the fraction of the squared unconditional mean returns described by the common factor model:

$$\text{Portfolio CS } R^2 = 1 - \frac{\sum_{i=1}^N \left(\bar{R}_i - \widehat{\bar{R}}_i \right)^2}{\sum_{i=1}^N \left(\bar{R}_i \right)^2}, \quad (16)$$

where $\widehat{\bar{R}}_i = \hat{\beta}_i^\top \tilde{\lambda}_f$. The risk premium estimation adopts the cross-sectional regression estimates of factors, and factor loadings are estimated from time series regressions.

Comparing asset pricing models. Table I.7 summarizes the performance of the P-Tree model under the asset pricing criterion. Panel A shows the performance of two P-Tree models on the market benchmark, with additional one and four P-Tree factors. The Sharpe ratio test from Barillas and Shanken (2017) selects the number of additional P-Tree factors generated through boosting. Panel B of Table I.7 compares these results with observable factor models such as CAPM, FF5, and the recent Q5 model of Hou et al. (2021). We also compare the performance of P-Tree with latent factor models from the ML finance literature, such as RP-PCA of Lettau and Pelger (2020b) and IPCA of Kelly et al. (2019). Panel C reports the results of the P-Tree model with time-series splits on market volatility, which is discussed in section III.D.

The total R^2 values for P-Trees are consistently positive in pricing individual stock returns, outperforming most observable factor models both in-sample and out-of-sample. The multifactor P-Tree models show comparable total R^2 to two strong latent factor models. Cross-sectional R^2 for individual stocks quantifies the aggregate pricing errors by the factor model. For the cross-sectional R^2 , the multifactor P-Tree models produce highly positive numbers for in-sample and out-of-sample analysis, which demonstrates the performance of the model-implied expected returns. These numbers are close to the strong IPCA benchmark.

Table I.7: Asset Pricing Performance

This table reports the asset pricing performances. “Tot” (total R^2 %) and “CS” (cross-sectional R^2 %) in equations (14) and (15) are measures for individual stock returns. The in-sample period is from 1981 to 2000, and the out-of-sample period is from 2001 to 2020. We also report cross-sectional R^2 % in (16), using the factor models in the rows to price the test asset portfolios in the columns. “Leaf20” indicates the 20 basis portfolios from Figure I.2, and “Leaf40” indicates the 40 basis portfolios from Figures I.2 and I.3. Panel A shows results for the asset pricing panel-tree models with #factors. Specifically, “P-Tree2” indicates a two-factor model of the market factor and a P-Tree factor. “P-Tree5*” indicates a five-factor model of the market factor and four P-Tree factors. “*” indicates the optimal number of factors selected by the squared Sharpe ratio test of Barillas and Shanken (2017). Panel B compares benchmark models introduced in section 4.3. Panel C reports for the asset pricing P-Tree applying the time-series split, which chooses market volatility (SMVAR) to split on the time-series dimension.

	Individual Stocks				Portfolios			
	In-Sample		Out-of-Sample		Entire Sample			
	Tot	CS	Tot	CS	FF25	Ind49	Leaf20	Leaf40
<u>Panel A: P-Tree</u>								
P-Tree2	11.1	25.5	11.1	10.4	93.1	93.1	98.5	68.6
P-Tree5*	13.0	22.7	13.7	16.5	94.2	99.2	94.0	85.0
<u>Panel B: Other Benchmark Models</u>								
CAPM	7.0	1.3	8.4	0.6	91.7	88.1	97.3	10.4
FF3	10.5	7.5	10.7	5.1	95.1	94.1	98.4	38.7
FF5	11.0	13.1	11.3	5.1	96.8	96.1	98.7	75.2
Q5	10.9	18.1	11.5	6.4	97.4	95.9	98.5	79.2
RP-PCA5	12.1	18.3	13.6	15.0	97.0	95.7	98.6	64.0
IPCA5	13.8	27.8	14.9	17.7	97.3	92.8	99.4	89.2
<u>Panel C: P-Tree with Time-Series Split on Market Volatility</u>								
TS-P-Tree2	11.7	24.7	11.3	9.3	87.6	81.8	92.1	26.6
TS-P-Tree5*	13.4	26.4	13.6	18.4	96.1	91.2	97.8	80.7

We also investigate P-Tree performance in pricing multiple sets of test portfolios, including 5×5 ME-BM and 49 industry portfolios. In addition, 20 leaf basis portfolios are generated in Figure I.2, and another 20 are generated without the market factor in Figure I.3. Generating basis portfolios is another data-driven output from our P-Tree framework, which resolves the high-dimensional characteristics challenge and follows a chosen economics-guided split criterion. Cross-sectional R^2 in (16) quantifies the aggregate unexplained pricing errors by the model-implied expected returns and is calculated using the entire sample of data.

The results suggest observable factor models, including CAPM, FF5, and Q5, explain the cross-sectional average excess returns for 5×5 ME-BM and 49 industry port-

folios. However, they cannot explain the leaf basis portfolios generated by P-Trees. In contrast, the P-Tree factor models in Panels A and C consistently price all sets of test assets. IPCA competes with robust positive performance among other latent factor models, but P-Trees offer more transparency and interpretability than other latent factor models.

Table I.8: Investing P-Tree Factors

This table reports the investment performance of the equity factor models. The monthly average return (%) and CAPM α (%) are reported, along with the annualized Sharpe ratio for the factors' mean-variance efficient (MVE) and equal-weighted ($1/N$) portfolios. "***" indicates the optimal number of factors selected by the squared Sharpe ratio test of [Barillas and Shanken \(2017\)](#). For t -statistics, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	In-Sample (1981-2000)						Out-of-Sample (2001-2020)					
	MVE			$1/N$			MVE			$1/N$		
	AVG	SR	α	AVG	SR	α	AVG	SR	α	AVG	SR	α
Panel A: P-Tree												
P-Tree2	1.75	1.58	1.51***	1.31	1.34	0.86***	0.30	0.31	0.29	0.45	0.56	0.17
P-Tree5*	1.26	3.47	1.20***	1.06	1.69	0.72***	0.80	1.93	0.76***	0.70	1.14	0.42***
Panel B: Other Benchmark Models												
FF5	0.45	1.48	0.38***	0.38	1.34	0.33***	0.27	0.64	0.13*	0.25	0.59	0.12
Q5	0.77	2.78	0.74***	0.63	2.10	0.53***	0.34	1.22	0.34***	0.31	1.10	0.25***
RP-PCA5	0.82	3.48	0.76***	1.07	1.77	0.75***	0.34	1.49	0.32***	0.50	1.00	0.27***
IPCA5	1.50	10.37	1.48***	0.90	3.15	0.80***	0.97	4.60	0.98***	0.73	2.14	0.61***
Panel C: Time-Series Split P-Tree - Market Volatility												
TS-P-Tree2	1.80	1.67	1.56***	1.32	1.36	0.86***	0.60	0.66	0.57***	0.60	0.77	0.31***
TS-P-Tree5*	1.20	3.65	1.14***	1.03	1.61	0.69***	0.78	1.69	0.73***	0.74	1.14	0.46***

Investment performance. To compare the risk-adjusted investment performances of observable or latent factors, we consider two common strategies for investing in multiple factors: the $1/N$ equal-weighted portfolio and the MVE portfolio. Table I.8 reports the monthly average return, CAPM alpha, and the annualized Sharpe ratio of the training and test samples.

P-Tree factor models show excellent asset pricing and investment performance. For the in-sample analysis, the MVE strategies of the five-factor model (market plus four boosted P-Tree factors) deliver a 3.47 annualized Sharpe ratio in Panel A. The monthly CAPM alphas are beyond 1%. These numbers are consistently high out-of-sample. If we focus on the long-only $1/N$ strategies (a strong benchmark), the numbers become lower but remain consistently positive. Overall, P-Tree factors and the IPCA

Table I.9: **Factor-Spanning Alpha Test**

This table reports the monthly alphas (bp) and significance for the factor-spanning test. Panel A shows the results for each factor of the P-Tree models in Table I.8. We regress each factor in the rows against a factor model in the columns. We name the P-Tree factors by the first two splitting characteristics selected in the interactive tree structure. We also show the mean-variance efficient (MVE) and 1/ N strategies investing these five factors. Notably, “FF25” and “IND49” are the 5 × 5 ME-BM and 49 industry portfolios. For t -statistics, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	In-Sample				Out-of-Sample			
	FF5	Q5	RP5	IP5	FF5	Q5	RP5	IP5
<u>Panel A: P-Tree factors</u>								
RVAR_FF3-EP	130***	101***	83***	107**	12	4	13	31
BM_IA-III	35**	33	19	-100***	107***	110***	81***	34
MOM12M-STD_DOLVOL	82***	53***	53	-24	25	22	22	-95***
ME-RDM	52***	48***	47***	109***	29**	27**	30**	13
MVE (4 factors + mkt)	58***	45***	58***	-21	36***	34***	36***	-11
1/N (4 factors + mkt)	60***	47***	60***	49*	35***	33***	35***	1
<u>Panel B: Other Test Assets</u>								
MVE-FF25	55***	42***	32***	27*	19***	15**	8	10
MVE-IND49	13*	20	-36**	14	10	8	20*	28*
1/N-FF25	-8***	-8	-68***	57***	3	8**	-23***	5
1/N-IND49	63*	30	56	62	-2	10	-13**	4

factors perform the best. P-Trees further offer transparency and interoperability.

We decompose each P-Tree factor on benchmark factor models to evaluate the risk-adjusted investment performances, alphas, in Table I.9. For example, the first investment P-Tree factor `RVAR_FF3-EP` in Panel A shows economically and statistically significant alphas against FF5 and Q5 models. A 5-factor IPCA model cannot explain this top P-Tree factor for either in-sample or out-of-sample evaluation. The MVE portfolio of four investment P-Tree factors plus the market factor still shows significant alphas everywhere. Moreover, the performances for P-Tree factors in Panel A are slightly lower but still consistently positive.

P-Tree factors are generated based on leaf portfolios, which are generated by P-Trees using multiple characteristics. In Panel B, we compare the investment performance of P-Tree factors with conventional clustering (5 × 5 ME-BM and 49 industry portfolios), which shows lower performance. We conclude that the positive performance of P-Tree factors comes from both the interaction and nonlinearity of characteristics and the data-driven approach used to identify them, as opposed to ad hoc

sorting.

III.D Time-Series Split

The panel data of asset returns has cross-sectional and time-series dimensions. Most literature focuses on explaining the cross-sectional return variation, as do the baseline P-Tree models, without considering potential structural breaks or regime shifts. Because the tree grows from top to bottom, the panel of assets is partitioned into many leaf nodes based on past characteristics. Each leaf node maintains the entire time series from period 1 to T but a subset of assets. However, empirical findings suggest that factor choices, risk premia, and betas can be time-varying under different macroeconomic states. For example, extensive literature examines regime changes in financial markets involving periods of high and low volatility or inflation, booms, or recessions (e.g., [Hamilton and Lin, 1996](#); [Buraschi and Jiltsov, 2005](#)).

With its ability to split nodes in the time-series dimension, P-Tree provides an alternative approach to constructing and estimating factor-beta with regime-switching. Moreover, the time-series information extracted from macro variables proves helpful in constructing the dynamic SDF. By splitting nodes (e.g., root node) by macro variables instead of firm characteristics, we can create time-varying factor models and estimate corresponding conditional betas under switching macroeconomic states. Each child node maintains the same cross section of assets as the parent node, but their time-series samples do not overlap. This approach improves P-Tree's flexibility in capturing cross-sectional and time-series variations, providing clustering over the two-dimensional panel.

The asset pricing split criterion for the first time-series split is

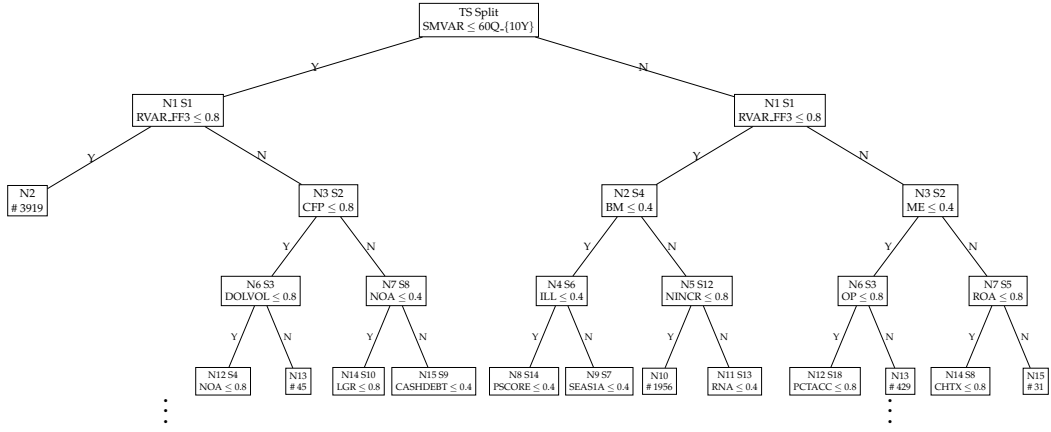
$$\mathcal{L}(\tilde{c}) = \sum_{t \in \mathcal{R}_A} \sum_{i \in N_{A,t}} (r_{i,t} - \beta_A(\mathbf{z}_{i,t-1})f_{A,t})^2 + \sum_{t \in \mathcal{R}_B} \sum_{i \in N_{B,t}} (r_{i,t} - \beta_B(\mathbf{z}_{i,t-1})f_{B,t})^2, \quad (17)$$

where the candidate split $\tilde{c}$ partitions the time series of data into two non-overlapped sets of periods $\mathcal{R}_A$ and $\mathcal{R}_B$, for example, high- or low-inflation states. Note both sets of periods are not necessarily consecutive. We use $N_{A,t}$ to represent the subsample of assets in the child node $\mathcal{R}_A$ during time period t , and $N_{B,t}$ for the corresponding

subsample in the child node $\mathcal{R}_B$. The factors $f_{A,t}$ and $f_{B,t}$ represent the P-Tree factor returns for the left and right child nodes, respectively. Note no overlapping periods exist between $\mathcal{R}_A$ and $\mathcal{R}_B$. Moreover, $\mathcal{R}_A$ and $\mathcal{R}_B$ may contain multiple periods for the changing macroeconomic states. In contrast to the cross-sectional split criteria in (4), the time-series split criterion is based on the *total* pricing loss of two time periods, with corresponding factors $f_{A,t}$ and $f_{B,t}$.

Figure I.4: **P-Tree Split on High/Low Market Volatility**

This figure depicts P-Tree, considering both cross-sectional and time-series variations. The most important macro predictor is stock market volatility $SMVAR$, and the first split is implemented when the current $SMVAR$ level is lower than the 60% quantile of the past decade. Two tree models are provided as two child leaves for high- and low- $SMVAR$ periods. The figure format follows Figure I.2.



Below is the two-period model with conditional betas:

$$\mathcal{T}(\mathbf{z}_{i,t-1}|\Theta) = \beta_A(\mathbf{z}_{i,t-1})f_{A,t}\mathbf{I}(t \in \mathcal{R}_A) + \beta_B(\mathbf{z}_{i,t-1})f_{B,t}\mathbf{I}(t \in \mathcal{R}_B). \quad (18)$$

Because the stock universe is much larger than the number of periods, after finding the best time-series split rule at the beginning, all subsequent split rules are selected based on cross-sectional characteristics only for demonstration purposes. Note any further split on either child of the root node depends only on the subsample on one side. This extension not only improves the modeling flexibility of P-Tree to capture cross-sectional and time-series variations, but also informs how macro variables interact with firm characteristics to explain asset returns, because they jointly define the space partition.

Figure I.4 shows a time-series split P-Tree on market volatility ($SMVAR$), which

identifies the regime-switching indicator as the levels of market volatility and might generate multiple regimes. Then, two separate cross-sectional P-Trees are developed on each regime. The P-Tree model varies under different regimes, though the idiosyncratic volatility remains one key indicator. Also, we show the asset pricing performance in Table I.7 and investment results in Table I.8, demonstrating robust performances over the market volatility regimes. All of these discoveries provide supportive proof for the time-varying factor structure of returns across different time periods.