Asset Pricing with Panel Tree under Global Split Criteria*

Lin William Cong Guanhao Feng Jingyu He Xin He

September 27, 2022

Abstract

We introduce a class of interpretable tree-based models (P-Tree) for analyzing (unbalanced) panel data, with iterative and global (instead of recursive and local) split criteria. We apply P-Tree to split the cross section of returns under the no-arbitrage condition, generating a stochastic discount factor model and diversified test portfolios for asset pricing. P-Tree visualizes nonlinear feature interactions, accommodates time-series splits, and allows interactions between macroeconomic states and firm characteristics. In an empirical study of U.S. equities, data-driven P-Tree reveals that long-term reversal, volume volatility, and industry-adjusted market equity interact to drive cross-sectional return variation, and that inflation constitutes the most critical regime-switching when interacting with firm characteristics. P-Tree models consistently outperform observable and latent factor models at pricing individual asset and portfolio returns, while delivering profitable and transparent trading strategies utilizing characteristic interactions. The methodology is broadly applicable in building trees with vectorized outcomes and economic restrictions as split criteria to guard against overfitting and improve model performance.

*We are especially grateful to Fahiz Baba-Yara (discussant), Svetlana Bryzgalova (discussant), Bing Han, Bob Jarrow, 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, Andrei Goncalves, P. Richard Hahn, Dashan Huang, Sophia Zhengzi Li, Asaf Manela, Stefan Nagel, Nick Polson, Adam Reed, Dacheng Xiu, Tsaig Yaki, Mao Ye, and Bobby Yu, and seminar and conference participants at Arison School of Business at Reichman University (IDC Herzliya), SC Johnson College of Business at Cornell University, University of Bath (AFL-QF & FinTech Research Seminar Series), University of Hawai'i at Mānoa Shidler College of Business, Mays Business School at Texas A&M University, Northern Finance Association Annual Meeting (Banff), University of Northern Carolina Kenan-Flagler Business School, Shanghai Jiao Tong University, Shanghai University of Finance and Economics, Southern University of Science and Technology, University of Science and Technology of China, Beihang University, Hunan University, Peking University School of Mathematical Sciences, 2022 UF Research Conference on Machine Learning in Finance, 2022 Western Finance Association Annual Conference, 2022 GSU-RFS FinTech Conference, 2022 China FinTech Research Conference, 2022 Asian Meeting of the Econometric Society, 2022 Asian Finance Association Annual Conference, 2021 INFORMS Annual Meeting, 6th PKU-NUS Annual International Conference on Quantitative Finance and Economics, Workshop on Machine Learning for Finance at Ca' Foscari University of Venice, Schroders Systematic Investments External Academic Seminar, 3rd Shanghai Financial Forefront Symposium, the HKAIFT-Columbia joint seminar, and World Online Seminars on Machine Learning in Finance for constructive discussions and feedback. This research was funded in part by INQUIRE Europe. Jingyu He acknowledges the GRF grant from Hong Kong Research Grants Council. Cong (E-mail: will.cong@cornell.edu) is at Cornell University SC Johnson College of Business; Feng (E-mail: gavin.feng@cityu.edu.hk) and He (E-mail: jingyuhe@cityu.edu.hk) are at the City University of Hong Kong; He (E-mail: xinh@hnu.edu.cn) is at Hunan University.

Key Words: CART, Cross-Sectional Returns, Interpretable AI, Latent Factor, Machine Learning.

Contents

1 Introduction

Decision tree algorithms, such as the well-known Classification and Regression Trees (CART, Breiman et al., 1984), are a class of highly adaptive nonlinear machine learning (ML) methods that partition the predictor space to visualize nonlinear and interaction effects. Essentially, CART fits data using high-dimensional step functions and performs well even in low signal-to-noise and small-sample environments. Meanwhile, despite the popularity of conventional factor models (e.g., Fama and French, 1993, 2015; Hou et al., 2015, 2021), empirical asset pricers recognize the need to apply nonlinear ML models to explain the cross-sectional variations in asset returns (e.g., Rossi and Timmermann, 2015; Freyberger et al., 2020; Gu et al., 2020; Nagel, 2021; Liu et al., 2020). Nevertheless, methods such as deep learning (e.g., Gu et al., 2021; Feng et al., 2021) often appear as black boxes, despite their superior pricing or prediction performances. Although decision trees potentially admit graphical representation, they are not tailored for financial panel data, and initial applications have focused on ensemble methods (Rossi, 2018) that are hard to interpret.[1]

Our new decision-tree approach balances model performance and interpretability when splitting the cross section of individual assets and addresses the high dimensionality, nonlinearity, and interactions in financial data. Specifically, we develop a unified P-Tree framework ("P" for "panel") for analyzing (unbalanced) panel data. In the application to asset pricing, P-Tree embeds a latent factor model with time-varying betas, to generate test assets and profitable trading strategies guided by a global asset pricing criterion, i.e., no-arbitrage, and recover the stochastic discount factor (SDF) that accurately prices asset returns. P-Tree offers an alternative top-down solution to security sorting for splitting the cross section, and partitions the data panel in hybrid (cross-sectional and time-series) dimensions. Furthermore, P-Tree inherits the interpretability from decision tree algorithms and offers a clear graphical display of the nonlinear variable interactions. The methodology is broadly applicable in building trees for panel data with vectorized or multi-period leaf parameters and global split criteria based on economic restrictions to guard against overfitting, improve model performance, and incorporate economic insights in ML. Because P-Tree is completely novel and has

[1]Applying CART or its ensemble methods (boosted trees or Random Forest) directly to financial panel data is problematic since these standard machine learning methods make independently and identically distributed (i.i.d.) assumptions and do not consider cross-sectional correlations, which renders the model specification essentially a single-period model and unsuitable for cross-sectional asset pricing.

off-the-shelf statistical packages, we develop a new package for explorations by other researchers in economics, finance, and other related disciplines.[2]

Our asset pricing application effectively illustrates two methodological innovations of P-Tree. First, the P-Tree nodes allow multi-period returns and make extracting information from panel data more effective. CART assumes data are independently and identically distributed (i.i.d.), ignoring the cross-sectional correlations. It searches for the optimal split rules by minimizing the sum of squared errors and then fitting a constant target value in each leaf node. In contrast, P-Tree admits the panel structure of the data, and incorporates a vectorized leaf parameter for multi-period returns. Furthermore, it implements a factor model and provides different vectorized target values to different assets in each leaf. An asset pricing P-Tree model resembles the sequential sorting scheme and solves the high-dimensional sorting challenge raised in Cochrane (2011) when using multiple characteristics.[3] The generated P-Tree leaf nodes are portfolios (in time-series vector forms) of (weighted) averages of individual asset returns within each period, constituting basis portfolios. This paper introduces the first multi-period tree in the statistics and ML literature, which has an ideal application in empirical asset pricing to approximate conditional security sorting.

Second, a P-Tree grows based on a global splitting criterion derived from the asset pricing theory. CART and other off-the-shelf ML methods minimize the sum of squared errors, which lacks the flexibility to incorporate domain knowledge in empirical asset pricing or other fields. More importantly, the CART algorithm grows recursively and optimizes split rules at each node without considering other sibling nodes. This greedy strategy focuses on local optimization and usually leads to overfitting since it operates on fewer observations as the tree grows. In contrast, P-Tree admits a linear factor model that prices all assets in the cross section. The tree splits iteratively to optimize the performance of the factor model globally. In other words, it utilizes data from all leaf nodes, which fits a common factor model to prevent overfitting while maintaining interpretability. The process is iterative, conditional on fixing prior splits, and therefore still

Empirically, we apply P-Tree to individual stock returns in the U.S. equity market from 1981

[2]We post the `R` package `TreeFactor` for P-Tree on the website `https://github.com/CityUHK-FinTech/TreeFactor`.

[3]The P-Tree framework is flexible enough to also accommodate simultaneous sorting scheme (i.e., Fama-French ME - B/M 5×5 equity portfolios); one just needs to split along ordered duplet or multiplet of characteristics, i.e., linear partitions not parallel to any single input variable.

to 2020, using the first 20 years as training and the recent 20 years as testing data. The P-Tree and market-adjusted P-Tree outperform most alternatives, including the known observable factor models (e.g., Fama and French, 2015; Hou et al., 2021) and latent factor models (e.g., Kelly et al., 2019; Lettau and Pelger, 2020), in terms of asset pricing, return prediction, and factor investment performance for both in-sample and out-of-sample analysis. In particular, P-Tree models price the cross section of both individual assets and of common, sorted portfolios well, while maintaining interpretability. In the test sample, investing in P-Tree factors also delivers a monthly alpha of 213 basis points against the Fama-French five factors and an annualized Sharpe ratio of 1.71. Further, leaf basis portfolios generated by P-Tree and market-adjusted P-Tree portfolios consistently show significant performance as test assets.

Our study of the out-of-bag variable importance reveals that a small set of characteristics, such as volume volatility (`STD_DOLVOL`) and industry-adjusted market equity (`ME_IA`), are significant in driving the cross-sectional return variation through their sequential interactions. P-Tree is also able to split the panel of return data on time-series horizons and identifies that inflation (`INFL`) is the most critical macroeconomic predictor for regime-switching, which is consistent with the findings in Boons et al. (2020). The long-term reversal (`MOM60M`) and one-year seasonality (`SEAS1A`) are more important during high inflation periods, whereas trading volume (`DOLVOL`) and volume volatility (`STD_DOLVOL`) are more important during the low inflation periods. Moreover, P-Tree displays all the split rules in sequence, which helps researchers understand the interactions among asset characteristics and between macroeconomic variables and asset characteristics to enhance strategy returns and resurrect anomalies. In particular, we allow the long and short legs of an anomaly portfolio to interact with different features which essentially loads the portfolio on different basis leaf portfolios. this is in contrast to the traditional treatment of a long-short portfolio as a single asset, complementing the pioneering work of Jarrow et al. (2021) to model the two legs of anomaly portfolios separately.

In summary, P-Tree has several advantages over the standard tree-based methods in ML. First, it incorporates bespoke global split criteria, such as the no-arbitrage condition, preventing overfitting and iteratively growing the tree guided by theoretical insights. Second, P-Tree prevents overfitting without relying on an ensemble of trees, and therefore a single P-Tree can be useful and maintains interpretability. It graphically displays the nonlinear predictor interactions by the data-driven sequential split rules. Third, it discards the i.i.d. assumption and extracts cross-sectional and time-series information from unbalanced panel data. Overall, P-Tree inherits the advantages of tree-based methods in accommodating high-dimensionality, nonlinearity, and interaction in the data, and at the same time, offers extra adaptability to the low signal-noise environment unbalanced panel data. Furthermore, P-Tree generates leaf basis portfolios that constitute natural basis portfolios for asset pricing applications. The leaf basis portfolios can recover the SDF and imply lower transaction costs of factor construction than principal component or deep learning models.

Literature. Machine Learning (hereafter ML), including tree-based methods, has been successfully applied in many fields.[4] Our paper contributes to the fast-emerging sub-field of ML in finance.[5] Prior attempts in applying regression trees (including Random Forest and boosting trees) to empirical asset pricing (e.g., Gerakos and Gramacy, 2012; Moritz and Zimmermann, 2016; Rossi and Timmermann, 2015; Rossi, 2018; Gu et al., 2020) typically do not use global split criteria or tailor the tree models for panel data analysis with distinctions between information in cross-sectional and time series dimensions. P-Tree incorporates dynamic patterns in the macroeconomic time series and their interactions with asset characteristics, as well as no-arbitrage conditions for asset pricing.

We add to the emerging studies of imposing economic restrictions on estimating and evaluating ML or statistical factor models (e.g., Feng et al., 2021; Chen et al., 2020; Chaieb et al., 2021; Avramov et al., 2021). Although the no-arbitrage condition have been applied to linear parametric models, the application to large and flexible ML models is rare. One recent attempt along this line is Creal and Kim (2021), which identifies variables that best explain assets' betas by splitting currency-return observations using a Bayesian factor model with additive regression trees (BART) priors. P-Tree differs from it in (i) considering vectorized outcomes (multi-period returns) in the

[4]CART (Breiman et al., 1984) and its variants are among the popular algorithms in various tree-based applications such as clinical studies (e.g., Arnett et al., 1988), ecology (e.g., Guisan and Zimmermann, 2000), and more recently, asset pricing (e.g,. Rossi and Timmermann, 2015; De Prado, 2018; Rossi, 2018; Li and Rossi, 2020).

[5]ML and AI models are shown to be effective in predicting equity returns (Freyberger et al., 2020; Gu et al., 2020; Cong et al., 2021, e.g.). Bianchi et al. (2021) and Feng et al. (2020) find predictability of treasury bond returns, whereas Bali et al. (2021) and He et al. (2021) find that of corporate bond returns. For direct portfolio construction, Cong et al. (2020) provide a reinforcement learning approach, whereas Feng and He (2022) and DeMiguel et al. (2020) provide regularization methods by optimization or Bayesian modeling. Han et al. (2019), Feng et al. (2020), Chinco et al. (2021), Dong et al. (2022), and Bryzgalova et al. (2021) apply variable selection methods to a large number of factors and characteristics.

nodes to suit panel data, (ii) growing one single tree (rather than drawing posterior samples of trees) that explicitly depicts predictor interactions and facilitates portfolio constructions, (iii) estimating a frequentist latent factor model rather than a Bayesian model over observable factors, and (iv) focusing on empirical analysis of equity returns instead of currencies.

This paper also contributes to the growing literature in empirical asset pricing that develops latent factor models for pricing cross-sectional returns. It effectively connects the literature on statistical factor models with the subsequent literature on economic factor models.[6]. Lettau and Pelger (2020), Kelly et al. (2019), Kelly et al. (2022), and Kim et al. (2021) show different principal component methods to produce latent factors, whereas Chen et al. (2020), Feng et al. (2021), and Gu et al. (2021) generate the SDF through various deep learning models. P-Tree instead constructs SDF and basis portfolio simultaneously by partitioning the cross section of the asset universe. Moreover, P-Tree's ability to price both individual assets and test portfolios add to recent effort for extracting common factors for explaining cross-sectional variations in multiple panels of data (Andreou et al., 2019, 2022).

Bryzgalova et al. (2020) also estimate the regularized SDF on a given set of characteristics-managed portfolios by a pruning algorithm with global criteria. P-Tree differs in growing a tree in a top-down fashion rather than bottom-up (pruning), and generating test portfolios endogenously. Enumerating all possible tree configurations to prune from is an NP-hard problem, meaning that it is not computationally feasible unless one manually specifies a small set of variables and shallow depth for initial trees as in Bryzgalova et al. (2020). Instead, our approach exploits a global iterative greedy algorithm to effectively search in a much larger space of tree structures than conventional sequential sorting to uncover nonlinear signals and predictor interactions.

Finally, this paper relates to asset pricing factor models in general and the construction of basis assets. We follow Avramov (2004), Gagliardini et al. (2016), and Feng and He (2022) to allow for time-varying, characteristics-driven betas for stock returns. Haddad et al. (2020) exploit predictability in equity factors for constructing the optimal factor timing portfolio to approximate the SDF. Zhu et al. (2021) employ the adaptive multi-factor (AMF) model to test the Generalized

[6]Latent factor models in empirical asset pricing starts with the arbitrage pricing theory of Ross (1976) and the empirical tests in Roll and Ross (1980) and Chen (1983). Early work apply principal component analysis and to extract latent factors with a purely statistical criterion, as done by Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1988).

Arbitrage Pricing Theory (Jarrow and Protter, 2016) and show that AMF with a rolling window is more consistent with realized asset returns than is the FF5 model. Our model estimation for the characteristics-driven betas is incorporated within the split criteria. Zhu et al. (2020) use high-dimensional methods such as prototype clustering to select basis assets with clear interpretation and prediction accuracy; Jarrow et al. (2021) adopt the methodology for understanding how the long and short legs of an anomaly portfolio behave as a function of changing group of basis assets. Ahn et al. (2009) use unsupervised clustering method to create basis assets based on correlation of asset returns, whereas our P-Tree provides a supervised solution of clustering basis assets. We also document clustering patterns and investment performances of leaf basis portfolios from both shallow and deep trees. The literature, including regularized linear models (Kozak et al., 2020; Bryzgalova et al., 2020) and the principal component rotation in (Kozak et al., 2018; Haddad et al., 2020; Lettau and Pelger, 2020), largely takes pre-specified or characteristics-managed portfolios as given. P-Tree generates these characteristics-managed portfolios, complementing the adversarial networks approach in Chen et al. (2020), among others.

2 P-Tree Factor Model for Asset Pricing

2.1 Tree-based Models

A decision tree consists of a sequence of split rules (or cutpoints) that define a rectangular partition of the predictor space. A split rule $\tilde{c} = (z_l, c)$ splits the data along the variable z_l using a threshold c. If there are J splits in total, the predictor space is partitioned in $J+1$ regions (leaf nodes), which we denote by $\mathcal{R}$. The regression tree, $\mathcal{T}$, with $J+1$ regions and parameters $\Theta_J = \{\mathcal{R}_j, \mu_j\}_{j=1}^{J+1}$ can then be written as

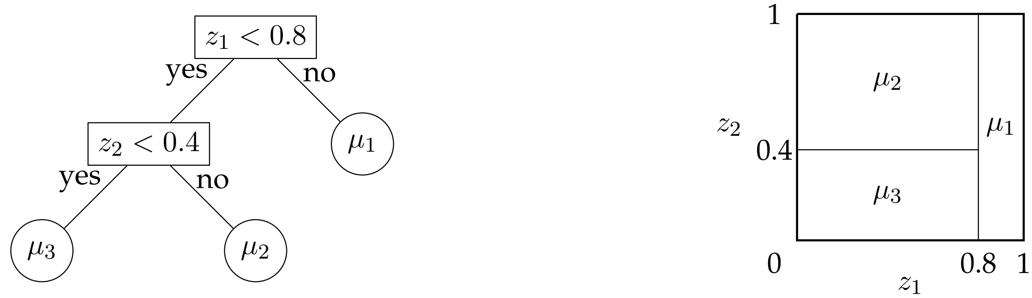
$$\mathcal{T}(z \mid \Theta_J) = \sum_{j=1}^{J+1} \mu_j I(z \in \mathcal{R}_j), \tag{1}$$

where $\{\mathcal{R}_j\}_{j=1}^{J+1}$ are disjoint leaves or regions defined by the sequence of split rules, z is a vector of all predictor variables, and μ_j is the target value for observations in region $\mathcal{R}_j$. Figure 1 illustrates a decision tree and the corresponding partition of the predictor space and their constant target values.

The essence for decision trees entails choosing split rules to grow the tree.

Figure 1: A Decision Tree Partitioning A Two Dimensional Predictor Space.

Left: A depth three decision tree with two splits. Right: Corresponding partition plot for the predictor space of the tree structure. $z = (z_1, z_2)$ is a vector of two variables.



Classification and Regression Trees (CART Breiman et al., 1984) are the most widely-adopted tree models and constitute the building block for ensemble methods such as Random Forest (Breiman, 2001), boosting trees (Chen and Guestrin, 2016) or Bayesian approaches such as Bayesian additive regression trees (Chipman et al., 2010), and the recent XBART (He et al., 2019; He and Hahn, 2021). Such tree models "grow" by recursively partitioning the space of predictors. Starting from the top node (root node), the best split is determined from all split rule candidates according to pre-specified split criteria, and then the dataset is divided into left and right child nodes. The growing process repeats until some pre-specified stopping conditions are met, such as a maximum tree depth or a minimal number of data observations in each node. When considering splitting a node, the recursive algorithm only processes data within that specific node without looking at other nodes—an ad hoc approach introduced in the early days, mainly for easy coding and fast computation.

Next we review the split criteria of CART when applied to asset pricing. Let $r_{i,t}$ denote the return of asset i at time period t. Treating the cross-sectional data as pooled data, the split criterion is sum of squared loss

$$\mathcal{L}(\tilde{c}) = \sum_{i,t \in \text{left node}} (r_{i,t} - \bar{r}_{\text{left}})^2 + \sum_{i,t \in \text{right node}} (r_{i,t} - \bar{r}_{\text{right}})^2, \tag{2}$$

where $\bar{r}_{\text{left}} = \frac{1}{\#_{\text{left node}}} \sum_{i,t \in \text{left node}} r_{i,t}$ and $\bar{r}_{\text{right}} = \frac{1}{\#_{\text{right node}}} \sum_{i,t \in \text{right node}} r_{i,t}$ are average returns in the left or right leaf nodes, respectively. Essentially, this split criterion measures the sum of squared error of fitting return data in a leaf using a single constant. The sum in Equation (2) does not distinguish cross section nor time series, which implies the assumption that data are i.i.d.

Once the algorithm terminates, the bottom nodes (leaf nodes) are associated with constant values as leaf parameters, which are averages of all leaf node observations. To predict the outcome for a new data observation, we search the decision rules along the tree to find the leaf node that this new data observation falls in. The corresponding leaf parameter serves as the prediction. Note that the prediction is identical regardless of the time period, and therefore, it is a single period prediction. Gu et al. (2020) and Bali et al. (2021) use the standard CART or its ensemble methods (Random Forest) to predict one-period-ahead returns without considering the panel data structure.

A few potential problems arise when applying the off-the-shelf CART algorithm directly to asset pricing. First, the i.i.d. assumption is not valid for asset pricing panel data in general. When one implements the standard CART model on single-period data, the constant leaf parameters can be interpreted as portfolios (equally-weighted portfolios, average returns of stocks in the same period) and serve as pricing kernels for the assets falling into the corresponding leaf node. But when one fits CART to multiple periods, the leaf parameters become the average of returns from multiple periods, which loses the interpretation as an equally-weighted portfolio. This i.i.d. assumption is common in many ML models, and therefore developing new models adapted to panel data structures is much needed. Specifically for tree models, it is desirable to replace the constant leaf parameter with a vector since we want to form a *basis portfolio*—a vector of returns over multiple periods. That way, we can split the data using firm characteristics (cross-sectional split) or macro variables (time-series split) to accommodate asset heterogeneities and the effect of regime switch.

Second, CART grows recursively such that when the algorithm splits a node, it only uses the data within that node without collecting information from its sibling nodes. This myopic and greedy strategy focuses on local optimization and is prone to overfitting. The major novelty of P-Tree lies in imposing a *global* split criterion defined on all leaf nodes and satisfying the no-arbitrage condition. It is a tree-based factor pricing model aiming to recover the SDF from *all* leaf basis portfolios created by tree partitions. As a result, P-Tree grows iteratively rather than recursively at each iteration. P-Tree searches for the all split rule candidates in all current leaf nodes to find the optimal one that improves the overall performance of SDF most.

Tree-based models work well with low signal-to-noise data and short observation history because (i) each node uses a small number of parameters for target values (CART uses one parameter for the constant target value in each node), and (ii) the global split criteria explicitly moderates in-sample overfitting, unlike deep neural networks that could amplify noise from layer to layer. Ensemble methods further help average out the noise in the data. While P-Tree inherits these excellent properties, it does not resort to ensemble methods or evolutionary algorithms that sacrifice a single tree's interpretability. Therefore, our global split criteria for guiding tree growth also provides a novel approach to guarding against overfitting while maintaining interpretability.

We next develop the P-Tree framework and illustrate its effectiveness through an application in asset pricing with conditional pricing kernels.

2.2 Conditional SDF Model Using P-Tree

Under the formal definition of no-arbitrage as contained in Schachermayer (1994), a conditional stochastic discount factor model explaining cross-sectional returns satisfies:

$$E_{t-1}\left[m_t r_{i,t}\right] = 0 \Longleftrightarrow E_{t-1}\left[r_{i,t}\right] = \underbrace{\frac{\text{Cov}_{t-1}\left(m_t, r_{i,t}\right)}{\text{Var}_{t-1}\left(m_t\right)}}_{\beta_{i,t-1}} \underbrace{\left(-\frac{\text{Var}_{t-1}\left(m_t\right)}{E_{t-1}\left[m_t\right]}\right)}_{\lambda_{t-1}}, \tag{3}$$

where $r_{i,t}$ is the excess return of individual asset i at time t, $\beta_{i,t-1}$ and λ_{t-1} are the conditional risk exposure of stock i and risk premium of an SDF, m_t, which is unique.[7] We can use the SDF as determined by the market and write it in the following form[8]:

$$m_t = 1 - w_{t-1}^{\intercal} r_t, \tag{4}$$

[7]In any discrete-time model with infinite horizon, even for risky assets without embedded price component analogous to that of fiat money (i.e., strictly positive liquidation value at time infinity), the assets' cumulative gains process may still fail to be a uniformly integrable martingale under the risk neutral measure, which gives room for potential bubbles. See. e.g., Jarrow (2015) to an overview of bubbles in asset pricing.

[8]In Cochrane (2009), SDF to excess returns is usually written as $m_t = 1 - \delta^{\intercal} f_t$, where f_t are risk factors and δ is the corresponding vector of risk price. Suppose risk factors are portfolios of individual assets, such as the market factor. In a general case, one can rewrite SDF as a linear function of individual assets and model it with time-varying weights.

where $r_t = (r_{1,t}, \cdots, r_{N_t,t})^\intercal$ denotes returns of all assets at time period t, and w_{t-1} is a vector of asset weights.[9] Substituting Equation (4) into Equation (3) backs out the weights in this SDF:

$$w_{t-1} = \text{Cov}_{t-1}\left[r_t r_t^\intercal\right]^{-1} E_{t-1}\left[r_t\right]. \tag{5}$$

However, the existing literature does not estimate w_{t-1} from many individual stock returns, mainly due to the large dimension of the return covariance and unbalanced panel structure. We use leaf basis portfolios generated by a P-Tree (as opposed to individual stocks) to overcome the high-dimensional estimation problem. In contrast to existing studies that pre-specify the basis portfolios, our P-Tree algorithm splits the cross section of asset returns into basis portfolios in a data-driven manner while imposing global no-arbitrage criteria.

A P-Tree splits the stock universe into non-overlapping basis portfolios by cross-sectional quantiles of past firm characteristics (and possibly by time-series macroeconomic variables). The number of basis portfolios increases one at a time when the tree splits a parent node into two child nodes. The j-th iteration of the growing process generates $j+1$ leaf portfolios. Therefore it reduces dimension from estimating weight of N_t individual assets to $j+1$ leaf portfolios. We denote the SDF $m_t^{(j)}$ generated by basis portfolios $R_t^{(j)}$ (length $j+1$ vector) as

$$m_t^{(j)} = 1 - W_{t-1}^{(j)} R_t^{(j)}, \tag{6}$$

where the vector of portfolio weights $W_{t-1}^{(j)}$ is

$$W_{t-1}^{(j)} = \text{Cov}_{t-1}\left[R_t^{(j)} R_t^{(j)\intercal}\right]^{-1} E_{t-1}\left[R_t^{(j)}\right]. \tag{7}$$

For the conditional factor model, the time-varying factor exposure $\beta_{i,t}$ is given by

$$\beta_{i,t-1}^{(j)} = \frac{\text{Cov}_{t-1}\left(f_t^{(j)}, r_{i,t}\right)}{\text{Var}_{t-1}\left(f_t^{(j)}\right)}, \quad \text{where } f_t^{(j)} = W_{t-1}^{(j)} R_t^{(j)}. \tag{8}$$

[9]This could mean that it is the SDF determined by the prices of traded derivatives, as discussed in Jarrow (2018). Our implicit assumption is that the SDF is traded, which is not that restrictive, as Jarrow (2018) also discusses. In addition, note that that the no-arbitrage condition is satisfied for the information set we use in our estimation. In general, one needs to use the largest information set consistent with ones estimation when defining conditional expectations.

In the j-th iteration of the tree growing process, we model the time-varying factor exposures in reduced form in terms of past firm characteristics, which is common in the literature (e.g., Avramov, 2004; Kelly et al., 2019; Feng and He, 2022). In other words,

$$\beta_{i,t-1}^{(j)} = b_0^{(j)} + [b^{(j)}]^\intercal z_{i,t-1}, \tag{9}$$

where $z_{i,t-1}$ are firm characteristics such as market equities or book-to-market ratios. This flexible and intuitive conditional beta formulation adopts the same parameters of $\{b_0^{(j)}, b^{(j)}\}$ for every $r_{i,t}$. When the tree grows and splits into more leaf basis portfolios, these parameters are updated iteratively along with the generated SDF.

2.3 Growing a P-Tree under No-Arbitrage

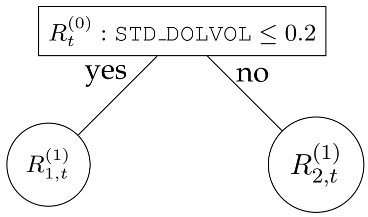
A P-Tree at birth corresponds to CAPM: a single leaf basis portfolio with returns of all assets. We gradually build the tree with additional splits by updating $\{R_t^{(j)}, f_t^{(j)}, \beta_{i,t}^{(j)}\}$ through an iterative scheme. First, leaf basis portfolios, $R_t^{(j)}$, are expanded one by one when the tree splits and grows one more leaf. Then, the latent factor, $f_t^{(j)}$, is re-estimated using the expanded leaf basis portfolios. Finally, time-varying factor exposures, $\beta_{i,t}^{(j)}$, are re-estimated by the updated SDF, $f_t^{(j)}$ when fitting the conditional factor model to individual asset returns.

Let $R_t^{(j)}$ denote the return of the leaf basis portfolios after the j-th iteration of the tree. The number of basis portfolios increases by one after each iterative split. As a result, after j-th iteration, the tree has $j+1$ basis portfolios, denoted by $R_t^{(j)} = [R_{1,t}^{(j)}, R_{2,t}^{(j)}, \cdots, R_{j+1,t}^{(j)}]$. Each $R_{p,t}^{(j)}$ represents a vector of returns of T time periods for the p-th leaf basis portfolio, which can be equal- or value-weighted portfolio of the asset returns in that specific leaf. Note that taking weighted portfolio works for unbalanced data, it is not necessary that each month contains the same number of individual assets in a leaf node.

Next, we introduce the tree growing process, starting from the first split of the observations along firm characteristics. Before the first split, the entire cross section of stock returns is in the top node (root node) of the tree, $R_t^{(0)}$. We consider several split rule candidates (pairs of firm characteristics and threshold values to split the assets). All firm characteristics are normalized cross-sectionally to $[-1, 1]$. We consider decile split threshold candidates at -0.6, -0.2, 0.2, and 0.6, a choice following conventional security sorting and for simple computation.[10]

Figure 2: Illustration of the First Split

To search for the best characteristic to split the optimal cutpoint, let us consider one split candidate, STD_DOLVOL≤ 0.2, for calculating the split criterion.



Each split rule candidate partitions the root node to the left and right child nodes consisting of a subset of assets and all months of the data. Each potential leaf can form a leaf basis portfolio, denoted by $R_{1,t}^{(1)}$ and $R_{2,t}^{(1)}$, which are equally weighted (or value weighted) portfolio returns of assets in corresponding leaf nodes. Note that the calculation below is invariant to the indexing of leaf portfolios. The latent factor (SDF) is estimated as a mean-variance efficient portfolio of the two leaf basis portfolios,

$$f_t^{(1)} = W^{(1)} R_t^{(1)}, \quad W^{(1)} \propto \Sigma_1^{-1} \mu_1, \tag{10}$$

where Σ_1 and μ_1 are the covariance matrix and average returns of two leaf portfolios $R_t^{(1)} = [R_{1,t}^{(1)}, R_{2,t}^{(1)}]$. [11] The sum of portfolio weights is restricted to one. The split criterion of this candidate $\tilde{c}_j$ is defined as follows:

$$\mathcal{L}(\tilde{c}_j) = \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(r_{i,t} - \beta(z_{i,t-1}) f_t \right)^2, \tag{12}$$

[10]The standard CART considers quantiles of predictors as candidate splits; conventional sorting in empirical asset pricing uses quintiles or deciles. We thus follow the convention to use a quintile grids and the results are likely strengthened with finer grid search, which our online package allows.

[11]On a separate note, to deal with the estimation errors in the sample mean and sample covariance matrix for estimating the mean-variance efficient portfolio, we include two small shrinkage parameters $\gamma_\Sigma = 10^{-4}$ and $\gamma_\mu = 10^{-4}$ in Equation (7). These two shrinkage parameters help stabilize the portfolio weight estimation and avoid over-leveraging. Larger shrinkage parameters imply a larger regularization. The similar regularized portfolio optimization problem is also addressed in Kozak et al. (2020) and Bryzgalova et al. (2020).

$$W^{(j)} = \left(\mathrm{Cov}(R_t^{(j)}) + \gamma_\Sigma \mathbf{I}_{k+1} \right)^{-1} \left(E(R_t^{(j)}) + \gamma_\mu \mathbf{1} \right). \tag{11}$$

where $\beta(z_{i,t-1}) = b_0 + b^\intercal z_{i,t-1}$ are conditional betas on the past firm characteristics $z_{i,t-1}$. Following the regression tree in Equation (1), the different constant target value $\{\mu_j\}$ for different leaves are replaced with one common factor model to all observations:

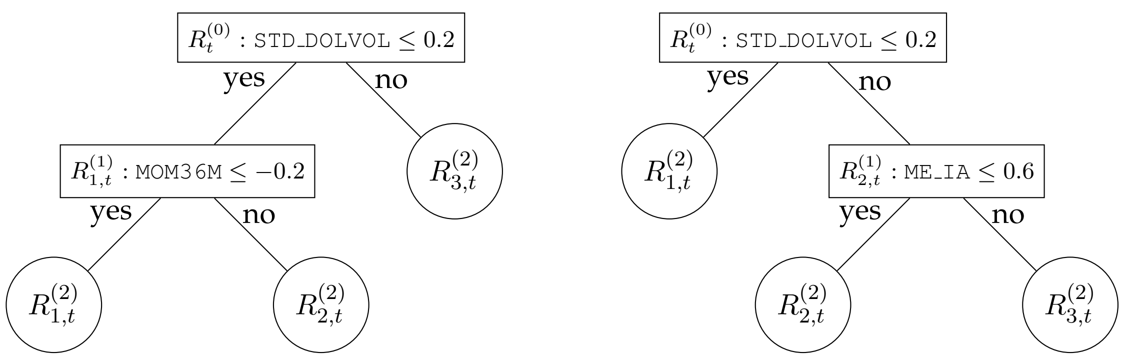
$$\mathcal{T}(z_{i,t-1}|\Theta) = \beta(z_{i,t-1}) f_t, \tag{13}$$

where $\Theta = \{\beta(\cdot)\}$ is a common function for all individual assets.

We estimate the regression coefficients using a pooled regression model for individual stock returns on f_t and $f_t \times z_{i,t-1}$ without an intercept. Note that different split candidates produce different partitions of the data, thus create different leaf basis portfolios, different corresponding SDFs, and different valuation of split criteria eventually. The split criteria in Equation (12) is evaluated for all split candidates $\tilde{c}_j \in \mathcal{C}$. The first split rule is the one that minimizes split criteria.

Figure 3: Illustration of the Second Split

This figure displays the candidates for the second split. Note that no matter which leaf node is split, the second (j-th) iteration of the tree model has three ($j+1$) leaf basis portfolios for the SDF construction.



(a) Splitting node $R_{1,t}^{(1)}$ at MOM36M.

(b) Splitting node $R_{2,t}^{(1)}$ at ME_IA.

Next, we proceed to searching for the second split. Note that two leaf nodes exist after the first split. The second split can happen at either the left or right child node of the root. We evaluate the split criteria for all candidates at *both* leaf nodes and pick the one that minimizes the split criteria as in Equation (12). Figure 3 depicts the tree of the candidates for the second split. In either of the two cases, one leaf node splits, becomes an internal node, and creates two new leaf nodes. The SDF is now evaluated based on the three basis portfolios:

$$f_t^{(2)} = W^{(2)} R_t^{(2)}, \quad W^{(2)} \propto \Sigma_2^{-1} \mu_2, \tag{14}$$

where $R_t^{(2)} = [R_{1,t}^{(2)}, R_{2,t}^{(2)}, R_{3,t}^{(2)}]$. The graphical position of the three basis portfolios depends on which node to split, as shown in Figure 3. Similarly, Σ_2, μ_2 are the return covariance matrix and averages for the basis portfolios $R_t^{(2)}$. The new SDF is plugged into the split criteria in Equation (12). It is worth noting that the split criterion is defined using *all* leaf basis portfolios to generate the SDF. Furthermore, we explore all candidates of all possible nodes to find the one that has the largest improvement on the global SDF performance, rather than focusing on a specific leaf node without looking at sibling nodes like CART. This strategy is less myopic and helps prevent overfitting since the optimization object is the pricing performance of the SDF on all assets and all time periods.

All subsequent splits are determined similarly. For the j-th split, each candidate of the existing j leaf nodes creates $j+1$ leaf basis portfolio, which adds to the generation of the latent factor $f_t^{(j)}$, and evaluates the global split criteria of Equation (12). A P-Tree stops growing when further splits do not improve the global split criterion. We pre-specify additional stopping conditions in terms of the max depth of the tree or the minimum number of assets in a leaf node (leaf size), whichever is met first. Algorithm 1 summarizes the above procedures.

In summary, the P-Tree model differs from the standard tree models such as CART in several important ways. First, CART aims to approximate a function by a step function represented by the tree, and each CART leaf is associated with a constant target value (one leaf parameter for each leaf). In contrast, the P-Tree has a vectorized leaf parameter $R_{p,t}^{(j)}$ for leaf basis portfolio returns, and it has a clear economic objective for the customized split criteria. Particularly, P-Tree fits a common factor model (regression) for assets and uses idiosyncratic fitted values as target values (several parameters for all leaves). Second, P-Tree exploits the cross-sectional correlations for the panel data analysis while CART assumes i.i.d. observations when fitting a decision tree. Finally, as opposite to CART, whose splits are local and greedy, our P-Tree grows iteratively by examining *all* assets to construct the *common* factor model. The split criterion is defined globally as the pricing errors of the generated latent factor for the returns of all assets, regardless of whether observations are in the current partition node or not.

Algorithm Algorithm of growing a single P-Tree

procedure GROWTREE(`root`)
outcome Grow the tree from the `root` node, form leaf basis portfolios
  for j from 1 to num_iter **do** ▷ Loop over number of iterations
    if current depth $\geq d_{\text{max}}$ **then**
      return.
    else
      Search the tree, find all leaf nodes $\mathcal{N}$
      for each leaf node N in $\mathcal{N}$ **do** ▷ Loop over all current leaf nodes
        for each split candidate $\tilde{c}_{N,k}$ in $\mathcal{C}_N$ **do**
          Partition data temporally in N according to $\tilde{c}_{N,k}$.
          if Either left or right child of N does not satisfy minimal leaf size **then**
            $\mathcal{L}(\tilde{c}_{N,k}) = \infty$.
          else
            Calculate leaf basis portfolios.
            Estimate SDF using *all* leaf basis portfolios as in Equation (10).
            Calculate the split criteria $\mathcal{L}(\tilde{c}_{N,k})$ in Equation (12).
          end if
        end for
      end for
      Find the best leaf node and split rule that minimizes split criteria

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

      Split the node selected at the j-th split rule of the tree $\tilde{c}_j$.
    end if
  end for
  return
end procedure

2.4 Nonlinearity, Interaction, and Interpretability

Observing Interactions in P-Tree. From root to leaf nodes, a tree learns a sequence of split rules which indicate nonlinearity and interactions of the split variables. While deep learning models can also learn nonlinearity and interactions, tree models graphically displays how predictors interact in the sequence of split rules. Despite this natural interpretability, a single tree tends to overfit noisy data because the standard recursive splitting scheme focuses on optimizing a local step and uses fewer and fewer observations with more splits. In the literature, ensemble methods, such as Random Forest or boosted regression trees, are often used to improve a model's out-of-sample performance at the expense of interpretability.

In P-Trees, we require that the generated tree factors must be common to all assets, and imposing this global asset pricing condition helps reduce model overfitting since it is less myopic. While a global sparse portfolio optimization can be used to prune the tree growth from bottom to top (Bryzgalova et al., 2020), P-Tree displays in a top-down manner the exact interactions of the input predictors in each layer of the tree growth without requiring exogenous specifications of the tree structure to start with. Furthermore, enumerating all exogenous specifications of trees is an NP-hard problem for a large number of firm characteristics, but P-Tree can search the space of all sequential interactions efficiently.

As we show in our empirical analysis later, Figure 4 plots a decision tree and Figure 5 displays the partition plot. The data-driven P-Tree allows us to observe predictor interactions and use the information to improve trading strategies significantly. They also provide information for recovering the SDF and building a factor pricing model that performs and, if not better than, the best competing models. More generally, global split criteria can be used to guard against the overfitting of a single tree and consequently presents direct empirical evidence for the nonlinearity and interaction in the data—a task other ML approaches or ensemble methods fall short of.

Factors and Portfolio Construction. P-Tree provides an alternative portfolio construction and strategy development solution that utilizes predictor interaction and nonlinearity information. For instance, a depth 2 P-Tree with two leaves splits the cross section and creates two left- and right-leaf portfolios. Assets within the same leaf tend to have similar risk-return dynamics, while assets between two splitting leaves have different characteristics-return directions for these published anomalies. Guided by economic theory, one can construct a univariate factor to long and short these two leaf basis portfolios. Researchers typically use quintile or decile sorted portfolios to create long-short portfolios (top-bottom or bottom-top) to generate higher return spreads between long and short portions. Researchers might also apply bivariate or triple-way sorting to include the interaction of the characteristics when creating long-short portfolios.[12]

[12]Different characteristics-sorted univariate, bivariate, or triple-way portfolios are available on Ken French's website. Commonly used long-short factors, such as Fama-French five factors, are linear combinations for these portfolios re-

A P-Tree searches for the optimal characteristics interaction for the second and probably third splits when growing the tree with the asset pricing goal. First, growing a tree model with depth three enables different interactions on the long and short portions from the first split. Second, by partitioning the stock universe, one can graphically display the nonlinear characteristics-return structures for the long and short portions. Third, the splitting order and cutpoints are data-driven and provide information on the sorting sequence and the interaction of characteristics.

3 An Empirical Implementation of P-Tree on U.S. Public Equities

3.1 Data

Equity data and asset characteristics. We apply the common filters (e.g., same as in Fama-French factor construction) to the universe of U.S. equities: (1) include only stocks listed on NYSE, AMEX, or NASDAQ for more than one year; (2) use those observations for firms with a CRSP share code of 10 or 11; and (3) exclude stocks with negative book equity or negative 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. Table A.1 lists these input variables whose computation follows Feng et al. (2021). We standardize the firm characteristics cross-sectionally in the range $[-1, 1]$ to construct regression trees for multiple data periods.[13]

The sample period ranges from January 1981 to December 2020. We use the first 20 years for training and the recent 20 years from 2001 to 2020 for testing. The average and median monthly numbers of stock observations are 4,667 and 4,450 in the training sample and 3,953 and 3,791 in the testing sample. Our P-Tree is flexible enough to handle such an unbalanced data panel.

Macro predictors. In addition, we use 10 macroeconomic variables as in Feng and He (2022) to demonstrate time-series splits. Table A.2 summarizes the macroeconomic variables, which include market timing macro predictors, bond market predictors, and aggregate characteristics for S&P

stricted by the economic direction.

[13]For example, the market equity in December 2020 is uniformly standardized to the range of $[-1, 1]$. The firm with the lowest market equity is -1, and the highest market equity is 1. All others are distributed uniformly in between. Therefore, this uniform standardization transforms the data onto $[-1, 1]$ every month. If a firm has missing values for some characteristics, the imputed values are 0, implying the firm is not important in security sorting.

500. We standardize these macro predictor data by the historical percentile numbers for the past ten years. For example, inflation greater than 0.7 implies the current inflation level is higher than 70% of observations during the past decade. This rolling window data standardization is useful when comparing the predictor level to detect different macroeconomic regimes.

3.2 Implementing P-Tree

We apply P-Tree to generate leaf basis portfolios and the SDF to fit cross-sectional individual asset returns. Figure 4 illustrates the structure of the baseline P-Tree. The tree structure shows the splitting orders S#, selected splitting characteristics, and the cross-sectional cutpoints. Before the first split, the tree grows from the root node, representing the market portfolio for all stocks (N1). The leaf basis portfolios at each depth of the tree are numbered with N#. The numbers printed in the leaves are the median number of stock observations in the monthly updated leaf basis portfolios.

Splits and stopping. Adding to the stopping criteria, we require the minimum number of assets in a leaf basis portfolio, one natural tuning parameter to limit the tree size to avoid overfitting, to be 50, and restrict the maximum depth of a tree to 6.[14] In the training sample of our equity study, the tree stops growing after 22 splits and generates 23 leaf basis portfolios. As Figure 4 shows, endogenously the P-Tree first splits along the trading volume volatility (`STD_DOLVOL`) at 0.2 (60% quantile), which is related to studies on market microstructure emphasizing volatility of liquidity (e.g., Chordia et al., 2001). After the split, 60% of the stocks go to the left leaf (labeled N2), and 40% go to the right (N3). The second split is on the industry-adjusted market equity (`ME_IA`) at 0.2 of the right leaf (N3), whereas the third split is on three-year long-term reversal (`MOM36M`) of the left one (N2). P-Tree learns interaction of characteristics appearing in the same path, since they jointly define the partition corresponding to the leaf node. For instance, `ME_IA` is valuable for high `STD_DOLVOL` stocks, whereas `MOM36M` is valuable for low `STD_DOLVOL` stocks.

[14]This consistent with the empirical literature that rarely sorts assets finer than deciles. Our findings are robust to such parameter tuning, and the online package allows users to have different specifications. As Figure 4 shows, the median portfolio size is well above 50 when the splitting stops, indicating that the improvement in the split criterion is the main driver for stopping.

Partition, clusters, and leaf portfolios. P-Tree provides a more precise mapping for generating these characteristics-managed portfolios relative to other ML methods. Figure 5 reports the corresponding partitions, together with the monthly average returns and the annualized Sharpe ratios of leaf basis portfolios. These performance gaps show the usefulness of splitting the cross section via the interaction of characteristics. The top right plot in Figure 5 further splits N2 by `STD_DOLVOL` and `MOM36M`. The bottom plot splits N6 by `ATO`.

We can also observe detailed clustering patterns for these leaf basis portfolios. The market portfolio is partitioned into four portfolios (N4 - N7) at a depth of three, with each group of individual assets sharing similar characteristic risk exposures. The return-risk relationship of the four portfolios, as well as their mean-variance efficient portfolio (`MVE3`), are plotted in Figure A.1. Further splitting these four portfolios, we obtain 23 leaf basis portfolios at depth 6. We find that the great-grandchild leaf basis portfolios (in a light color) labeled with (N4+ - N7+) cluster around their great grandparent nodes (in dark color). The mean-variance efficient portfolio from a 6-depth tree produces a higher Sharpe ratio than that from a 3-depth tree, whereas both are higher than the Sharpe ratio of the market portfolio. These performances are robust as in the out-of-sample plot in Figure A.2.

Factors in pricing kernel and boosting. P-Tree models offer practical ways to build a factor model. The baseline single P-Tree can generate the single-factor SDF; to generate a multi-factor model for asset pricing, one can use boosting, an ML approach to combine an ensemble of weak learners for better fitting and prediction (Freund and Schapire, 1997). It grows a list of additive trees sequentially where each tree fits the residual of all previous trees. Together, they form a strong learner, and the prediction is tree outcomes' (weighted) sum. For example, Rossi and Timmermann (2015) use boosted regression trees to effectively estimate conditional covariance for ICAPM.

Because standard boosting does not involve latent factor generation or portfolio returns, we propose new boosting split procedures. The boosting P-Tree creates additional factors sequentially to fit the unexplained pricing errors of all previous factors. The steps are as follows:

1. The first factor $f_{1,t}$ is generated by the standard P-Tree on excess returns $\{r_{i,t}\}$, as discussed in Section 2.3. The residual of the first factor is $\epsilon_{i,t}^{(1)} = r_{i,t} - \beta(z_{i,t-1})f_{1,t}$.

2. The second factor is generated from fitting the P-Tree with a boosting split criteria. The tree growing steps are the same as in Section 2.3 except we use the boosting split criteria:

$$\mathcal{L}(\tilde{c}_j) = \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(\epsilon_{i,t}^{(1)} - \beta_2(z_{i,t-1}) f_{2,t} \right)^2. \tag{15}$$

Note that the factor model in the second tree focuses on explaining the residual yielded from the first factor model. After fitting the second factor, the new residual is taken with respect to the first two factors as $\epsilon_{i,t}^{(2)} = \epsilon_{i,t}^{(1)} - \beta_2(z_{i,t-1}) f_{2,t}$.

3. The third factor $f_{3,t}$ proceeds similarly while focus on explaining residuals unexplained by the first two factors.

$$\mathcal{L}(\tilde{c}_j) = \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(\epsilon_{i,t}^{(2)} - \beta_3(z_{i,t-1}) f_{3,t} \right)^2. \tag{16}$$

4. Repeat the above process K times to generate K factors, with each tree factor model focusing on fitting residual from all previous factors.

Note that all factors are generated using the original asset returns. However, the subsequent factors focus on generating new leaf portfolios and thus new factors to explain the pricing errors from previous factors. Therefore, we separate the leaf basis portfolios and pricing target for subsequent factors $k = 2, 3, \cdots$. Suppose we fit the P-Tree model with $K = 3$ factors. The three generated factors can be listed with a decreasing order of importance as $[f_{1,t}, f_{2,t}, f_{3,t}]$. Generating additional P-Tree factors is similar to generating additional principal components and each subsequent factor explains some residual variations in asset returns. A three-factor P-Tree model with conditional betas simply requires re-estimating the betas:

$$\mathcal{T}(z_{i,t-1}|\Theta) = \beta_1(z_{i,t-1}) f_{1,t} + \beta_2(z_{i,t-1}) f_{2,t} + \beta_3(z_{i,t-1}) f_{3,t}. \tag{17}$$

Market-Adjusted P-Tree. Another natural application of boosting is to create an augmented factor model. One can start with a benchmark model such as CAPM or Fama-French factor models and then use the boosting P-Tree (first tree in the ensemble) to directly fit the benchmark model's pricing errors. Tree factors developed beyond the benchmark model try to explain "unexplained" or "orthogonal" information from factor model residuals. Complementing our standard P-Tree in Figure 4, we also provide a market-adjusted P-Tree in Figure 6. We show in Section 4.1 that this set of market-adjusted leaf basis portfolios and the resulting factor have different pricing and investment performances. In particular, the market-adjusted P-Tree factor delivers a higher Jensen's alpha than the standard one, which demonstrates the market hedging performance for boosted P-Trees.

3.3 Interpretability and Robustness without Overfitting

Characteristic importance and interactions. Sequential or independent asset sorting schemes primarily focus on two or three characteristics in practice because of the high-dimensional challenge raised in Cochrane (2011). The popular sequential sorting in empirical asset pricing can be shown to be equivalent to using a decision tree to generate basis portfolios (see, also, Bryzgalova et al., 2020). However, enumerating all possible tree structures (i.e., sequential sorts with many characteristics) is computationally impractical for statistical testing and leaves no precise evaluation for characteristics' importance. P-Tree's iterative growth represents an effective greedy algorithm to get around this NP-hard problem.

Besides boosting, Random Forest (Breiman, 2001) is another prominent ensemble method that fits multiple trees to improve performance. Plain-vanilla Random Forest grows many trees on bootstrap training samples, which come from random samples of the observations with replacement and sampling a subset of variables from the complete training set. The idea behind bagging or Random Forest is that sampling data from empirical distribution approximate sampling from the true underlying distribution, thus enabling quantifying the uncertainty of model estimation. We deploy this strategy to gauge variable importance under the P-Tree framework. We refer to our strategy as Random P-Tree Forest in the following text.

First, we bootstrap the data on the time-series horizon with replacement. We preserve the complete cross section of the panel data for the time periods selected, to exploit the low serial correlations among asset returns. Second, we randomly draw ten characteristics. P-Trees grow on each bootstrap data independently. We repeat the procedure 500 times to form a forest consisting of 500 P-Trees.

Our empirical exercises focus on the out-of-bag (OOB) variable importance and define two measurements. The first measurement of variable importance counts the frequency of one variable being selected as splits. Intuitively, the more often a characteristic is selected as tree split rules, the more important it is. For each bootstrap sample, a subset of characteristics is randomly drawn to grow the tree, and only a fraction of them are actually selected as split rules. We count the number of times a particular $l-$th characteristic z_l is being used in the first J splits and the total number of appearances in bootstrap subsamples. We define the first measurement of importance as

$$\text{Selection Probability}(z_l) = \frac{\#\,(z_l \text{ is selected at first } J \text{ splits})}{\#\,(z_l \text{ appears in the bootstrap subsamples})}. \tag{18}$$

The second measurement of importance aims to capture a feature's "treatment effect." For the 500 bootstrap subsamples, we know a characteristic is randomly included in part of the bootstrap subsamples following a Bernoulli distribution. Note that, even if one characteristic is included in the subsample, it is not guaranteed to be selected as a split rule when building the tree. The with- and without-sampling scheme for a particular characteristic creates the treatment effect evaluation and allows one to perform the significance test of its importance. We compute the Total R^2 in Equation (20) as the asset pricing fitting measure for this evaluation:

$$\text{Characteristic Importance}(z_l) = \left[\frac{E\,(\text{P-tree fit} \mid \text{with } z_l)}{E\,(\text{P-tree fit} \mid \text{without } z_l)} - 1\right] \times 100. \tag{19}$$

Figure 7 summarizes the predictor importance in a Random Forest of P-Trees. According to Equation (19), a negative value implies that the model has an average high asset pricing performance value when we include this characteristic rather than dropping it. We conduct a two-sample t-test using the bootstrap samples and plot those significant characteristics in a deeper color. Only 4 out of 61 characteristics show statistical significance. The in-sample results identify the significant characteristics such as volume volatility (`STD_DOLVOL`), long-term reversal (`MOM60M`), market-adjusted market equity (`ME_IA`), and one-year seasonality (`SEA1A`). These four characteristics also show consistent asset pricing improvement for the test sample of the recent 20 years. Panel A of Table A.3 summarizes the selection frequency (Equation (18)) of characteristics as splits in the root node. We find that volume volatility (`STD_DOLVOL`) has a 73% chance of being the first splitting characteristic once the bootstrap sample contains it. Other important characteristics are long-term reversal (MOM60M), trading volume (DOLVOL), and market-adjusted market equity (ME_IA).

Overall, the two different metrics reveal the same subset of important predictors for splitting the cross section. We note that these features are important for capturing the interactions, and they do not necessarily overlap with the significant predictors in a linear regression model.

Robustness and overfitting. Observing predictor interactions enables P-Tree models to have great interpretability. A nonlinear tree structure and the interactions of characteristics are displayed when the decision tree continues splitting further. However, these patterns gleaned from individual P-Trees are only helpful if the P-Tree model can avoid overfitting. We use the aforementioned characteristic importance metrics to examine if a P-Tree indeed splits on similar characteristics and is not merely reflecting noise in the data. Specifically, we compare important characteristics from the Random P-Tree Forest with those revealed in our single P-Tree model. If they are similar, the P-Tree model with global split criteria behaves similarly to the less interpretable Random P-Tree Forest that typically achieves better out-of-sample performance.

The results are indeed consistent. Those important characteristics from Random P-Tree Forest, namely volume volatility (STD_DOLVOL), long-term reversal (MOM60M), and market-adjusted market equity (ME_IA) are also the characteristics used in the first split in the single-factor P-Tree (Figure 4 and 8) and the boosted multi-factor P-Trees. We deduce that P-Tree models behave similarly to Random P-Tree Forest ensembles and are not subject to overfitting. Yet, the nonlinearities and interactions revealed in P-Tree models then provide helpful information that is impossible to glean directly from ensemble models. Conventional non-ensemble tree model tends to overfit because the splits are guided by observations in local nodes, which become smaller in number as the tree grows. In contrast, the global split criteria in P-Tree utilize observations across nodes and preserve the interpretability of trees while preventing overfitting.

4 Advantages and Applications of P-Tree

The P-Tree framework is particularly effective when applied to asset pricing for the following reasons. First, P-Tree generates (multi-factor) SDF for pricing the cross section of assets. Second, P-Tree creates tradable factors for investment and test assets in the leaves for testing asset pricing models that better span the mean-variance efficient frontier. Third, relative to other ML models, including tree ensembles, P-Tree graphically displays nonlinearities and interactions in a direct and interpretable way, which can help enhance or resurrect well-known anomalies through sequential sorting. Finally, P-Tree offers the possibility of splitting the time series and fully extracting information from financial panel data through potential interactions of cross-sectional asset characteristics with time-series macroeconomic variables, which is hitherto absent in extant ML and tree-based models. The framework also admits applications beyond asset pricing.

4.1 Asset Pricing Performance

Performance metrics. We follow Feng et al. (2021) to include multiple performance metrics. Total R^2 (for individual assets) and Cross-Sectional R^2 (for portfolios) evaluate economic asset pricing performance for variation in time-series and cross-sectional dimensions, while Predictive R^2 (for individual assets) measures statistical model fitness and the model forecasting power. Specifically,

$$\text{Total } R^2 = 1 - \frac{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\left(r_{i,t} - \widehat{r}_{i,t}\right)^2}{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N_t} r_{i,t}^2}, \tag{20}$$

where $\widehat{r}_{i,t} = \hat{\beta}_{i,t}^{\intercal} f_t$. Total R^2 represents the fraction of realized return variation explained by the factor model-implied contemporaneous return, aggregated over all assets and all periods.

$$\text{Predictive } R^2 = 1 - \frac{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N_t}\left(r_{i,t} - \widehat{r}_{i,t}\right)^2}{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N_t} r_{i,t}^2}, \tag{21}$$

where $\widehat{r}_{i,t} = \hat{\beta}_{i,t}^{\intercal} \lambda_f$. The risk premia λ_f is the risk premia estimate, which can be the average returns of the tradable factor. Predictive R^2 summarizes the predictive performance by the factor model-implied return forecasts, aggregated over all assets and all periods.

$$\text{Cross-Sectional } R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}(R_{i,t} - \widehat{R}_{i,t})\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} R_{i,t}\right)^2}, \tag{22}$$

where $\widehat{R}_{i,t} = \hat{\beta}_{i,t}^{\intercal} f_t$. Cross-Sectional R^2 represents the fraction of the squared unconditional mean returns that is described by the common factor model, aggregated over all assets.[15]

Comparing Asset Pricing models. Table ?? summarizes the individual asset pricing performances in terms of Total R^2 and Predictive R^2. Panel A shows P-Tree models with different numbers of factors using the boosting strategy, and Panel B shows the market-adjusted P-Tree models. Panel C of Table ?? compares these results with observable factor models such as CAPM, Fama-French factor models (Fama and French, 1993, 2015), and the recent Q5 model of Hou et al. (2021). We also implement latent factor models from the ML finance literature, such as Risk Premium PCA (RP-PCA) of Lettau and Pelger (2020) and Instrumented PCA (IPCA) of Kelly et al. (2019). Panels D and E provide results for time-series split P-Tree models, which are discussed in Section 4.4.

Total R^2 quantifies a model's success in describing the common risks in cross-sectional returns. P-Tree and the market-adjusted P-Tree show positive and consistent performance in pricing individual equity returns. Adding additional boosted factors further improves the performance. Furthermore, the five-factor P-Tree and market-adjusted P-Tree outperform almost all models at pricing individual stock returns in terms of Total R^2. Predictive R^2 is equivalent to the out-of-sample R^2 used in the return predictability literature, such as Gu et al. (2020), but demonstrates the return predictability power out of an asset pricing factor model. The five-factor P-Tree and the market-adjusted P-Tree produce positive numbers for in-sample and out-of-sample analysis. Notably, another latent factor model, IPCA (Kelly et al., 2019) shows strong performance for pricing and predicting individual assets because of their direct objective function.

Next, we investigate the asset pricing performance on multiple sets of test portfolios. In this case, cross-Sectional R^2 in Equation (22) is calculated using the entire sample of data. We consider Fama-French ME-B/M 5×5 and 49 industry portfolios. In addition, there are 23 leaf basis portfolios generated by P-Tree in Figure 4 and 18 ones generated by market-adjusted P-Tree in Figure 6.

Observable factor models, such as FF5 and Q5, show great explanatory power for the cross-sectional average returns for the first three sets of test portfolios. Among the latent factor models, P-Tree, market-adjusted P-Tree, and Risk Premium PCA (RP-PCA) are as effective. These factor

[15] This measure is equivalent to XS-R^2 used in Chen et al. (2020) and pricing error R^2 in Feng et al. (2021).

models can explain about 80% of equity average returns, resulting in economically minor alpha scales. The factors spanning the SDF should reside at the intersection of individual asset returns and sorted portfolio returns (Andreou et al., 2022). Although IPCA performs well in pricing individual assets, it does not work well on pricing portfolios, because IPCA is trained with individual stocks with stock-specific time-varying betas and might lose these advantages when evaluating test portfolios. In contrast, P-Tree models work well pricing both individual assets and test portfolios.

The last column in Table ?? reveals that the 18 market-adjusted leaf basis portfolios are difficult to price using well-known observable or latent factor models, as evidenced by the large negative Cross-Sectional R^2. These 18 portfolios are generated by controlling the market factor and cannot be explained by CAPM. However, our market-adjusted P-Tree models explain the the cross section of leaf portfolio returns.

4.2 Test Portfolios, Investments, and Trading Factors.

P-Tree factors for investment. Since P-Tree factors are tradable, we can assess the out-of-sample investment performance of P-Tree factor portfolios. Figure 9 shows investing performance of two single P-Tree factors, P-Tree and time-series split P-Tree, as well as the portfolios constructed from value-weighted averages of multiple P-Tree factors. All factor portfolios are rebalanced monthly. The top panel shows the subsample robustness for investing P-Tree factors, where we find that almost all models deliver higher returns than the market portfolio in all subsample periods. In other words, any potential look-ahead bias is mitigated because the input characteristics have already been known in the later sample periods, and yet P-Trees utilizing them can generate excess returns. The bottom panel plots the cumulative returns for twenty years in the test period. Even single P-Tree factors double the cumulative return over the market factor, and the multi-factor model delivers extremely high compounded returns.[16]

We next compare the risk-adjusted investment performances of these observable or latent factors in various models. We consider two common investment strategies for each factor model:

[16]We note that the investment performance entails raw returns, not excess returns. We also leave out transaction costs and short-sale constraints and allow high leverages with individual asset weights in the $[-5, 5]$ range. In practice, the cumulative returns are likely significantly lower. Nevertheless, they would remain higher than other common ML strategies, which rely on high leverages. Moreover, the leverages do not affect asset pricing performance or risk-adjusted performance.

the $1/N$ equal-weighted portfolio and the mean-variance efficient portfolio.[17] Table 2 reports the monthly average return, Jensen's alpha, and the annualized Sharpe ratio of the training and test samples. The significance levels for alphas are also provided.

The P-Tree and the market-adjusted P-Tree models deliver positive and increasing performance with additional boosted factors. For the in-sample analysis, the $1/N$ strategy of the five factors model (P-Tree5) in Panel A generates a 2.42% monthly average return, a highly significant 1.84% alpha, and a 1.98 annualized Sharpe ratio, where the MVE strategy even provides higher values. For the out-of-sample analysis, this $1/N$ strategy of P-Tree5 delivers a 1.44% monthly average return, a highly significant 1.00% alpha, and a 1.25 annualized Sharpe ratio. These numbers are higher than the $1/N$ strategies of all other factor models, including the Fama-French five factors and IPCA.

Moreover, given that adding boosted factors is helpful for pricing, prediction, and investment, we investigate the additional investment information they possess beyond that in observable factor models. We generate these boosted factors by shallow trees with a maximum of depth three, and thus, we can interpret these factors as interaction factors by the first two splits. These interaction factors are similar to sequentially bivariate- or triple-sorted factors because the second and third splits can be implemented on different characteristics. The row names are the first two spitting characteristics of the corresponding tree. We regress each P-Tree factor on benchmark factor models to evaluate their alphas and summarize the results in Table 3.

Almost all P-Tree factors show economically and statistically significant alphas against the benchmark models. `BM_IA-ME` in Panel B is generated by the market-adjusted P-Tree and shows extremely positive alphas against Fama-French five factors and the Q5 model. The `MOM12M-ME` factor in Panel A shows much higher FF5-adjusted alphas over the standard momentum factor of Carhart (1997). These boosted performances come from both the interaction and nonlinearity of characteristics and the data-driven way of identifying these characteristics (as opposed to ad hoc sorting. We further investigate these issues in Section 4.3.

[17] We follow Equation (11) to calculate portfolio weights and rescale their sum to 1.

Leaves as new test assets. The empirical literature primarily uses portfolios (R_t in Equation (6)) as test assets instead of individual assets when evaluating asset pricing models. These test assets are usually non-overlapping portfolios that partition the asset universe. However, standard test assets such as Fama-French ME-B/M 5×5, and other characteristics-managed portfolios may fail to span the efficient investment frontier. For data-driven test asset construction, Ahn et al. (2009) use clustering to construct test assets based on return correlations, such that similar assets are grouped in the same cluster. If one wants to apply decision trees to split the cross section and generate test assets, it must be the same tree for every period to the cross section. This issue motivates our P-Tree design that allows dissection of the cross-sectional relationship for multiple periods from panel data of asset returns. The resulting leaf portfolios, for which the 18 market-adjusted leaf basis portfolios constitute one example, naturally serve as test assets for various models.

We gauge the performance of a set of test assets by their asset pricing and investment performances. Given that almost all models fail the asset pricing test, we consider the investment performance of test assets to see if they can span the efficient investment frontier. In Panel C of Table 3, we report the $1/N$ and MVE strategies constructed by different test assets, where only P-Tree and market-adjusted P-Tree portfolios deliver consistently significant performances in out-of-sample studies. Because the market-adjusted P-Tree successfully clusters those stocks with large negative returns, the MVE strategy alphas are very high, while the $1/N$ strategy alphas are significantly negative. These findings present strong evidence that these leaf basis portfolios are valid and useful test assets that span the efficient frontier of the asset return space. Moreover, these leaf portfolios produced by P-Tree are diversified test portfolios because the leaves include a sufficient number of assets, different from the security sorting with multiple characteristics.

4.3 Nonlinearity, Interactions, and Interpretability

CART graphically displays predictor interactions and nonlinear patterns in the partition plot. P-Tree perfectly inherits these advantages, see Figures 4 and 5 show. The decision tree approximates the sequential sorting scheme, where the long-short portfolio is to have a long position on one leaf and a short one on another. The partition plot demonstrates the nonlinear interaction patterns of asset characteristics when splitting the cross section. Economic theory and extant empirical literature determine the long and short directions for anomalies, such as small-minus-big or high-minus-low.

Trading strategies using nonlinearity and interaction. We create four different long-short characteristics-based portfolios, as shown in Figure A.4, following the directions for long-short positions in the literature. Specifications (a) and (b) are the standard univariate sorting schemes. Specifications (c) and (d) are interaction portfolios generated by our P-Tree and market-adjusted P-Tree models, which are essentially sequential sorting schemes.[18] We restrict all splitting cutpoints to be 0 (median of [-1, 1]) for this example, and the training sample fits these interaction portfolios from the training sample.

We summarize the number of significantly positive cases of monthly average returns and Jensen's alphas in Table 4 Panel A and B at 5% and 10% significance levels. First, "Uni-Sort 4×1" has more significant portfolio performance than "Uni-Sort 2×1" in the training, test, and the entire samples. Second, fewer profitable portfolios are in the test sample than in the training sample, which reflects either variance-bias tradeoff in the training or the post-publication bias (McLean and Pontiff, 2016). Third, our interaction portfolios have more "average-significant" cases than "Uni-Sort 4×1" in the recent 20 years and the entire sample. Our market-adjusted interaction portfolios also have more "alpha-significant" factors than "Uni-Sort 4×1." Out of the 61 factors or anomalies, the interaction specification increases the profitable portfolios from 27 to 36, and the market-adjusted interaction specifications increase from 36 to 48. These empirical findings confirm the advantages of the P-Tree that extract more information from nonlinearity and interactions beyond traditional asset sorting.

Panel C of Table 4 reports cross-sectional quantile numbers of average returns and Jensen's alphas. The "Uni-Sort 4×1" portfolios produce high median (50%) average returns and Jensen's alphas in the first 20 years. However, the returns from these portfolios have decreased dramatically in the recent 20 years. Conversely, the interaction portfolios have higher median average returns and alphas in the recent 20 years and the entire sample. The market-adjusted interaction portfolios show consistently positive 25% values for alphas in every case.

[18] For each characteristic, we train a P-Tree and restrict the first split on that characteristic, the second and third split on other characteristics, and the maximum depth is three.

Enhancing and resurrecting anomalies through interactions. To further understand how these interaction portfolios behave, six examples are reported in Table 5 and Figure A.5. The "Uni-sort 4x1" portfolios show significant average returns and alphas, and interaction portfolios improve these results. For example, the "Uni-sort 4x1" portfolio along Standardized Unexpected Earnings (SUE) of Rendleman Jr et al. (1982) has persistent excess returns. When we create an interaction portfolio with R&D to market (RMD) and market equity (ME), the performance almost doubles. RMD is useful for the short portion of SUE, and ME is helpful for the long portion of SUE. We show the average returns for leaf portfolios during splitting in Figure A.5 and find decreasing and increasing numbers in the corresponding direction. This important insight reveals the necessity to model the two legs of portfolio construction separately. While Jarrow et al. (2021) point out this necessity based on the argument that long- and short-legs have different underlying risk factors, and thus reflect different risks, we complement by highlighting how the long- and short-legs can interact with other characteristics differently.

Many factors or anomalies do not show significant risk premia or alphas and are deemed not replicable. Utilizing the information in interactions might solve this post-publication decay problem. In other words, once we include the "correct" control or interaction, some of these factors or anomalies can be resurrected.[19] For instance, we find "Uni-sort 4x1" factor change in profit margin (CHPM) of Soliman (2008) has almost zero premium. Interacting with R&D to sales (RD_SALE) on the short portion and dollar trading volume (DOLVOL) on the long portion, this factor earns 47 basis points for monthly average returns and 51 basis points for alpha. Unsurprisingly, the greedy algorithm of P-Tree helps lower the short portion and raise the long portion for the in-sample analysis. Consistent results for the test sample of the recent 20 years in Table 5 strongly support our P-Tree model. P-Tree helps recover some well-known characteristics and anomalies (not all of them), allows asymmetric interactions, and takes a data-driven approach to identify the predictors for the interactions.

[19] A related study is Asness et al. (2018), which resurrects the size factor by controlling for the junk factor.

4.4 P-Tree for Panel Data: Cross Section + Time Series

A panel data of asset returns have both cross-sectional and time-series dimensions. Most of the literature focuses on explaining the cross-sectional return variation, and so do the P-Tree models hitherto discussed. As 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 show that the factor risk premia and/or betas may differ substantially under different macroeconomic states (e.g., high and low inflation or interest rate states). There is a long literature studying regime changes in financial markets involving, e.g., periods of high and low volatility, booms and recessions (e.g., Maheu and McCurdy, 2000; Ang and Timmermann, 2012), and bull versus bear markets (Fabozzi and Francis, 1977). In addition, Boons et al. (2020) document that inflation risk is priced in stock returns, and the inflation risk premia vary over time.

A major innovation of P-Tree is that it allows splitting the time-series dimension and provides an alternative investigation to factor-beta estimation with regime-switching. Moreover, the time-series information thus extracted from macroeconomic variables proves helpful in constructing new factors and trading strategies. Specifically, although betas are time-varying based on past characteristics, the latent factor generation is static for the fixed decision tree in a P-Tree with only cross-sectional splits. A natural generalization is to create different factor models and estimate corresponding conditional betas under switching macroeconomic states. This can be achieved by splitting nodes (e.g., the root node) by macroeconomic variables (the time-series dimension) instead of asset characteristics. Each child node maintains the same cross section of the assets as the parent node, while their time-series samples are non-overlapped. This setting improves the flexibility of the P-Tree to capture both cross-sectional and time-series variations, which is impossible in the standard tree algorithms.

Given the short time series of individual stock returns relative to the large cross section, we only implement one time-series split at the root node (i.e., the first split rule of the tree) for illustration.[20] The P-Tree model searches over all possible split candidates of the macro variables. The split

[20] It is possible to implement multiple time-series splits and cross-sectional splits for panel data analysis. However, for the dimension that involves the latent factor generation, one needs to ensure there are more periods than the number of leaves for generating a factor model.

criteria of the first time-series split are defined as

$$\mathcal{L}(\tilde{c}_j) = \sum_{t \in \mathcal{R}_A} \sum_{i \in N_{A,t}} \left(r_{i,t} - \beta_A(z_{i,t-1}) f_{A,t}\right)^2 + \sum_{t \in \mathcal{R}_B} \sum_{i \in N_{B,t}} \left(r_{i,t} - \beta_B(z_{i,t-1}) f_{B,t}\right)^2, \tag{23}$$

where the split candidate $\tilde{c}_j$ partitions the time series of data to two non-overlapped sets of time periods $\mathcal{R}_A$ and $\mathcal{R}_B$, for example, high or low inflation states. Let $N_{A,t}$ denote the set of asset-return observations in the child node $\mathcal{R}_A$ in time period t, and $N_{B,t}$ proceeds similarly. Compared with the cross-sectional split criteria in (12), the time-series split criterion is the *total* pricing loss of two time periods, with two corresponding factors $f_{A,t}$ and $f_{B,t}$. Below is the two-period model with conditional betas:

$$\mathcal{T}(z_{i,t-1}|\Theta) = \beta_A(z_{i,t-1}) f_{A,t} I(t \in \mathcal{R}_A) + \beta_B(z_{i,t-1}) f_{B,t} I(t \in \mathcal{R}_B). \tag{24}$$

In any period, there is a one-factor model applied to the cross section. Notice that, $f_{A,t}$ does not exist in period $\mathcal{R}_B$, and $f_{B,t}$ does not exist in period $\mathcal{R}_A$.

After searching for the optimal time-series split rule at the root, all subsequent split rules are chosen from cross-sectional characteristics only. Note that any further split on either child of the root node only depends on asset return observations on one side. This extension also informs how macroeconomic variables interact with firm characteristics to price assets since they jointly define the partition of the space to create leaf basis portfolios and the SDF.

We consider growing a P-Tree with both cross-sectional and time-series splits. Given the short history of monthly returns, we only consider the time-series split for the first split at the root node.[21] The first split determines the optimal macro predictor for regime-switching. After a complete search among all ten macro predictors and split deciles, P-Tree splits inflation (`INFL`) at the median (50% quantile) in the training sample. Figure 8 shows the time-series split and the two branches of a cross-sectional split. The left and right branches have different tree structures in high and low inflation periods, implying that the tree model adapts to different macro conditions.

Figure A.3 reports important equity characteristics for high and low inflation periods. Ta-

[21] It is possible to mix multiple time-series and cross-sectional splits for growing the tree, which is to partition the predictor space in three dimensions.

ble A.3 presents variable importance in terms of selection probability for the time-series split. We discover that the important characteristics are different during high and low inflation periods, consistent with the findings in Boons et al. (2020). The long-term reversal (MOM60M) and one-year seasonality (SEAS1A) are more important during high inflation periods, whereas trading volume (DOLVOL) and volume volatility (STD_DOLVOL) are more important during low inflation periods. In conclusion, inflation is a lead regime-switching indicator, but these equity characteristics are valuable for differentiating the cross section.

Finally, the asset pricing results for the time-series split P-Tree model are summarized in Panels C and D of Table **??**, and their investment gains are reported in Table 2. Compared with standard P-Tree models, the 5-factor time-series P-Tree models in Panels C and D have higher Total and Predictive R^2. Also, the time-series P-Tree models provide higher and more stable Cross-Sectional R^2 to different test portfolios. In the recent 20 years, the $1/N$ equal-weighted portfolio for these five time-series tree factors provides a monthly average return of 2.78%, a highly significant alpha of 2.5%, and an annualized Sharpe ratio of 1.78. Time-series split models for high and low inflation periods tremendously improve the equity investment performance. By incorporating regime-switching and time-series splits, our P-Tree model can guide market-timing in equity trading and characteristics-based portfolio construction in practice.

5 Conclusion

Panel Tree (P-Tree)—a class of tree-based models for panel data analysis—complements traditional classification and regression tree models by relaxing their i.i.d. assumption and implementing a vectorized leaf parameter. In addition, P-Tree growth is guided by bespoke global criteria (e.g., based on the latent factor model) instead of local recursive criteria. When applied to asset pricing, these innovations help extract panel data information to generate basis portfolios and recover a robust stochastic discount factor while preserving the widely recognized advantages of tree-based models: graphically displaying nonlinear interactions and accommodating noisy data with a short history of observations.

In our empirical study of U.S. equities, we find that P-Tree models consistently outperform the known observable and latent factor models for pricing individual assets and portfolio returns. The generated P-Tree factor portfolios also show robust positive risk-adjusted investment performance. At the same time, the leaf portfolio can serve as diversified test portfolios for other asset pricing models. P-Tree creates profitable strategies based on characteristic interactions and can enhance or resurrect several known anomalies. Furthermore, P-Tree learns regime-switching across macroeconomic states and splitting in both the time-series dimension and cross-sectional asset characteristics. For example, P-Tree identifies inflation as the most useful macroeconomic variable for regime-switching, consistent with previous literature.

P-Tree models are applicable to other asset classes with a large cross section of data, such as corporate bonds. Future studies can integrate the time-series and cross-sectional splits in the split criteria for every possible partition. One can also split the cross section of stocks according to portfolio optimization objective functions to create a minimal variance portfolio in a data-driven way. Incorporating Bayesian regularization priors to develop a Bayesian P-Tree constitutes another interesting direction. More broadly, the P-Tree framework can be applied to general panel data structures with customized split criteria beyond asset pricing and finance.

References

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

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2019). Inference in group factor models with an application to mixed-frequency data. *Econometrica 87*(4), 1267–1305.

Andreou, E., P. Gagliardini, E. Ghysels, and M. Rubin (2022). Three common factors. *Available at SSRN 4080215*.

Ang, A. and A. Timmermann (2012). Regime changes and financial markets. *Annu. Rev. Financ. Econ. 4*(1), 313–337.

Arnett, F. C., S. M. Edworthy, D. A. Bloch, D. J. Mcshane, J. F. Fries, N. S. Cooper, L. A. Healey, S. R. Kaplan, M. H. Liang, H. S. Luthra, et al. (1988). The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 31*(3), 315–324.

Asness, C., A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen (2018). Size matters, if you control your junk. *Journal of Financial Economics 129*(3), 479–509.

Avramov, D. (2004). Stock return predictability and asset pricing models. *The Review of Financial Studies 17*(3), 699–738.

Avramov, D., S. Cheng, and L. Metzker (2021). Machine Learning versus Economic restrictions: Evidence from stock return predictability. *Management Science, Forthcoming*.

Bali, T. G., D. Huang, F. Jiang, and Q. Wen (2021). Different Strokes: Return Predictability Across Stocks and Bonds with Machine Learning and Big Data. Technical report, Georgetown University.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *The 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 (2021). Bayesian solutions for the factor zoo: We just ran two quadrillion models. Technical report, London Business School.

Bryzgalova, S., M. Pelger, and J. Zhu (2020). Forest through the trees: Building cross-sections of stock returns. Technical report, London Business School.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), 57–82.

Chaieb, I., H. Langlois, and O. Scaillet (2021). Factors and risk premia in individual international stock returns. *Journal of Financial Economics 141*(2), 669–692.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, 1281–1304.

Chen, L., M. Pelger, and J. Zhu (2020). Deep learning in asset pricing. Technical report, Stanford University.

Chen, N.-F. (1983). Some empirical tests of the theory of arbitrage pricing. *The Journal of Finance 38*(5), 1393–1414.

Chen, T. and C. Guestrin (2016). XGboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chinco, A., A. Neuhierl, and M. Weber (2021). Estimating the anomaly base rate. *Journal of Financial Economics 140*(1), 101–126.

Chipman, H. A., E. I. George, R. E. McCulloch, et al. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics 4*(1), 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. *The Journal of Finance 66*(4), 1047–1108.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2020). AlphaPortfolio: Direct Construction Trough Deep Reinforcement Learning and Interpretable AI. Technical report, Cornell University.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021). Deep sequence modeling: Development and applications in asset pricing. *The Journal of Financial Data Science 3*(1), 28–42.

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 (2021). Empirical asset pricing with Bayesian regression trees. Technical report, University of Notre Dame.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies 33*(5), 2180–2222.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou (2022). Anomalies and the expected market return. *The Journal of Finance 77*(1), 639–681.

Fabozzi, F. J. and J. C. Francis (1977). Stability tests for alphas and betas over bull and bear market conditions. *The Journal of Finance 32*(4), 1093–1099.

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 (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Feng, G., A. Fulap, and J. Li (2020). Real-time macro information and bond return predictability: Does deep learning help? Technical report, City University of Hong Kong.

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

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

Feng, G., N. Polson, and J. Xu (2021). Deep Learning of Characteristics-Sorted Factor Models. 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.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*(3), 985–1046.

Gerakos, J. J. and R. B. Gramacy (2012). Regression-Based Earnings Forecasts. Technical report, Dartmouth College.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The 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.

Guisan, A. and N. E. Zimmermann (2000). Predictive habitat distribution models in ecology. *Ecological modelling 135*(2-3), 147–186.

Haddad, V., S. Kozak, and S. Santosh (2020). Factor timing. *The Review of Financial Studies 33*(5), 1980–2018.

Han, Y., A. He, D. Rapach, and G. Zhou (2019). Expected stock returns and firm characteristics: E-lasso, assessment, and implications. Technical report, Washington University in St. Louis.

He, J. and P. R. Hahn (2021). Stochastic tree ensembles for regularized nonlinear regression. *Journal of the American Statistical Association, Forthcoming*.

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.

He, X., G. Feng, J. Wang, and C. Wu (2021). Predicting individual corporate bond returns. Technical report, City University of Hong Kong.

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. *The Review of Financial Studies 28*(3), 650–705.

Jarrow, R. and P. Protter (2016). Positive alphas and a generalized multiple-factor asset pricing model. *Mathematics and Financial Economics 10*(1), 29–48.

Jarrow, R. A. (2015). Asset price bubbles. *Annual Review of Financial Economics 7*, 201–218.

Jarrow, R. A. (2018). *Continuous-time asset pricing theory*. Springer.

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. T., D. Palhares, and S. Pruitt (2022). Modeling corporate bond returns. *Journal of Finance, Forthcoming*.

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. *The Review of Financial Studies 34*(6), 2813–2856.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting factor models. *The 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.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies 33*(5), 2274–2325.

Li, B. and A. G. Rossi (2020). Selecting mutual funds from the stocks they hold: A machine learning approach. Technical report, Wuhan University.

Liu, Y., G. Zhou, and Y. Zhu (2020). Maximizing the sharpe ratio: A genetic programming approach. Technical report, Hunan University.

Maheu, J. M. and T. H. McCurdy (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics 18*(1), 100–112.

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

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. Technical report, Ludwig Maximilian University Munich.

Nagel, S. (2021). *Machine Learning in Asset Pricing*. Princeton University Press.

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.

Roll, R. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance 35*(5), 1073–1103.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Rossi, A. G. (2018). Predicting stock market returns with machine learning. Technical report, Georgetown University.

Rossi, A. G. and A. Timmermann (2015). Modeling covariance risk in merton's icapm. *The Review of Financial Studies 28*(5), 1428–1461.

Schachermayer, W. (1994). Martingale measures for discrete-time processes with infinite horizon. *Mathematical Finance 4*(1), 25–55.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review 83*(3), 823–853.

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.

Zhu, L., R. A. Jarrow, and M. T. Wells (2021). Time-invariance coefficients tests with the adaptive multi-factor model. *The Quarterly Journal of Finance 11*(04), 2150019.

Figure 4: Panel Tree for the Period from 1981 to 2000 (new)

The P-Tree trained from the period from 1981 to 2000, with minimum leaf size as 50, is displayed in this figure. We show splitting characteristics and split rule values for each parent node. The node numbers (N#) and splitting order numbers (S#) are also printed on each parent node. We have included the median monthly number of assets in the final leaves basis portfolios. The description of equity characteristics are listed in Table A.1.

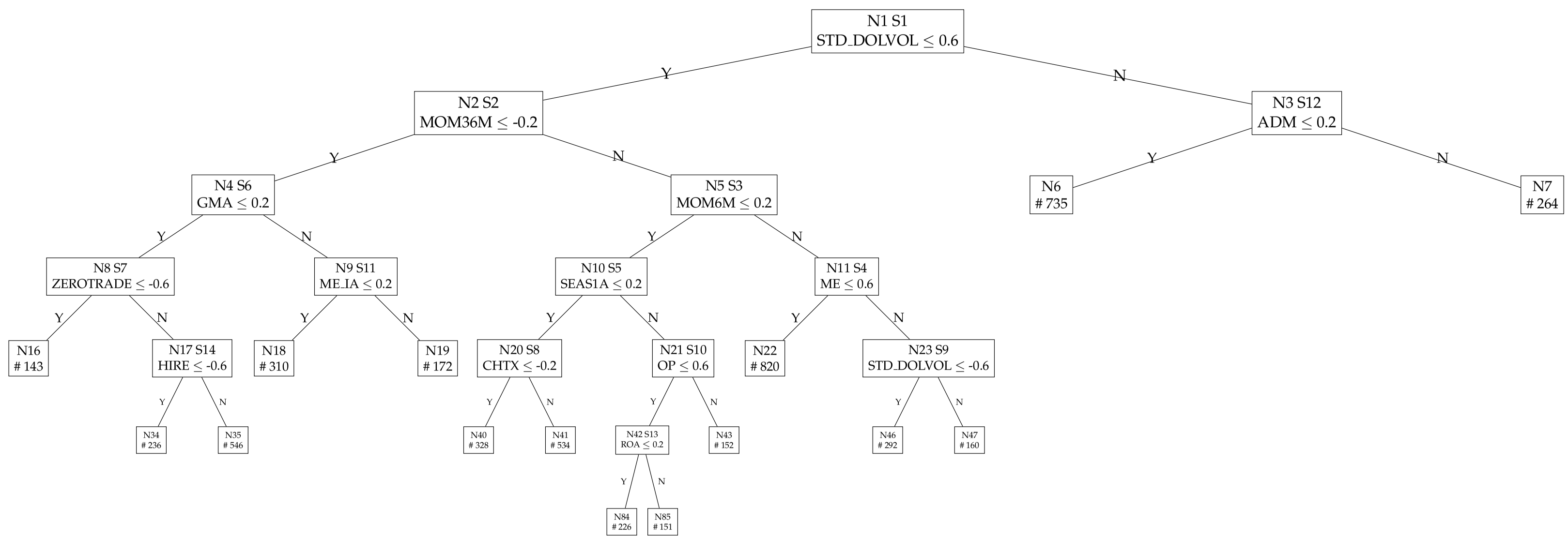


Figure 5: Partition Plots for Figure 4 (old)

This diagram illustrates the partition for the first five splits of the tree structure in Figure 4. For example, the first split (S1) is implemented with `STD_DOLVOL` on the entire stock universe, and the second split is implemented with `ME_IA` on the high `STD_DOLVOL` portfolios. The space area for each partition represents the corresponding proportion of the stock universe. We also provide the monthly average return and annualized Sharpe ratio for each leaf. The overlaid arrows represent the next split is implemented on the partitioned area from the previous partition.

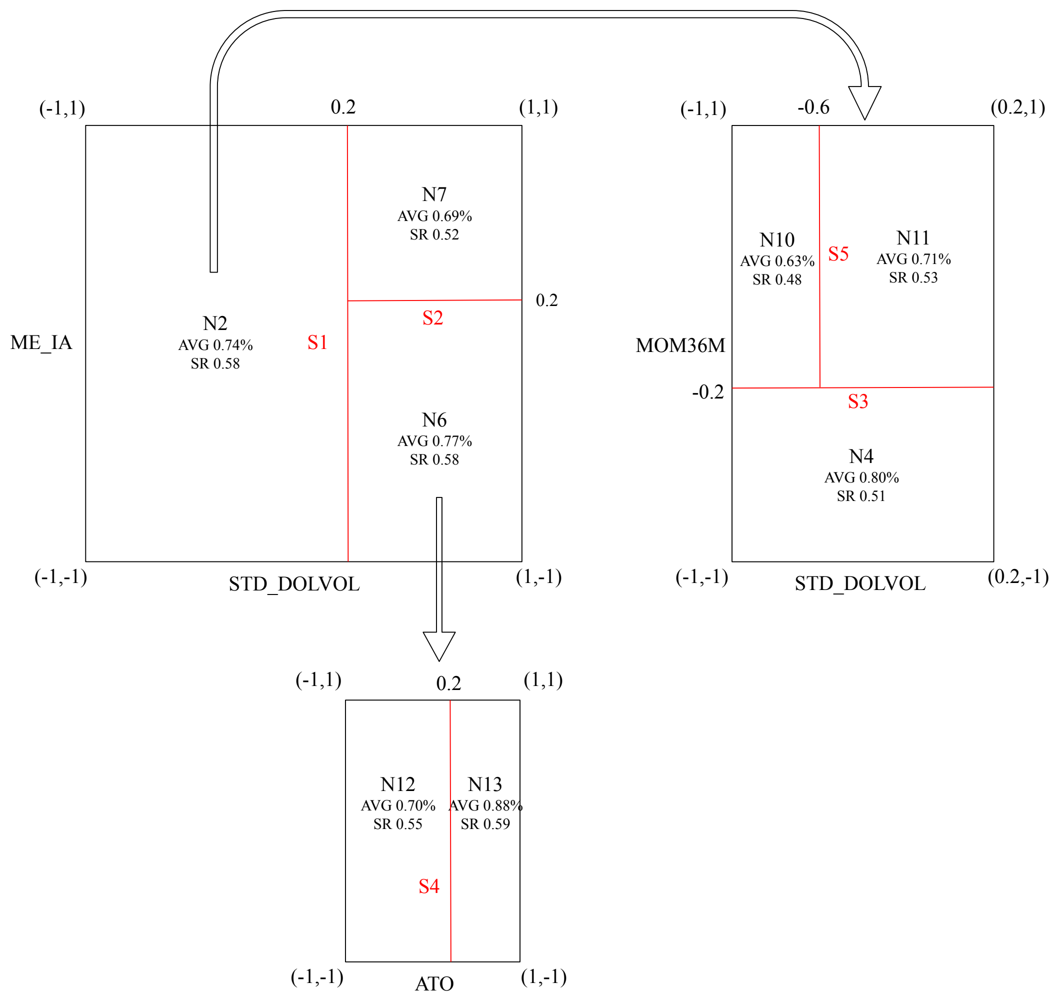


Figure 6: Market-Adjusted P-Tree for the Period from 1981 to 2000 (new)

The market-adjusted P-Tree trained from the period from 1981 to 2000, with minimum leaf size as 50, is displayed in this figure. The figure format follows Figure 4.

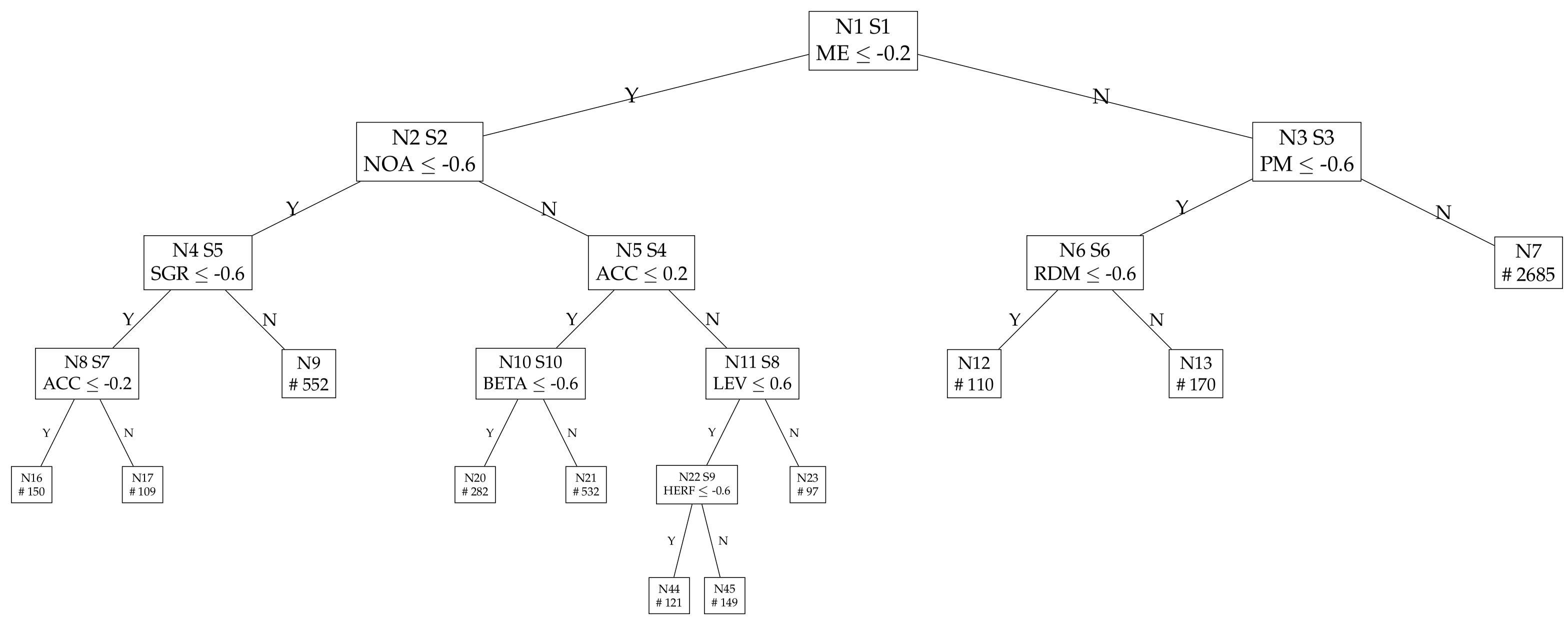
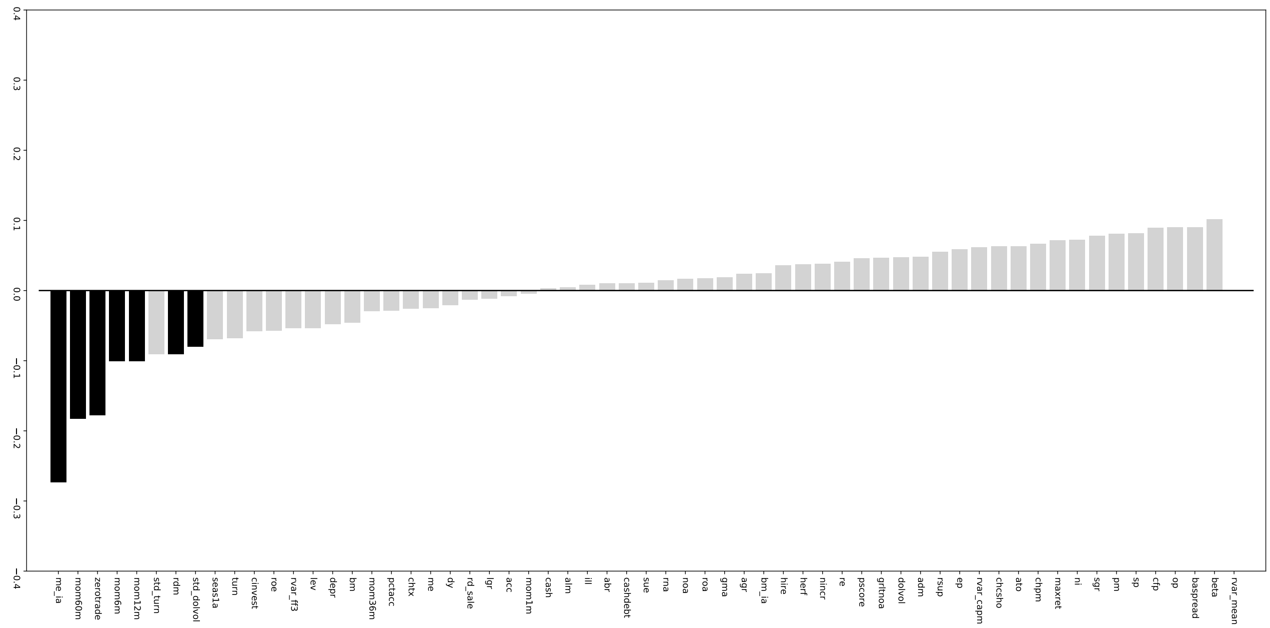
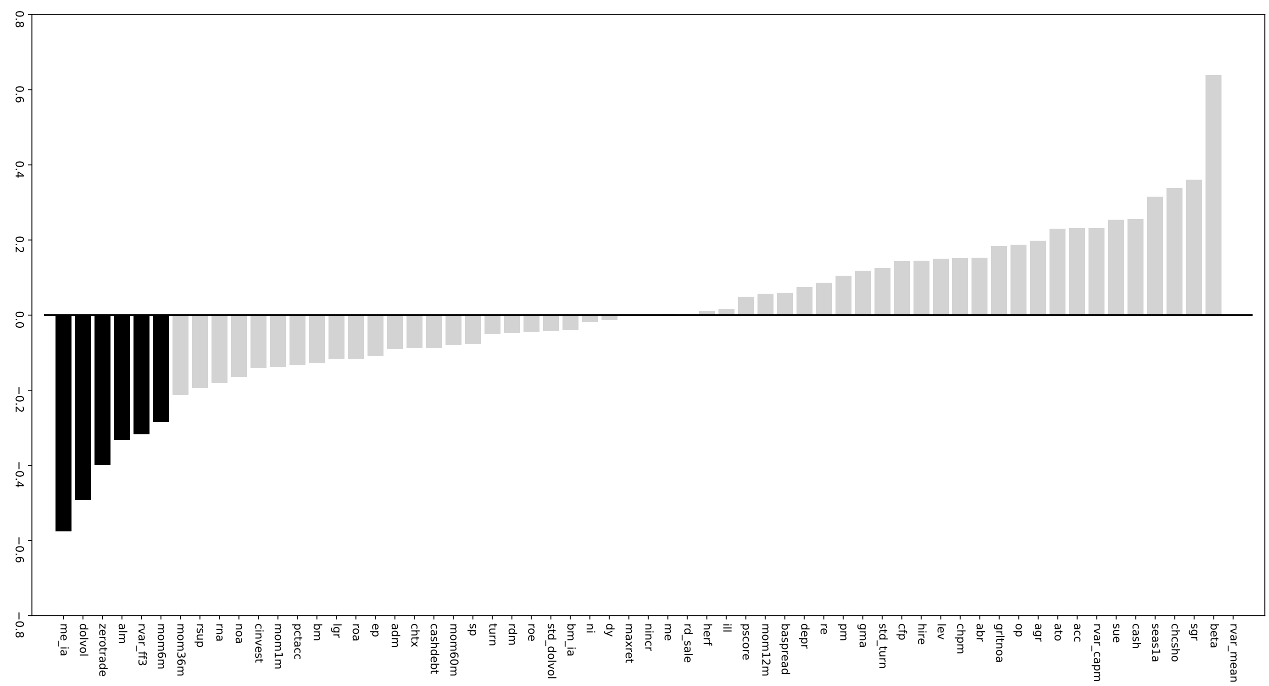


Figure 7: Out-of-Bag Characteristics Significance (new)

This figure reports the characteristic significance by the out-of-bag ensembles from the Random P-Tree Forest of 500 trees. The training sample period is 1981-2000, and the testing sample period is 2001-2020. The variable importance measure is defined as the average increase percentage of loss function by including a characteristic in a tree model. A negative value implies that including this characteristic reduces loss and is useful. The dark color bars on the left are significant characteristics at the 5% level by the two-sample t-test. The description of characteristics are listed in Table A.1.



1981-2000



2001-2020

Figure 8: Panel Tree for the Period from 1981 to 2000: High/Low Inflation (new)

This figure shows the panel tree by considering both cross-sectional and time-series variation. The most important macro predictor is inflation, and the first split is implemented when the current inflation level is lower than the median of the past decade. For high and low inflation periods, two different tree models are provided as two child leaves. The figure format follows Figure 4.

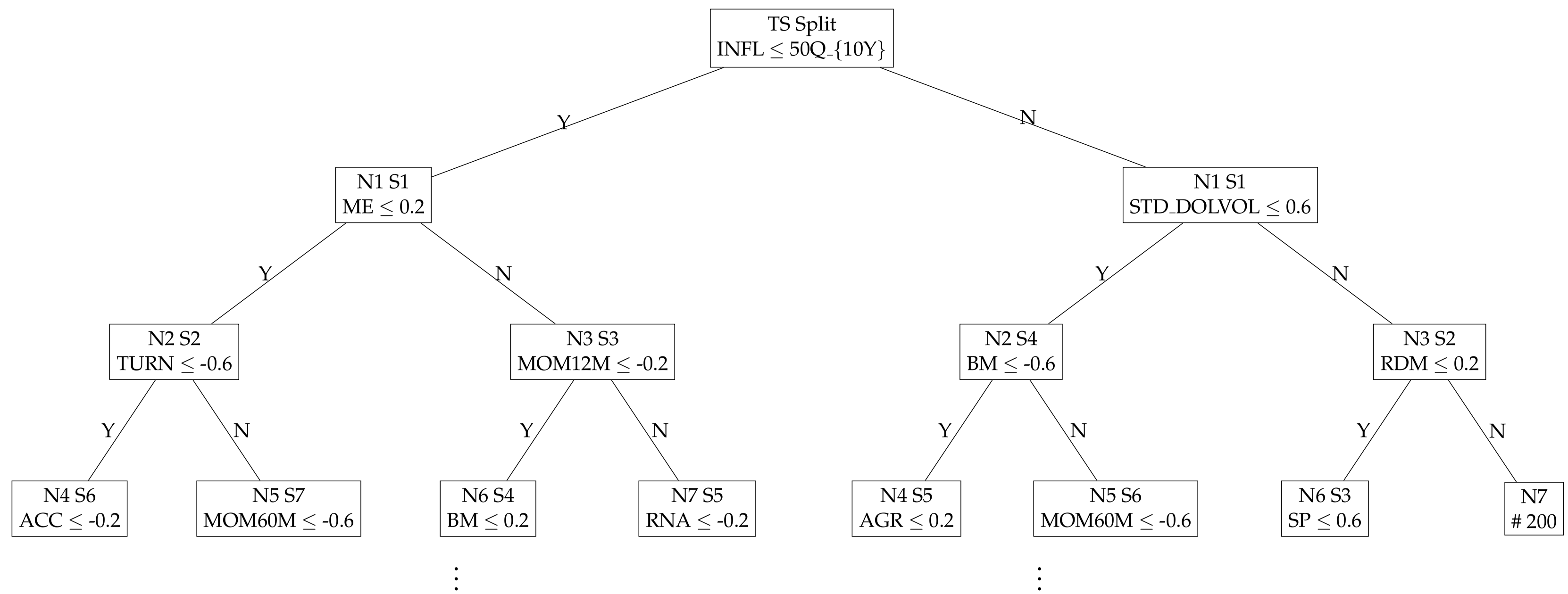
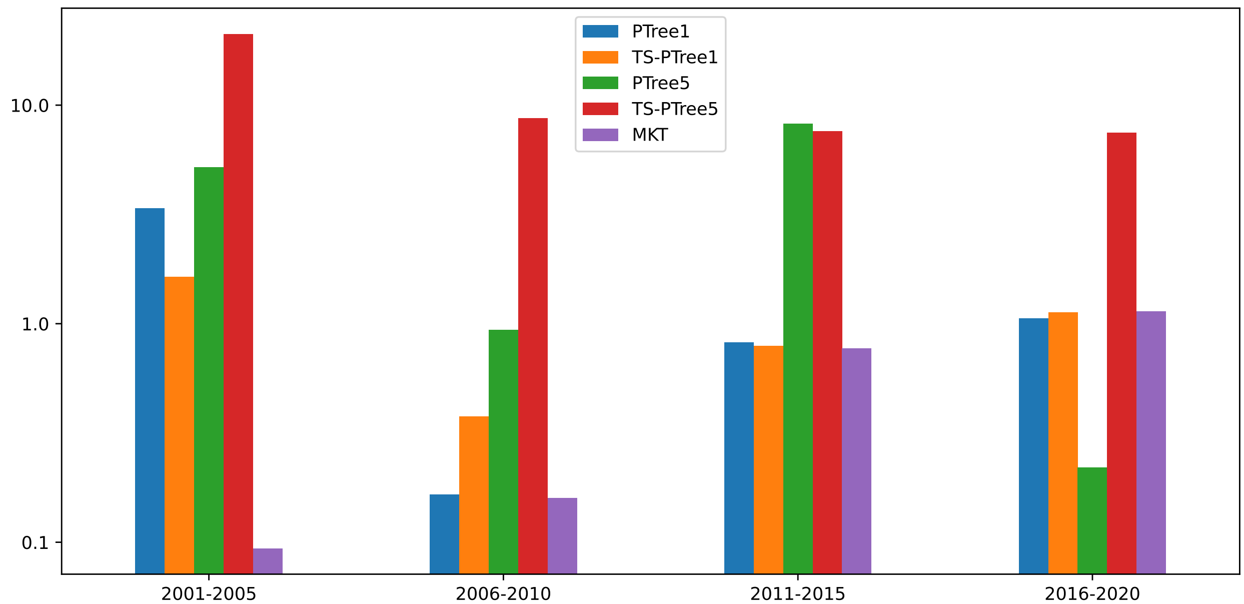
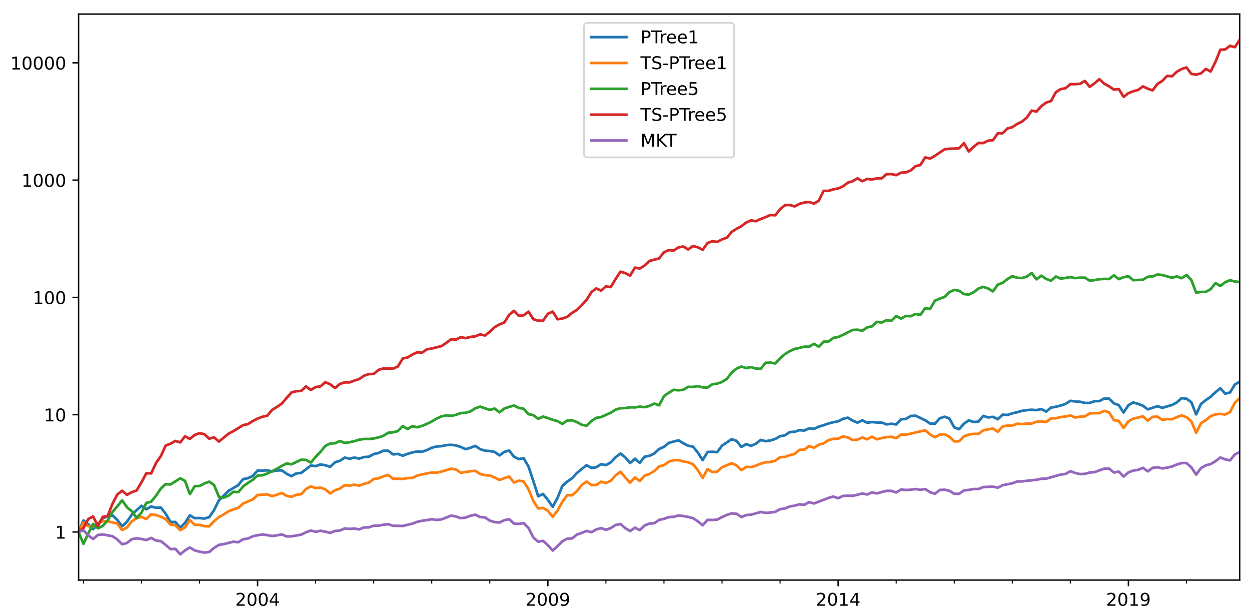


Figure 9: Subsample Robustness: Investing P-Tree Factors (old)

This figure reports *out-of-sample* investment performance for single P-Tree factors and efficient portfolios of five-factor models. The label "PTree1" and "PTree5" corresponds to PTree1 and PTree5 in Table 2 Panel A, "TS-PTree1" and "TS-PTree5" corresponds to TS-Ptree1 and TS-PTree5 in Panel D, and "MKT" is the market portfolio. With the log scale y-axis, the top figure shows the bar plots of five-year cumulative returns, and the bottom one plots the cumulative returns over the 20-year test sample.



Five-Year Subsample Returns



Cumulative Returns

Table 1: Asset Pricing Performance

This table reports the asset pricing performances for factor models. The 'Tot' (Total R^2 %) and 'Pred' (Predictive R^2 %) in Equation (20) and (21) 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. Panel A shows results for the panel tree model with #factors, and Panel B shows the market-adjusted P-Tree model. Panels C provides comparisons for benchmark models introduced in Section 4.1. Panels D and E demonstrate the same performance when applying the time-series splits. We also report Cross-Sectional R^2 % in Equation (22), using the factor models in the rows to price the test asset portfolios in the columns. Cross-Sectional R^2 values are estimated using the entire sample. "P-Tree23" indicates the 23 basis portfolios in Figure 4, and "Ma-PT18" indicates the 18 basis portfolios in Figure 6.

		Individual Stocks						Portfolios			
		In-Sample			Out-of-Sample						
	mLS	Tot	PE	Pred	Tot	PE	Pred	FF25	Ind49	P-Tree23	Ma-PT18
Panel A: Panel Tree Factors											
PTree1	50	9.14	1.70	-0.10	10.29	5.64	0.26	89.6	83.5	92.8	-83.1
PTree3		11.58	24.67	0.25	12.71	12.97	-0.10	93.5	78.3	90.8	36.5
PTree5		12.08	24.68	0.36	13.24	14.91	0.05	96.9	81.7	86.2	79.5
Panel B: MKT Adj. Panel Tree Factors											
MaPTree2	50	10.70	20.06	0.20	11.14	8.57	-0.01	90.6	91.2	89.9	39.8
MaPTree5		12.27	21.79	0.38	13.74	15.69	0.10	90.4	88.7	88.7	64.5
Panel C: Time-Series split - Panel Tree Factors											
TS-PTree1	50	10.85	14.44	0.40	12.01	13.08	0.33	88.9	81.0	92.9	53.7
TS-PTree3		13.47	29.55	1.06	13.24	14.58	0.36	76.9	62.0	77.5	31.4
TS-PTree5		14.65	32.44	1.55	14.53	16.80	0.54	69.2	41.0	61.6	58.0
Panel D: Time-Series split - MKT Adj. Panel Tree Factors											
TS-MaPTree2	50	12.38	30.35	0.74	12.78	12.30	-0.43	75.9	85.8	81.7	45.7
TS-MaPTree5		14.32	32.35	1.54	14.11	17.88	0.51	93.5	65.4	82.8	73.7
Panel E: Other Benchmark Models											
CAPM		7.07	1.57	-0.05	8.26	0.71	0.23	91.4	88.1	92.8	-79.1
FF3		10.10	7.26	0.14	10.54	4.73	0.29	94.9	85.4	87.5	-58.4
FF5		10.50	11.82	0.21	11.10	4.81	0.28	96.1	78.5	93.0	44.0
Q5		10.36	17.26	0.38	11.36	5.67	0.18	96.1	88.7	94.3	90.2
RP-PCA3		11.10	18.53	0.20	11.71	9.55	0.11	87.3	71.9	79.2	34.2
IPCA3	5z	12.21	20.09	0.23	13.51	14.75	0.17	-42.8	-95.8	16.7	-608.2
	61z	14.87	26.42	1.27	16.47	17.43	0.54				

Asset Pricing Performance, Continue

		Individual Stocks						Portfolios			
		In-Sample			Out-of-Sample						
	mLS	Tot	PE	Pred	Tot	PE	Pred	FF25	Ind49	P-Tree23	Ma-PT18
Panel F: Panel Tree Factors minLeafSize20											
PTree1	20	9.33	1.85	-0.10	10.07	4.78	0.25	89.9	85.2	93.9	-310.1
PTree3		11.99	22.48	0.25	13.73	18.23	0.09	80.1	45.8	65.3	7.5
PTree5		12.61	25.85	0.41	14.67	17.88	0.07	66.3	32.5	69.1	26.1
Panel G: Panel Tree Factors minLeafSize100											
PTree1	100	9.06	2.02	-0.09	9.79	4.83	0.25	90.4	85.3	95.8	-45.5
PTree3		11.72	26.69	0.25	12.90	15.18	-0.20	93.7	85.0	91.7	39.8
PTree5		12.33	26.09	0.41	13.72	16.96	0.05	94.5	72.5	87.1	85.3
Panel H: MKT Adj. Panel Tree Factors minLeafSize20											
MaPTree2	20	7.08	1.65	-0.05	8.20	0.72	0.23	89.6	91.4	82.2	16.2
MaPTree5		12.00	21.74	0.39	13.45	14.85	0.12	88.6	87.5	88.0	55.6
Panel I: MKT Adj. Panel Tree Factors minLeafSize100											
MaPTree2	100	10.71	20.30	0.20	11.23	8.75	-0.01	89.4	91.7	88.6	71.7
MaPTree5		12.06	22.84	0.37	13.80	14.07	0.09	86.1	89.1	87.1	96.5

Table 2: Investment Performance of Factors (new)

This table reports the investment performance of the equity factor models. The table format is similar to Table ??. We report the monthly average return and CAPM alpha (%), and the annualized Sharpe ratio for the equal-weighted ($1/N$) and mean-variance efficient (MVE) portfolios. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

		In-Sample (1981-2000)						Out-of-Sample (2001-2020)					
		$1/N$			MVE			$1/N$			MVE		
	mLS	AVG	SR	α	AVG	SR	α	AVG	SR	α	AVG	SR	α
Panel A: Panel Tree Factors													
PTree1	50	0.89	0.55	0.10	0.89	0.55	0.10	0.81	0.44	-0.02	0.81	0.44	-0.02
PTree3		3.68	1.97	3.02***	3.95	2.06	3.39***	1.30	0.62	1.17**	1.32	0.59	1.30***
PTree5		2.99	1.70	2.27***	3.76	2.53	3.47***	0.88	0.50	0.55	1.78	0.77	2.01***
Panel B: MKT Adj. Panel Tree Factors													
MaPTree2	50	1.57	1.00	0.87***	2.24	1.05	1.56***	0.42	0.32	-0.08	0.24	0.14	-0.14
MaPTree5		1.81	1.30	1.10***	3.64	2.35	3.31***	0.89	0.79	0.48***	1.92	0.94	2.13***
Panel C: Time-Series split - Panel Tree Factors													
TS-PTree1	50	1.74	0.82	1.64***	1.74	0.82	1.64***	0.96	0.27	1.51**	0.96	0.27	1.51**
TS-PTree3		3.08	1.96	2.77***	4.41	2.29	4.01***	2.21	0.93	2.66***	4.43	1.51	4.78***
TS-PTree5		6.99	0.35	8.29***	9.01	3.78	8.64***	4.19	0.16	0.54	9.10	2.49	9.27***
Panel D: Time-Series split - MKT Adj. Panel Tree Factors													
TS-MaPTree2	50	6.73	1.94	6.00***	9.02	1.88	8.50***	1.19	0.38	1.37**	2.01	0.38	2.76**
TS-MaPTree5		4.46	2.49	3.95***	6.36	3.81	6.03***	1.97	1.52	1.90***	4.01	2.10	4.21***
Panel E: Other Benchmark Models													
FF3		0.38	0.85	0.20***	0.53	1.16	0.40***	0.28	0.40	0.01	0.22	0.30	-0.06
FF5		0.38	1.34	0.33***	0.45	1.48	0.38***	0.25	0.59	0.12	0.27	0.64	0.13*
Q5		0.63	2.10	0.53***	0.77	2.78	0.74***	0.31	1.10	0.25***	0.34	1.22	0.34***
RP-PCA3		14.85	1.27	8.42***	10.11	1.52	7.56***	6.50	0.88	3.05***	2.35	0.53	3.16***
IPCA3		1.91	2.25	2.26***	3.15	5.54	3.13***	0.84	0.88	1.22***	2.16	3.40	2.10***
Panel F: Panel Tree Factors MinLeafSize20													
PTree1	20	0.81	0.54	0.03	0.81	0.54	0.03	0.82	0.52	0.10	0.82	0.52	0.10
PTree3		2.32	1.70	1.74***	3.14	1.87	2.69***	1.33	0.88	0.82***	1.69	0.90	1.26***
PTree5		3.60	2.20	2.91***	5.12	3.21	4.82***	1.45	0.96	1.24***	2.89	1.22	2.98***
Panel G: MKT Adj. Panel Tree Factors MinLeafSize20													
MaPTree2	20	1.57	1.00	0.87***	2.23	1.05	1.56***	0.50	0.38	0.02	0.38	0.22	0.04
MaPTree5		1.80	1.30	1.10***	3.63	2.35	3.30***	0.92	0.82	0.52***	2.00	0.98	2.22***
Panel H: Panel Tree Factors MinLeafSize100													
PTree1	100	0.79	0.54	0.04	0.79	0.54	0.04	0.77	0.48	0.05	0.77	0.48	0.05
PTree3		4.78	2.05	4.07***	5.08	2.14	4.49***	1.33	0.48	1.47**	1.38	0.46	1.65**
PTree5		3.56	2.07	2.94***	4.70	3.01	4.46***	1.10	0.61	0.98**	2.11	0.77	2.70***
Panel I: MKT Adj. Panel Tree Factors MinLeafSize100													
MaPTree2	100	1.54	0.99	0.85***	2.20	1.05	1.53***	0.53	0.41	0.05	0.44	0.26	0.08
MaPTree5		1.70	1.31	1.01***	4.31	2.45	3.91***	1.05	0.98	0.62***	2.61	1.18	2.65***

Table 3: Factor Spanning Alpha Test (old)

This table reports the monthly alphas in basis points and statistical significance for the equity factor spanning test. Specifically, we regress the trading strategies (portfolios) in the rows against the factor models in the column. "FF5" is the five-factor model in Fama and French (2015), and "Q5" is the five-factor model in Hou et al. (2021). Panel A shows results for each factor of the panel-tree model in Table ??, and Panel B shows the market-adjusted tree factors. We regress each factor in the rows against a factor model in the columns. We name the factors by the first two splitting characteristics in the tree structure for the P-Tree factor to emphasize the interactions. We also show the five factors' mean-variance efficiency (MVE) and $1/N$ portfolios. Notably, "FF25" and "INF49" are the value-weighted 25 portfolios on size and value of stocks and the 49 industry portfolios. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	In-Sample		Out-of-Sample		Entire-Sample	
	FF5	Q5	FF5	Q5	FF5	Q5
Panel A: Panel Tree Factors						
P-Tree1	54***	54***	61***	60***	54***	60***
OP-ME	134***	91*	69*	52	98***	71**
BM_IA-DOLVOL	34	57*	115***	136***	75***	115***
MOM12M-ME	450***	371***	177***	142***	321***	224***
ME-STD_DOLVOL	33**	11	-10	-35**	8	-18
MVE(5-factor)	349***	303***	192***	183***	276***	229***
$1/N$(5-factor)	141***	117***	83***	71***	111***	90***
Panel B: MKT Adj. Panel Tree Factors						
RVAR_FF3-STD_TURN	377***	265***	49	12	196***	120***
BM_IA-ME	51***	77***	106***	114***	83***	109***
MOM12M-ME_IA	217***	162***	62	25	140***	56*
ME-CASH	118***	94***	21	3	68***	43***
MVE(5-factor)	409***	357***	213***	182***	309***	258***
$1/N$(5-factor)	153***	120***	47***	31***	97***	65***
Panel C: Other Test Assets						
MVE-FF25	270***	213***	93***	77**	181***	141***
MVE-IND49	79	101	63	45	74*	81
MVE-PT23	54***	54***	61***	60***	54***	60***
MVE-Ma-PT18	377***	265***	49	12	196***	120***
$1/N$-FF25	-8***	-9	3	7*	-2	4
$1/N$-IND49	65*	30	-2	10	26	37*
$1/N$-PT23	-5	-3	16***	22***	6	13***
$1/N$-MaPT18	-74***	-56***	-17*	-8	-46***	-25***

Table 4: Uni-Sort Factors vs. Interaction Factors

This table summarizes the significant counts for long-short factors for average returns and Jensen's alphas. We count the number of significant average returns and alphas at the 5% and 10% level in Panels A and B. We report the cross-sectional quantile values of average returns and alphas in Panel C among the 61 characteristics. The four specifications are (1) 4x1 long-short portfolio, (2) 2x1 long-short portfolio, (3) interaction factors, and (4) market-adjusted interaction factors. Our panel tree creates specifications (3) and (4), which follow the same models in Table ??.

Panel A: # of Significant Cases at 5% level

	Uni-Sort 4x1		Uni-Sort 2x1		Interaction		Mkt-Adj Interaction	
	# Mean	# Alpha	# Mean	# Alpha	# Mean	# Alpha	# Mean	# Alpha
81-00	17	30	6	22	11	27	10	30
01-20	4	22	5	11	24	28	11	21
81-20	22	33	12	28	32	31	16	42

Panel B: # of Significant Cases at 10% level

	Uni-Sort 4x1		Uni-Sort 2x1		Interaction		Mkt-Adj Interaction	
	# Mean	# Alpha	# Mean	# Alpha	# Mean	# Alpha	# Mean	# Alpha
81-00	26	34	10	31	18	30	19	39
01-20	8	27	7	18	29	30	14	33
81-20	27	36	18	34	36	41	21	48

Panel C: Cross-Sectional Quantiles for Average and Alpha

		Uni-Sort 4x1		Uni-Sort 2x1		Interaction		Mkt-Adj Interaction	
	q	Avg	Alpha	Avg	Alpha	Avg	Alpha	Avg	Alpha
	25	0.12	0.06	0.07	0.06	0.09	0.09	0.1	0.22
81-00	50	0.33	0.46	0h.18	0.21	0.22	0.28	0.19	0.29
	75	0.61	0.69	0.23	0.31	0.33	0.43	0.27	0.41
	25	0.02	0.07	-0.02	0.05	0.06	0.08	-0.02	0.13
01-20	50	0.18	0.29	0.06	0.13	0.41	0.36	0.09	0.25
	75	0.36	0.58	0.2	0.31	0.56	0.65	0.25	0.41
	25	0.13	0.14	0.04	0.08	0.08	0.12	0.05	0.17
81-20	50	0.28	0.34	0.11	0.16	0.30	0.29	0.14	0.28
	75	0.41	0.61	0.19	0.26	0.43	0.48	0.25	0.39

Table 5: Examples of Interaction Factors

This table follows the 6 examples in Figure A.5. We report the monthly average returns (%) and Jensen's alpha (%) of the 4x1 long-short factors and the interaction factors. The interaction factors are created with the train sample period from 1981 to 2000, and we have provided the corresponding interaction characteristics. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The description of characteristics are listed in Table A.1.

	Panel A: Stand. Unexp. Earnings				Panel B: Profit Margin			
	Uni-Sort 4x1		+RDM+ME		Uni-Sort 4x1		+RDS+DOLVOL	
	Mean	Alpha	Mean	Alpha	Mean	Alpha	Mean	Alpha
81-00	0.68***	0.63***	0.95***	0.91***	0.10	0.06	0.39***	0.46***
01-20	0.53***	0.67***	1.43***	1.36***	0.18	0.28*	0.55***	0.56***
81-20	0.61***	0.66***	1.19***	1.13***	0.14	0.17	0.47***	0.51***

	Panel C: Cash Holdings				Panel D: Dollar Trading Volume			
	Uni-Sort 4x1		+NOA+BAS		Uni-Sort 4x1		+RVAR_FF3+ME	
	Mean	Alpha	Mean	Alpha	Mean	Alpha	Mean	Alpha
81-00	0.39	0.05	1.12***	1.10***	-0.20	0.02	0.11	0.33**
01-20	0.43*	0.21	-0.04	0.02	0.33*	0.42**	0.34*	0.45***
81-20	0.41**	0.14	0.54***	0.56***	0.06	0.22	0.23*	0.39***

	Panel E: Seasonality				Panel F: R&D to Sales			
	Uni-Sort 4x1		+RDM+DOLVOL		Uni-Sort 4x1		+RVAR_FF3+EP	
	Mean	Alpha	Mean	Alpha	Mean	Alpha	Mean	Alpha
81-00	0.54**	0.37*	0.13	0.21	0.43	0.08	1.11***	1.29***
01-20	0.01	0.05	0.47**	0.46**	-0.06	-0.28	0.04	0.49*
81-20	0.27	0.22	0.30*	0.33**	0.19	-0.10	0.58***	0.90***

Appendices

Table A.1: Equity Characteristics (61 in total)

This table lists the description for characteristics used in the empirical study.

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

Continue: Equity Characteristics (61 in total)

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

Table A.2: Macro Predictors for Market Timing

This table lists the description for macro predictors used in the empirical study.

No.	Variable Name	Description
1	EP	Earnings-to-price of S&P 500 over past 12 months
2	DY	Dividend yield of S&P 500 over past 12 months
3	LEV	Leverage of S&P 500 over past 12 months
4	NI	Net equity issuance of S&P 500 over past 12 months
5	SVAR	Stock Variance of S&P 500 over past 12 months
6	ILL	Pastor-Stambaugh illiquidity aggregate level
7	INFL	Annual inflation change rate
8	TBL	Three-month treasure bill rate
9	DFY	Default yield BAA and AAA corporate bond yields spread
10	TMS	Term spread: BAA and AAA corporate bond yields spread

Table A.3: Characteristics Importance by Top Splits (old)

This table reports the most frequently selected characteristics from the Random P-Tree Forest of 500 trees, which can be used to assess the variable importance. The Random P-Tree Forest ensemble design is discussed in section 3.3. The "Top 1" rows only count the first split for 500 trees. The "Top 2" or "Top " rows only count the first two or three splits. The numbers reported are the selection frequency for these top characteristics selected out of the 500 ensembles. The description of characteristics is listed in Table A.1.

Panel A: Equity Entire Training Sample (1981-2000)

	1	2	3	4	5
Top1	STD_DOLVOL	MO60M	DOLVOL	ME_IA	LGR
	0.73	0.64	0.61	0.33	0.29
Top2	STD_DOLVOL	DOLVOL	MMO60M	ME_IA	BETA
	0.78	0.73	0.81	0.61	0.44
Top3	STD_DOLVOL	MMO60M	DOLVOL	ME_IA	ATO
	0.78	0.76	0.73	0.64	0.52

	Panel B: Equity High Inflation			Panel C: Equity Low Inflation		
	1	2	3	1	2	3
Top 1	MOM60M	DOLVOL	ME_IA	ME_IA	DOLVOL	ME
	0.53	0.44	0.42	0.56	0.42	0.41
Top 2	MOM60M	DOLVOL	ME_IA	ME_IA	ME	DOLVOL
	0.59	0.56	0.53	0.68	0.55	0.49
Top 3	MOM60M	DOLVOL	ME_IA	ME_IA	MOM60M	ME
	0.60	0.56	0.53	0.63	0.50	0.47

Figure A.1: Return-Risk Clustering for Leaf Basis Portfolios (old)

Following the tree structure in Figure 4, this figure plots the monthly average return and standard deviation for leaf basis portfolios. Labels N4, N5, N6, and N7 represent those four leaf basis portfolios at depth 3. Labels N4+, N5+, N6+, and N7+ represent corresponding leaf basis portfolios at depth 6. We also show our generated SDF portfolios at depth 3 and depth 6 labeled with MVE-3 and MVE-6. For the reference lines, we have included the annualized Sharpe ratios 0.3, 0.6, and 0.9.

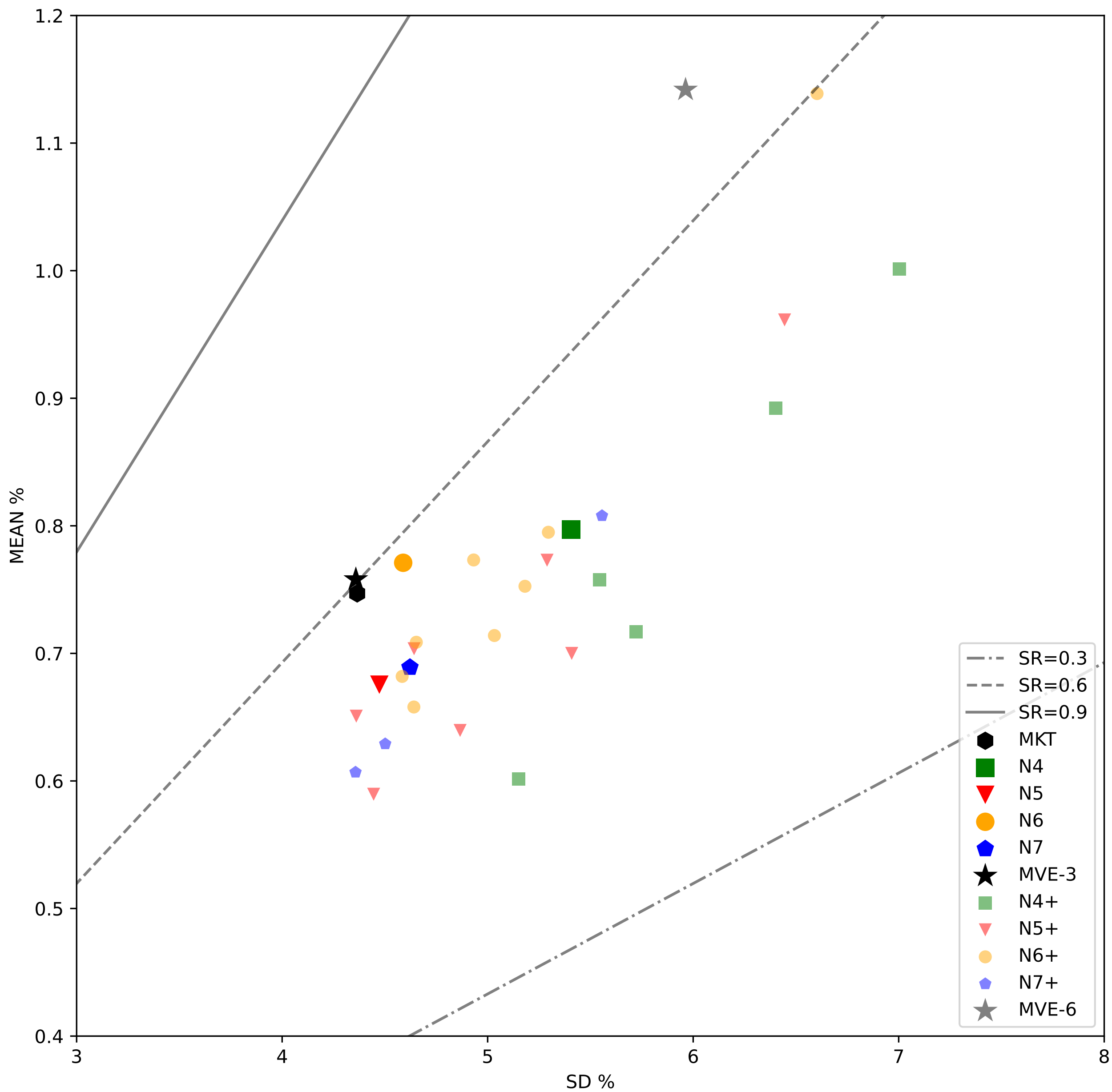


Figure A.2: Out-of-Sample: Return-Risk Clustering for Leaf Basis Portfolios (old)

This figure shows the out-of-sample performance for all those portfolios plotted in Figure A.1. The figure format follows Figure A.1.

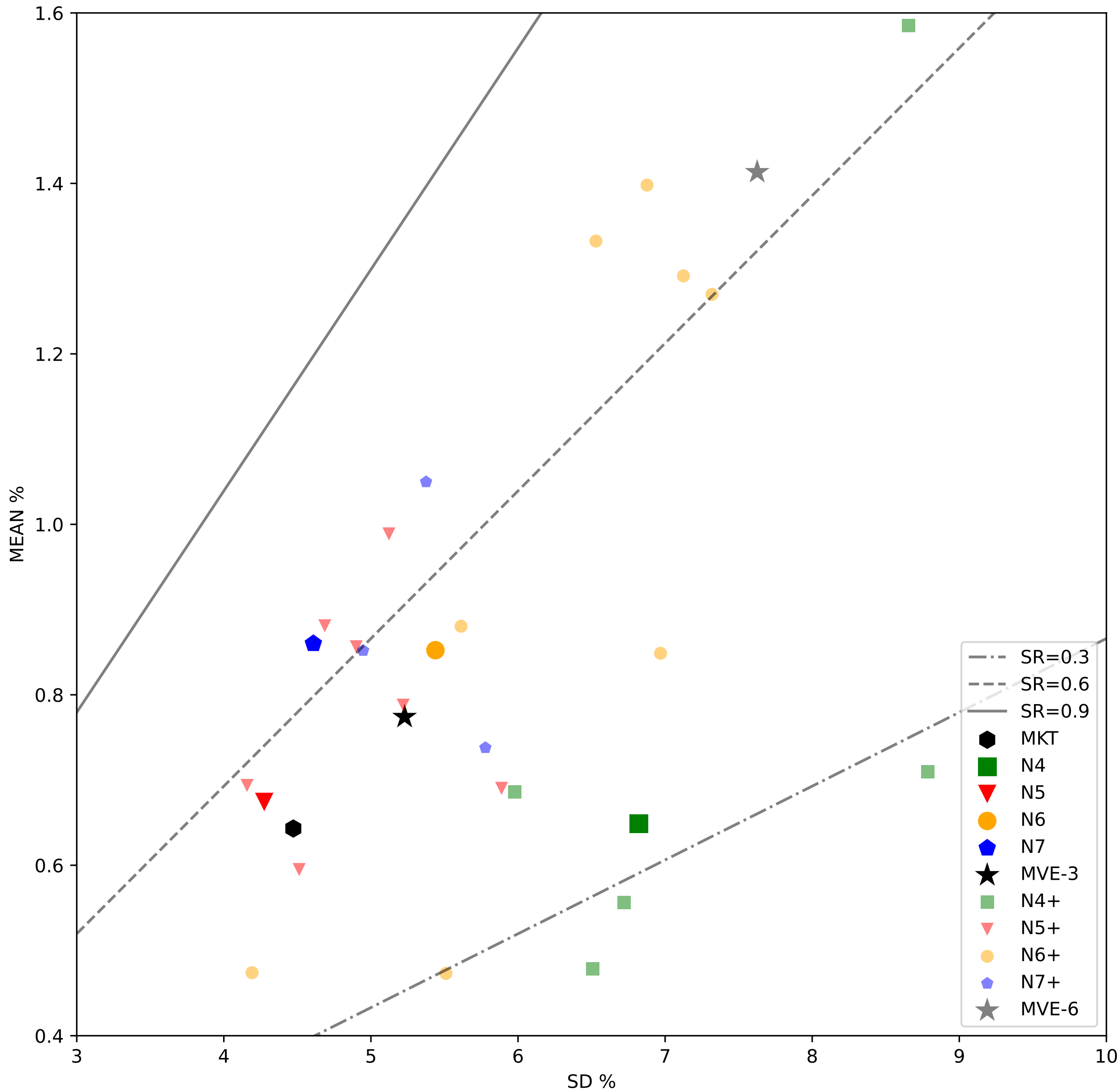
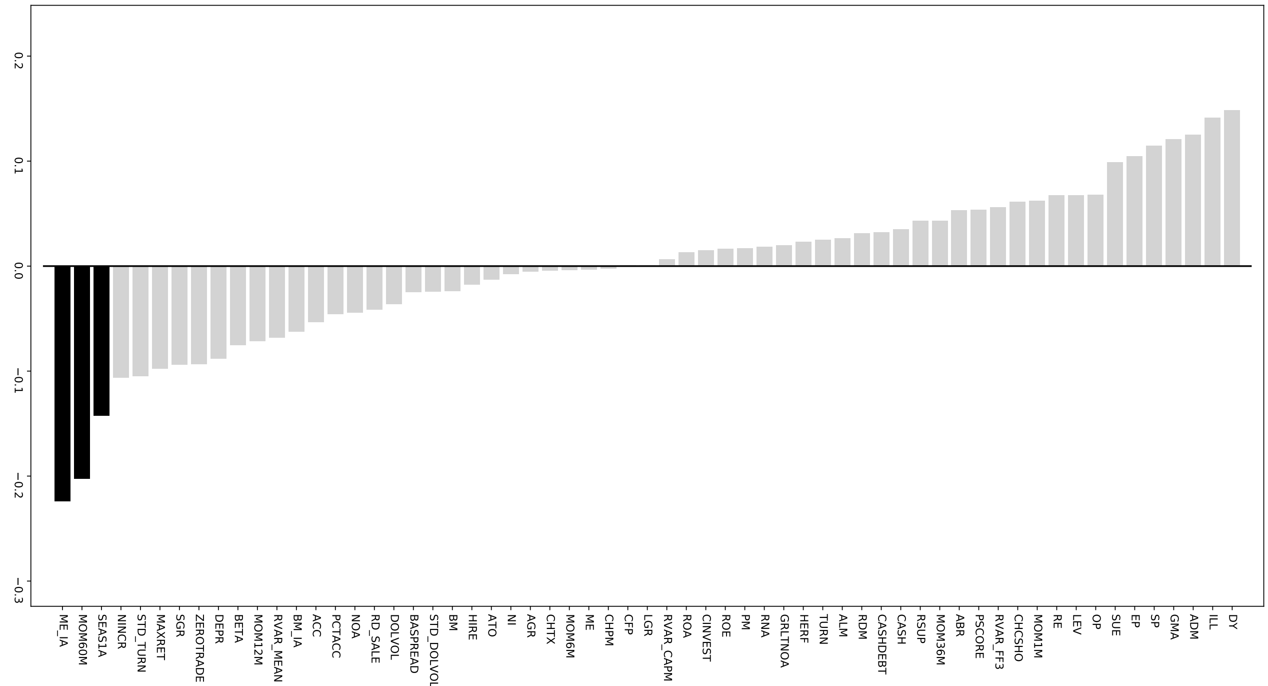
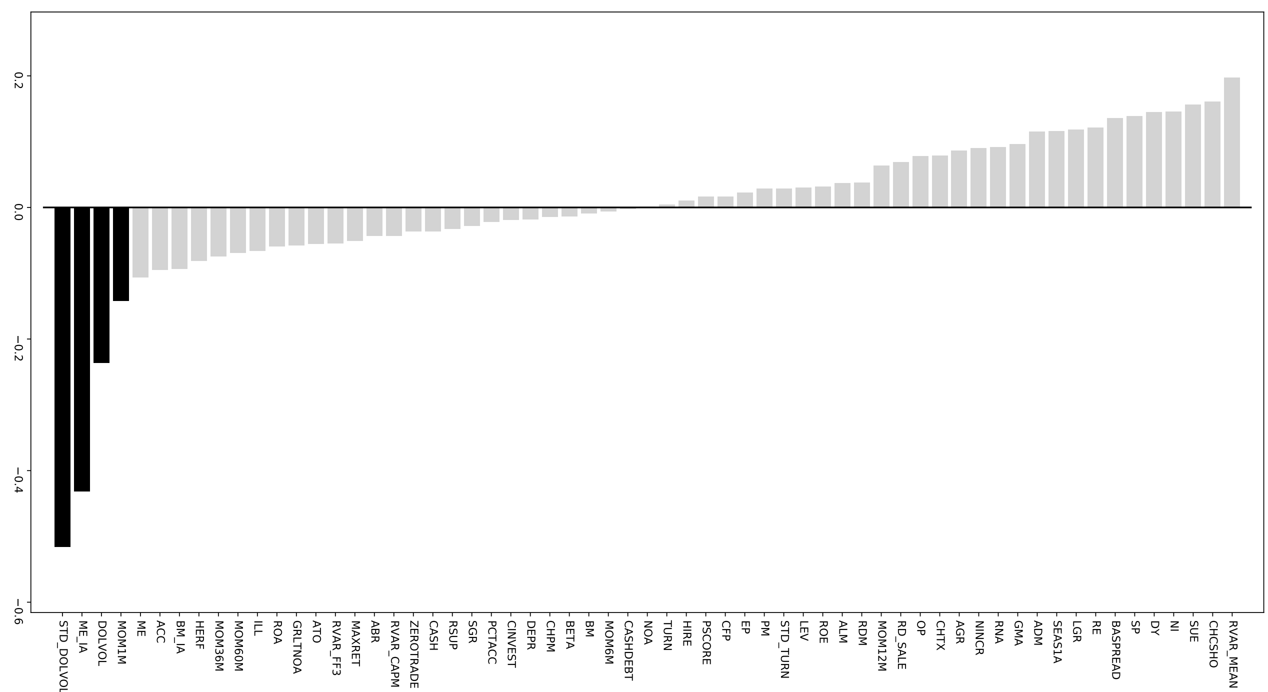


Figure A.3: Out-of-Bag Characteristics Significance: High/Low Inflation (old)

This figure reports the characteristics significance by the out-of-bag ensembles from the Random P-Tree Forest of 500 trees. We report results for high and low inflation periods in the training sample 1981-2000. This figure's details follow Figure 7. The left dark columns indicate significant characteristics.



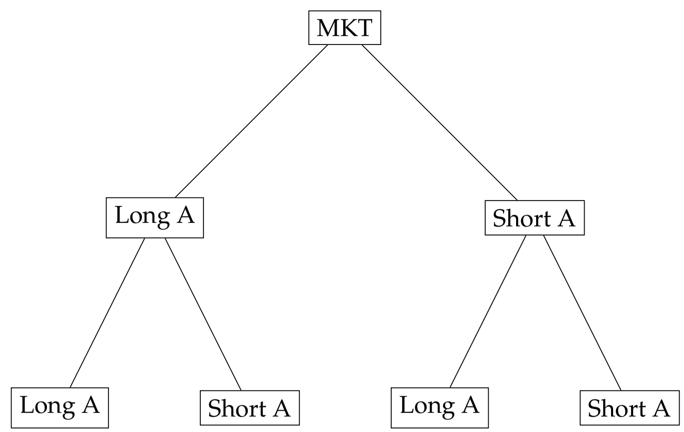
High INFL



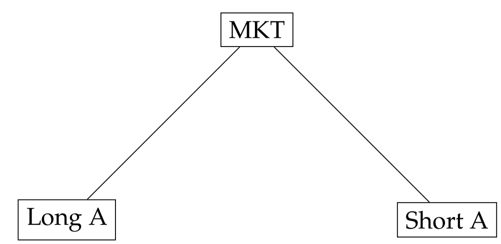
Low INFL

Figure A.4: Four Specifications of Characteristic-Sorted Factors

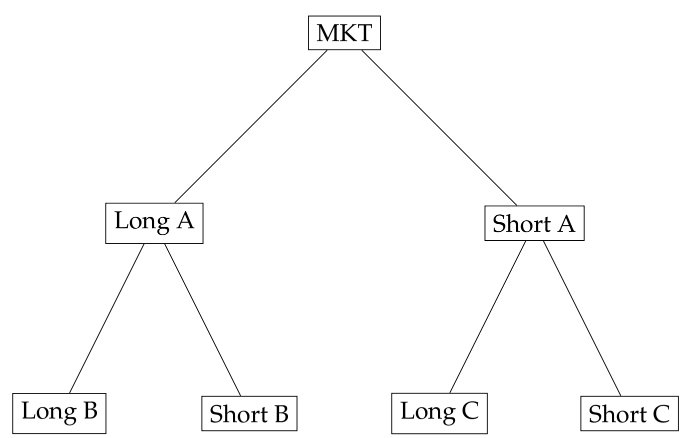
We show four different specifications for creating long-short factors by the tree model. Specification (a) is the classic sorting with a single characteristic and creates four 4x1 sorted portfolios. Specification (b) is similar and creates two 2x1 sorted portfolios. Specification (c) is trained by the panel-tree model with interactions of characteristics. To each characteristic, we fit the panel tree model to search optimal characteristics for its interaction by fixing the first split. Specification (d) is similar to (c) but fits with market-adjusted returns.



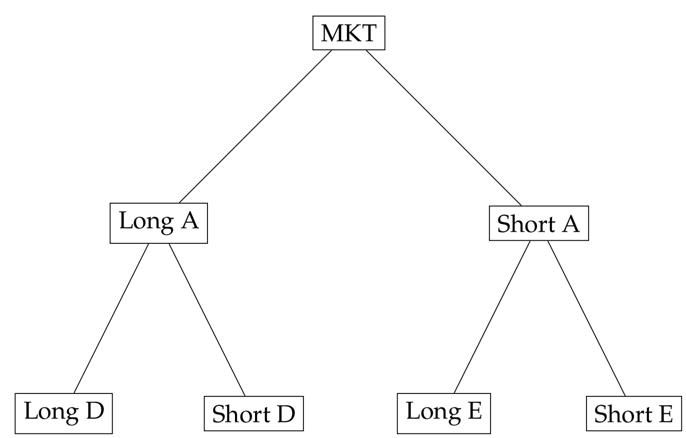
(a) Uni-Sort 4x1



(b) Uni-Sort 2x1



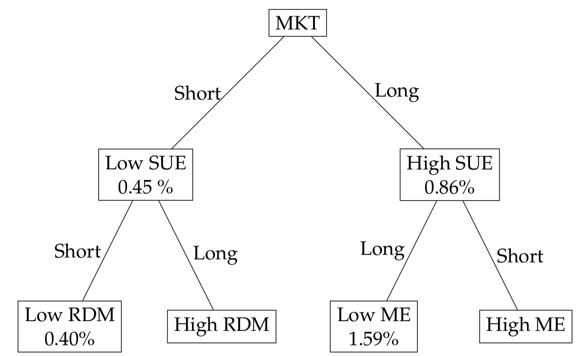
(c) Interaction



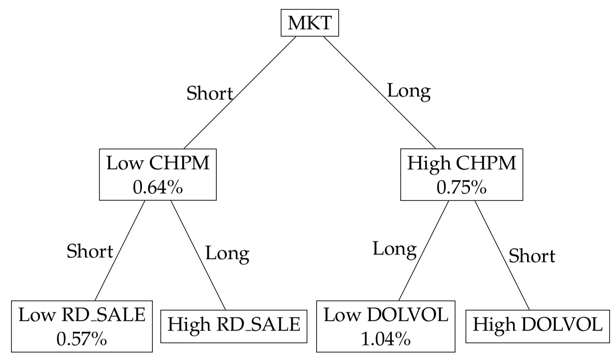
(d) Market-Adjusted Interaction

Figure A.5: Examples of Interaction Factors

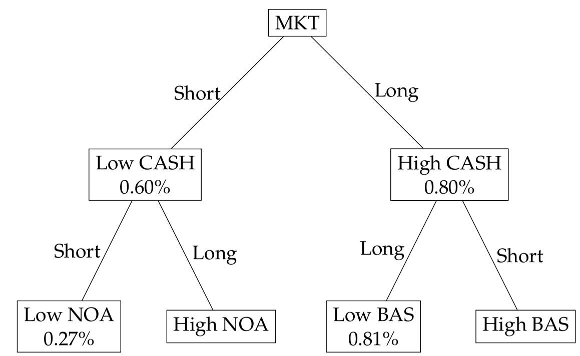
This figure shows six examples of the interaction factors. More details are reported in Table 5. In the second depth, the numbers report the average returns of the long portfolio and short portfolio. In the third depth, the numbers are the average returns of the long-long portfolio and short-short portfolio. One can see further splitting helps create a higher return spread.



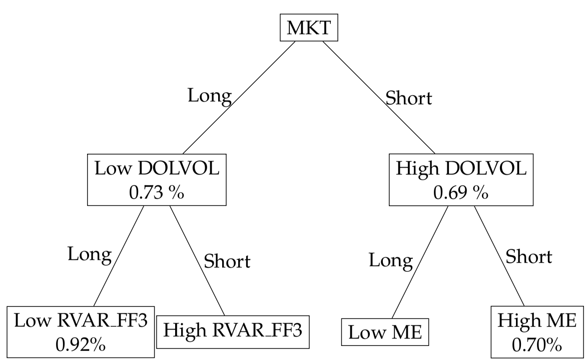
(a) Standardized Unexpected Earning



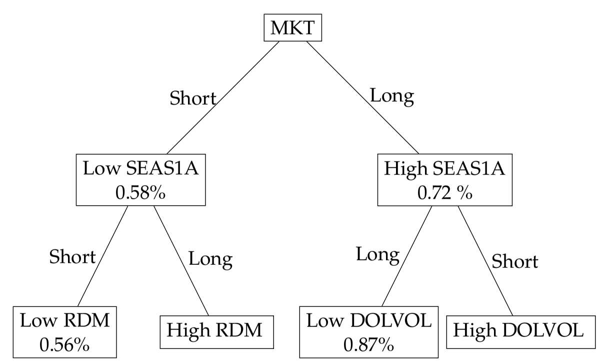
(b) Profit Margin



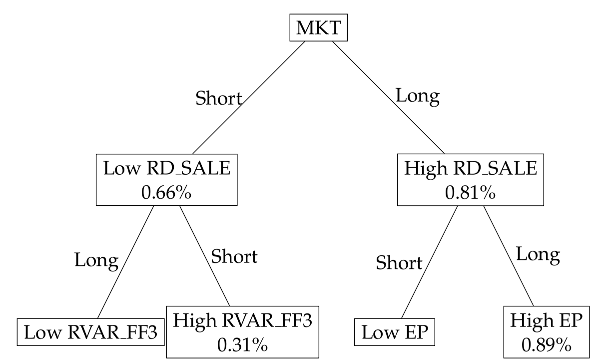
(c) Cash Holdings



(d) Dollar Trading Volume



(e) Seasonality



(f) R&D-to-Sales