Forest through the Trees: Building Cross-Sections of Stock Returns *

Svetlana Bryzgalova † Markus Pelger ‡ Jason Zhu §

October 31, 2019

PRELIMINARY AND INCOMPLETE
Please do not circulate.

ABSTRACT

We show how to build a set of basis assets that captures complex information contained in a given list of stock characteristics. Our cross-section of portfolios is a small number of long-only strategies that (a) fully reflect the information in the cross-sectional return predictors, allowing for conditional interactions and non-linearities, (b) provide a small set of interpretable test assets for evaluating asset pricing models, (c) are substantially harder to price than conventional double or triple sorted portfolios constructed from the same information set, and (d) are the building blocks for a stochastic discount factor (SDF) projected on the characteristic space. We use decision trees to generalize the concept of conventional sorting, and develop a novel approach to the robust recovery of a sparse set of the SDF basis assets. Empirically, we show that traditionally sorted portfolios and factors present a too low hurdle for candidate models as they miss the complex information structure of the original returns. Our results have important implications for evaluating asset pricing models, and modeling expected returns.

Keywords: Asset pricing, sorting, portfolios, cross-section of expected returns, decision trees, elastic net, stock characteristics, machine learning.

JEL classification: G11, G12, C55, C58.

---

*We thank Caio Almeida (discussant), Victor DeMiguel, Kay Giesecke, Serhiy Kozak, Andreas Neuhierl (discussant), Serge Nyawa (discussant), Juan Passadore, Ken Singleton, Dacheng Xiu (discussant) and seminar and conference participants at Stanford, Chicago Asset Pricing Conference, Fourth International Workshop on Financial Econometrics, Toulouse Financial Econometrics Conference, SIAM Financial Mathematics and INFORMS. Svetlana Bryzgalova and Markus Pelger gratefully acknowledge financial support from the Stanford Lillie Fund.

†London Business School, Department of Finance, sbryzgalova@london.edu

‡Stanford University, Department of Management Science and Engineering, mpelger@stanford.edu

§Stanford University, Department of Management Science and Engineering, jzhu121@stanford.edu

I. Introduction

We develop a new way of building cross-sections of asset returns that are based on a set of characteristics. Our method is rooted in the idea of decision trees, and builds up on the appeal of standard double and triple sorts. Unlike the classical approach, however, it is able to handle a larger number of characteristics and their interactions at a time, and deliver a set of *basis assets* that reliably capture the underlying information, while staying interpretable and easy to construct.

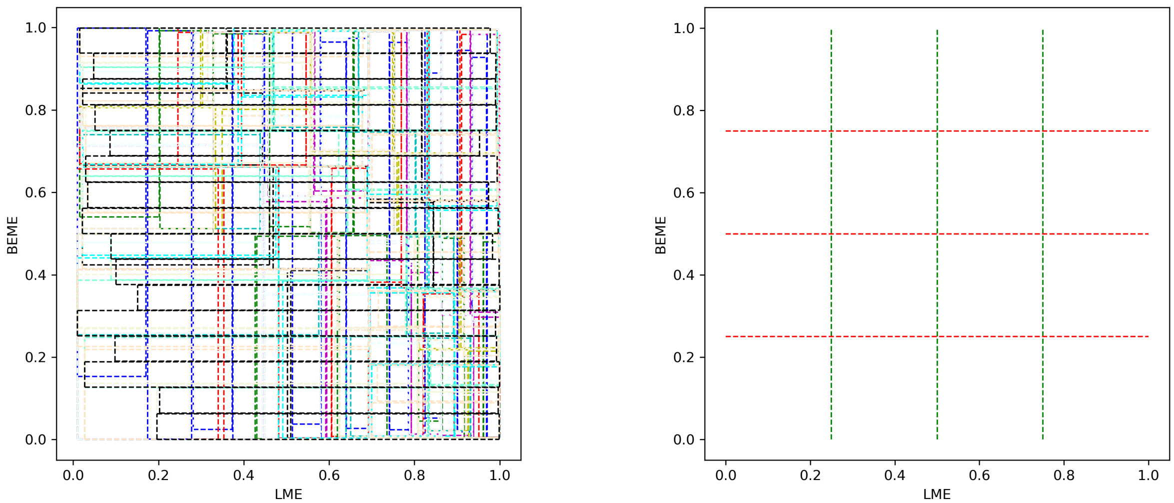
We generalize the notion of sorting by building Asset-Pricing Trees (AP-Trees) that capture all the important information contained in the characteristics. A simple decision tree is constructed based on a sequence of consecutive splits. For example, one could start by dividing the universe of stocks into two groups based on the individual stock's market cap, then within each group, - by their value, and so on. Naturally, the nodes of such a tree correspond to managed portfolios, created based on stock characteristics, and what is important, reflecting their *conditional* impact in a simple and easy measurable way. Furthermore, relying on a different order of the splits, or list of the variables employed, one could end up with a completely different composition of a portfolio. As a result, a collection of all the possible decision trees forms a high-dimensional and diverse set of possible investment strategies, with their features providing a direct mapping into the pattern of expected returns. In other words, they form a potential set of basis assets that span the SDF in an efficient and easily interpretable manner.

Inspired by the empirical success of machine learning techniques, we develop a new approach to *pruning* this large set of portfolios, to end up with an easily identified cross-section that is sparse, well-diversified, and truly informative. Grounded in the economic theory, this second step generalizes the approach of Kozak, Nagel, and Santosh (2018) to SDF recovery, and relies on dual shrinkage in the variance *and* mean to find an optimal solution.

In a large scale empirical application, we build a set of 36 cross-sections for different combinations of firm-specific characteristics, and compare their performance and informational content to that achieved by triple sorts. While AP-Trees in principle can easily accommodate a larger set of characteristics, this is empirically a limit for conventional triple-sorted portfolios that are comparable in depth, and provide a natural benchmark to judge their performance.

We find that for every single cross-section, relative to triple sorts, the test assets constructed with AP-Trees have dramatically higher Sharpe ratios, sometimes up to a factor of three. Importantly, this does not come from higher loadings on conventional sources of risk: the difference in these returns is not spanned by leading asset pricing models, and remains significant (both statistically and economically), even when the cross-section is pitted against 11 candidate risk factors. Resulting portfolios are well-diversified, and do not load on the extreme deciles of the cross-sectional sorts. In fact, their construction with decision trees enforces a stable and balanced composition by default, making them fully comparable, and often more diversified than triple sorts of the same *depth*.

All our results are obtained out-of-sample. We show that the main driving force behind the superior performance of AP-Trees is a) their ability to efficiently capture the interactions among characteristics, and b) creating a set of optimally pruned overlapping *basis functions* to better



(a) Conditional trees **(b)** Double-sorting

Figure 1. Cross-sectional quantiles of the portfolio splits based on conditional trees and double sorting with size and value.

reflect the patterns in expected returns. We find that triple sorts, while largely reflecting some of the true patterns in the data, present a very coarse structure of the cross-section, and often fail to reflect the underlying stocks.

Sorting plays a dual role in asset pricing, providing not only the building blocks for the candidate risk factors, but also *test portfolios* that are used to evaluate their performance and discriminate among the models. Many structural asset pricing models are defined through a set of moment conditions. This makes their estimation on the unbalanced noisy panel of the individual stocks with their time-varying exposure to the SDF, often infeasible. As a result, relying on a low-dimensional, interpretable, and informative cross-section of asset returns, is often the only way to test the model and identify its pitfalls. However, the test of the model can only be as good as the assets that represent the underlying information set. For example, portfolios sorted by size and value, are only the right benchmark to evaluate the setting aiming to capture these cross-sectional effects, if they really reflect the information contained in the underlying characteristics. After all, if the cross-section itself does not represent these patterns, there is little hope that the SDF, built from these portfolios, or spanning them, will do a better job.

So, how do we build a cross-section of portfolios that reflect multiple characteristics at once, remain a feasible testing playground for structural models, and retain interpretability at the same time? In other words, what are the *basis assets*?

The first key observation lies in using conditional sorts to span the universe of managed portfolios and the diversity of "building blocks" it naturally provides. This is particularly true for characteristics that are cross-sectionally correlated, as changing the order of the variables used for splits will lead to potentially different sets of basis assets.

Figure 1 illustrates the difference between the kind of portfolios one could get with conditional trees and double sorting based on the information contained in probably the most well-known pair of characteristics: size and value. We build simple conditional trees of the same depth as the double-sorted portfolios (that is, each node constructed with AP-Trees contains at least 1/16 of all the stocks at any given point of time). The outcome is staggering: the type of patterns and strategies one could span with conditional splits, using the same level of granularity as double sorts, is unquestionably richer. Furthermore, it directly incorporates the joint cross-sectional distribution of characteristics: since AP-Trees rely on sequential splits, the cutoffs are internally chosen to give the balanced, equally populated set of assets, and could never end up with empty portfolios. Building the whole universe of potential basis assets is important, but so is reducing their dimension to eliminate redundant strategies and identify the informative ones.

We introduce a novel way to extract a small set of basis assets capturing the key pricing information from a large set of portfolios by carefully selecting the basis functions that span the SDF. Finding the weights of these securities is equivalent to finding the tangency portfolio in the mean-variance space, i.e. the portfolio with the highest Sharpe ratio. We find a sparse representation of the SDF weights by including shrinkage in the mean-variance optimization problem. As our optimal mean-variance portfolio with shrinkage could potentially include any final or intermediate nodes of the trees, we prune a large set of assets both in the choice of splits and their depth. Intuitively, applying our estimator to tree portfolios corresponds to choosing an optimal bandwidth in a non-parametric regression, with the decision to use a finer set of nodes in spanning the SDF being equivalent to choosing a smaller bandwidth in the characteristic space. Instead of the conventional bias-variance trade-off for parameter estimation, however, we set up a global objective motivated by the fundamental nature of the problem.

Our shrinkage estimator is grounded in economic theory. We first construct a robust mean-variance efficient frontier that accounts for uncertainty in the mean and variance estimation, and then we trace it out to find the tangency portfolio. This approach combines three crucial features: (a) it diminishes the contribution of the assets that do not explain enough of common variation, (b) in includes a lasso-type term to obtain a sparse representation of the SDF, i.e. selecting a small number of managed portfolios, and (c) it shrinks the estimated mean of the tree-based sorts towards the average return. The last feature is crucial, as sample means are characterized by massive estimation errors, and their large absolute values are often likely to be due to noise, rather than reflecting a fundamental property of the data.

Our paper contributes to the growing literature in asset pricing that tackles the “multidimensional challenge”, as formulated by Cochrane (2011) in his AFA Presidential Address. Most of these models apply a version of PCA to characteristic-sorted or characteristic-projected data, which does not offer the same interpretability as AP-Trees. Our shrunk mean-variance optimization generalizes the approach suggested by Kozak, Nagel, and Santosh (2018), who estimate a small number of SDF basis assets by solving an efficient frontier optimization problem with a lasso penalty that selects their subset, and a ridge component that shrinks the contribution of lower order principal component assets. Our approach generalizes their estimator by applying an additional shrinkage in the mean, the degree of which we choose optimally. As a result, we can show empirically that the special case of Kozak, Nagel, and Santosh (2018) can be improved. The second differentiating factor is the use of trees instead of simple sorted portfolios, linear long-short factors, or their principal components. Shrunk tree basis functions offer an interpretable alternative to PCA. For example, consider building a cross-section based on size and value using PCA on double-sorted portfolios. The resulting factors are typically a market portfolio, a long-short size factor, a long-short value factor, and some higher-order PCs capturing convexity effects in the two characteristics. Instead, we will select long-only portfolios based on different tree cuts. One basis asset can be the market portfolio, which is just the original node of the tree. One cut in the size dimension results in small cap and large cap portfolios, and similar with value and growth. However, if the relevant pricing information requires the interaction between size and value an additional cut in the interaction dimension would result in a small value stock portfolios and other combinations (see Figure 1). As a result, one can aways trace a particular asset back to stock fundamentals.

Lettau and Pelger (2018) extend principal component analysis to include a cross-sectional pricing restriction that helps to identify weak factors, and our paper is based on a similar intuition, relying on a no-arbitrage criterion to select the optimal tree portfolios. Kelly, Pruitt, and Su (2019) and Fan, Liao, and Wang (2016) explicitly model stock loadings on the SDF as a function of its characteristics, and as a result apply PCA to managed portfolios that represent linear (and nonlinear) projections of asset returns on characteristics.

A lot of pathbreaking contributions have recently been made to studying the impact of characteristics on returns directly, without imposing an underlying risk model or a non-arbitrage condition. Freyberger, Neuhierl, and Weber (2019) estimate conditional expected returns as a function of characteristics with adaptive group lasso, allowing for a high-dimensional structure with non-linearities, but rule out the crucial interaction effects between characteristics. DeMiguel, Martin-Utrera, Nogales, and Uppal (2019) and Feng, Giglio, and Xiu (2019) focus on characteristics based factor selection with a lasso-type penalty, while Gu, Kelly, and Xiu (2019) and Chen, Pelger, and Zhu (2019) use machine learning techniques, like deep neural networks, to estimate asset pricing models that account for general functional forms including interactions. None of these papers, however, offer the same portfolio interpretability as our AP-Trees.

To the best of our knowledge, Moritz and Zimmerman (2016), Gu, Kelly, and Xiu (2019), and Rossi (2018) are the only papers relying on decision trees in estimating conditional moments of stock returns. Moritz and Zimmerman (2016) apply tree-based models to studying momentum, while Gu, Kelly, and Xiu (2019) use random forest to model expected returns on stocks as a function of characteristics. Rossi (2018) uses Boosted Regression Trees to form conditional mean-variance efficient portfolios based on the market portfolio and the risk-free asset. Since we use decision trees not for a direct prediction of returns, but for constructing a set of basis assets that span the efficient frontier, none of the standard pruning algorithms available in the literature are applicable in our setting because of its global optimization nature. Section III highlights this difference further, and introduced an alternative criterion we develop for pruning.

Naturally, our paper extends the literature on constructing optimal test assets. Lewellen, Nagel, and Shanken (2010) argue that conventional double-sorted portfolios, exposed to a small number of characteristics, often present a low hurdle for asset pricing models due to their strong embedded factor factor structure, and recommend mixing them with other cross-sections. For a given set of characteristics we offer an alternative set of test assets that is harder to price, as they extract additional information due to inherent interactions and nonlinearities in both the impact of characteristics on asset returns, as well as the distribution of the assets in the cross-section. Nagel and Singleton (2011) introduce optimal managed portfolios based on a General Method of Moment (GMM) argument that essentially builds a set of optimal instruments for a particular pair of the null and alternative models. Our test assets are long-only portfolios chosen as a robust span of the SDF, projected on a given space of characteristics. As a result, they are not designed to improve the power of the test when making inference for *a particular* parameter and/or factor, but rather try to answer the question of whether a given set of basis assets is representative of the information contained in the original stocks and their characteristics.

Finally, our pruning approach also contributes to the econometrics literature of shrunk mean-variance estimation and non-parametric mean estimation. We offer a new perspective of mean-variance optimization by solving the Markowitz optimization problem, i.e. minimizing the variance for a specific target mean return, with an elastic net penalty and treating the target mean as a tuning parameter. We show that this corresponds to tracing out the efficient frontier and finding the portfolio with the highest Sharpe ratio. As a result, our estimator has three different statistical interpretations. First, it can be interpreted as robust mean-variance optimization under uncertainty bounds on the estimated mean and variance. Second, it has the flavor of a regression with a ridge type shrinkage in the variance and mean, and a lasso penalty to obtain a sparse set of coefficients. In particular, a conventional elastic net regression (Zou and Hastie (2005)) is a special case of our approach with a different weighting in the ridge penalty and without the mean shrinkage. Last but not least, by extending the findings of Kozak, Nagel, and Santosh (2018) our estimator can be viewed from the Bayesian perspective as imposing a specific prior on the mean.

The rest of the paper is organized as follows. Section II highlights the role of sorting as a simple nonparametric conditional estimator of expected returns, and introduced conditional trees as the collection of alternative basis assets. Section III we describe the methodology behind Asset Pricing Pruning based on the shrunk mean-variance optimization. We illustrate the basic properties of our approach with a simple simulation setup in Section IV, and provide empirical results in Section VI. Finally, Section VII concludes. All the additional empirical results, as well as alternative simulation setups, are delegated to the Appendix.

II. Sorting and Trees

A. *Sorting as a Conditional Estimator*

A conventional view of characteristics impact on the asset returns is that they proxy for the underlying exposure to the systematic sources of risk (Fan, Liao, and Wang (2016), Kelly, Pruitt, and Su (2019)):

$$\underbrace{r_t^{ex}}_{N\times 1} = \underbrace{\beta_{t-1}}_{N\times K}\underbrace{F_t}_{K\times 1} + \epsilon_t = g(C_{t-1})F_t + \epsilon_t$$

where F_t are $K \times 1$ common factors and conditional factor loadings $\beta_{t-1} = g(C_{t-1})$ are functions of $N \times L$ matrix of characteristics C_{t-1}.

While the characteristics of individual companies (and hence, their risk exposure) can rapidly change over time, sorting-based portfolios provide a time-varying rotation from individual securities to their baskets, not only diversifying the idiosyncratic risk, but also providing (hopefully) a time-invariant exposure to the underlying risk factors. While it certainly restricts the functional dependency of expected returns on characteristics, the ultimate goal is to move away from the individual assets and focus on the stable underlying factor structure in the characteristic space. In this setting sorting can also be viewed as a nonparametric estimator of expected returns on individual stocks, with a particular choice of a kernel that corresponds to cross-sectional quantiles (see Cochrane (2011)). Cattaneo, Crump, Farrell, and Schaumburg (2019) formalize this intuition further and derive the optimal portfolio splits for estimating expected returns in the case of a single characteristic.

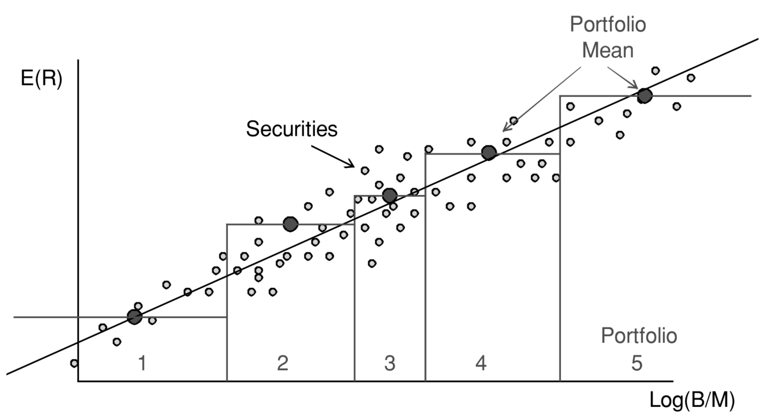


Figure 2. Cochrane (2011): BM-sorted portfolios as a nonparametric estimator of the stock returns.

What are the shortcomings of conventional sorting-based procedures? Traditional sorting-based methods to create portfolios rarely capture more than 2 or 3 firm-specific characteristics at the same time. Indeed, simple intersections of unconditional sorts that form a set of non-overlapping base assets, quickly leads to a curse of dimensionality. For example, 25 Fama-French portfolios are based on the intersection of 2 unconditional sorts, each into into five groups (by size and value), and then take the intersections to generate 25 groups. For 3 characteristics (keeping the depth constant), the same approach would attempt to create already 125 portfolios, some of which are poorly diversified, or even empty. In practice, this type of method never goes beyond triple sorting since the number of stocks in each group decreases exponentially with the number of sorting variables. The only feasible alternative is to stack a set of double-sorted cross-sections against each other, which rapidly increases the dimensionality, and has no fundamental basis behind it. In other words, the standard methodology finds it challenging to create a set of assets that not only adequately reflects the information contained in a pre-specified list of characteristics, but also highlight their *joint* impact on the underlying structure of expected returns.

There are two main elements to our methodology that alleviates the curse of dimensionality and providse an alternative, more informative set of basis assets. First, we use a large set of conditional sorts to define all the potential 'building blocks' that form the base of the stochastic discount factor. This is a natural extension of the standard sorting-based methodology, and works particularly well when the underlying characteristics have a complex joint distribution, characterized by substantial cross-sectional dependence. Then, we use a shrinkage-based approach to *globally prune* the trees and find an interpretable, sparse, and stable subset of the assets spanning the SDF. The second step is crucial: an optimal cross-section of portfolios depends not only on the spread of the expected returns identified through characteristic-based sorting, but also their joint dynamics and comovement. This interplay between the first and second moments is fundamental for our objective, and sets it apart from the standard prediction-based literature.

B. Trees and Conditional sorts

Conditional portfolio sorts, based on a selected set of characteristics, are intuitive to build and interpret within the setting of decision trees. First, we categorize firms into two equally sized groups according to the high/low value of a characteristic variable, for example, size. Then within each group, stocks are further split into two smaller groups of the same dimension based on the value of some other variable, for example, value, creating four groups in total. This step can be repeated until the desired depth of the tree d is reached, obtaining 2^d groups of stocks in total, each of the size $\frac{N}{2^d}$, where N is the total number of stocks. The stocks in each group can then be combined into an equally or value-weighted portfolio.

Figure 3 gives an example of a conditional tree of depth 3, where stocks are first sorted by size, then value, and then size again. Conditional splits are built using the information at time t, and hence, similar to double sorting, can easily handle an unbalanced panel of stocks. At time t, all the stocks that have valid size and value information from $t-1$ (or previous periods), are first sorted into group 11 and 12 based their size at $t-1$. Then group 11 is further split into 111 and 112, and group 12 is split into 121 and 122. The key point here is that the marginal splitting value for group 11 and 12 might not be equal to the unconditional median, since they are based on the conditional information of the previous size split. Lastly, the four groups 111,112,121, and 122 are each further split into two portfolios each, to form 8 level-3 portfolios. The notation of each node is therefore reflecting a chosen path along the tree with a specified list and *order* of the split criteria.

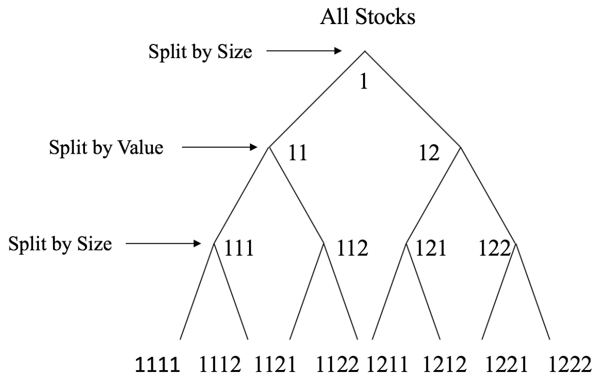


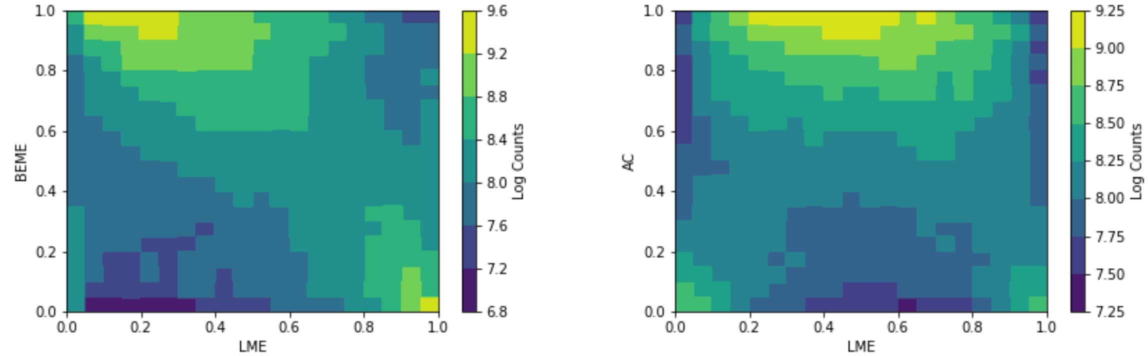
Figure 3. Example of a conditional tree based on size and value

If stock-specific characteristics are independent, the order of the variables used for splits does not matter, and we end up with the same quantiles for splits as double sorting. Unconditionally, the cumulative density function of each characteristic has a uniform distribution, however, it is well known empirically that characteristics have a complicated joint relationship that question the validity of coarse double sorting as an appropriate tool to reflect expected returns. Figure 4 shows the sample distribution of characteristics in the cross-section of stocks and their conditional and unconditional impact on expected returns for two example: size/value, and size/accrual. On average, there seems to be a negative cross-sectional correlation between size and book-to-market with a clear clustering around the north-west and south-east corners.[1] As a result, double-sorted portfolios are heavily unbalanced across the characteristic spectrum. Depending on the pair of characteristics, their joint distribution could exhibit very different shapes and relationships that are far from a simple linear correlation structure, as the right top panel of Figure 4 illustrates for size and accruals.

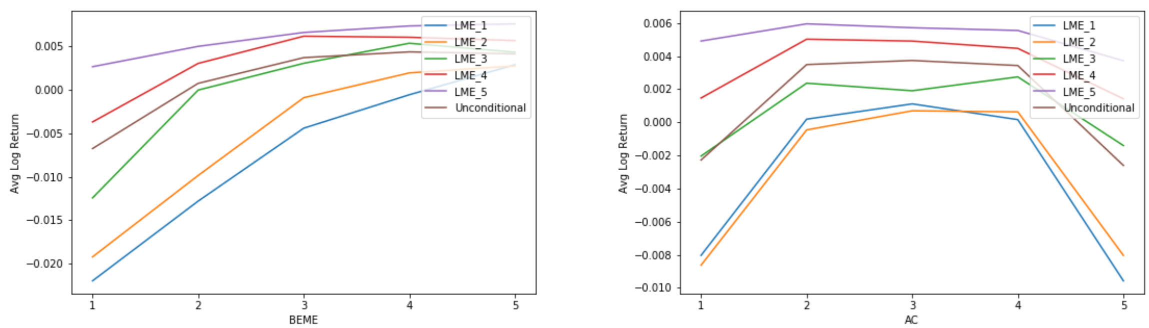
However, it is not only the dependence structure of the characteristics that substantially affects sorting-based portfolios, but also their price impact, which is often highly nonlinear and full of interactions[2]. For example, it is well known that the value effect is not homogenous across the different size deciles of the securities, and is particularly strong for the smallest stocks (Figure 4, left side of Panel B). The impact of accruals is almost flat unconditionally, however, controlling for the stock size reveals a striking inverted U-shape pattern, which is particularly pronounced for for the smallest quintiles of securities. Figures B.1 and B.2 in the Appendix further demonstrate different cross-sectional distributions of characteristics and their conditional and unconditional effect on expected returns, which sometimes are parallel to each other, speaking to the additive

[1] In part, this is mechanically arising due to the market size of the company being in the denominator of the B/M fraction.

[2] See, e.g. Freyberger, Neuhierl, and Weber (2019) for the use of adaptive group lasso to estimate the nonlinear impact of characteristics on expected returns, Gu, Kelly, and Xiu (2019) and Chen, Pelger, and Zhu (2019) for the machine learning approach



Panel A. Empirical distribution of the scaled characteristics.



Panel B. Conditional and unconditional characteristic impact on expected returns

Figure 4. Joint empirical distribution of (scaled) characteristics in the cross-section of stocks
The graphs represent the pairwise empirical cross-sectional distribution of the characteristic quantiles across the stocks. Empirical distribution is computed on a uniform 20x20 grid. Panel B presents the impact of characteristics (book-to-market on the left, accruals on the right) on the stock expected returns, conditional on the size quintile and unconditionally.

nature of the data generating process, but often are not.

Since characteristics are generally dependent, and have a nontrivial joint impact on expected returns, the order of the variables used to build a tree and generate conditional sorts, matters. In fact, each sequence used for splitting the cross-section generates another set of 2^d portfolios. If we denote the number of sorting variables by M, then there are M^d different combinations of splitting choices, and we end up with $M^d \cdot 2^d$ (overlapping) portfolios, each consisting of $\frac{N}{2^d}$ stocks (which does not depend on M). Naturally, these portfolios (both final and intermediate leaves of the tree) can capture at most d-way interactions between sorting variables.

To sum up, in case of independent characteristics, conditional trees lead to creating portfolios in line with the double sorting methodology. However, if the underlying distribution is in fact dependent, the collection of all possible conditional splits delivers a multitude of basis assets, that are fundamentally different from those created with double sorting, that, as a result, could span a different type of the SDF.

III. Asset Pricing Pruning

AP-Trees form the set of basis assets that reflect the relevant information conditional on characteristics, and could be used to build the stochastic discount factor. However, using all the potential portfolios is often not feasible due to the curse of dimensionality: their number grows exponentially with the depth of the tree. For example, 2 characteristics in a tree of depth 3 produce 2^3 subtrees, each having 2^3 portfolios, and a total set of 64 overlapping basis functions. Using 3 characteristics with the depth of 3 (4) results in $3^3 \cdot 2^3 = 216$ (1296) nodes. Finally, with 10 characteristics and depth of 3 the total number of basis portfolios explodes to 8,000, which makes it impossible to use in some applications and creates a lot of redundancy in others. Hence, we introduce a technique to shrink the dimension of the basis assets, with the key goal of retaining *both* the relevant information contained in characteristics, and portfolio interpretability.

Despite the existence of many conventional ways to prune a tree, available in the machine-learning literature, they are not applicable in our setup. Fundamentally, the key reason for that is that asset pricing is a *global* problem and cannot be handled by *local* decision criteria. For example, in a mean-variance optimization problem the optimal weights can only be found by considering the complete covariance matrix of assets and not just expected returns and correlation between two individual securities. The conventional way to prune a tree, a bottom-up approach, would only work if the current split does not affect the process of decision-making in other nodes, or in the language of mean-variance optimization if tree portfolios from different parts of the tree are uncorrelated, which is, of course, generally not the case.

Our new approach,"Asset Pricing Pruning", selects the AP-tree basis functions with the most non-redundant pricing information that could span the SDF. Since the problem is generally equivalent to finding the tangency portfolio with the highest Sharpe ratio in the mean-variance space, we focus on its sparse representation. Importantly, we consider both final and intermediate nodes of the trees, which leads to an SDF representation that adapts its degree of sparsity in both depth and types of the splits.

A. *Stochastic Discount Factor Construction with Elastic Net*

We find SDF weights by solving a mean-variance optimization problem with elastic net shrinkage applied to all final and intermediate nodes of AP-trees. This approach combines three crucial features:

1. it shrinks the contribution of the assets that do not help in explaining variation, implicitly penalizing the impact of lower order principal components;
2. it shrinks the sample mean of tree portfolios towards their average return, which is crucial, since estimated means with large absolute values are likely to be very noisy, introducing a bias;
3. it includes a lasso-type shrinkage to obtain a sparse representation of the SDF, selecting a small number of AP-tree basis assets.

The search of the tangency portfolio can be effectively decomposed into 2 separate steps. First, we construct the whole mean-variance efficient frontier using the standard optimization with shrinkage terms. Then, we select the optimal portfolio located on the frontier, finding the tangency point.

We show that our approach is a generalization of the SDF estimation approach of Kozak, Nagel, and Santosh (2018), who also shrink the covariance matrix of the assets and select a sparse set of weights to span the tangency portfolio. Our methodology generalizes their approach by also including shrinkage in the mean, the latter being crucial for robust estimation. Empirically, we solve the Markowitz optimization problem by finding the minimum variance portfolio on a grid of target expected returns, and then selecting the one that leads to the tangency portfolio. We show that tracing out the whole efficient frontier is generally equivalent to different levels of shrinkage on the mean expected return, and generally does not have to be zero, which is imposed in Kozak, Nagel, and Santosh (2018). In fact, using cross-validation to find the optimal value of this shrinkage for a set of 36 different cross-sections we build in the empirical application, we find that it is almost never equal to 0.

Splitting the original optimization problem into two steps has three different statistical interpretations. First, there is the actual shrinkage in the estimated mean relative to the naive solution of the tangency problem. Second, it can be interpreted as a robust estimation under parameter uncertainty. Finally, since our approach generalizes the logic of Kozak, Nagel, and Santosh (2018), we also benefit from their Bayesian interpretation of building the SDF.

Consider the whole cross-section of excess returns on the portfolios built with AP-trees, and denote their sample estimates of mean and variance-covariance matrix by $\hat{\mu}$ and $\hat{\Sigma}$. For each target expected return μ_0, we find the minimum variance portfolio weights $\hat{\omega}$ with an elastic net penalty. This is the classical Markowitz problem with additional shrinkage. All the tuning parameters, i.e. the target mean μ_0, the lasso weight of λ_1 and ridge with λ_2 are treated as fixed at this step, and chosen separately on the validation data set. In other words, the estimation proceeds as follows:

1. *Mean-variance portfolio construction with elastic net*: For a given set of values of tuning parameters μ_0, λ_1 and λ_2, use the training dataset to solve

$$\begin{aligned}
\text{minimize} \quad & \frac{1}{2} w^\top \hat{\Sigma} w + \lambda_1 ||w||_1 + \frac{1}{2}\lambda_2 ||w||_2^2 \\
\text{subject to} \quad & w^\top \mathbb{1} = \mathbb{1} \\
& w^\top \hat{\mu} \geq \mu_0.
\end{aligned}$$

   where $\mathbb{1}$ denotes a vector of ones, $\|\omega\|_2^2 = \sum_{i=1}^N \omega_i^2$ and $\|\omega\|_1 = \sum_{i=1}^N |w_i|$, and N is the number of assets.

2. *Tracing out the efficient frontier*: Select tuning parameters μ_0, λ_1 and λ_2 to maximize Sharpe ratio on a validation sample of the data.

Note that the two-step approach for finding a robust tangency portfolio is a method that is not specific to using tree-based basis assets, but can be applied to any potential cross-section.

However, we argue that it is particularly appealing when using AP-trees, as it obtains a small number of interpretable basis assets which are selected from a larger set of potentially highly correlated portfolios. While there are other techniques that could efficiently handle a case of highly correlated securities, like PCA, they often lack interpretability, and hence, investigating a failure of a given candidate model to price them could be challenging at best.

B. *Shrinkage Perspective*

Without imposing any shrinkage on the portfolio weights for the SDF, the problem has an explicit solution, $\hat{\omega}_{naive} = \hat{\Sigma}^{-1}\hat{\mu}$. To see that our estimator is a shrinkage version of this estimator, consider first only the impact of ridge penalty, i.e. setting $\lambda_1 = 0$:

$$\min_{\omega} \omega^\top \hat{\Sigma} \omega + \lambda_2 \|\omega\|_2^2 \qquad \text{subject to } \omega^\top \hat{\mu} = \mu_0 \text{ and } \omega^\top \mathbb{1} = 1.$$

It is easy to notice that this problem is equivalent to the conventional Markowitz problem, where the sample covariance matrix $\hat{\Sigma}$ is replaced by $\left(\hat{\Sigma} + \lambda_2 I_N\right)$:

$$\min_{\omega} \omega^\top \left(\hat{\Sigma} + \lambda_2 I_N\right) \omega \qquad \text{subject to } \omega^\top \hat{\mu} = \mu_0 \text{ and } \omega^\top \mathbb{1} = 1,$$

the solution to which is well known in closed form:

$$\hat{\omega} = \delta \omega_\mu + (1 - \delta)\omega_1,$$

where

$$\begin{aligned}
\omega_\mu &= \frac{1}{\mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \hat{\mu}} \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \hat{\mu} \\
\omega_1 &= \frac{1}{\mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \mathbb{1}} \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \mathbb{1} \\
\delta &= \frac{\mu_0 - \hat{\mu}^\top \omega_1}{\hat{\mu}^\top \omega_R - \hat{\mu}^\top \omega_1}.
\end{aligned}$$

Relaxing the normalization of $\omega^\top \mathbb{1} = 1$ (which can be always enforced ex post), and plugging in the formulas yields

$$\begin{aligned}
\hat{\omega} &= \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \hat{\mu} + \underbrace{\frac{\hat{\mu}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right) \hat{\mu} - \mu_0 \mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right) \mathbb{1}}{\mu_0 \mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right) \mathbb{1} - \hat{\mu}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right) \mathbb{1}}}_{\gamma} \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \mathbb{1} \\
&= \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} (\hat{\mu} + \gamma \mathbb{1})
\end{aligned}$$

up to the normalization of $\omega^\top \mathbb{1} = 1$. Note, that the shrinkage parameter γ is a decreasing function of μ_0, i.e. a lower target expected return implies more shrinkage towards the mean. Naturally, there is a direct mapping between these parameters as $\mu_0 \in \left[\frac{\mu^\top(\Sigma+\lambda_2 I_N)^{-1}\mathbb{1}}{\mathbb{1}^\top(\Sigma+\lambda_2 I_N)\mathbb{1}}, \frac{\mu^\top(\Sigma+\lambda_2 I_N)\mu}{\mathbb{1}^\top(\Sigma+\lambda_2 I_N)\mathbb{1}}\right)$ corresponds to $\gamma \in [0, +\infty)$. In particular, using $\gamma = 0$ is equivalent to setting the target expected return to $\mu_0 = \frac{\mu^\top(\Sigma+\lambda_2 I_N)^{-1}\mu}{\mathbb{1}^\top(\Sigma+\lambda_2 I_N)^{-1}\mathbb{1}}$, which corresponds to the optimal portfolio weights of

$$\omega = \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \hat{\mu},$$

This estimator is identical to the one used in Kozak, Nagel, and Santosh (2018) without the lasso term,

$$\hat{\omega} = \arg\min_{\omega} \left(\hat{\mu} - \hat{\Sigma}\omega\right)^\top \hat{\Sigma}^{-1} \left(\hat{\mu} - \hat{\Sigma}\omega\right) + \lambda_2 \|\omega\|_2^2.$$

Next, we consider the case of adding the lasso penalty , i.e. $\lambda_1 \neq 0$. If $\hat{\Sigma}$ is a diagonal matrix denoted by $\hat{D}$, this is conceptually equivalent to using our estimator in the PCA space, as advocated by Kozak, Nagel, and Santosh (2018), and has a closed form solution[3]:

$$\omega = \left(\hat{D} + \lambda_2 I\right)^{-1} \left(\hat{\mu} + \gamma \mathbb{1} - \lambda_1 \mathbb{1}\right)_+,$$

with $(x)_+ = max(x, 0)$ elementwise. Note that the estimator of Kozak, Nagel, and Santosh (2018) that includes a lasso term, also has an explicit solution[4]:

$$\hat{\omega} = \left(\hat{D} + \lambda_2 I\right)^{-1} \left(\hat{\mu} - \lambda_1 \mathbb{1}\right)_+$$

which coincides with our estimator for a particular choice of λ_1, as a solution to the following

[3]The analytical solution for a diagonal covariance matrix is based on the following argument. We solve the optimization problem with a Lagrange multiplier:

$$L = \frac{1}{2}\omega^\top \hat{D}\omega + \frac{1}{2}\lambda_2\|\omega\|_2^2 + \lambda_1\|\omega\|_1 - \tilde{\gamma}_1\left(\omega^\top\hat{\mu} - \mu_0\right) - \tilde{\gamma}_2\left(\omega^\top\mathbb{1} - 1\right).$$

The first order condition on the active set, i.e. for the non-zero values of ω_i equals

$$(D_i + \lambda_2)\,\hat{\omega}_i = \tilde{\gamma}_1\mu_i + \tilde{\gamma}_2 1 - \lambda_1 \text{sign}(\hat{\omega}_i) \qquad \text{for } i \text{ in the active set,}$$

which yields

$$\hat{\omega} = \left(\hat{D} + \lambda_2 I_N\right)^{-1} \left(\tilde{\gamma}_1\hat{\mu} + \tilde{\gamma}_2\mathbb{1} - \lambda_1\right)_+$$

As both Lagrange multiplier are functions of μ_0 we can reformulate the problem as:

$$\hat{\omega} = \left(\hat{D} + \lambda_2 I\right)^{-1} \left(\hat{\mu} + \gamma\mathbb{1} - \tilde{\lambda}_1\mathbb{1}\right)_+$$

Here we have relaxed the constraint $\hat{\omega}^\top\mathbb{1} = 1$ which can be enforced ex post and substituted $\gamma = \frac{\tilde{\gamma}_2}{\tilde{\gamma}_2}$ and $\tilde{\lambda}_1 = \frac{\lambda_1}{\tilde{\gamma}_1}$.

[4]See Lettau and Pelger (2019) for a derivation.

problem:

$$\hat{\omega} = \arg\min_{\omega} \frac{1}{2}\left(\hat{\mu} - \hat{\Sigma}\omega\right)^{\top} \hat{\Sigma}^{-1} \left(\hat{\mu} - \hat{\Sigma}\omega\right) + \lambda_1\|\omega\| + \frac{1}{2}\lambda_2\|\omega\|_2^2.$$

Hence, our estimator is equivalent to the one used by Kozak, Nagel, and Santosh (2018) in the case of uncorrelated assets.

In the general case of a non-diagonal sample covariance matrix $\hat{\Sigma}$, however, the impacts of ridge and lasso penalties cannot be separated, and hence the lasso penalization cannot subsume the mean shrinkage. We, therefore, have to solve the general optimization problem with Lagrange multipliers:

$$L = \frac{1}{2}\omega^{\top}\hat{\Sigma}\omega + \frac{1}{2}\lambda_2\|\omega\|_2^2 + \lambda_1\|\omega\|_1 - \tilde{\gamma}_1\left(\omega^{\top}\hat{\mu} - \mu_0\right) - \tilde{\gamma}_2\left(\omega^{\top}\mathbb{1} - 1\right).$$

The first order condition on the active set, i.e. the non-zero values of ω_i imply

$$\left[\left(\hat{\Sigma} + \lambda_2\right)\omega\right]_i = \tilde{\gamma}_1\mu_i + \tilde{\gamma}_2 1 - \lambda_1 \text{sign}(\hat{\omega}_i) \qquad \text{for } i \text{ in the active set,}$$

which in turn can be formulated as follows:

$$\left[\left(\hat{\Sigma} + \lambda_2\right)\omega\right]_i = \hat{\mu}_i + \gamma 1 - \lambda_1 \text{sign}(\hat{\omega}_i) \qquad \text{for } i \text{ in the active set,}$$

Note, that the corresponding first order condition for Kozak, Nagel, and Santosh (2018) equals

$$\left[\left(\hat{\Sigma} + \lambda_2\right)\omega\right]_i = \hat{\mu}_i - \lambda_1 \text{sign}(\hat{\omega}_i) \qquad \text{for } i \text{ in the active set.}$$

This implies that our estimator is a general case of Kozak, Nagel, and Santosh (2018), and will coincide with their solution, if instead of $\hat{\mu}$ one does the optimization, using $\hat{\mu} + \gamma\mathbb{1}$.

C. Robust Estimation Perspective

Our estimator can also be interpreted as a robust approach to the mean-variance optimization, when there is uncertainty about the mean and variance-covariance matrix of returns.

Consider the case when the true covariance matrix Σ and the vector of expected returns belong to the following uncertainty sets:

$$\Sigma \in S_{\Sigma} = \left\{\tilde{\Sigma} : \tilde{\sigma}_{i,j} = \widehat{\Sigma}_{i,j} + e_{i,j}; \quad \|e\|_2^2 \leq \bar{\Delta}; \quad \tilde{\Sigma} \text{ is p.s.d. and } \bar{\Delta} \geq 0\right\}$$
$$\mu \in S_{\mu} = \left\{\tilde{\mu} : \tilde{\mu}_i = \hat{\mu}_i + \psi_i; \quad |\psi_i| \leq \bar{\psi} \quad \text{and } \bar{\psi} \geq 0\right\}.$$

The robust version of the classical mean-variance portfolio optimization is equivalent to finding the solution under the worst case scenario:

$$\min_{w} \max_{\tilde{\Sigma} \in S_\Sigma,\, \tilde{\mu} \in S_\mu} \frac{1}{2}\omega^\top \tilde{\Sigma}\omega - \tilde{\gamma}_1 \left(\omega^\top \tilde{\mu} - \mu_0\right) - \tilde{\gamma}_2 \left(\omega^\top \mathbb{1} - 1\right).$$

Concentrating out $\tilde{\mu}$:

$$\begin{aligned} &\min_{w} \max_{\tilde{\Sigma} \in S_\Sigma} \frac{1}{2}\omega^\top \tilde{\Sigma}\omega - \tilde{\gamma}_1 \sum_{i=1}^{N} \left(\omega_i \left[\hat{\mu}_i - \bar{\psi} sgn(\omega_i)\right] - \mu_0\right) - \tilde{\gamma}_2 \left(\omega^\top \mathbb{1} - 1\right) \\ =& \min_{w} \max_{\tilde{\Sigma} \in S_\Sigma} \frac{1}{2}\omega^\top \tilde{\Sigma}\omega - \tilde{\gamma}_1 \left(\omega^\top \hat{\mu} - \mu_0\right) + \tilde{\gamma}_1 \sum_{i=1}^{N} \bar{\psi} |\omega_i| - \tilde{\gamma}_2 \left(\omega^\top \mathbb{1} - 1\right). \end{aligned}$$

where $sgn(\cdot)$ is the sign function. Concentrating out $\tilde{\Sigma}$ yields

$$\min_{w} \frac{1}{2} tr\left(\omega^\top \hat{\Sigma}\omega\right) + \bar{\Delta}\omega^\top \omega - \tilde{\gamma}_1 \left(\omega^\top \hat{\mu} - \mu_0\right) + \tilde{\gamma}_1 \sum_{i=1}^{N} \bar{\psi}_i |\omega_i| - \tilde{\gamma}_2 \left(\omega^\top \mathbb{1} - 1\right).$$

Finally, taking $\lambda_1 = \tilde{\gamma}_1 \bar{\psi}$ and $\lambda_2 = \bar{\Delta}$ the problem becomes equivalent to the standard regularized optimization with the sample mean estimates of expected returns and the covariance matrix:

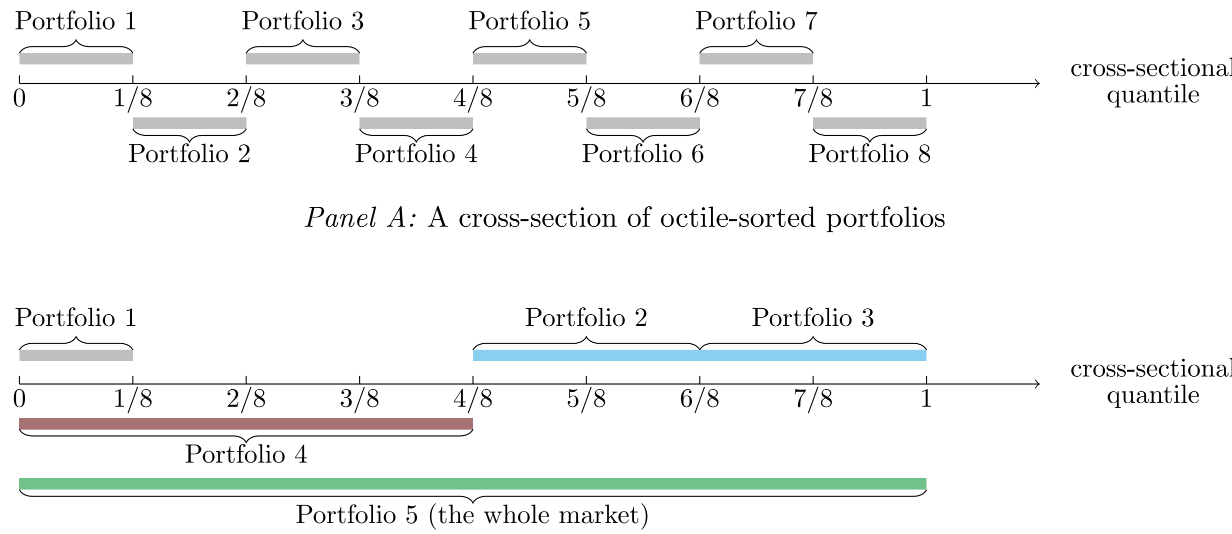
$$\min_{w} \frac{1}{2}\omega^\top \hat{\Sigma}\omega + \lambda_2 \|\omega\|_2^2 + \lambda_1 \|\omega\|_1 - \tilde{\gamma}_1 \left(\omega^\top \hat{\mu} - \mu_0\right) - \tilde{\gamma}_2 \left(\omega^\top \mathbb{1} - 1\right).$$

Hence, a larger penalty λ_1 corresponds to higher uncertainty in the mean and larger λ_2 accounts for more uncertainty in the covariance matrix. Therefore, the first stage regularized optimization results in a robust estimation of the whole mean variance efficient frontier, which we trace out in the second step to find the tangency portfolio.

D. *Pruning in Depth*

Our portfolio selection approach prunes the AP-Tree in both depth and width, since we include all the intermediate and final nodes in the robust mean-variance portfolio selection. Hence, our selected tree portfolios can be higher level nodes without further splits. As an example consider a univariate sorting with only one characteristic and depth two. If only high values of the characteristics have an effect on mean returns, the low level of characteristics can be grouped together in one large portfolio. In this example, a possible selection could be a subtree of depth one for lower values that has 50% of the observations and two subtrees of level four for high values that each have 25% of the stocks.

Figure 5 further illustrates the importance of pruning in depth of the tree, which amounts to selecting larger, denser portfolios. In a case of a single characteristic, simple univariate sort with the same population density as the tree of depth 3, would yield 8 portfolios, that are equally spaced from each other in the characteristic space, similar to the standard decile-sorted portfolios. In contrast, depending on the data generating process, pruning the tree could yield a simple cross-section of a smaller set of assets, some of which are retained from original octiles, while others come from the intermediate nodes of the tree, essentially merging higher level nodes, as long as there is not enough informational gain from doing such a split. In the particular example of the selection in Figure 5, the algorithm ended up including only one node based on the single octile of the stock distribution, while the rest of the portfolios were much denser, including 2, 4, or 8 times more stocks than the original selection. Why does it happen?



Panel B: A cross-section of octile-sorted portfolios after pruning

Figure 5. Selection of basis portfolios through pruning

The problem of split selection fundamentally reflects the trade-off between the estimation error and bias. Tree portfolios at higher (intermediate) nodes are more diversified, naturally leading to a smaller variance of their mean estimation, etc, while more splits allow to capture a more complex structure in the returns at the cost of using investment strategies with higher variance. To mitigate this trade-off, we use the weighting scheme inspired by the properties of the GLS estimator. The idiosyncratic noise in each of the tree portfolio is diversified at the rate $\frac{1}{N_i}$ where N_i is the number of stocks in tree portfolio i. Hence, the optimal rate to weight each portfolio is $\sqrt{N_i}$. An equivalent approach would be to simply weight each portfolio by $\frac{1}{\sqrt{2^{d_i}}}$ where d_i is the depth of portfolio i in the tree.[5].

The first node in our AP-Tree is always a value-weighted market portfolio. The first split results in portfolios with 50% of the stocks whose returns are multiplied by $\frac{1}{\sqrt{2}}$ to account for the higher variance. Note, that this re-weighting relies on the similar arguments as the PCA weighting in Kozak, Nagel, and Santosh (2018), where all the assets are multiplied by the eigenvectors of the covariance matrix, and hence portfolio selection is done in the PCA space. Note, that the first PC is usually an equally weighted market factor and is scaled by $\sqrt{N}$ as the market affects by

[5]Note that the number of stocks in each month is time-varying and hence, the depth is the most natural invariant statistic to account for differences in diversification.

construction most of the N assets in the sample. A PC loading on only half of the stocks is then naturally scaled by $\sqrt{\frac{N}{2}}$ similar to our tree portfolios of depth 1. In other words, the presence of higher order nodes have the same effect as higher order PCs: they offer a chance at achieving high rate of return, however, at a cost of larger estimation error and noise. In contrast to the PCs, however, tree-based asset returns are long-only portfolios, that are easy to trace back to the fundamentals and interpret. The depth selection can also be interpreted as a intuitive way to optimally select the bandwidth of a non-parametric estimator based on the criterion, relevant for asset pricing.

In summary, rescaling tree portfolios by the efficient weight ensures an optimal trade-off between the bias and variance, and the selection of higher order nodes, whenever additional splits are not beneficial overall, and it works by grouping together stocks that have the same exposure to the SDF.

IV. Simulation

We illustrate the benefits of using tree-based portfolios in uncovering the patters of expected returns in the characteristic space and the efficiency of our pruning approach with a simple simulation that is designed to capture some of the stylized features of the data.

Suppose there is a single factor that drives expected returns on a cross-section of stocks, with loadings being a function of 2 stock-specific characteristics:

$$R^e_{t+1,i} = \beta_{t,i} F_{t+1} + \epsilon_{t+1,i}.$$

In our simple model the factor follows $F_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_F, \sigma_F^2)$ and the idiosyncratic component $\epsilon_{t,i} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. Motivated by a wide range of empirical patterns in the joint empirical distribution of characteristics and their impact on expected returns (see, e.g. Figures B.1 – B.3 in the Appendix), we consider two different formulations for the risk-loadings as a function of characteristics $C^{(1)}_{i,t}$ and $C^{(2)}_{i,t}$:

1. *Additively linear*, with stock beta being simply the sum of two characteristics:

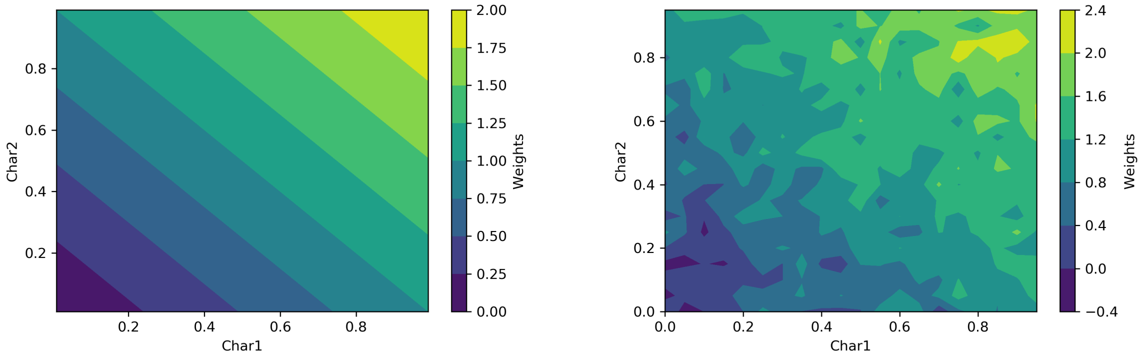
$$\beta_{t,i} = C^{(1)}_{t,i} + C^{(2)}_{t,i}.$$

2. *Nonlinear*, with stock beta being a nonlinear and nonadditive function of characteristics:

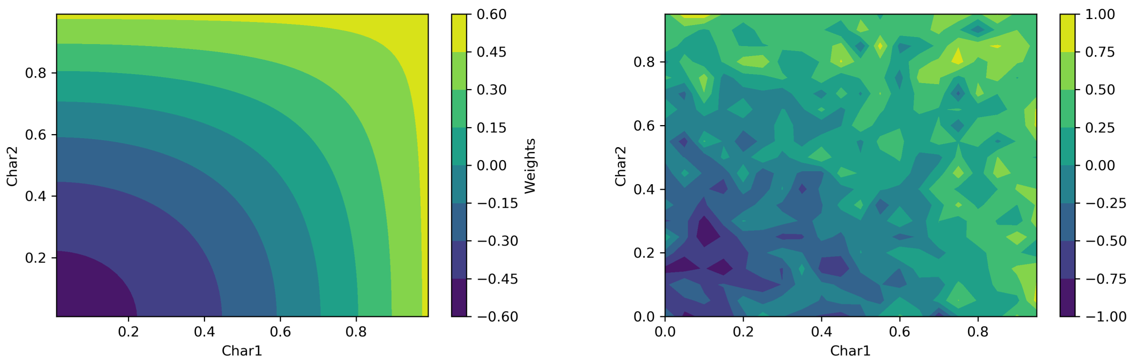
$$\beta_{t,i} = \frac{1}{2} - [1 - (C^{(1)}_{t,i})^2] \cdot [1 - (C^{(2)}_{t,i})^2].$$

We model stock characteristics as quantiles of the cross-sectional distribution, and allow for their potential dependence:

$$C^{(1)}_{t-1,i}, C^{(2)}_{t-1,i}) \overset{i.i.d.}{\sim} \text{Corr-Uniform}[0, 1, \rho]$$



(a) Additively linear model



(b) Nonlinear model

Figure 6. True factor loadings and sample average returns for a cross-section of stocks, as a function of characteristics.

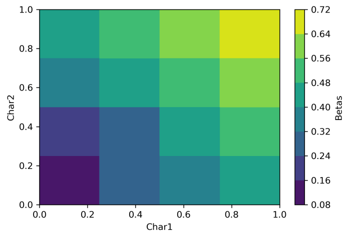
where Corr-Uniform$[0, 1, \rho]$ denotes a pair of uniformly distributed random variables that have correlation ρ, and marginal densities $U[0, 1]$. To model the impact of characteristic dependence on the portfolio structure, we consider three cases: $\rho \in \{0, 0.5, 0.9\}$. The risk factor has $\mu_F = 1, \sigma_F = 2$ implying a Sharpe ratio of 0.5, and the idiosyncratic noise has a standard deviation of $\sigma_e = 8$ to mimic a noisy and volatile process for stock returns. Since $\mathbb{E}(f_t) = 1$, stock's beta is exactly equal to its expected return. Finally, we set cross-sectional dimensions to $N = 800, \quad T = 600$. Figure 6 illustrates the resulting empirical distribution of stock returns in the characteristic space and their true loadings on the SDF which is equal to the expected returns up to a proportionality constant.

We build double-sorted portfolios (DS), based on two characteristics distributed on a 4×4 grid, which creates a cross-section of 16 portfolios. Similarly, we build AP-Trees with a depth of 4, so that each portfolio consists of at least 1/16th of all the stocks and has a similar granularity as the double-sorted cross-section. In line with the strategy used for our empirical applications, we rely on the first $T_{tr} = 240$ time series observations for training, the next $T_{val} = 120$ for parameter validation, and the last $T_{test} = 240$ observations to form the actual data set for studying model performance. Once all the portfolios have been constructed (and in the case of AP-Trees, pruned to the set of 20 potentially overlapping basis assets) we estimate stock betas spanned by these basis assets.

Our focus is the estimation of the stock loadings $\beta_{t,i}$ on the SDF, because they incorporate all the relevant information for asset pricing. Naturally, they are equal to the conditional expected return (up to a constant of proportionality, since the mean of estimated SDF is not identified), and could be used to predict future stock returns. Second, a cross-sectional projection of asset returns on their betas is a valid estimate of the SDF itself. Third, projecting out the cross-sectional space spanned by the stock betas results in the residual component, which captures the pricing error. Hence, the recovery of almost all the objects of interest in asset pricing, i.e. return prediction, SDF estimation, and pricing errors, crucially relies on the precise estimates of these betas.

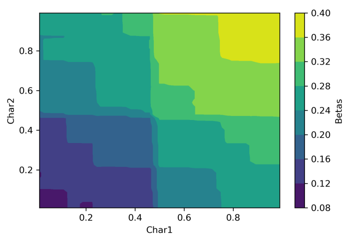
Figure 7 presents the scaled version of the estimated SDF betas (and hence, conditional expected returns) for different basis assets. Similar to the way portfolio buckets are used to estimate expected returns, we compute security betas by averaging the stock betas that belong to the same portfolios (and in case of the trees, averaging across the overlapping portfolios). For AP-Trees, we plot all the potential basis assets (both final and intermediate nodes of the trees). The pattern in expected returns is quite striking. Ideally, it should be close to the one used to generate the data (Figure E.1, Panel A). However, in practice there is a substantial difference between the type of shapes and figures one could get with double-sorted portfolios relative to those reflected by conditional trees. The difference remains substantial even in the simplest case of independent characteristics, where both tree-based portfolios and double-sorted ones are most similar to each other, having 1/16 of all the stocks (in case of the AP-trees, there are also intermediate nodes that contain 1/2, 1/4 and 1/8 of all the securities). Clearly, averaging betas across conditional basis asset allows to track the underlying SDF loadings in a the characteristic space.

Panel A: Double-sorted portfolios, spanned by DS-SDF



(a) Correlation = 0 **(b)** Correlation = 0.5 **(c)** Correlation = 0.9

Panel B: Conditional sorts, spanned by pruned AP-Trees



(d) Correlation = 0 **(e)** Correlation = 0.5 **(f)** Correlation = 0.9

Figure 7. Tracking expected returns: SDF loadings as a function of characteristics in a linear model.

Why does the correlation between characteristics matter? Double-sorted portfolios are based on the unconditional quantiles of the cross-sectional distribution of characteristics, and hence could naturally lead to rather unbalanced composition of the basis assets. For example, for the case of $\rho(C_i^{(1)}, C_i^{(2)}) = 0.9$, the joint density of stocks located in the north-west and south-east corners of the characteristic space is particularly low, therefore averaging expected returns of these securities produces a much noisier estimate compared to the other areas of the grid. In contrast, conditional quantiles, similar to the nearest neighbor predictor in nonparametric econometrics, adapt their bandwidth to the density of the data. This explains the difference in shapes across the AP-Trees basis assets as we change the correlation between characteristics.

For the sake of brevity, we have discussed here results for the case of linear factor loadings and present additional findings in the Appendix. In particular, Table E.1 in Appendix reports the Sharpe ratios of the SDFs spanned by different assets. The AP-Tree SDF has clearly higher out-of-sample Sharpe ratios compared to the DS approach. Note, that for this particular simulation setup, obtaining just a high SR is a too low bar. Indeed, as we mentioned earlier, in this particular setting the data is coming from a very simple one factor model, that drives both the time series of returns and their cross-sectional spread. As a result, almost any well-diversified stock portfolio (equally weighted, or long-short in a given characteristic, or a simply first principal component) would also be loading on the same latent factor, and, as a result, asymptotically recover both the highest SR, and an accurate SDF projection on returns. The true test of the basis assets lies in their ability to span expected returns in the characteristic space, and match its empirical counterpart, and this is where AP-Trees really shine.

To summarize, even for the case of factor loadings, linear in characteristics we find that: (1) Overall, conditional sorts build portfolios that provide a substantially more accurate reflection of expected returns in the characteristic space, and remain fully interpretable. (2) A small-dimensional, sparse set of the basis assets formed from AP-Trees, is able to span both the whole set of tree-based portfolios, and those constructed by double-sorting (the reverse, however, is not true). (3) The higher is the correlation among characteristics, the larger is the benefit of using conditional sorts.

V. Data

A. Returns and Firm Specific Characteristics Variables

We obtain monthly equity returns data for all securities on CRSP and construct decision tree-sorted portfolios using firm specific characteristics variables from January 1964 to December 2016, yielding 53 years in total. One-month Treasury bill rates are downloaded from the Kenneth French Data Library as the proxy for the risk-free rate to calculate the excess returns.

To build the sorting variables for the decision trees, we have constructed 10 firm-specific characteristics as defined on the Kenneth French Data Library. Note, that we have at least one anomaly for each major characteristic category.[6] All these variables are constructed from either accounting variables from CRSP/Compustat database or past returns from CRSP. Monthly updated variables are updated at the end of each month for use in the next month. Yearly updated variables are updated at the end of each June following the Fama-French convention. The full details on the construction of these variables are in the Appendix.

Table I Stock-specific characteristics

Past Returns	**Investment**	**Profitability**	**Intangibles**	**Value**	**Trading Frictions**
Momentum	Investment	Operating profitability	Accrual	Book to Market Ratio	Size
Short-term Reversal					Turnover
Long-term Reversal					Idiosyncratic Volatility

AP-Trees can naturally deal with missing values and do not require a balanced panel of firm characteristics or returns. Since we have collected all available data, it is not surprising that there will be missing values either due to database errors or other technical issues. For our tree portfolios sorted on a set of M characteristics variables, any stock that has a valid return on time t, market capitalized at the end of $t-1$ (for value-weighting purpose), and all M characteristic variables observable by the end of $t-1$ are included to construct the portfolio return at time t. For example, the tree portfolios return at time t sorted on two characteristics BEME and OP requires return information on time t and valid LME, BEME, and OP at $t-1$. A stock missing Investment information at $t-1$ will still be included in such tree portfolios. By this construction

[6] We are currently working on extending results to 45 characteristics.

with unbalanced panel data, we avoid not only the bias introduced from imputation, but also partially alleviate the survivorship bias.

Our main analysis focuses on triple interacted characteristics, because this is the only case where we have a natural counterpart in creating a cross-section: triple-sorted portfolios. We report the results for size interacted with any of the two other characteristics which results in total in 36 cross-sections[7].

VI. Empirical Results

A. *Estimation and Hyperparameter Tuning*

To minimize the possibility of overfitting, we divide all the data into 3 samples: training, validating, and testing data sets. By fixing the portfolio structures estimates from the training sample, and optimally choosing the tuning parameters on the validating data set, we focus on the out-of-sample behavior of triple-sorted portfolios vs those produced by pruned AP-Trees, and the SDFs, spanned by different basis assets. We consider a cross-section of portfolios, constructed with AP-Trees with depth 4, and its closest analogue in the standard methodology: 32 and 64[8] triple-sorted portfolios. All the portfolios are value-weighted. For additional robustness, we also exclude all the level 4 nodes that are sorted only on 1 characteristic to avoid going into extreme tails of the distribution without interaction with other variables. As an example, from the set of tree portfolios generated for size, value, and turnover, we exclude all the nodes corresponding to a single $\frac{1}{16}$th quantile of any of these characteristics.

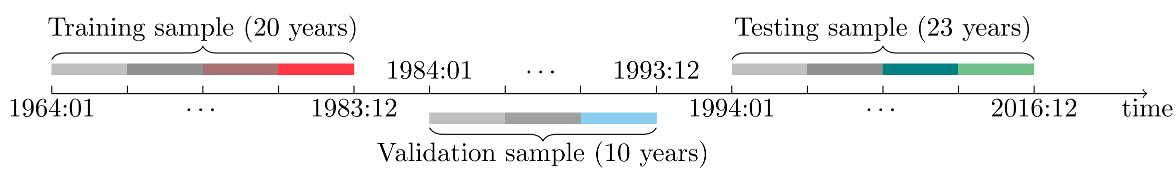


Figure 8. Timeline of the empirical strategy

Training sample. We use the first 20 years of data (1964 - 1983, or 240 monthly observations) to form AP-Trees of depth 4, and estimate the vector of average returns and covariance matrix for both final and intermediate nodes that are then used to construct an efficient frontier with elastic net. For different levels of shrinkage (lasso and ridge) and target expected return, we find the optimal test assets and the SDF spanned by them with the corresponding weights. The same is done for triple-sorted portfolios, where we also use elastic net to construct robust portfolio weights to mitigate the impact of estimation noise and sample variation. In other words, we allow triple-sorted portfolios to also benefit from the stability of shrinkage. This makes the overall comparison

[7]We are currently working on extending the results to all possible interactions with 45 characteristics.

[8]32 triple-sorted portfolios consider only a 50/50 split based on size, while 64 portfolios reflect all 3 characteristics in a similar way, and hence could provide a somewhat more justified benchmark for AP-Trees.

more fair, and allows us to focus on the direct impact of using different basis assets to span the SDF.

Validating sample. The middle 10 years (1984 - 1993, 120 monthly observations) of returns serve as the validation set for hyper-parameter tuning: we pick the model based on Sharpe Ratio of the tangency portfolio on the validation dataset, fixing the SDF weights at their training values. Table C.1 reports the hyperparameters we used for the SDF construction. For each combinations of (μ_0, K, λ_2) the lasso penalty λ_1 is chosen such that the number of non-zero weights reaches the target number K. In particular, we tune the value of λ_1 for AP-Trees to select 40 portfolios, which makes the dimension of pruned trees comparable to Fama-French triple-sorted 32 and 64 portfolios

Testing sample. The last 23 years of monthly data are used to compare basis assets, recovered by AP-Trees and triple sorting. We fix portfolio weights and their selection at the values, estimated on the training sample, and tuning parameters chosen with the validation, making, therefore, all the performance metrics effectively out-of-sample. We focus on Sharpe ratios and portfolio performance in the spanning tests, and compute all the statistics on the testing sample only.

B. Evaluation Metrics

Our main goal is to compare the performance of tree-based portfolios with the information spanned by triple sorted portfolios. Therefore, we focus on the following cross-sections:

- *AP-Trees*: The pruned version of AP-Trees, consisting of 40 basis assets, selected from the final and intermediate nodes, based on conditional sorting with depth 4;
- *TS (32)*: The 32 triple-sorted portfolios are constructed the same way as Fama-French triple sorting portfolios for the combinations of three characteristics, that is by combining a single cut on size and two splits in the other two characteristics.
- *TS (64)*: The 64 triple-sorted portfolios that have two splits on all three characteristic dimensions, leading to a set of $4 \times 4 \times 4$ basis assets.

We evaluate the pricing information of these basis assets in two different dimensions: First, how well can these basis assets be priced by conventional factor models? Second, how well can the SDF constructed with these basis assets price alternative cross-sections? We run the standard asset pricing tests of different portfolios, build on the same characteristics, against the most popular reduced form models:

- *FF3*: Fama-French 3 factor model with a market, size and value factor;
- *FF5*: Fama-French 5 factor model which adds an investment and profitability factor to *FF3*;
- *XSF*: a cross-section specific model, that includes the market factor, and three long-short portfolios, corresponding to the three characteristics used to build a cross-section. In order to build the factors, we use the same approach as used in constructing HML, momentum, or other conventional long-short portfolios.
- *FF11*: an 11-factor model, consisting of the market factor and all 10 long-short portfolios, based on the full list of characteristics.

For each combination of characteristics (36 cross-sections) we report:

1. *SR*: The out-of-sample Sharpe ratio of the SDF constructed with *AP-Trees*, *TS(32)* and *TS(64)*. In each case we use the mean-variance efficient portfolio with optimal shrinkage.
2. α: The t-statistics of the SDF pricing error α, obtained from an out-of-sample time-series regression of the corresponding SDF on different factor models. Note that since the mean of the SDF is generally not identified, we focus on the t-statistics, corresponding to its alpha, which leads to a more balanced comparison across different cross-sections.
3. α_i: Pricing errors for individual basis assets, that are estimated from a time series regression of portfolio return on a set of candidate factors. Note, that these pricing errors are specific to a choice of cross-section, portfolio, and a test model.
4. *XS-*R^2 The cross-sectional R^2 is the adjusted cross-sectional pricing error for the basis assets defined as:

$$R^2 = 1 - \frac{N}{N-K} \frac{\sum_{i=1}^{N} \alpha_i^2}{\sum_{i=1}^{N} E[R_i]^2}.$$

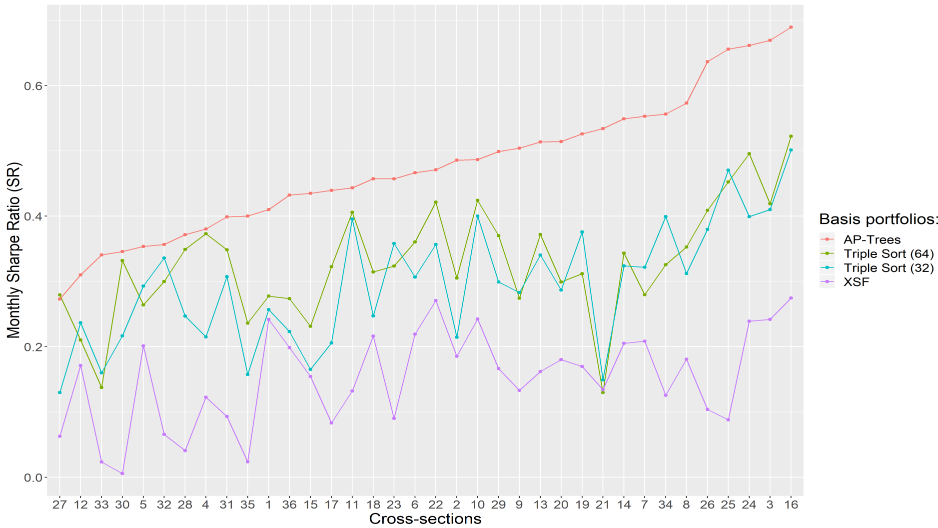
   We use the uncentered version of R^2 to make sure it captures the pricing errors that are not only *relative* to the other assets, but also specific to the overall cross-section of securities, so it reflects both common and asset-specific levels of mispricing.

C. 36 Cross-Sections of Expected Returns

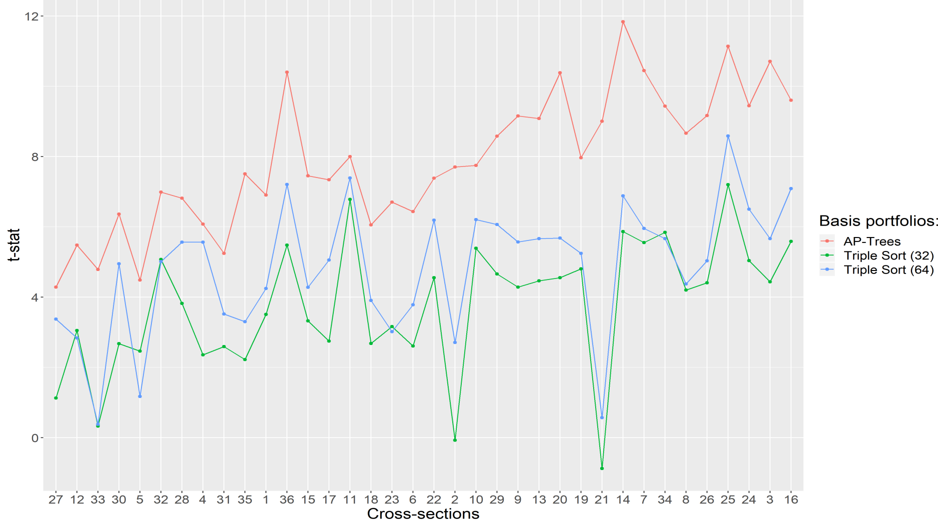
We start by comparing the SDFs spanned by cross-sections with different basis assets. Figure 9a summarizes their Sharpe ratios, when using *AP-Trees*, *TS(32)*, or *TS(64)* as the 'building blocks' of the SDF, along with the set of cross-section-specific long-short portfolios, accompanied by the market factor.

AP-Trees obtain considerably higher out-of-sample Sharpe ratios compared to the triple-sorted portfolios or conventional long-short factors. These 36 cross-sections are arranged according to the out-of-sample Sharpe ratio achieved with *AP-Trees*, with their labels and corresponding values reported in Table C.2. The differences in Sharpe ratios are striking. Compared to the case of simple long-short factors, our basis assets are able to deliver SR that are up to three times higher. As the Sharpe ratio measures the mean-variance efficiency of the SDF constructed with different basis assets, it implies that our AP-Trees extract more pricing information compared to the conventional benchmark basis assets. These results are particularly strong for such characteristics as investment, idiosyncratic volatility, and profitability.

Where does this superior performance come from? First, note that cross-section-specific factors have around half of the Sharpe ratios spanned by triple sorts, i.e. the linear factors already miss information. This is expected, since even by construction the long-short factors cannot efficiently account for the interactions between the characteristics, while triple sorting can reflect their impact at least to a partial extent. Having one or two splits in the size dimension, i.e. 32 or 64 triple sorted portfolios, does not seem to have a substantial impact either.



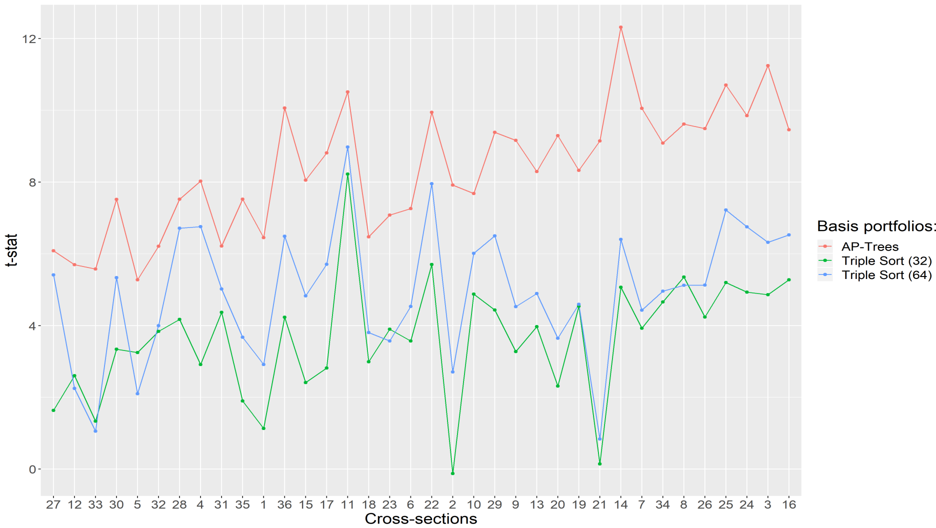
(a) Monthly out-of-sample Sharpe ratios of mean-variance efficient portfolios spanned by pruned AP-Trees, triple sorts, and XSF. Cross-sections are sorted by the SR achieved with AP-Trees.



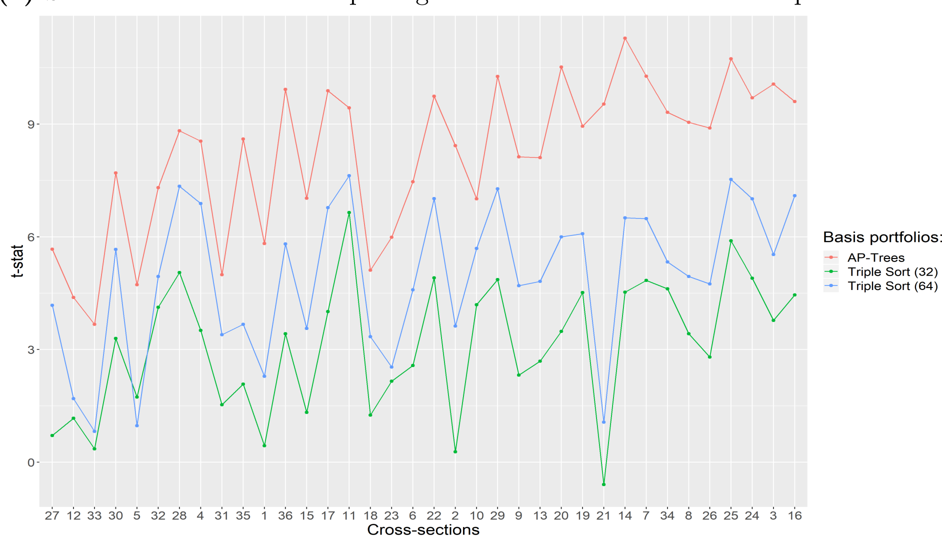
(b) SDF α: t-statistics of the pricing errors relative to the Fama-French 5 factor model.

Figure 9. Sharpe ratios and pricing errors of the SDFs spanned by AP-Trees and triple sorts.

Could it be that the difference in SR is simply driven by a higher loading on conventional risk factors? This does not seem to be the case: tree-based portfolios are substantially harder to price using any standard factor model. Indeed, Figure 9b confirms that the pricing errors of the different SDF portfolios have much higher t-statistics for AP-Trees. In fact, the pattern in these pricing errors aligns *exactly* with the total SR achieved for different characteristics. While Fama-French 5-factor model successfully spans some of the cross-sections built with triple sorts, it fails to capture the information reflected in AP-trees. Consider, for example, the case of size, value, and profitability (cross-section 2). The SDF, spanned by triple-sorted portfolios, does not have a significant alpha, when pitted against Fama-French 5 factors, while the one build from AP-trees has a t-statistics of 8. In other words, pricing conventional triple-sorted cross-sections, or spanning the factors that explain them, may be too low of a hurdle, and has a real chance of missing important information, contained in the original universe of stocks.



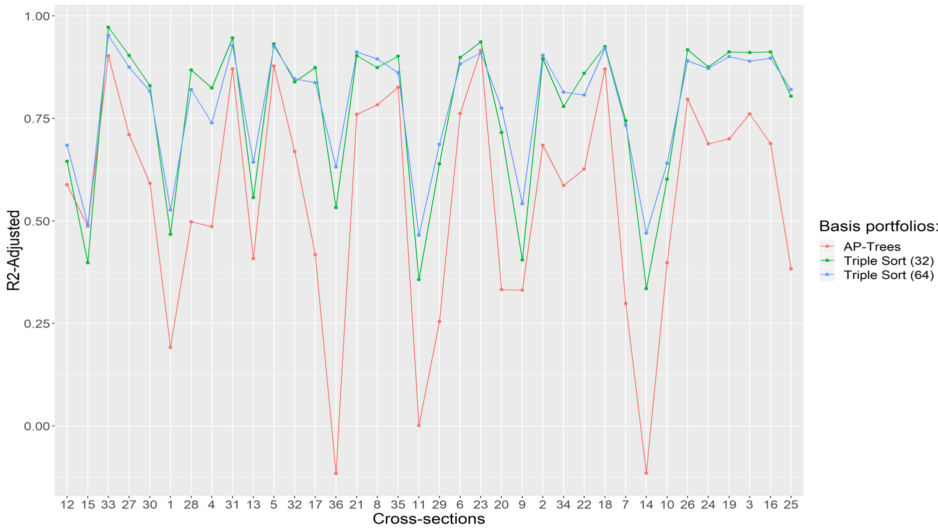
(a) SDF α: t-statistics of the pricing error relative to cross-section-specific factors.



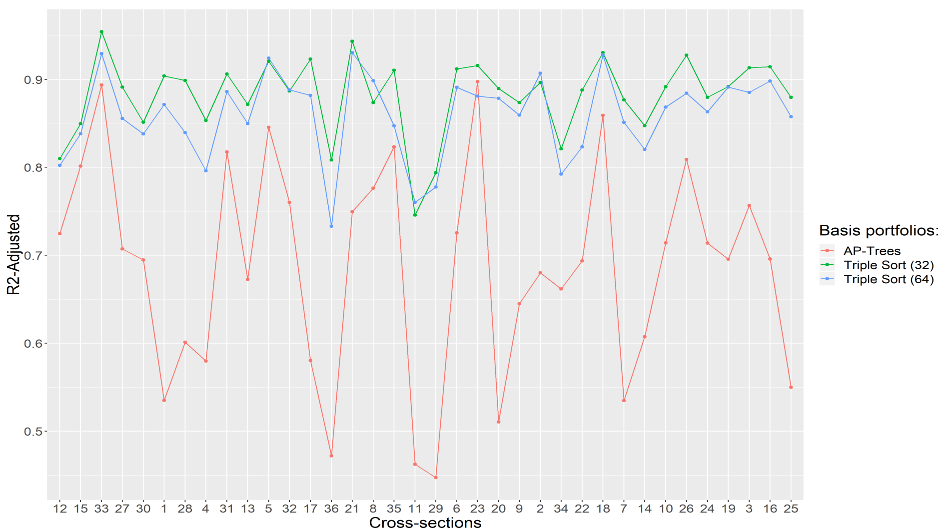
(b) SDF α: t-statistics of the pricing error relative to the market factor and 10 long-short portfolios.

Figure 10. Pricing errors of the SDFs spanned by AP-Trees and triple sorts.

These findings are robust to the choice of benchmark factors. They are almost identical to the t-statistics obtained with cross-section-specific factors (see Figure 10a), and survive even when faced with a whole set of 10 long-short portfolios, built from all the characteristics used in our application (Figure 10b). Interestingly, even when using all 11 cross-sectional factors, our AP-Trees have uniformly significant pricing errors. In contrast, around one third of the triple sorts can be explained by the large set of factors. This again provides a strong support for using AP-trees as more informative test assets.



(a) *XS-*R^2: pricing portfolios with 5 Fama-French factors.



(b) *XS-*R^2: pricing portfolios with XS-specific factors.

Figure 11. Cross-sectional fit of popular asset pricing models.

Cross-sectional R^2 is another benchmark often used to check whether the candidate model spans asset returns, and they confirm our alpha-based findings. Figures 11a and 11b demonstrate the pricing ability of the Fama-French 5 and cross-section specific factors to span different sets of basis assets. Typically, cross-sections that obtain an R^2 of more than 80% would be considered as being well-explained by a linear factor model, and this is largely the case for triple sorts. In contrast, one hardly gets the fit of over 50% on AP-Trees, if at all positive. This again illustrates that conventional sorting does not provide a sufficiently high hurdle for asset pricing models.

D. How Many Portfolios?

In the previous section we chose the same number of portfolios to prune the AP-Trees in order to make results comparable across different cross-sections, without the additional contamination by the degrees of freedom. However, in practice, increasing the number of portfolios may not always be beneficial: as always, it represents the trade-off between the bias and variance: a larger number of portfolios could generally yield a higher rate of return, highlighting the areas of the SDF in the portfolio space that are particularly challenging to price, or generally heterogenous in their implied risk exposure. However, doing so often leads to a more fragmented nature of the SDF, potentially unnecessary duplicating some of the data features. Moreover, depending on the information, reflected in a given set of characteristics, a simple and rather parsimonious structure could often be enough to capture all the properties of the SDF projected on them. This naturally raises the question of sparsity, - how many basis assets are enough to capture most of the stylized features of the data? The use of lasso penalty as a tuning parameter, governing the number of chosen basis assets, provides a natural way to investigate this question.

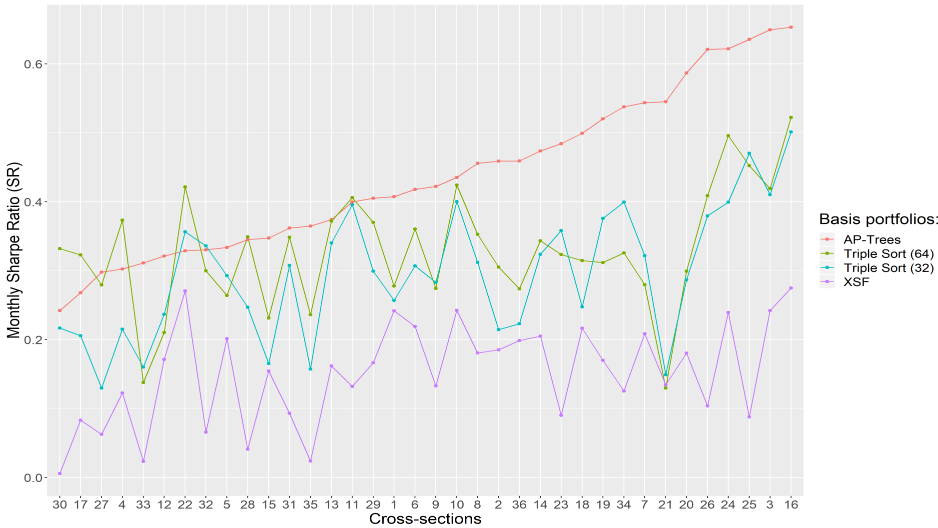
We find that for most of the cross-sections all the empirical results largely carry through even for a relatively small number of basis assets. Figure 12 demonstrates monthly out-of-sample Sharpe ratios, and pricing errors of the SDFs spanned by only 10 portfolios, based on trees, relative to those of triple sorts (32/64 portfolios) and the corresponding long-short factors. For most of the cross-sections using only 10 portfolios is enough to retain roughly 90% of the original Sharpe ratio and its alpha relative to standard Fama-French 5 factor model[9]. Of course, some characteristics are more heterogenous in their impact than others, and generally the optimal number of portfolios will differ, depending on the complexity of the conditional SDF, projected on these characteristics, and could be chosen optimally based on the validation sample, or other data-driven techniques. The general point we would like to highlight, is that the number of portfolios used to build a cross-section and successfully priced by a given model, is often a poor reflection of the underlying information captured by that set of assets, with *effective* degrees of freedom quite often different from the sheer number of such portfolios. Naturally, simple measures of fit are not robust to recombining assets into larger, denser portfolios, and as a result, are prone to a substantial bias.

The key reason behind the ability of AP-Trees to retain a large amount of information in a small number of assets, lies in the pruning methodology outlined in Section III.D. Since our method is designed to select portfolios not only in the type and value of the characteristics used for splitting, but also the *depth* of the latter, it will naturally yield portfolios that contain a larger number of securities (both in count and market cap), effectively merging smaller assets together, as long as the reduction in variance is large enough to compensate for a potential heterogeneity in returns. Indeed, we find that intermediate nodes of the trees, and often the market itself, constitute a substantial fraction of the chosen basis assets, and contribute to the expected return, its variance, and the pricing error of the tangency portfolio spanned by these combinations of stocks.

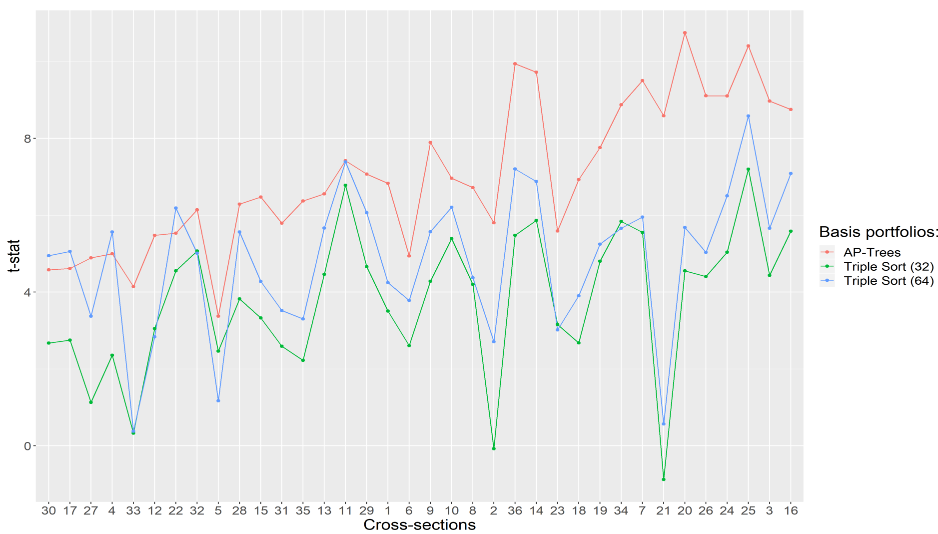
We now turn to a particular example of the cross-section to investigate further the structure

[9] These results largely remain the same for other asset pricing models, and are presented in the Appendix

of the SDF, spanned by AP-Trees, the characteristics of the optimal basis assets, and the patterns they reveal in expected returns.



(a) Monthly out-of-sample Sharpe ratios of mean-variance efficient portfolios spanned by pruned AP-Trees, triple sorts, and XSF. Cross-sections are sorted by the SR achieved with AP-Trees.



(b) SDF α: t-statistics of the pricing errors relative to the Fama-French 5 factor model.

Figure 12. Sharpe ratios and pricing errors of the SDFs spanned by AP-Trees (10) and triple sorts.

E. Zooming into the Cross-Section: Size, Operating Profitability, and Investment

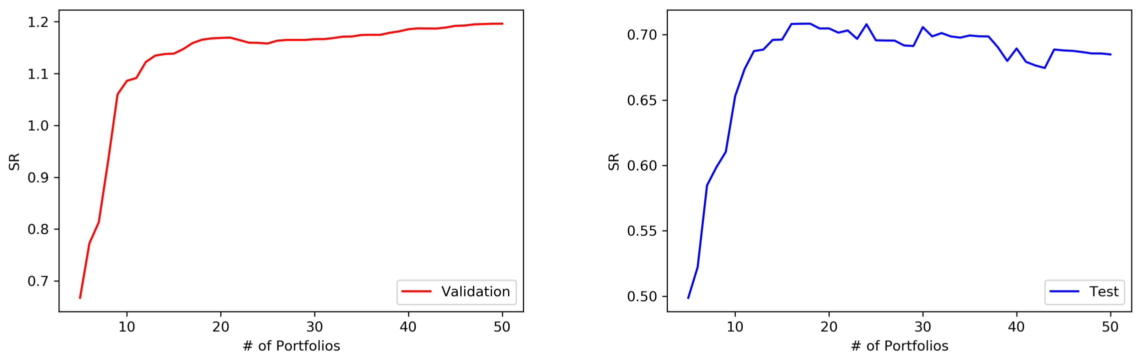
We focus on a representative cross-section to better understand the source of the pricing performance of AP-Trees. Consider the set of portfolios, that could be built to reflect size, investment, and operating profitability. Table II presents the summary statistics for cross-sections, built with AP-Trees and triple sorts. As we have already observed previously, the AP-Trees have higher SR and generally larger pricing errors α (or equivalently, lower cross-sectional R^2_{adj}) than the conventional portfolios, based on the unconditional quantiles. Interestingly, almost all of the information, relevant for asset pricing, is already contained in the set of 10 portfolios. A small number of assets is intuitively appealing, and allows to easily analyze the source of the model performance.

Table II Cross-sections based on Size, Operating Profitability and Investment

		Type of the cross-section			
		AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)
SDF SR		0.65	0.69	0.50	0.52
α	FF3	0.942	0.903	0.705	0.717
		[10.11]	[11.033]	[7.466]	[8.733]
	FF5	0.813	0.762	0.420	0.496
		[8.76]	[9.603]	[5.588]	[7.090]
	XS-F	0.813	0.759	0.416	0.491
		[8.77]	[9.457]	[5.278]	[6.528]
	FF11	0.886	0.803	0.344	0.512
		[9.12]	[9.601]	[4.455]	[7.092]
$XS\ R^2_{adj}$	FF3	3%	52%	82%	82%
	FF5	2%	65%	91%	90%
	XSF	18%	66%	91%	90%
	FF11	–	44%	92%	87%

The table presents the aggregate properties of the cross-section based on Size, Investment, and Operating Profitability, created from AP-Trees (pruned to 10 and 40 portfolios correspondingly), and triple sorts (32 and 64 portfolios). For each of the cross-section the table reports its monthly Sharpe ratio on the test sample, along with the alpha of the SDF spanned by the corresponding basis portfolios. Alphas are computed relative to Fama-French 3 and 5 factor model, cross-section-specific factors (market and long-short portfolios, reflecting size, investment, and operating profitability), and the composite FF11 model that includes the market portfolio, along with the 10 long-short portfolios, based on the cross-sectional characteristics.

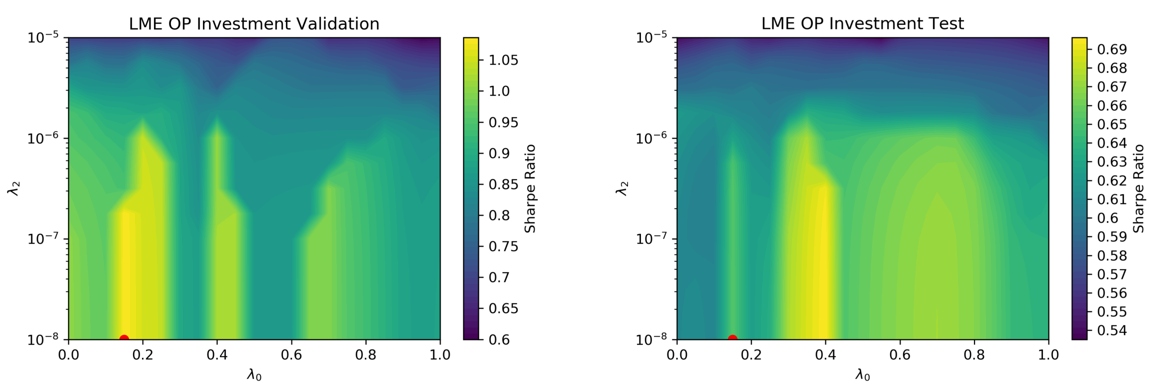
The selection of 10 portfolios, while seemingly surprising, is supported by the data, with both the validation and testing sample suggesting that having 10 portfolios is enough to capture most of the variation in asset returns, without introducing the danger of overfitting (see Figure 13). This decision is therefore in line with the traditional approach of choosing the optimal tuning parameter with one standard deviation of the one maximizing cross-validation of other criteria, - inherently, both methods aim at selecting a sparse set of parameters/portfolios that give a performance very similar to that of the first best, but relatively denser, selection, and present an almost an unbiased estimate of the optimal shrinkage.



(a) Validation sample **(b)** Testing sample

Figure 13. SR as a function of portfolio number, ex ante (validation)and ex post (oracle)

The heatmap of the tuning parameters indicates the range of shrinkage that was chosen as optimal on the validation dataset (Figure 14). In this particular case the impact of L2 penalty (ridge) on the composition of the cross-section was rather small, however, the model performance crucially depended on the choice of the shrinkage to the mean, λ_0. This value, which out-of-sample, could have been even higher, corresponds to the shrinkage of extreme sample returns towards the average in the cross-section, reflecting the intuition that returns both too high, and too low relative to the benchmark, are probably due to the statistical error, and will revert back at some point in the future. In this particular case, the chosen value of λ_0 was 0.15, corresponding to an important, but not excessive shrinkage towards the minimum variance portfolio induced by the estimation error in expected returns.



(a) Validation sample **(b)** Testing sample

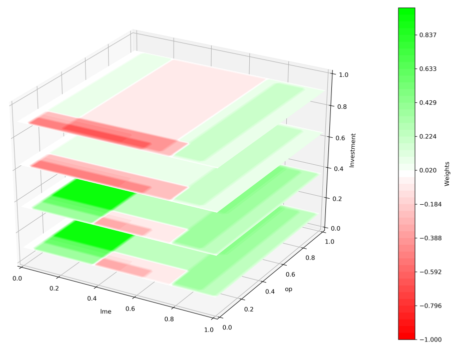
Figure 14. SR as a function of tuning parameters, ex ante and ex post (oracle)

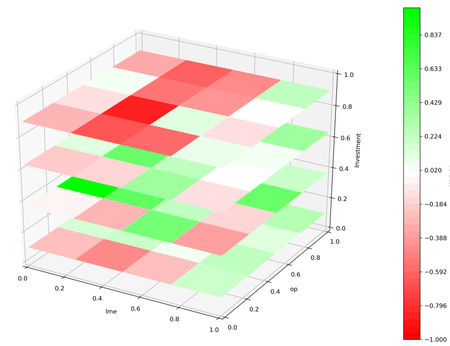
What are these 10 portfolios, left after pruning the tree? Table III, Panel A, provides the description of these basis assets, and their main properties: the relative fraction of stocks, that goes into their creation, and their value-weighted quantiles, based on the market cap of the securities that go into a corresponding portfolio, as well the pricing errors, associated with the leading asset pricing models. Out of 10 selected portfolios, only 5 correspond to the final nodes of the trees, and contain just over 6% of all the stocks, - other assets include trading strategies that can be constructed from only one or two splits based on the characteristics, and even the market itself. For example, 1221.1111 is created by taking bottom 50% of the stocks based on their size (LME), and within them the lowest quartile based on investment, as a result, containing 12.5% of the stocks at all the time periods. This portfolio presents a challenge to all the baseline tradable asset pricing models, which is significant not only statistically, but economically as well, yielding a monthly alpha of about -30 b.p. Similarly, a portfolio 3331.12221, which is constructed by taking 1/8 of the stocks, highest in investment, and then choosing the half of them, smallest by the market cap, has an alpha of 41-86 bp, depending on the underlying model.

Figure 15 displays the structure of the SDF, conditional on characteristics, and the pricing errors for each of the individual basis assets with respect to XS-specific factors, with the candlestick denoting 5% confidence intervals. Overall 6 of the chosen 10 tree-based portfolios have consistently significant alphas. These portfolios are not just reflecting extreme quantiles of the underlying characteristics, but are in fact characterized by a complex interaction structure. The prevalence of pricing errors, and the diversity of the stocks that go into such portfolios, are robust to the choice of risk factors, as almost the same pattern persist, when basis assets are priced with the Fama-French 5, or 11 cross-sectional factors (see Table III, Panel A). Examining conventional triple sorts, in turn, reveals a completely different pattern (see Figure 15, Panel (d)) with generally substantially smaller pricing errors, that are largely subsumed by standard risk factors.

Table III, Panel B lists top 10 portfolios from the cross-section of 64 triple sorted assets, that are most difficult to price, according to the most comprehensive, FF11 model. While there are obviously substantial alphas, associated with some of these portfolios, the sheer number of them is somewhat misleading: most of these portfolios consist of a small number of stocks (usually about 1.5%), and often reflect securities that are very similar to each other in terms of characteristics: small in size and profitability, high in investment. Since our pruning algorithm selects the basis assets in both characteristics and depth, these portfolios are actually often grouped together by AP-Trees, bundling these securities together. This pattern is particularly visible if you compare the structure of the SDF, spanned by different portfolios (see Figure 15, Panels a) and b)). For example, the stocks that are generally high in investment (the top layer of the 3-dimensional graph in Panel a)) are roughly grouped together and present the same exposure to the SDF, compared to 6 separate portfolios spanning the same characteristic space in triple sorting. Naturally, since the tree-based assets are constructed with both conditional and unconditional quantiles, the actual space of returns could be somewhat different, but the general pattern remains, - there is often not enough signal in the data to warrant such a granular split.

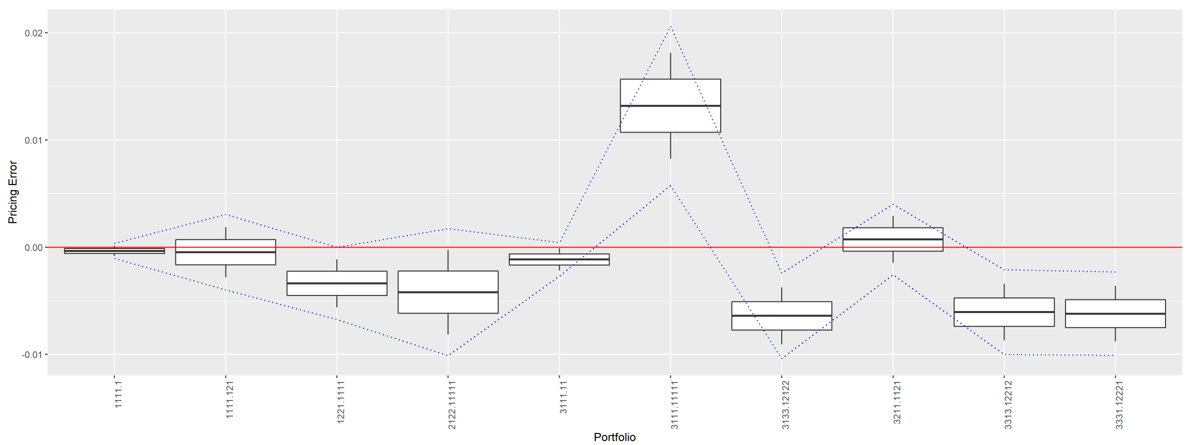
While some of the general patterns of loadings are shared across the cross-sections, it is immediately clear that AP-Trees reflect a more sophisticated data generating process that triple sorts aim to capture only with a very coarse grid. The ability of conditional sorts to map the finer resolution of returns in the characteristic space without heavily loading on poorly-diversified portfolios, allows us to uncover different long-short patterns in the data, and could present a new challenge for structural models. Overall, it seems that the triple-sorted portfolios aim to capture roughly the same underlying patterns in returns, but lack the ability to flexibly adapt the weights, based on the incremental information.



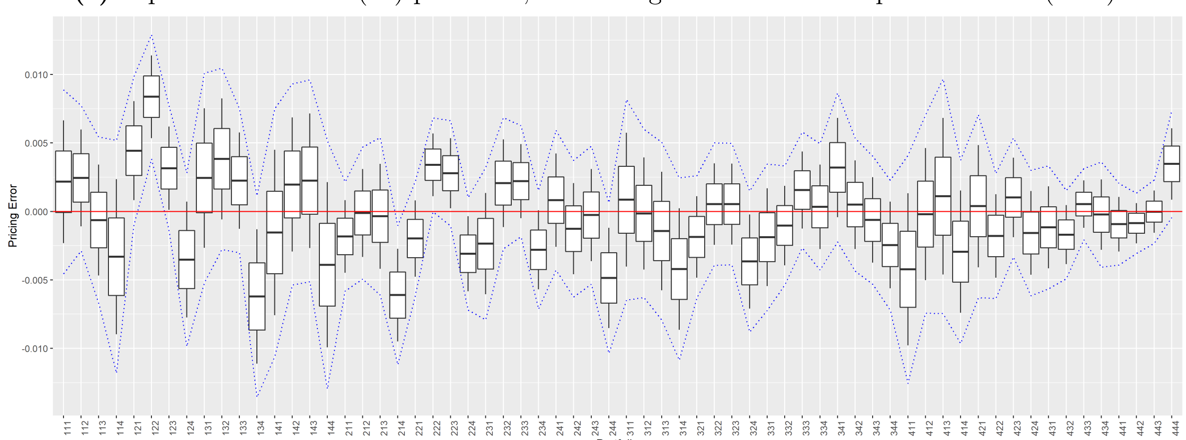


(a) SDF spanned by AP-Trees (10)

(b) SDF spanned by TS (64)



(c) Alphas of AP-Trees (10) portfolios, controlling for cross-section specific factors (XSF)



(d) Alphas of 64 triple-sorted portfolios, controlling for cross-section specific factors (XSF)

Figure 15. Composition of conditional SDF with 10 AP-Tree and 64 TS basis assets.

Table III Portfolios in the cross-sections

Portfolio ID	Portfolio construction	% of stocks	WV q_{LME}	α_{FF3}	α_{FF5}	α_{XSF}	$\alpha_{FF_{11}}$
Panel A: 10 AP-Tree portfolios							
1111.1	Market	100%	0.945	–	–	–	–
1111.121	LME(0.5-0.75)	25%	0.707	-0.034	-0.366	-0.046	-0.026
				[-0.340]	[-0.356]	[-0.396]	[-0.259]
3111.11	OP(0-0.5)	50%	0.941	0.030	-0.111**	-0.114**	-0.109**
				[-0.536]	[-2.331]	[-2.198]	[-2.455]
3111.11111	Inv(0-0.5)→LME(0-0.125)	6.25%	0.070	1.487***	1.316***	1.318***	1.631***
				[6.144]	[5.328]	[5.328]	[6.474]
3211.1121	Inv(0-0.5)→OP(0.5-1)→LME(0-0.5)	12.5%	0.505	0.227***	0.087	0.073	0.0628
				[2.953]	[1.193]	[0.661]	[0.831]
1221.1111	LME(0-0.5)→OP(0-0.25)	12.5%	0.372	-0.365***	-0.336***	-0.337***	-0.285**
				[-2.809]	[-3.000]	[-3.010]	[-2.413]
2221.11111	OP(0-0.25)→LME(0-0.5)	6.25%	0.192	-0.162	-0.215	-0.207	-0.050
				[-0.717]	[-0.949]	[-0.892]	[0.213]
3331.12221	Inv(0.875-1)→LME(0-0.5)	6.25%	0.452	-0.815***	-0.616***	-0.620***	-0.460***
				[-5.858]	[-4.814]	[-4.783]	[-3.746]
2122.11111	OP(0-0.5)→LME(0-0.5)→OP(0-0.25)	6.25%	0.266	-0.374*	-0.427**	-0.419**	-0.200
				[-1.948]	[-2.207]	[-2.126]	[-0.997]
3133.12122	Inv(0.5-1)→LME(0-0.5)→Inv(0.75-1)	6.25%	0.475	-0.842***	-0.636***	-0.640***	-0.495***
				[-5.918]	[-4.843]	[-4.814]	[-3.922]
Panel B: Top 10 portfolios from TS (64)							
122	LME(0-0.25)&OP(0.25-0.5)&Inv(0.25-0.5)	1.91 %	0.173	0.974***	0.841***	0.837***	0.948***
				[6.609]	[5.603]	[5.541]	[6.012]
214	LME(0.25-0.5)&OP(0-0.25)&Inv(0.75-1)	1.55 %	0.396	-0.93***	-0.61***	-0.611***	-0.659***
				[-5.272]	[-3.602]	[-3.613]	[-3.768]
132	LME(0-0.25)&OP(0.5-0.75)&Inv(0.25-0.5)	1.03 %	0.172	0.393*	0.396*	0.383*	0.769***
				[1.932]	[1.897]	[1.736]	[3.651]
121	LME(0-0.25)&OP(0.25-0.5)&Inv(0-0.25)	1.9 %	0.165	0.594***	0.447**	0.443**	0.663***
				[3.378]	[2.482]	[2.45]	[3.562]
112	LME(0-0.25)&OP(0-0.25)&Inv(0.25-0.5)	1.66 %	0.17	0.243	0.251	0.244	0.486***
				[1.449]	[1.451]	[1.38]	[2.786]
123	LME(0-0.25)&OP(0.25-0.5)&Inv(0.5-0.75)	1.47 %	0.173	0.379**	0.318**	0.316**	0.428***
				[2.557]	[2.099]	[2.077]	[2.71]
222	LME(0.25-0.5)&OP(0.25-0.5)&Inv(0.25-0.5)	1.95 %	0.399	0.462***	0.35***	0.34***	0.28**
				[4.693]	[3.521]	[2.975]	[2.687]
244	LME(0.25-0.5)&OP(0.75-1)&Inv(0.75-1)	1.18 %	0.408	-0.567***	-0.467***	-0.486**	-0.342**
				[-3.435]	[-3.183]	[-2.653]	[-2.278]
334	LME(0.5-0.75)&OP(0.5-0.75)&Inv(0.75-1)	1.08 %	0.656	0.345**	0.337**	0.321*	0.367**
				[2.325]	[2.179]	[1.768]	[2.251]
111	LME(0-0.25)&OP(0-0.25)&Inv(0-0.25)	5.04 %	0.16	0.42*	0.207	0.216	0.483**
				[1.892]	[0.953]	[0.968]	[2.192]

The table presents the properties of the portfolios, spanning the impact of size, investment and operating profitability on asset returns. Panel A it presents the set of 10 portfolios, created from trees of depth 4 (that exclude the extreme portfolios, representing 1/16th of the stocks sorted by the same characteristic), and their features: the average percentage of currently available stocks, included in the portfolio, their value-weighted quantile based on size, and alphas with respect to Fama-French 3(5) models, the model that includes cross-section-specific factors in addition to the market, and the FF11 model (market and 10 long-short portfolios based on all the available characteristics), with the corresponding t-stats in the brackets. *, **, and *** correspond to 10%, 5%, and 1% significance levels. Panel presents the same statistics for 10 portfolios out of 64, created by triple sorting, that are the most challenging to price, according to the composite FF11 model.

The excessive granular nature of the triple sorted portfolios can also mask the true fit of the leading asset pricing models. Suppose that there is a group of portfolios that are perfectly spanned by a given set of risk factors, and do not command a separate risk premia, or provide an alternative exposure to these risk factors. Treating them as a separate group of assets does not yield better investment opportunities, and does not reveal an informative pattern in returns. Yet, the sheer number of these perfectly priced portfolios will substantially increase the quality of cross-sectional fit, leading to higher R^2, based on the simple OLS estimates. This is precisely why for many popular cross-sections using GLS to evaluate the model performance often leads to a substantially lower measure of fit (Lewellen, Nagel and Shanken (2010)), since treating separately these assets does not provide incremental Sharpe ratio, which is reflected in the GRS statistic and other quantities that target investment opportunities, rather than a linear measure of fit.

Our empirical results are unlikely to be driven by micro caps. First, in constructing the trees, we specifically eliminated extreme groups of stocks, that are heavily loading on a single characteristic, for example, size. We excluded all the splits that use the same characteristic to make the splits (e.g. 16 portfolios, sorted only by size). Second, not only tree-based portfolios are generally composed of a larger number of stocks, efficiently diversifying the idiosyncratic noise, and endogenously grouping similar securities together, they are often comparable, if not better, to the triple sorts in terms of the actual market cap of the stocks that drive most of the difference in performance. Table III presents the value-weighted (since the portfolios are value-weighted themselves) size quantile of the stocks, that comprise each of the basis assets. Compared to the triple sorts, there is only one portfolio that loads heavily on the small caps (bottom 50% on investment, and bottom 12.5% on the size within), and the rest have the same market cap as those based on triple sorts, or larger. Lastly, the benefit of the tree-based approach lies in it's intuitive and adaptive structure: if the market cap of two portfolios after the split is too low to be considered reliable for inference, one could explicitly impose such a restriction on constructing the trees, making all the potential nodes balanced not only in terms of the number of stocks, but also their market size.

Finally, we can check whether most of the superior performance of the SDFs based on AP-Trees is coming at the expense of an unrealistically high turnover, that could prove it impossible to implement the strategies empirically (see Table IV).

For the baseline case of monthly rebalancing, an SDF portfolio, spanned by AP-Trees, achieves a monthly turnover in the range of 20-35%, depending on the leg, which is generally comparable to investing in the corresponding long-short factors, while providing a higher Sharpe ratio. With annual rebalancing, Sharpe ratio naturally declines, but remains substantial, contrary to many passive long-short strategies. Pricing errors, cross-sectional fit, and other empirical results remain similar to the baseline case, and are available upon request.

Table IV Portfolio turnover.

		AP-Trees	XSF	Prof	Invest
SR	Monthly	0.66	0.30	0.10	0.08
	Yearly	0.38	0.29	-0.07	-0.10
Turnover+	Monthly	0.30	0.24	0.35	0.35
	Yearly	0.15	0.13	0.11	0.12
Turnover-	Monthly	0.25	0.24	0.28	0.28
	Yearly	0.09	0.11	0.09	0.09

Sharpe ratios and portfolio turnovers are converted to a monthly scale. The table compares the performance of the SDF, spanned by AP-Trees, with the one captured by cross-sectional factors (used individually, or as a combination). For AP-Trees and XSF, we fix the weights at the portfolio level, and track the performance of the strategy with monthly or yearly rebalancing. Portfolio turnover is reported for the long and short legs separately.

VII. Conclusion

We propose a novel way to build basis assets for asset pricing that capture the complex information of a large number of cross-sectional stock return predictors. Our Asset-Pricing Tree portfolios are a small number of long-only portfolios that (1) reflect the information in many stock-specific characteristics allowing for conditional interactions and non-linearities, (2) provide test assets for asset pricing that are considerably harder to price than conventional cross-sections, and (3) act as the building blocks for a stochastic discount factor (SDF) that performs well out-of-sample in various empirical applications. Our approach generalizes the idea of characteristic-based sorting to decision trees to better capture complex interactions between many characteristics, and select a sparse set of portfolios with the most relevant and non-redundant information. We show that conventional cross-sections do not fully reflect the information contained in the characteristics of the underlying stocks, and often present a rather crude, if not misguided, description of the expected returns.

We encourage the use of novel test assets to discipline the discovery of asset pricing factors, and provide a better evaluation and diagnostics of the structural model performance. Existing double-, or triple-sorted cross-sections, as well as their typical combinations, not only are a poor reflection of the underlying stock returns, but also have been substantially overstudied in much of the empirical literature, also contributing to the current problems of publication bias and factor zoo (e.g. Harvey, Liu, and Zhu (2015)).

Our cross-sections can be used as alternative test assets, providing both a more reliable representation of the underlying stock returns, as well as a fresh way of testing whether existing models are really capable of explaining the impact of a given set of characteristics on returns. Compared to PCA and other related dimension reduction techniques, our test assets are easily interpretable, and therefore, could be particularly useful as a diagnostic tool for the areas of model mispricing, whether the SDF is in a linearized reduced form or coming from a structural model. We are planning to make a public library of the new cross-sections built for the most relevant stock characteristics.

REFERENCES

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2004): "The Cross-Section of Volatility and Expected Returns," Working Paper 10852, National Bureau of Economic Research.

Cattaneo, M. D., R. Crump, M. H. Farrell, and E. Schaumburg (2019): "Characteristic-Sorted Portfolios: Estimation and Inference," *Review of Economics and Statistics*.

Chen, L., M. Pelger, and J. Zhu (2019): "Deep Learning in Asset Pricing," *Working paper*.

Cochrane, J. H. (2011): "Presidential Address: Discount Rates," *The Journal of Finance*, LXVI(4), 1047–1108.

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *Journal of Finance*, 63(4), 1609–1651.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998): "Liquidity and stock returns: An alternative test," *Journal of Financial Markets*, 1(2), 203 – 219.

DeMiguel, V., A. Martin-Utrera, F. Nogales, and R. Uppal (2019): "A Transaction-Cost Perspective on the Multitude of Firm Characteristics," *The Review of Financial Studies*.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47(2), 427–465.

Fama, E. F., and K. R. FrencH (1996): "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51(1), 55–84.

Fama, E. F., and K. R. French (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1 – 22.

Fan, J., Y. Liao, and W. Wang (2016): "Projected Principal Component Analysis in Factor Models," *Annals of Statistics*, 44(1), 219–254.

Feng, G., S. Giglio, and D. Xiu (2019): "Taming the Factor Zoo: a Test of New Factors," *The Journal of Finance*.

Freyberger, J., A. Neuhierl, and M. Weber (2019): "Dissecting Characteristics Nonparametrically," *Review of Financial Studies*.

Gu, S., B. T. Kelly, and D. Xiu (2019): "Empirical Asset Pricing Via Machine Learning," *Review of Financial Studies*.

Harvey, C. R., Y. Liu, and H. Zhu (2015): "... and the Cross-Section of Expected Returns," *The Review of Financial Studies*, 29(1), 5–68.

Jegadeesh, N., and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance*, 48(1), 65–91.

——— (2001): "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations," *The Journal of Finance*, 56(2), 699–720.

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

Kozak, S., S. Nagel, and S. Santosh (2018): "Shrinking the Cross-Section," *Journal of Financial Economics*.

Lettau, M., and M. Pelger (2018): "Factors that Fit the Time-Series and Cross-Section of Stock Returns," *Working paper*.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics*, 96, 175–194.

Moritz, B., and T. Zimmerman (2016): "Tree-based conditional portfolio sorts: The relation between past and future stock returns," *Working paper*.

Nagel, S., and K. J. Singleton (2011): "Estimation and Evaluation of Conditional Asset Pricing Models," *The Journal of Finance*, 66, 873–909.

Rossi, A. G. (2018): "Predicting Stock Market Returns with Machine Learning," *Working paper*.

Sloan, R. G. (1996): "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *The Accounting Review*, 71(3), 289–315.

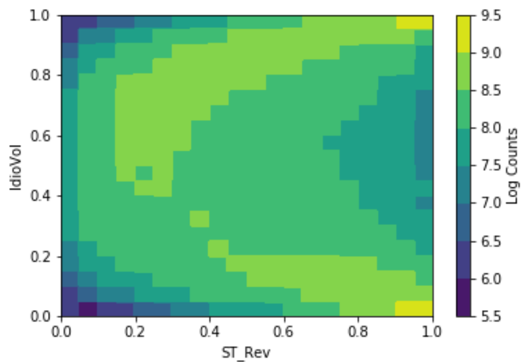
Zou, H., and T. Hastie (2005): "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(2), 301–320.

Appendix A. List of the Firm-Specific Characteristics

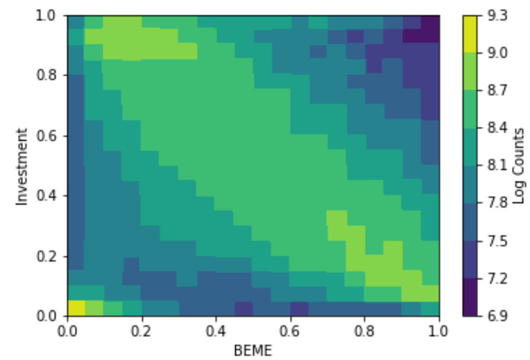
Acronym	Name	Definition	Reference
AC	Accrual	Change in operating working capital per split-adjusted share from the fiscal year end t-2 to t-1 divided by book equity (defined in BEME) per share in t-1. Operating working capital per split-adjusted share is defined as current assets (ACT) minus cash and short-term investments (CHE) minus current liabilities (LCT) minus debt in current liabilities (DLC) minus income taxes payable (TXP).	Sloan (1996)
BEME	Book to Market Ratio	Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholdersÄÅ equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity (PRC*SHROUT) is as of December t-1.	Fama and French (1992)
IdioVol	Idiosyncratic volatility	Standard deviation of the residuals from aregression of excess returns on the Fama and French three-factor model	Ang, Hodrick, Xing, and Zhang (2004)
Investment	Investment	Change in total assets (AT) from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets	Cooper, Gulen, and Schill (2008)
LME	Size	Total market capitalization at the end of the previous month defined as price times shares outstanding	Fama and French (1992)
LT_Rev	Long-term reversal	Cumulative return from 60 months before the return prediction to 13 months before	Jegadeesh and Titman (2001)
Lturnover	Turnover	Turnover is last month's volume (VOL) over shares outstanding (SHROUT)	Datar, Naik, and Radcliffe (1998)
OP	Operating profitability	Annual revenues (REVT) minus cost of goods sold (COGS), interest expense (TIE), and selling, general, and administrative expenses (XSGA) divided by book equity (defined in BEME)	Fama and French (2015)
r12_2	Momentum	To be included in a portfolio for month t (formed at the end of month t-1), a stock must have a price for the end of month t-13 and a good return for t-2. In addition, any missing returns from t-12 to t-3 must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of month t-1.	Fama and FrencH (1996)
ST_Rev	Short-term reversal	Prior month return	Jegadeesh and Titman (1993)

Note: Characteristic Variables as listed on the Kenneth French Data Library.

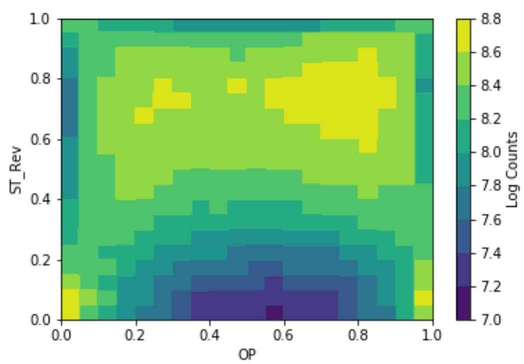
Appendix B. Additional stylized empirical facts



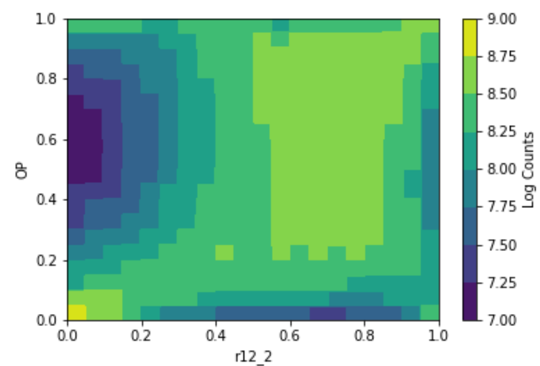
(a) Short-term reversal and idiosyncratic vol.



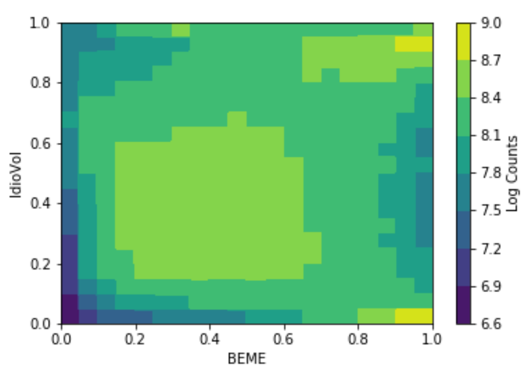
(b) Book-to-market and investment



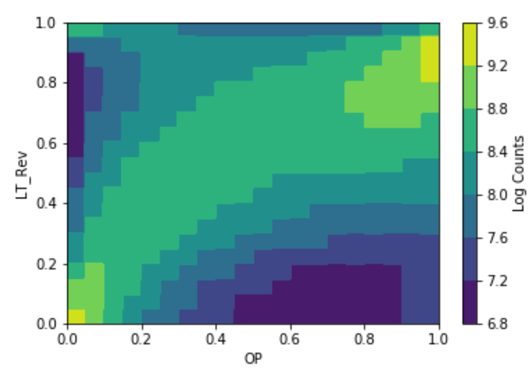
(c) Operating profitability and short-term reversal



(d) Momentum and operating profitability



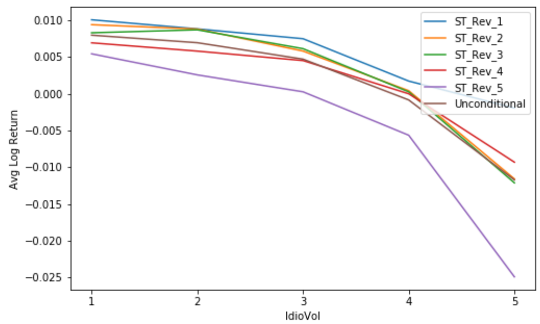
(e) Book-to-market and idyosyncractic volatility



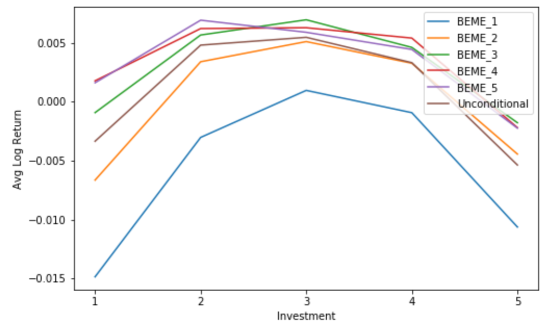
(f) Operating profitability and long-term reversal

Figure B.1. Joint empirical distribution of (scaled) characteristics in the cross-section of stocks

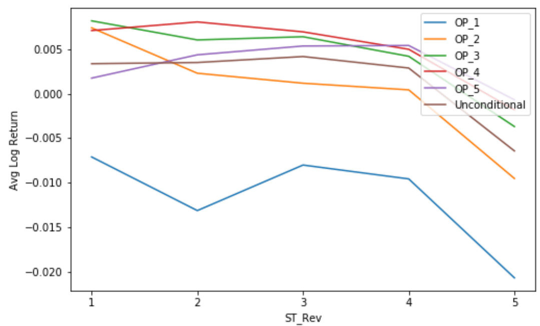
The graphs represent the pairwise empirical cross-sectional distribution of the characteristic quantiles across the stocks. The frequency is computed on a quantile 20x20 grid.



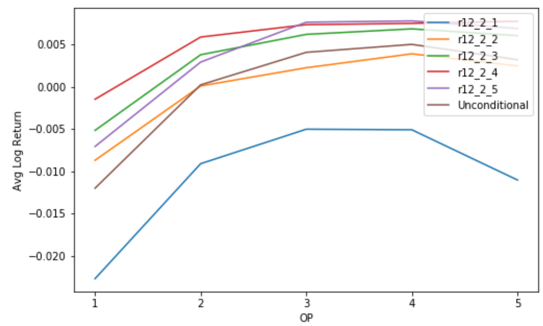
(a) Short-term reversal and idiosyncratic vol.



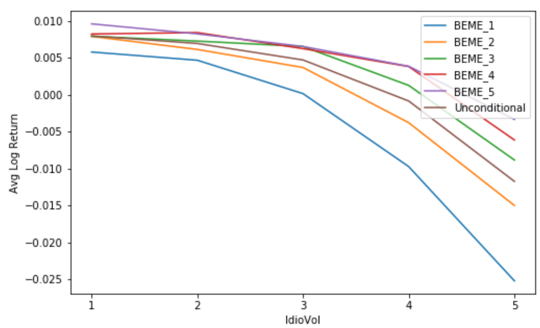
(b) Book-to-market and investment



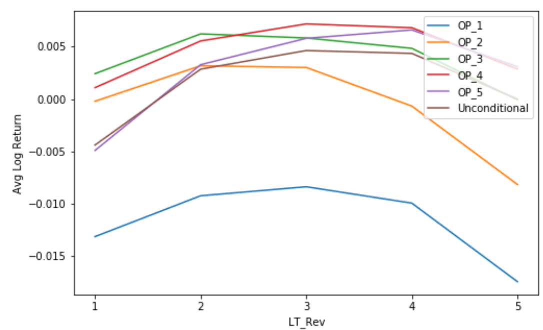
(c) Operating profitability and short-term reversal



(d) Momentum and operating profitability



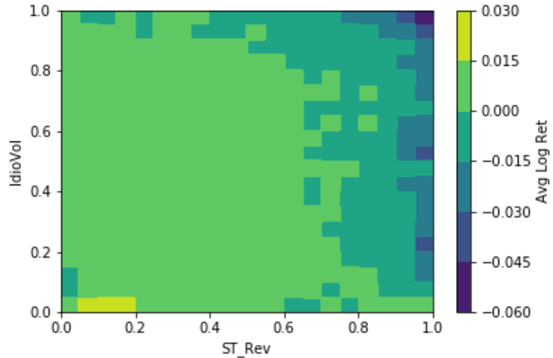
(e) Book-to-market and idyosyncractic volatility



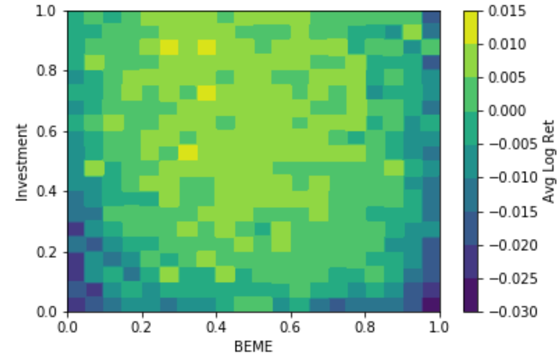
(f) Operating profitability and long-term reversal

Figure B.2. Conditional and unconditional characteristic impact

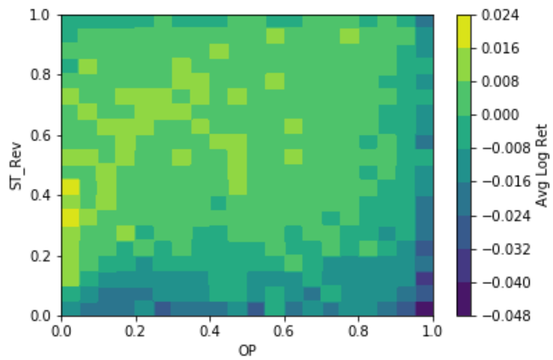
The graphs represent the impact of a characteristic quintile on expected returns unconditionally, and conditionally on the second characteristic. For example, Panel (a) describes the average returns on stocks sorted by their idiosyncractic volatility (Quintile 1-5), both unconditionally, and conditionally on belonging to quintile 1-5 based on short-term reversal.



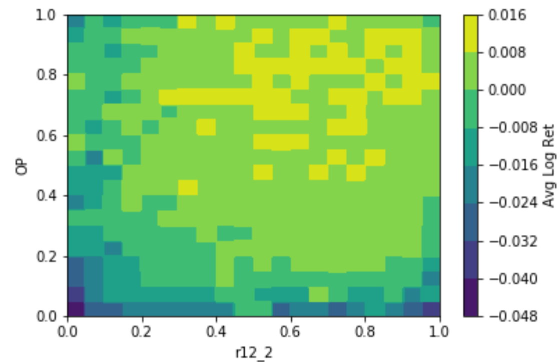
(a) Short-term reversal and idiosyncratic vol.



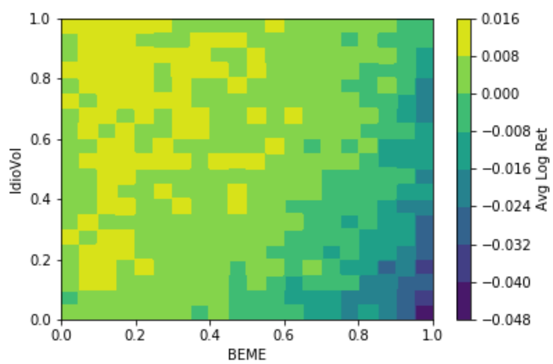
(b) Book-to-market and investment



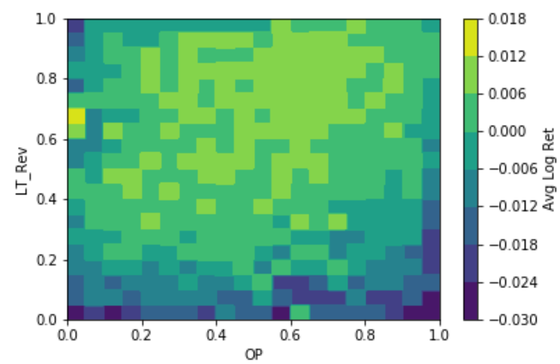
(c) Operating profitability and short-term reversal



(d) Momentum and operating profitability



(e) Book-to-market and idyosyncractic volatility



(f) Operating profitability and long-term reversal

Figure B.3. Expected return as a function of stock characteristics

The graphs represents empirical distribution of expected returns in the pairwise characteristic space. Yellow areas denote high expected return, while dark blue corresponds to the areas with the lowest expected returns. All the quantities are computed on a grid on 20 $\times$ 20 unconditional quantiles by averaging historical returns on the stocks belonging to that portfolio.

Appendix C. Additional Empirical Results for Cross-Sections built on 3 Characteristics

Table C.1 Selection of hyperparameters

Notation	Hyperparameters	Candidates
γ	Mean shrinkage	$\{0, 0.05, 0.1, ..., 0.9\}$
K	Number of portfolios	$\{10, 11, ..., 40\}$
λ_2	Variance shrinkage	$\{0.1^5, 0.1^{5.25}, ..., 0.1^8\}$

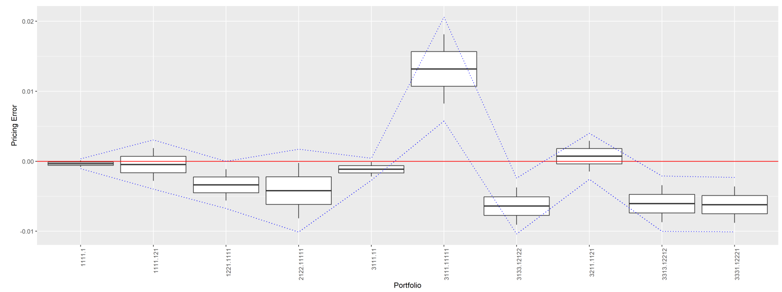
The table presents the list of hyperparameters for pruning AP-Trees, and the actual range of their values used for simulation and empirics. The mean shrinkage is normalized as a convex combination between sample means and the cross-sectional average: $(1-\gamma)\hat{\mu} + \gamma\bar{\hat{\mu}}\mathbb{1}$.

Table C.2 Cross-sections of three characteristics sorted by out-of-sample AP-Tree Sharpe ratios

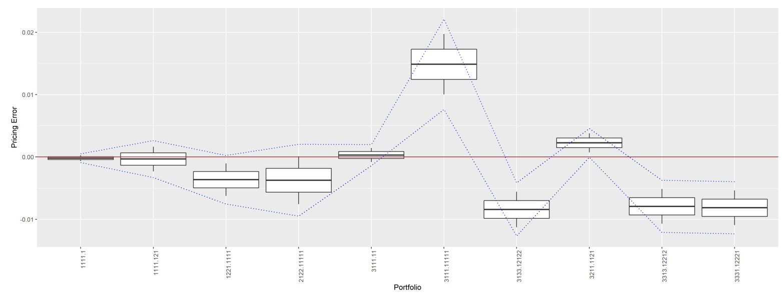
Id	Char 1	Char 2	Char 3	SR (AP-Tree)	Id	Char 1	Char 2	Char 3	SR (AP-Tree)
12	Size	Mom	LRev	0.24	6	Size	Val	Acc	0.47
15	Size	Mom	Turn	0.25	23	Size	Inv	LRev	0.48
33	Size	LRev	Turn	0.28	20	Size	Prof	IVol	0.48
27	Size	SRev	LRev	0.31	29	Size	SRev	IVol	0.48
30	Size	SRev	Turn	0.31	9	Size	Mom	Prof	0.50
1	Size	Val	Mom	0.34	2	Size	Val	Prof	0.51
28	Size	SRev	Acc	0.37	22	Size	Inv	SRev	0.51
4	Size	Val	SRev	0.40	18	Size	Prof	LRev	0.52
36	Size	IVol	Turn	0.40	7	Size	Val	IVol	0.52
31	Size	LRev	Acc	0.41	14	Size	Mom	IVol	0.52
13	Size	Mom	Acc	0.41	34	Size	Acc	IVol	0.52
5	Size	Val	LRev	0.41	10	Size	Mom	Inv	0.54
32	Size	LRev	IVol	0.42	26	Size	Inv	Turn	0.56
17	Size	Prof	SRev	0.42	24	Size	Inv	Acc	0.57
21	Size	Prof	Turn	0.44	19	Size	Prof	Acc	0.57
8	Size	Val	Turn	0.44	3	Size	Val	Inv	0.61
35	Size	Acc	Turn	0.45	16	Size	Prof	Inv	0.66
11	Size	Mom	SRev	0.45	25	Size	Inv	IVol	0.70

The table summarises potential cross-sections built on 3 characteristics from the overall selection of 10 candidate variables. For each case we present the identification number of the cross-section (used throughout the figures for summarizing the overall results), the list of the characteristics involved, and monthly Sharpe ratio, achieved on the testing subsample of the data from the portfolios and SDF weights built with AP-Trees.

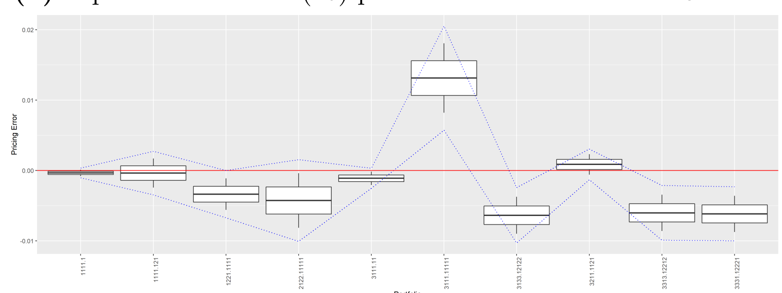
Appendix D. Zooming into the Cross-Section: Size, Profitability, and Investment



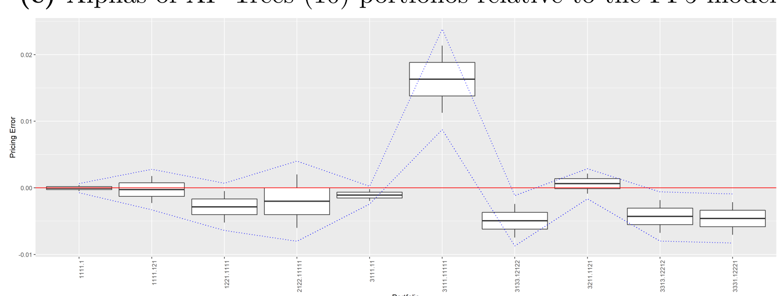
(a) Alphas of AP-Trees (10) portfolios relative to the XSF model



(b) Alphas of AP-Trees (10) portfolios relative to the FF3 model

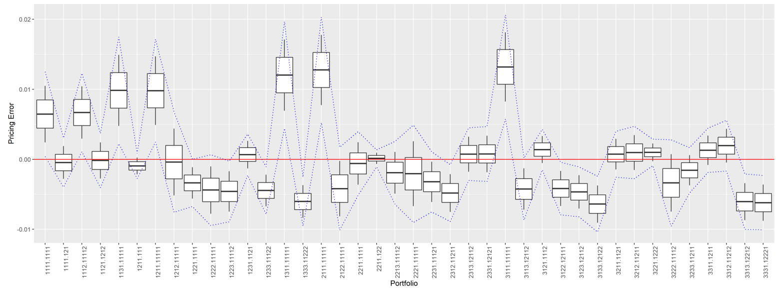


(c) Alphas of AP-Trees (10) portfolios relative to the FF5 model

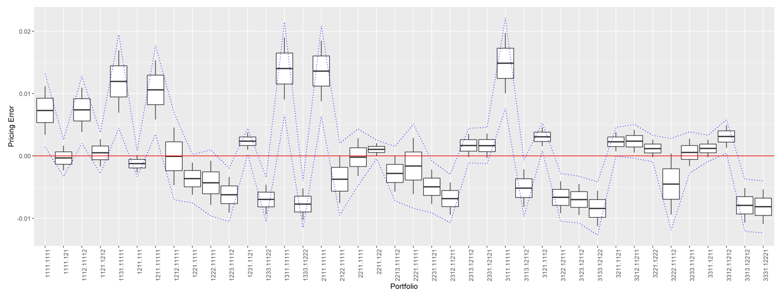


(d) Alphas of AP-Trees (10) portfolios relative to the FF11 model

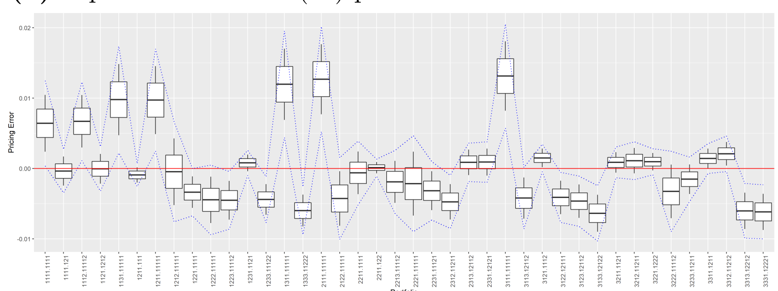
Figure D.1. Pricing errors α_i for AP-Trees (10) with FF3, FF5, FF11, and XSF models for Size, Operating Profitability, and Investment



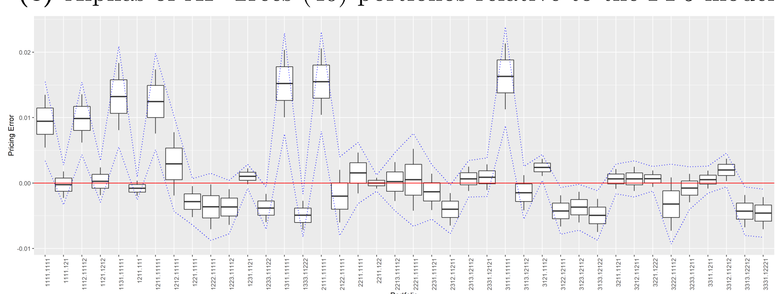
(a) Alphas of AP-Trees (40) portfolios relative to the XSF model



(b) Alphas of AP-Trees (40) portfolios relative to the FF3 model

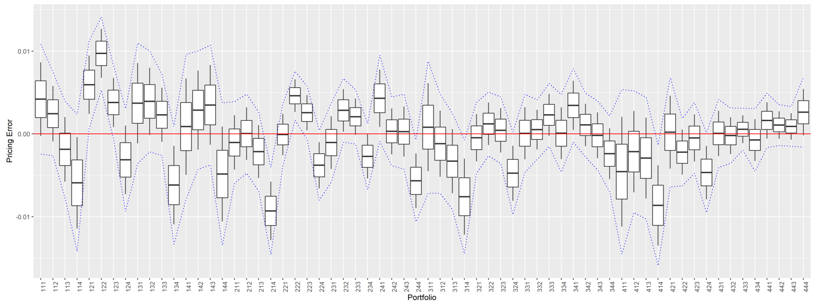


(c) Alphas of AP-Trees (40) portfolios relative to the FF5 model

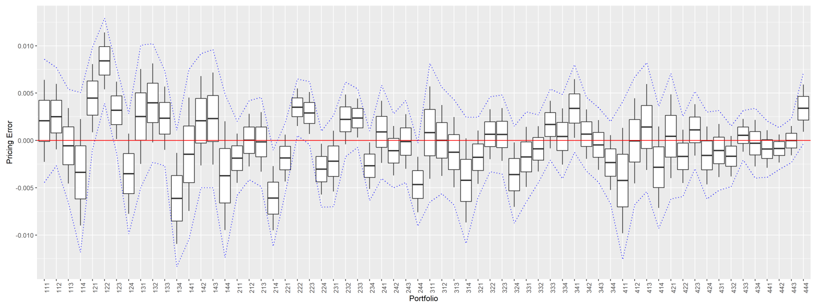


(d) Alphas of AP-Trees (40) portfolios relative to the FF11 model

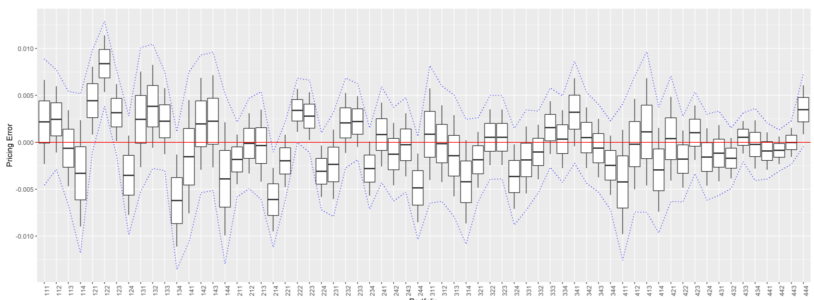
Figure D.2. Pricing errors α_i for AP-Trees (40) with FF3, FF5, FF11, and XSF models for Size, Operating Profitability, and Investment



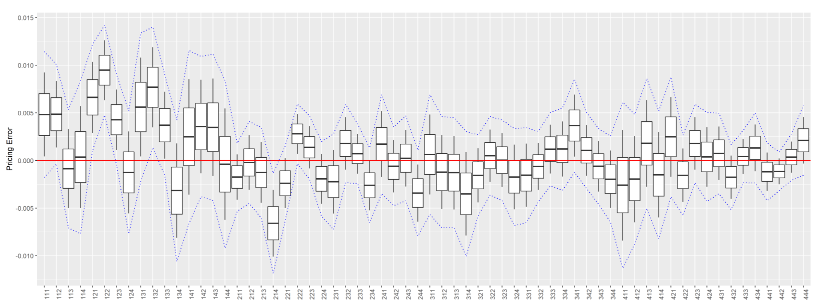
(a) Pricing errors of 64 triple-sorted portfolios, controlling for the Fama-French 3 factors



(b) Pricing errors of 64 triple-sorted portfolios, controlling for the Fama-French 5 factors



(c) Pricing errors of 64 triple-sorted portfolios, controlling for the cross-section-specific factors (market + size, profitability, investment)



(d) Pricing errors of 64 triple-sorted portfolios, controlling for 11 cross-sectional factors

Figure D.3. Portfolio-specific pricing errors, α_i, of triple sorts (64 assets) relative to leading reduced-form asset pricing models

Appendix E. Additional Simulation Results

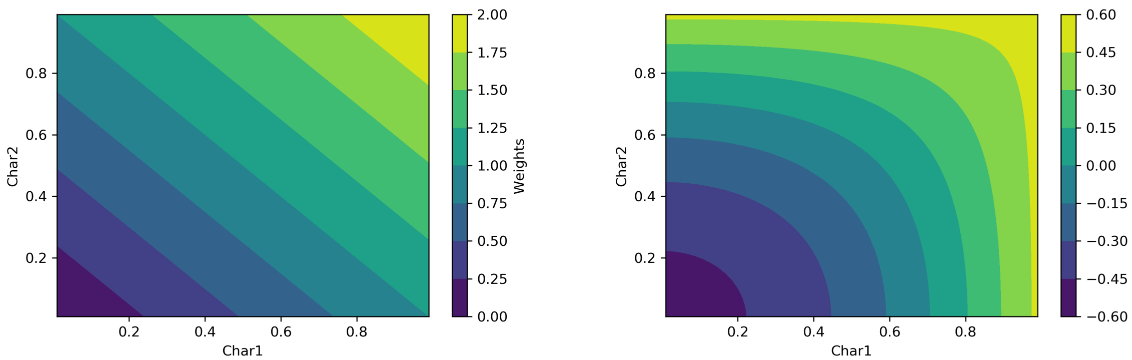


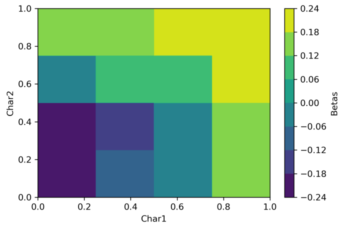
Figure E.1. Loading functions of Simulation

Table E.1 SR and SDF recovery in a simulated environment

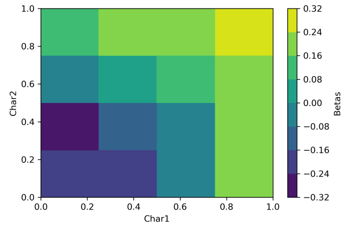
	$\text{Corr}(C_{i,t}^{(1)}, C_{i,t}^{(2)}) = 0$				$\text{Corr}(C_{i,t}^{(1)}, C_{i,t}^{(2)}) = 0.5$				$\text{Corr}(C_{i,t}^{(1)}, C_{i,t}^{(2)}) = 0.9$			
Portfolios	SR_{tr}	SR_{val}	SR_{test}	$\rho(F_t, \hat{F}_t)$	SR_{tr}	SR_{val}	SR_{test}	$\rho(F_t, \hat{F}_t)$	SR_{tr}	SR_{val}	SR_{test}	$\rho(F_t, \hat{F}_t)$
Additively linear model												
AP-Trees	0.74	0.55	0.45	0.94	0.74	0.55	0.43	0.92	0.76	0.50	0.42	0.92
DS	0.67	0.47	0.38	0.88	0.69	0.50	0.39	0.87	0.69	0.52	0.38	0.87
Non-linear model												
AP-Trees	0.69	0.38	0.37	0.81	0.72	0.45	0.34	0.81	0.71	0.42	0.37	0.81
DS	0.65	0.38	0.34	0.78	0.64	0.42	0.32	0.78	0.64	0.46	0.30	0.75

Sharpe Ratio (SR) of the SDF factor for the tree portfolios and double sorting 16 portfolios under two different loading functions and three different correlations of the two characteristics variables. The data is generated with an SDF factor with Sharpe Ratio $SR = 0.5$. $N = 800, T = 600$, $T_{tr} = 240$, $T_{val} = 120$ and $T_{test} = 240$.

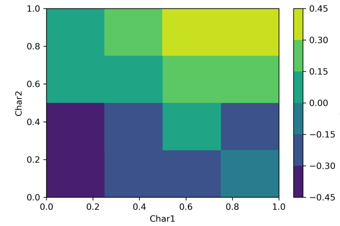
Panel A: Double-sorted portfolios, spanned by DS-SDF



(a) Correlation = 0

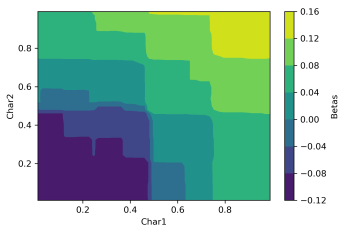


(b) Correlation = 0.5

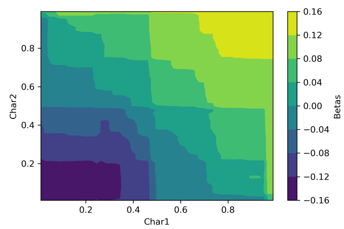


(c) Correlation = 0.9

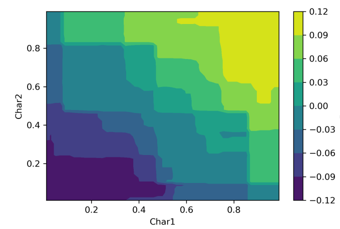
Panel B: Conditional sorts, spanned by pruned AP-Trees



(d) Correlation = 0



(e) Correlation = 0.5



(f) Correlation = 0.9

Figure E.2. Tracking expected returns: SDF loadings as a function of characteristics in a nonlinear model.