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

Svetlana Bryzgalova† Markus Pelger‡ Jason Zhu§

September 2021

ABSTRACT

We build cross-sections of asset returns for a given set of characteristics, that is, managed portfolios that serve as test assets for asset pricing models and building blocks for new risk factors. We use decision trees to endogenously group similar stocks together by selecting optimal portfolio splits to span the Stochastic Discount Factor. Our portfolios are interpretable, and reflect many characteristics and their interactions. Compared to combinations of traditional sorts and machine learning prediction-based portfolios, our cross-sections have up to three times higher out-of-sample Sharpe ratios and pricing errors, and do not suffer from excessive repackaging/duplication of the original stocks.

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, Christian Julliard, Bryan Kelly (discussant), Serhiy Kozak, Andreas Neuhierl (discussant), Serge Nyawa (discussant), Juan Passadore, Andrew Patton (discussant), Ilaria Piatti (discussant), Alberto Rossi (discussant), Ken Singleton, Georgios Skoulakis, Raman Uppal, Michael Weber, Dacheng Xiu (discussant), Paolo Zaffaroni, Goufu Zhou, and seminar and conference participants at Chicago, Stanford, Duke, London Business School, Imperial College London, UBC Sauder, Rice University, Purdue University, Singapore Management University, University of Edinburgh, AFA Annual Meeting, EFA Annual Meeting, Chicago Asset Pricing Conference, SFS Cavalcade, Fourth International Workshop on Financial Econometrics, Toulouse Financial Econometrics Conference, SIAM Financial Mathematics, Adam Smith Junior Workshop, Yale Junior Finance Conference, London CFE, Machine Learning Day at Georgetown McDonough, 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

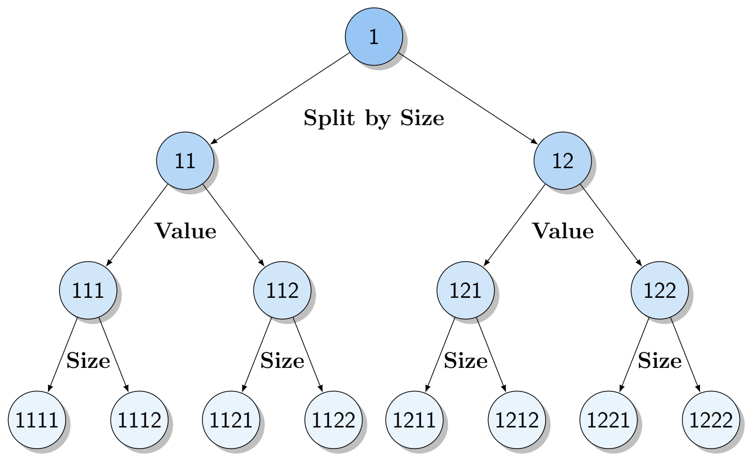
Introduction

What explains expected returns? This simple question has two unknowns: The returns of *test assets* – typically cross-sections of portfolios – and the model that succeeds in pricing them. While most of the literature has historically focused on the second problem, choosing the right test assets is equally crucial, and perhaps, even more so. First, unless the test assets span the underlying Stochastic Discount Factor (SDF), conditional on a given information set, there is no reason to expect that a well-fitting model is anywhere close to the truth, as it could easily fail at pricing omitted assets. Second, informative test assets are not only crucial for evaluating existing models but also serve as *basis assets* for building new models. Since tradable risk factors are also usually constructed from the same building blocks as a typical cross-section, they inherit all the problems of the latter.

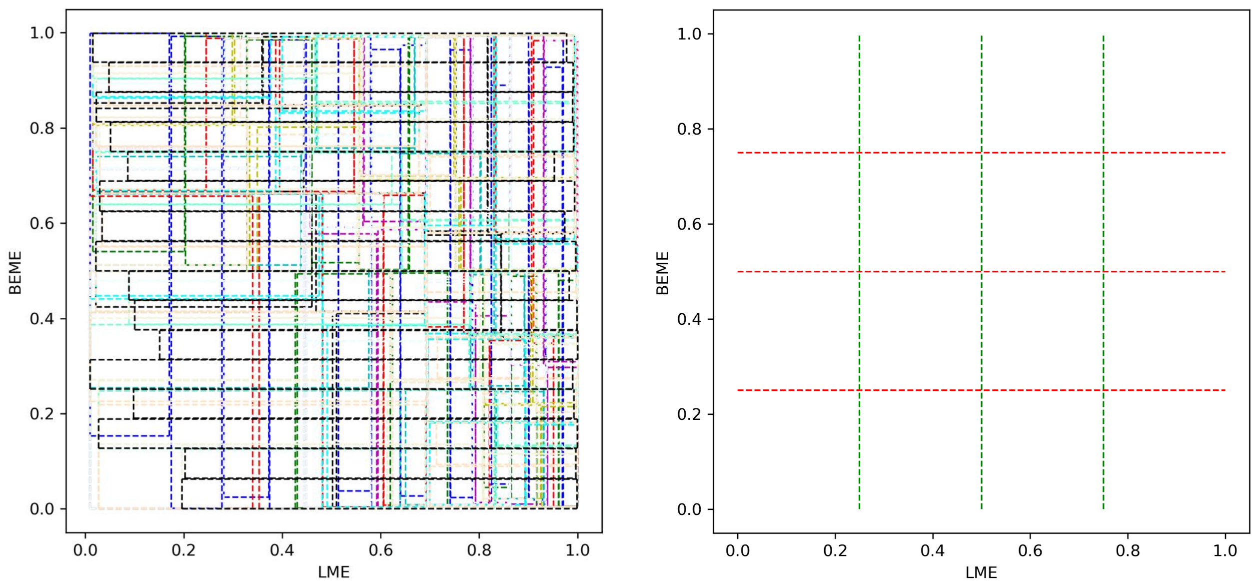
Building a cross-section of asset returns conditional on asset-specific characteristics is equivalent to forming *managed portfolios*, which can be used for unconditional modeling. The current standard are single- or double-sorted portfolios, for example the 25 size-and value sorted-portfolios of Fama and French (1993). We show that the conventional sorting-based portfolios drastically fail to span the SDF and hence lead to the wrong conclusions when used to evaluate or construct asset pricing models. These conventional cross-sections do not reflect the joint effect of multiple characteristics neglecting their interactions. Stacking multiple single/double sorts against each other often makes the problem worse: Still missing important interaction effects, these large-dimensional composite cross-sections are full of duplicated portfolios and noisy repackaging of the same underlying economic risks. Our solution, *Asset Pricing Trees* (AP-Trees) addresses all the crucial problems of conventional sorts: complex interactions, the curse of dimensionality, repackaging and duplication, thus providing a new way of building better cross-sections of portfolios, to be used in structural and reduced-form models.

AP-Trees deliver a small cross-section of interpretable, well-diversified portfolios that provide a robust span of the SDF, conditional on many characteristics. AP-Trees provide a novel approach to grouping individual stocks into managed portfolios that reflects information in a given set of characteristics. Our method is rooted in the idea of decision trees and builds up on the appeal of popular standard double and triple sorts. It has two key elements: (1) the construction of conditional tree portfolios and (2) the *pruning* of the overall portfolio set based on the SDF spanning requirement. Importantly, the second step yields a set of long-only portfolios with an endogenous decision of *what characteristics* to use in splits, and *how granular* they should be so that their overall combination spans the SDF.

AP-Trees are intuitive and simple, while capturing complex dependencies. Figure 1a



(a) Example of a conditional tree based on size and value



(b) Conditional trees

(c) Double-sorting

Figure 1: Size and Value: Portfolio splits with AP-Trees and double sorts

Panel (a) shows an example of an AP-Tree of depth 3 based on size and book-to market. In this subtree the first 50/50 split is done by size, the second by value, and the last one by size again. The portfolio label corresponds to the path along the tree that identifies it, with "1" standing for going left, while "2" stands for going right. A full AP-Tree is the combination of all possible orders and splits.

Panels (b) and (c) illustrate the grouping of individual stocks in the space of size and book-to-market quantiles that define portfolios in AP-Trees (left, both final and intermediate nodes) and the double-sorted cross-section (right). AP-Trees are constructed up to depth 4, which is allowing for at most four consecutive splits; i.e., all the portfolios have at least 1/16 of all the stocks. The number of double-sorted portfolios (16) is chosen to match the depth of the trees based on the granularity/diversification of the splits.

shows a simple decision tree based on a sequence of *conditional* consecutive splits. First, we divide the universe of stocks into two groups based on the stocks' market cap, then within each group – by their value, then by size again, and so on. The nodes of such a tree correspond to managed portfolios and reflect the conditional impact of characteristics in a simple and transparent way: Using different variables and changing the order and depth of the splits produces a very diverse and rich set of portfolios compared to the conventional unconditional sorting as illustrated in Figure 1b. As a result, all the decision trees' final and intermediate nodes represent a high-dimensional set of possible investment strategies, while easily ensuring that each portfolio is well-diversified. Importantly, each portfolio (tree node) can always be traced back to economic fundamentals.

We consider the whole set of potential managed portfolios offered by AP-Trees and develop a new approach to reduce them to a small number of interpretable test assets, the process we refer to as *pruning*. Assets are combined together in higher-level nodes, making the original portfolios redundant, only if their combination spans the SDF as well as the original, granular trading strategies. For example, in Figure 1a we split the portfolio with 50% of the smallest stocks (node 11), only if including the small-value and small-growth portfolios (in addition to the other assets) in the SDF results in a higher total Sharpe ratio than including the combined small cap portfolio. Targeting the global Sharpe ratio, spanned by a subset of final and intermediate nodes, is completely different from standard pruning criteria in either the classic decision-tree-based literature, or related applications in machine learning. The key difference lies in the fact that split decisions cannot be done with only local information (i.e., by simply comparing average returns or volatilities of parent and children nodes) but also how they co-move with *all the other* potential basis assets. By including all the intermediate and final nodes and exploiting the recursive structure of tree portfolios, we map the pruning into a robust SDF recovery within a mean-variance framework. As a result, we group individual stocks into the same managed portfolios, if they have the same conditional contribution to the SDF. Finally, to eliminate the danger of overfitting and to produce an empirically reliable solution, we rely on the dual shrinkage in the variance *and* mean. By doing so, we generalize the robust SDF recovery of Kozak, Nagel, and Santosh (2020), and focus only on the out-of-sample performance of our test assets.

In a large-scale empirical application, we build cross-sections based on the 10 most used firm-specific characteristics and compare their performance (Sharpe ratios, alphas, cross-sectional fit) to conventional sorts. All our results are achieved out-of-sample. First, we start with the simpler problem of building cross-sections based on only 3 characteristics, yielding a set of 36 different cross-sections which we compare with conventional triple-sorted portfolios. We find that for every single cross-section, relative to triple sorts, test assets based on AP-Trees have dramatically higher Sharpe ratios, sometimes up to a factor of three. This means that standard cross-sections and long-short factors, created from them, do not span the SDF, conditional on characteristics, and present a wrong benchmark for models as both test assets *and* factors. Importantly, this does not come from higher loadings on conventionally perceived 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 an 11-factor model. Our portfolios are well-diversified and do not load on extreme quantiles or microcaps. In fact, the tree-based construction enforces a stable and balanced composition by default, making them fully comparable and often more diversified than triple sorts of the same depth. The AP-Tree basis assets contain strictly more information than conventional sorts, as the implied SDF of AP-Trees explains the expected returns of conventional sorts, but not the other way around. We conclude that even for a small number of characteristics, conventional sorts neglect asset pricing information and AP-Trees are strictly superior.

Second, we create small-dimensional cross-sections of portfolios (i.e., 20–40 assets) that reflect the information contained in *all 10 characteristics* and their interactions. Compared to an optimal use of decile-sorted portfolios (100 assets) we *double* both the Sharpe ratio achieved out-of-sample and the alphas associated with the SDF spanned by these portfolios relative to traditional models. This implies that even using 100 decile-sorted portfolios, based on 10 separate characteristics, presents too low of a bar to judge model success or failure. Relative to the best combinations of size-adjusted double-sorted portfolios,[1] AP-Trees offer a 30% increase in both Sharpe ratios and the SDF alpha out-of-sample. This again confirms that standard sorts present a relatively inefficient and crude way of spanning the SDF. Factors based on them are likely to be heavily misspecified as well.[2]

Our empirical findings have important implications for asset pricing. First, we demonstrate the importance of taking into account the joint distribution of multiple characteristics. We show that the main driving forces behind the superior performance of AP-Trees is their ability to efficiently capture interactions among characteristics. Shutting down the interaction channel reduces the asset pricing performance in terms of SDF spanning and pricing errors by around a half. Second, characteristic-based cross-sections should be constructed with the right objective function, which is the SDF spanning criterion. Using the same flexibility and information, but either only return spreads or only the variation as the criterion to select test assets drops the asset pricing performance to around 20–50%. Third, we confirm that robust estimation is important. While the effect of interactions and the asset pricing objective dominates the construction of informative test assets, the inclusion of mean or variance shrinkage in addition to sparsity shrinkage improves almost all out-of-sample results by

[1] That is the combination of 25 size-value portfolios, 25 size-momentum portfolios, etc., *all* stacked together in a single large cross-section.

[2] Intuitively, a model that matches the average size and value premium on the corresponding long-short portfolios could still fail at pricing many other well-diversified simple strategies based on size and value as characteristics. Furthermore, consistent with Daniel and Titman (2012), we find that standard unconditional sorts have a relatively low power in discriminating models and develop alternative test assets based on AP-Trees.

at least 5%.

Our paper also makes several methodological contributions. Our approach to modeling the mapping between characteristics and stock returns and to use it to group individual stocks into managed portfolios is conceptually different from existing ones and is optimally tailored to the problem. While there are other ways to either explicitly include interactions (e.g., ad-hoc products of characteristic terms) or to reduce the dimensionality of a set of given portfolios sorts, for example, Principal Component Analysis (PCA), we combine both steps with a clear asset pricing objective – spanning the SDF. Our tree-based portfolios are special in that they present a mix of recursive portfolios with parents and children nodes, which are optimally weighted to favor more diversified larger parent nodes. Our pruning selects a sparse set of these tree portfolios, which implies a classification of individual stocks by combining similar assets in larger portfolios. Hence, we ensure economic continuity of the resulting portfolios, allowing us to trace back each portfolio to its economic fundamentals. This unique feature has three advantages. First, it allows us to efficiently capture the impact of interaction and nonlinear effects in a parsimonious way. Second, we retain economic interpretability for the basis assets, which is crucial for understanding the areas of model failure, or the type of modeling needed to improve existing frameworks. Last, but not least, it is straightforward to incorporate many economic constraints on the desired cross-section of portfolios, such as the number of test assets, the degree of interactions among characteristics, restrictions on the minimum number of shares in portfolios, their market cap or liquidity.

Our paper also contributes to the evaluation of asset pricing models. The issues of robust out-of-sample evaluation, repackaging of redundant information and spanning tests lie at the core of the construction of AP-Trees and carry over to evaluation metrics for asset pricing. Martin and Nagel (2019) recently highlighted a stark difference between the in-sample predictability reflected by traditional statistical tests and its feasible, out-of-sample counterpart. In the spirit of their work, we fully focus on the out-of-sample SDF Sharpe ratios and alphas. We propose evaluating asset pricing models based on how well they can explain the out-of-sample robust SDF constructed from a cross-section of test assets. This metric accounts for potential redundancies in portfolios and allows for a large number of test assets. In comparison, conventional performance measures, such as average absolute pricing errors, are inflated in the presence of highly correlated and repackaged test assets, which is a particularly severe problem for conventional sorts that are stacked against each other. We show empirically that traditional sorts and evaluation metrics can fail to discriminate between different asset pricing models and tend to inflate the performance of models due to duplication in the test assets. AP-Trees and model evaluation based on robust SDF spanning, successfully address these concerns.

More generally, although the focus of our paper is on finding interpretable basis assets for the SDF, we also contribute to a broader discussion on the role of machine learning in economics and finance. So far, most papers have separated the construction of profitable investment strategies with machine learning or other techniques into two steps. First, advanced methods are used to extract signals for predicting future returns. Second, these signals are used to form profitable portfolios, which are typically long-short investments based on total predicted returns, or their mean-variance efficient combinations. However, we strongly believe that these two steps should be merged together; that is, machine learning techniques should extract the signals that are the most relevant for the overall economic problem, not just return prediction. Following this insight, we compare the out-of-sample performance of pruned AP-Trees to that of long-short prediction portfolios, formed by leading tools from the machine learning literature, namely random forest and neural networks from Gu, Kelly, and Xiu (2020b). We find that while retaining easy interpretability of the portfolios, AP-Trees are superior in their empirical performance, based on both the Sharpe ratio and SDF alpha. Intuitively, this happens because our approach uses information contained in *both* expected returns and variance to form building blocks of the SDF, and, as such, it solves a problem fundamentally different from just an optimal prediction of returns.[3] In other words, we show that a carefully designed problem formulation can combine the best of both worlds: the flexibility and robustness of machine learning methods and the relevance of economic restrictions.

Closely Related Literature

Our paper contributes to the growing literature on asset pricing that tackles the "multi-dimensional challenge," as formulated by Cochrane (2011) in his AFA Presidential Address. On a fundamental level this literature extracts basis assets with statistical and economic models that span the SDF. Most of the papers can be assigned to one of the following three streams. The first direction takes a set of characteristic-sorted or characteristic-projected portfolios as given and applies a dimension reduction to either obtain the SDF or a small number of factors, with different versions of latent factor models, such as PCA, being among the most popular approaches.[4] While these approaches deal with the large dimensionality

[3]Conversely, a researcher interested in pure prediction of returns as a function of characteristics, should rely on techniques that specifically target prediction only.

[4]Lettau and Pelger (2020) extend the standard 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 latent factors as a function of characteristics, and as a result apply PCA to managed portfolios that represent linear (and nonlinear) projections of asset returns on characteristics. Kozak, Nagel, and Santosh (2020) estimate the SDF by solving a regularized mean-variance

with some form of regularization, they can only model the information in the pre-specified given portfolio sorts. Our work is complementary, as we construct novel characteristic-based portfolios, which can be used as input to any of those unconditional models, including PCA and its modifications, generally without further dimension reduction.

The second strand of literature models the impact of characteristics on returns directly, without imposing an underlying risk model or a no-arbitrage condition. Freyberger, Neuhierl, and Weber (2020) and Gu, Kelly, and Xiu (2020b), among others, use non-parametric regularized techniques (adapted group lasso on kernels and machine learning tools) to estimate conditional mean returns as a function of characteristics. Moritz and Zimmerman (2016), Gu, Kelly, and Xiu (2020b), and Rossi (2018) use decision trees for estimating conditional moments of stock returns in the context of a prediction problem. Our paper is complementary to this literature, as we are not focused on pure return prediction, but instead try to find the optimal "building blocks" to span the projection of the SDF, conditional on these characteristics. 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 is applicable in our setting because of its global optimization nature.[5]

The third stream is an emerging literature that estimates a conditional SDF without prespecifying the mapping between stocks and characteristics. Chen, Pelger, and Zhu (2019) estimate a conditional SDF by solving the conditional general method of moment problem implied by no-arbitrage with neural networks. Gu, Kelly, and Xiu (2020a) estimate a conditional latent factor model with a deep learning autoencoder. Our goal is complementary: We not only estimate a feasible SDF, conditional on a range of characteristics, but also recover a set of interpretable building blocks that can be used as test assets in models with nontradable SDF. From this perspective, AP-Trees combine the elements of both prediction and classification problems.

Our paper stands out by combining the non-parametric mapping between stocks and characteristics with the economic objective of a robust SDF recovery. For the second element, we generalize the robust SDF recovery of Kozak, Nagel, and Santosh (2020) by applying an additional degree of optimal shrinkage to the mean. This 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. We formally show that this problem is equivalent to finding an optimal portfolio allocation of an ambiguity-averse investor who is facing a *joint* uncertainty in expected returns, volatilities,

optimization for a given set of portfolios.

[5]Section II highlights this difference further and introduces an alternative criterion we develop for pruning.

and Sharpe ratios of investment opportunities.[6] As a result, the special case of Kozak, Nagel, and Santosh (2020) can be improved upon.

Naturally, our paper expands the literature on constructing optimal test assets. We introduce the fundamental insight that informative test assets should span the conditional SDF. 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 structure, and they recommend mixing them with other cross-sections. We confirm their findings, and offer an alternative set of test assets that is small in dimension, allows for flexible interactions and nonlinearities, and does not suffer from duplication after repackaging the original stocks. Furthermore, we explicitly show that tree-based portfolios are able to isolate more diverse/idiosyncratic areas of mispricing relative to even a large combination of traditional sorts.

Nagel and Singleton (2011) consider managed portfolios as optimal instruments for hypothesis testing in a conditional asset pricing model. Our test assets are not designed to improve the power in making inference for *a particular* parameter and/or factor but rather try to answer the question of whether the set of portfolios spans the conditional SDF overall (and as a result, whether it could be used to discriminate models). They are "all purpose" generic test assets constructed for a given set of characteristics (similar to those available in Kenneth French's data library) that could then be applied to assess the performance of both linear and nonlinear models regardless of their structure and/or a particular hypothesis tested by the researcher, or they could be used to construct new reduced-form risk factors.

I. Test Assets, Sorting, and Trees

1. Test Assets and SDF

Conditional on a set of characteristics C_{t-1}, a valid SDF M_t is constructed by its projection on the space of individual stock returns R_t as follows:

$$M_t = 1 - \sum_{i=1}^{N} b_{t-1,i} R_{t,i} \qquad \text{with } b_{t-1,i} = f\left(C_{t-1,i}\right),$$

where C_{t-1} is an $N \times K$ matrix of K characteristics observed for N stocks and $f(\cdot)$ is a general, potentially nonlinear and non-separable function. Reduced-form asset pricing

[6] In a related paper, Garlappi, Uppal, and Wang (2007) find that estimation uncertainty has a first order impact on the empirical performance of the portfolio strategy and show a unique map between different types of shrinkage and the corresponding priors on the return distribution (for a Bayesian investor), and parameter uncertainty sets (for a robust version of the mean-variance approach).

models approximate this dependence by a (potentially infinite) set of basis functions $f_j(\cdot)$, such that $f\left(C_{t-1,i}\right) \approx \sum_{j=1}^{J} f_j\left(C_{t-1,i}\right) w_j$. This allows us to express the problem as an unconditional model based on simple managed portfolios:

$$M_t = 1 - \sum_{j=1}^{J} w_j R_{t,j}^{\text{managed}} \qquad \text{with } R_{t,j}^{\text{managed}} = \sum_{i=1}^{N} f_j\left(C_{t-1,i}\right) R_{t,i}, \tag{1}$$

where R_t^{managed} are the returns of managed portfolios whose portfolio weights correspond to the different basis functions $f_j(\cdot)$ in the characteristic space.[7]

Since the pricing kernel is spanned by these managed portfolios, pricing them would be equivalent to spanning the SDF itself. The SDF spanning requirement of eq. (1) implies that we should find a set of managed portfolios such that their combination has the highest Sharpe ratio, which is equivalent to minimizing the distance between the true SDF and a candidate one. For example, if $R_{t,j}^{\text{managed}}$ were represented by 25 size-and-value-sorted portfolios (or their rotations), a model that successfully prices them would be a good candidate SDF. However, if a cross-section of these managed portfolios does not span the SDF, a model that explains it well could still be far from the actual pricing kernel. In fact, it is even possible that a model *failing* to price some of 25 size-and-value portfolios is actually closer to the true SDF than the one that succeeds in explaining them perfectly.

Importantly, the maximum Sharpe ratio property for these basis assets needs to be feasible for the investor, that is, achievable out-of-sample, as in-sample Sharpe ratios are prone to being overfitted. Furthermore, as the functional dependency of the SDF on the characteristics can be complex, we need basis assets that can reflect potential nonlinearities and interaction effects. Therefore, trees, allowing for flexible conditional and unconditional modeling of characteristic impact, emerge as a natural candidate.

Note that there could in principle be multiple rotations of the managed portfolios $R_{t,j}^{\text{managed}}$ from eq. (1), all spanning the same SDF (see, e.g., Kozak, Nagel, and Santosh (2020)). Intuitively, however, optimal basis functions should form managed portfolios that ensure economic continuity in the underlying basis assets: They should group together securities, which are similar in firm-specific characteristics. This is also naturally provided by AP-Trees and allows us to represent a complex SDF with a small number of portfolios, each of which we can trace back to its fundamentals. This small set of interpretable portfolios allows researchers to understand the sources of potential model failure. Finally, our managed portfolios easily allow for economic restrictions (e.g., conditions on the liquidity, diversification, and other

[7] Kelly, Pruitt, and Su (2019) consider linear basis functions resulting in R_t^{managed} being standard long-short factors, and conventional risk factors (as in Fama and French (1992)) rely on quantile-based indicators, while Kozak, Nagel, and Santosh (2020) employ polynomial functions.

features of the underlying portfolios). As a result, tree-based conditional sorts are a natural choice, generalizing concepts well familiar to a wide research community, trivial to interpret, and powerful in the empirical applications.

In summary, optimal portfolios spanning the SDF, projected on a certain group of characteristics, should possess the following properties:

(a) They should be able to reflect the impact of multiple characteristics at the same time.
(b) They should *span the SDF*, that is, achieve the highest possible Sharpe ratio *out-of-sample* when combined together.
(c) They should allow for flexible nonlinear dependence and interactions.
(d) They should be a relatively small number of well-diversified managed portfolios feasible for investors.
(e) They should provide an interpretable link to fundamentals.

Since our approach is a generalization of the conventional sorting-based procedure, we start our analysis by discussing the properties of the latter.

2. Conditional and Unconditional Sorting

Single- and double-sorted portfolios form popular convenient, interpretable, and diversified characteristic-based cross-sections.[8] Unfortunately, this approach does not allow us to capture more than two or three firm-specific characteristics at the same time, quickly leading to a curse of dimensionality. For example, 25 Fama-French portfolios are based on the intersection of two unconditional sorts, with five groups each (by size and value), with the intersection yielding 25 groups. With three characteristics, the same approach would attempt to make already 125 portfolios, often 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 so far has been to stack a set of double-sorted cross-sections against each other, which rapidly increases the overall number of test assets without any fundamental basis behind it, and often leads to unnecessary duplication/repackaging of the same underlying securities. Both small- and large-dimensional sorts are constructed in a fully ad-hoc manner, and there is no guarantee that the underlying test assets span the SDF regardless of the size of the cross-section.

Conditional sorts (trees) can model interactions among many characteristics without suffering from the curse of dimensionality. They are a natural extension of the standard

[8] Another popular choice of directly interpretable test assets, a set of long-short anomaly portfolios, formed on a wide range of characteristics, is a special case of the standard sorts.

sorting-based technique and work particularly well when the underlying characteristics have a complex joint distribution, characterized by substantial cross-sectional dependence. Similar to usual sorts, they group together securities that have similar economic fundamentals and are directly interpretable.

Figure 1a gives an example of *a conditional tree of depth 3*, where stocks are first sorted by size, then value, and then size again (i.e., all the potential portfolios – final and intermediate nodes – rely on up to three consecutive splits). At time t, all the stocks that have valid size and value information from $t-1$ are first sorted into groups 11 and 12 based on 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 groups 11 and 12 might not be equal to the unconditional median, since they are based on the conditional information of the previous size split. Finally, the four groups 111, 112, 121, and 122 are further split into two portfolios each, to form eight level-three portfolios. The notation of each node, therefore, reflects a unique chosen path along the tree with a specified list and order of the split criteria.[9] The conditional splits have two effects: First, they capture the conditional dependency between characteristics and second, they ensure diversification by controlling the number of stocks in each portfolio.

If stock-specific characteristics are independent, the order of the variables used for splits does not matter, and we end up with the same portfolios as double-sorts. 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. For example, Figure C.1 in the Appendix, Panel A, shows the sample cross-sectional distribution of standardized characteristics and their conditional and unconditional impact on expected returns for the pairs of size/value and size/accruals.[10] On average, there seems to be a negative correlation (in part, mechanical) between size and book-to-market, with a clear clustering around the north-west and south-east corners. As a result, double-sorted portfolios are heavily unbalanced across the characteristic spectrum. Furthermore, the conditional impact of characteristics on expected returns is also known to be highly nonlinear and full of interaction effects.[11] For example, it is well-known that the value effect is not homogenous across the different size deciles of the securities, and it is particularly strong for the smallest stocks. At the same time, the impact of accruals is almost flat for large stocks. However,

[9]Denoting size for 1 and value for 2, the bottom-left node of the tree becomes 121.111, where "121" stands for the order of the variables used in splits (size, value, size) and "111" maps into the corresponding portfolio choice (bottom 50% of stocks in each of the decision points).

[10]The Online Appendix includes additional examples in Section IA.5.

[11]See, e.g., Freyberger, Neuhierl, and Weber (2020) for the use of adaptive group lasso to estimate the nonlinear impact of characteristics on expected returns, and Gu, Kelly, and Xiu (2020b) and Chen, Pelger, and Zhu (2019) for the machine learning approach.

medium and small securities reveal a striking inverted U-shape pattern (see Figure C.1 in the Appendix).

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. Each sequence used for splitting the cross-section generates another set of 2^d portfolios, where d is the depth of the trees. If we denote the number of sorting variables by M, this implies 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. Naturally, these portfolios (both final and intermediate leaves of the tree) can capture at most d-way interactions between sorting variables.

In sum, if the underlying distribution of characteristics is cross-sectionally dependent, and/or characteristic effects on managed portfolio properties are conditional and nonlinear, AP-Trees deliver a multitude of basis assets that are fundamentally different from those created by double sorting, that, as a result, could lead to a better SDF spanning.

3. Recursive Portfolios and Split Choice

The recursive overlapping structure of portfolios in an AP-Tree is a unique feature that makes them a particularly appealing choice for test assets. AP-Trees create managed portfolios in both final and intermediate nodes that could be used as interpretable set of test assets spanning the SDF. The children nodes in a tree provide more granular splits based on the underlying information. Choosing only the portfolio of a parent node but removing the children nodes, provides a natural way of merging two portfolios if they are similar to each other. Keeping the children nodes is equivalent to a split in the tree, which is beneficial if it provides additional asset pricing information. Hence, the selection of a sparse set of portfolios from the intermediate and final nodes of an AP-Tree corresponds to the selection of splits in the tree, a process we call *pruning*. Importantly, such a decision should be based on the overall objective function in mind: An optimal set of portfolios used for return prediction could be very different from the one that targets spanning the SDF. Our selection procedure for a sparse set of portfolios is discussed in detail in Section II and based on an asset pricing objective.

The selection of a sparse set of AP-Tree portfolios provides an economically meaningful grouping of stocks. This feature of the AP-Trees can be particularly useful in the case of many characteristics that have a large asset pricing impact only in the tails of the cross-sectional distribution (e.g., deciles 1 and 10). The selection of both final and intermediate nodes of AP-Trees provides a simultaneous choice of what characteristics to use for splits

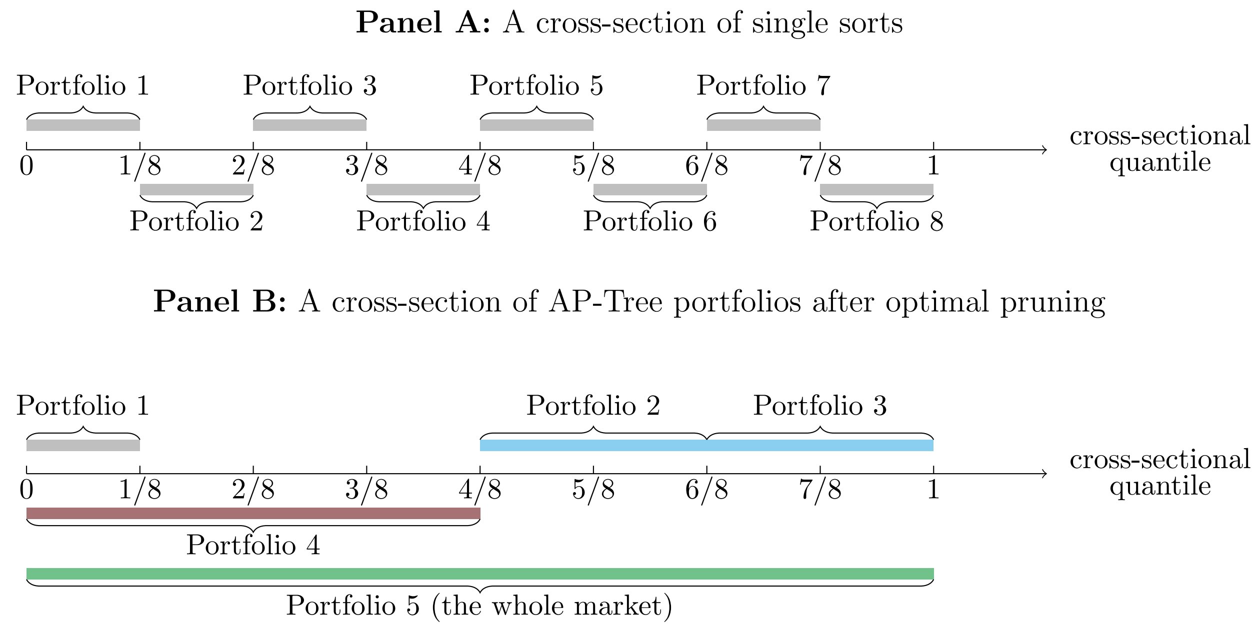


Figure 2: Selection of basis portfolios through pruning

The figure presents an example of the depth of pruning for a set of eight portfolios, sorted by a single characteristic. Standard double-sorting (Panel A) corresponds to the quantiles that form eight equispaced portfolios and eight test assets correspondingly. Panel B presents one of the potential outcome from applying pruning, resulting in five portfolios with portfolio sizes that range from 1/8 to 100% of all the stocks.

and how granular to go. Figure 2 presents a simple case of univariate sorting, which is equivalent to the final nodes of a tree of depth tree. In addition, the AP-Tree also includes the intermediate nodes. Pruning the tree selects a subset of all the nodes, which essentially merges lower level nodes, as long as there is not enough informational gain from doing such a split. In the particular example of Figure 2, 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 two, four, or eight times more stocks than the original selection. Figure 3 illustrates the pruning process with multiple characteristics. In this example, we select one intermediate node, which includes 50% of all stocks with a book-to-market ratio below the median, and a final node which corresponds to the smallest 50% of stocks whose book-to-market ratio is in between the 50%–75% quantile.

The AP-Pruning is necessary for its use as test assets. AP-Trees form a large set of diverse characteristic-based portfolios, which capture very complex dependencies. However, using all the potential portfolios is often not feasible: Their number grows exponentially with the depth of the tree. Using three characteristics with the depth of 3 results in $3^3 \cdot 2^3 = 216$ final and intermediate nodes. With 10 characteristics, the total number of portfolios explodes to over 8,000. Our optimal selection among both final and intermediate nodes of the trees, allows us to create a cross-section manageable in size with endogenous choice of the types of

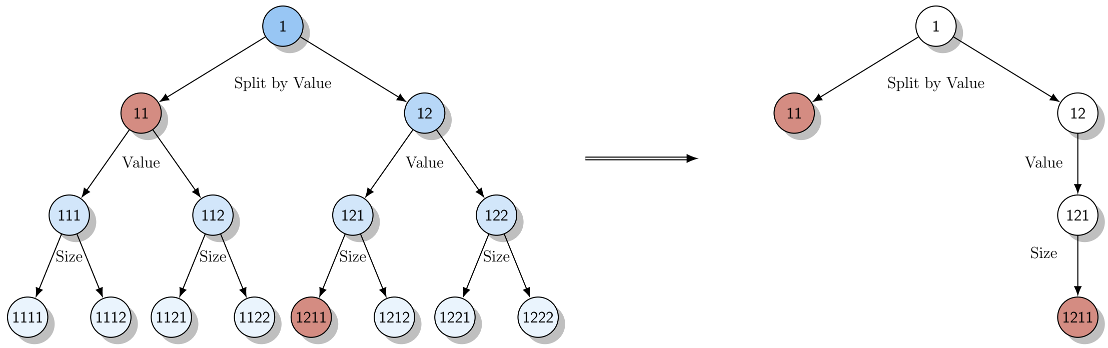


Figure 3: Illustration of pruning with multiple characteristics

The figure shows sample trees, original and pruned for portfolios of depth 3, and constructed based on size and book-to-market as the only characteristics. The fully pruned set of portfolios is based on eight trees, where the right figure illustrates a potential outcome for one tree.

the splits and their depth. Notably, an optimal set of portfolios could also be overlapping, if this is the best sparse combination of the final and intermediate nodes that spans the SDF.

The AP-Pruning is statistically desirable. The problem of split selection fundamentally reflects the bias-variance trade-off. Tree portfolios at higher nodes are more diversified, naturally leading to a smaller variance of their mean estimation, and so forth, while more splits allow us to capture a more complex structure in the returns at the cost of using investment strategies with higher variance. Our selection method will choose the parent nodes, unless the children nodes add asset pricing information that outweighs the increase in variance.[12]

Importantly, the selection of a sparse set of tree portfolios implies a classification of individual stocks. Because of its recursive overlapping structure, the selection of a small number of AP-Trees is conceptually different from sparsity on some generic portfolios. The AP-Pruning estimates the mapping between individual stocks and their characteristics by combining similar assets in larger portfolios. While trees allow us to model the problem as a portfolio selection problem, our sparsity operates directly on assigning individual stocks into groups. This is not the case for generic basis portfolios, which do not have a recursive structure.

[12]The selection of depth in an AP-Tree can be interpreted as the choice of bandwidth in a non-parametric estimation. Our AP-Tree chooses the optimal bandwidth based on an asset pricing criterion. If lower-level splits are removed in favor of a higher-level node, it implies that the split creates redundant assets. It can be interpreted as choosing a larger bandwidth, which reduces the variance. Removing higher-level nodes and splitting the portfolios into finer level, corresponds to a smaller bandwidth, which reduces the bias.

II. Pruning AP-Trees and Portfolio Selection

Our AP-Pruning is a novel approach to selecting a small set of AP-Tree nodes with the most non-redundant pricing information for spanning the SDF. Despite the existence of many conventional ways to prune a tree, those are not applicable to our case. The key issue lies in the fact that finding an optimal set of portfolios that span the SDF and maximize Sharpe ratio is a *global* problem and cannot be handled by *local* decision criteria. Similar to a mean-variance optimization problem, optimal tangency portfolio weights can only be found by considering the complete covariance matrix of assets and not just expected returns and correlations between two individual securities. As a result, conventional approaches to pruning, for example, by comparing average returns of portfolios obtained in a split, would work only if this particular decision does not affect others – a requirement obviously violated if the goal is to optimize the overall risk-return trade-off. Furthermore, none of the existing criteria to tree pruning actually guarantees that the resulting set of portfolios spans the SDF or delivers a feasibly high Sharpe ratio: They simply have a different objective function.

Our selection procedure is based on a global robust asset pricing objective. Since the issue of SDF spanning is generally equivalent to finding the tangency portfolio with the highest Sharpe ratio in the mean-variance space, our procedure builds up directly on this intuition. Importantly, we consider both final and intermediate nodes of *all the trees simultaneously*, which leads to portfolio selection, and a robust feasible SDF representation, reflecting sparsity in both depth and types of the splits. In more detail, we solve the mean-variance optimization problem applied to all the final and intermediate nodes of AP-Trees. Naturally, due to a high-dimensional nature of the problem, we rely on shrinkage to deliver robust and reliable portfolio selection. Here we provide a summary of the procedure, while the formal mathematical statements, detailed derivations, and description of the implementation could be found in the Online Appendix.

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 $\hat{\Sigma}$. Without imposing any shrinkage on the portfolio weights for the SDF, the problem has an explicit solution, $\hat{\omega}_{naive} = \hat{\Sigma}^{-1}\hat{\mu}$; it is likely to suffer from a substantial degree of overfitting. Instead, for each target expected return μ_0, we find the minimum variance portfolio weights $\hat{\omega}_{\text{robust}}$ with an elastic net penalty. Note that all the tuning parameters (the target mean μ_0, the lasso weight of λ_1, and the 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:

DEFINITION 1: *Divide original sample of data into three non-intersecting subsamples for use in training, validation, and testing of the model. A procedure for pruning AP-Trees for* *out-of-sample performance is defined by the following steps:*

1. *For a given set of values of tuning parameters μ_0, λ_1 and λ_2, use training data to solve*

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

   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. *Use validation data to estimate the Sharpe ratio of the optimal portfolio from eq. (2) as a function of tuning parameters $(\mu_0, \lambda_1, \lambda_2)$. Select parameters that maximize Sharpe ratios, store resulting portfolio selection and their weights in the SDF.*
3. *A previously unused testing subsample of data is used to trace the fully out-of-sample performance of the SDF and the properties of the individual portfolios in the chosen cross-section.*

The search of the tangency portfolio is effectively decomposed into two separate steps. First, we construct a robust mean-variance efficient frontier using standard optimization with shrinkage terms. Second, we select the optimal portfolio located on the robust frontier, that is, we find the tangency portfolio on the validation data set that has not been used in the first step, which is guiding the choice for the optimal level of shrinkage. The resulting approach combines the following three crucial features:

1. It shrinks the contribution of the assets that do not help span the SDF. Although the selection is done in the space of original nodes of the trees, implicitly this is similar to penalizing the impact of lower-order principal components that portfolios load on.
2. It shrinks the sample average portfolio returns toward their cross-sectional average value, since the highest/lowest sample returns in a large cross-section are likely to be over/underestimated.
3. It includes a lasso-type shrinkage to obtain a sparse representation of the SDF, selecting a small number of AP-Tree basis assets.

Proposition 1 further demonstrates that using target mean return μ_0 to trace out the efficient frontier in Step 2 is equivalent to shrinking average sample returns of the portfolios toward their cross-sectional mean, adding another layer of robustness to the standard portfolio optimization with elastic net. To demonstrate the solution in closed-form, we consider optimization in the absence of a lasso penalty.

PROPOSITION 1 (Target Return and Shrinkage to the Mean): *Tracing out the efficient frontier (without an elastic net penalty,* $\lambda_1 = \lambda_2 = 0$*) out-of-sample is equivalent to using in-sample mean-variance optimization with a vector of sample mean returns shrunk toward their cross-sectional average. The solution to this problem,*

$$\hat{\omega}_{robust} = \hat{\Sigma}^{-1} \left(\hat{\mu} + \lambda_0 \mathbb{1}\right),$$

has a one-to-one mapping between the target mean μ_0 *and mean shrinkage* λ_0*. Therefore, the robust portfolio is equivalent to a weighted average of the naive tangency and the minimum-variance portfolio.*

Tracing out the robust efficient frontier out-of-sample with a additional ridge penalty, is equivalent to conventional in-sample mean-variance optimization with (a) a vector of sample means shrink toward its cross-sectional average and (b) a sample covariance matrix shrunk toward a diagonal one. The resulting portfolio weights are given by the following expression:

$$\hat{\omega}_{robust} = \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \left(\hat{\mu} + \lambda_0 \mathbb{1}\right).$$

Our approach generalizes the SDF recovery of Kozak, Nagel, and Santosh (2020) by including a mean shrinkage to the standard elastic net formulation. Empirically, shrinkage to the mean is rarely found to be zero and often has a significant effect on the chosen portfolio spanning the SDF (see the related discussion in Section IV). Intuitively, since estimated expected returns have severe measurement errors, it is likely that extremely high or low rates of return (relative to their peers) are actually overestimated/underestimated simply due to chance and, hence, if left unchanged, would bias the SDF recovery.[13]

PROPOSITION 2 (Robust SDF Recovery): *The robust mean-variance optimization in Definition 1 generalizes the robust SDF recovery of Kozak, Nagel, and Santosh (2020), and is identical to it in the case of zero mean shrinkage (*$\lambda_0 = 0$*) or for a diagonal covariance matrix of returns. In the latter case, the robust SDF weights are given by*

$$\hat{\omega}_{robust} = \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1} \left(\hat{\mu} - \lambda_1 \mathbb{1}\right)_+,$$

where $(x)_+ = \max(x, 0)$ *and sign*(x) *is the sign of* x*. In the general case of a non-diagonal*

[13] The same reasoning famously underlines the use of *adjusted* stock betas by Bloomberg that shrink their sample estimates toward 1, which is the average in the overall cross-section.

sample covariance matrix, the solution on the active set of ω_i *(non-zero values), is given by*

$$\left[\left(\hat{\Sigma} + \lambda_2 I_N\right)\hat{\omega}_{robust}\right]_i = \hat{\mu}_i + \lambda_0 - \lambda_1 sign(\hat{\omega}_{robust,i}) \qquad \textit{for } i \textit{ in the active set.}$$

Our pruning procedure can also be interpreted as a robust approach to the mean-variance optimization problem that explicitly takes into account estimation uncertainty about the sample means and variance-covariance matrix of returns. Individually, lasso, ridge, and shrinkage to the mean have been successfully motivated with both robust control and a Bayesian approach to portfolio optimization with a specific choice of uncertainty sets (priors), leading to substantial empirical gains in a number of applications (see, e.g., Garlappi, Uppal, and Wang (2007) and Fabozzi, Huang, and Zhou (2010) for a review).

PROPOSITION 3 (Robust-Control Interpretation of Pruning): *The robust mean-variance optimization in Definition 1 is equivalent to finding the mean-variance efficient solution under a worst case outcome for estimation uncertainty. Given uncertainty sets for the achievable Sharpe ratio* S_{SR}*, estimated mean* S_μ *and estimated variance* S_Σ*, the robust estimation solves*

$$\min_{w} \max_{\mu,\Sigma \in S_{SR} \cap S_\mu \cap S_\Sigma} w^\top \Sigma w \qquad s.t.\ \ w^\top \mathbb{1} = 1,\ \ w^\top \hat{\mu} = \mu_0.$$

Mean shrinkage provides robustness against the overall Sharpe ratio uncertainty, variance shrinkage governs robustness against variance, while reflecting uncertainty on the mean estimation.

In summary, 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 (2020), we also benefit from their Bayesian interpretation of building the SDF. Empirically, the optimum amount of shrinkage (toward the mean, as well as lasso and ridge) are chosen on the validation dataset, keeping all the parameter estimates (portfolio weights, as a function of shrinkage) constant. Importantly, our approach to pruning is directly driven by the problem at hand: finding a small interpretable set of test assets that efficiently span the SDF out-of-sample.

AP-Pruning achieves an optimal bias-variance trade-off by properly weighting the tree portfolios. Tree portfolios at higher nodes are more diversified, resulting in a smaller variance, while lower nodes capture a more complex structure in the returns at the cost of 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 portfolios 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 $\sqrt{N_i}$. Therefore, we weight each portfolio by $\frac{1}{\sqrt{2^{d_i}}}$, where d_i is the depth of portfolio i in the tree, which is the most natural invariant statistic to account for differences in diversification for a time-varying number of stocks in each month. The sparse selection on tree portfolios is based on a similar insight as PCA but offers alternative basis assets that are interpretable while sharing the key features of PCA. 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 arguments similar to those of the PCA weighting in Kozak, Nagel, and Santosh (2020), where all the assets are multiplied by the eigenvectors of the covariance matrix, and, hence, portfolio selection is done in the PCA space. The first PC is usually an equally weighted market factor and is scaled by $\sqrt{N}$, as the market affects by 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 to achieve a 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, which are easy to trace back to the fundamentals and interpret. Section IA.2 in the Online Appendix provides several illustrative simulations.

III. Empirical Design

1. Data

We use standard data from CRSP/Compustat to construct portfolios based on firm-specific characteristics from January 1964 to December 2016, and the one-month Treasury bill rate as a proxy for the risk-free rate, yielding 53 years of monthly observations.

To build sorting variables for decision trees and standard portfolio sorts, we have constructed 10 firm-specific characteristics from the Kenneth French Data Library, based on accounting and market data. Market-based measures are updated monthly, while all the accounting variables are updated at the end of June, following the convention. A full description of the data can be found in Appendix A.

Tree-based portfolios rely only on the characteristic data used in the construction of a particular tree, which is usually a small subset of the overall list of predictors. Therefore, AP-Trees (like standard sorts) do not require a balanced panel of stocks having observations

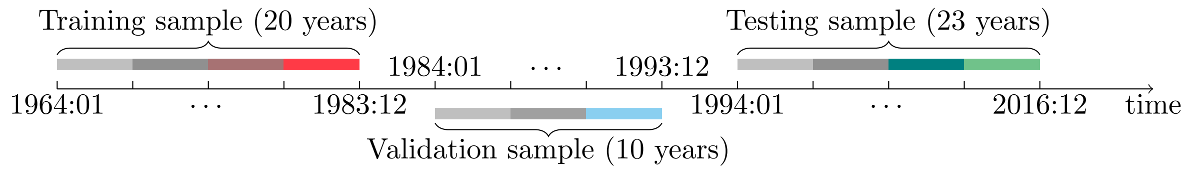


Figure 4: Timeline of the empirical strategy
The SDF weights and portfolio components are estimated on the training data (first 20 years). The shrinkage parameters are chosen on the validation data (10 years). All performance metrics are calculated out-of-sample on the testing sample (23 years).

for *all* the characteristics – a very demanding, yet typical prerequisite for many machine-learning methods. Empirically, this would require either a substantial sample reduction, or ad-hoc characteristic imputation, resulting in further selection bias.[14] In comparison, our approach requires minimum data filters.

2. Estimation and Hyperparameter Tuning

To establish baseline empirical results closest to the current literature, we start with a triplet of stock-specific characteristics: size and two other variables, which implies a set of 36 cross-sections based on three variables each. We consider a cross-section of AP-Trees with depth four and its closest analogue in the literature: 32 and 64 triple sorts.[15] All the portfolios are value-weighted, and we exclude level-four nodes based on a single characteristic to avoid what could be considered an unfair advantage of AP-Trees going into extreme tails.[16] We select 40 AP-Tree portfolios in the baseline empirical application, which makes the dimension of pruned trees comparable to Fama-French triple-sorted 32 and 64 portfolios, and later investigate this choice further. Our main results are based on measuring out-of-sample forward-looking performance of portfolios. We divide all the data into three samples as described in Figure 4:

Training sample. We use the first 20 years of data to estimate portfolio characteristics and fix their selection for a range of tuning parameters values. We form AP-Trees of depth four and estimate average returns and covariances for both final and intermediate nodes that are then used to construct an efficient portfolio frontier with elastic net regularization. For different levels of shrinkage (lasso, ridge, and target expected return), we select optimal test assets and form the corresponding SDF spanned by them. The same procedure is applied

[14] For example, whenever a characteristic is missing, it is usually set to the cross-sectional average of the variable, regardless of related market/accounting characteristics.

[15] 32 triple-sorted portfolios consider a single split on size, while 64 portfolios reflect all three characteristics in a similar way and hence could provide a somewhat more justified benchmark for AP-Trees.

[16] Since many anomaly returns tend to be concentrated in the tails of the cross-sectional distribution, including such portfolios would only improve the performance of AP-Trees. The results are available upon request.

to triple sorts and other benchmark test assets, allowing us to focus on the effect of using different basis assets.

Validating sample. 10 years of returns are used for selecting tuning parameters: We track the performance of the portfolio frontier (formed in the training sample) based on the validation data, and select optimal shrinkage to maximize its Sharpe ratio. Lasso penalty is used to set the number of non-zero weights to the target number of portfolios, K, delivering portfolio selection. We then fix the set of chosen portfolios and their weights in the SDF.

Testing sample. The last 23 years of data are used to compare the performance of different basis assets, achieved fully out-of-sample.

3. Baseline Evaluation Metrics

We run standard asset pricing tests of different cross-sections built on the same characteristics against the most popular reduced form models:

- *FF3*: Fama-French three-factor model with market, size, and value factors;
- *FF5*: Fama-French five-factor model, adding investment and profitability factors;
- *XSF*: A cross-section-specific model with market and three long-short portfolios, corresponding to the three characteristics used in the cross-section.
- *FF11*: An 11-factor model, consisting of the market factor and all 10 long-short portfolios, based on the full set of 10 characteristics.

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

1. *SR*: Out-of-sample Sharpe ratio of the SDF constructed with *AP-Trees*, *TS(32)*, and *TS(64)*, using cross-section-specific optimal shrinkage in each case.
2. α: SDF pricing error α (and its t-statistic), obtained from a linear regression against different factor models.
3. α_i: Pricing errors for individual basis assets, relative to a set of candidate factors.
4. *XS-*R^2: Relative magnitude of pricing errors in a given cross-section, that is,

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

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

IV. AP-Trees vs. Triple Sorts

1. 36 Cross-Sections of Expected Returns

We start by evaluating out-of-sample SDFs spanned by cross-sections of different basis assets, based on three characteristics. Figure 5a presents monthly Sharpe ratios (SR), achievable with AP-Trees, triple sorts, and the cross-section-specific long-short portfolios XSF (accompanied by the market factor). Cross-sections are sorted based on SR achievable with 40 AP-Tree portfolios. The full estimation results are reported in Table B.1 in the Appendix.

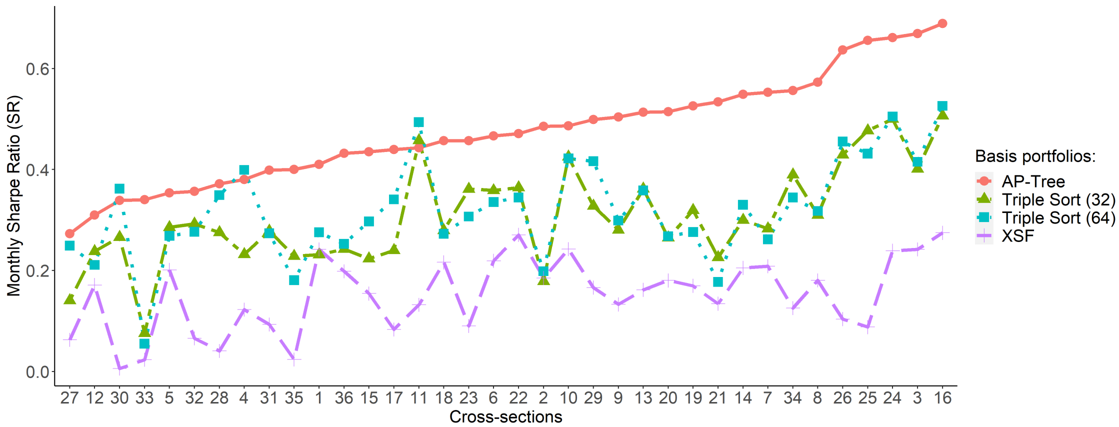
AP-Trees deliver considerably higher out-of-sample Sharpe ratios compared to triple sorts or conventional long-short factors, and the difference is often striking. Compared to simple long-short factors, our basis assets deliver SR up to three times higher, implying that AP-Trees reflect considerably more underlying pricing information. These results are particularly strong for cross-sections including investment, idiosyncratic volatility, or profitability.[17]

Cross-section-specific factors have around half of the Sharpe ratios spanned by triple sorts, that is, the standard long-short factors already miss important cross-sectional information. This is expected, since even by construction the long-short factors cannot efficiently account for the interactions between the characteristics, while triple-sorting could reflect their impact at least to a partial extent. However, 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.

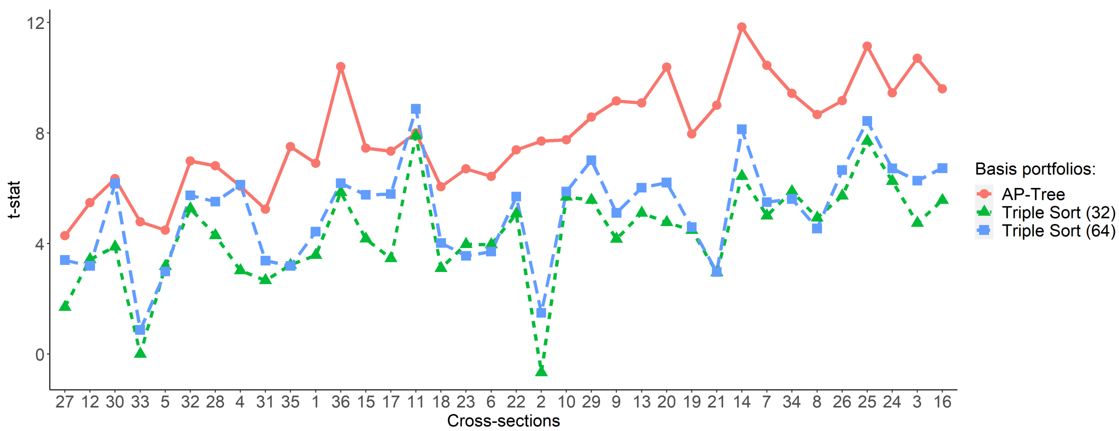
Could it be that the difference in SR is simply driven by a higher loading on conventional risk factors? This does not seem to be the case: Figure 5b confirms that the pricing error of the robust SDF, constructed from AP-Trees, is the highest among different basis assets. In fact, the pattern in these pricing errors aligns *exactly* with the total SR achieved for different characteristics. While the Fama-French five-factor model successfully spans some of the cross-sections built with triple sorts, it fails to capture the information reflected in the AP-Tree portfolios. Consider, for example, the case of size, value, and profitability (cross-section 2). The SDF, spanned by triple sorts, does not have a significant alpha, when pitted against Fama-French five factors, while the one built from AP-Trees has a t-statistics of 8.

Conventional factor models, as measured by $XS\text{-}R^2$, also fail to explain the spread of average returns within AP-Tree cross-sections (see Figure 5c). Typically, cross-sections that obtain an R^2 of more than 80% would be considered as being well-explained by the set of

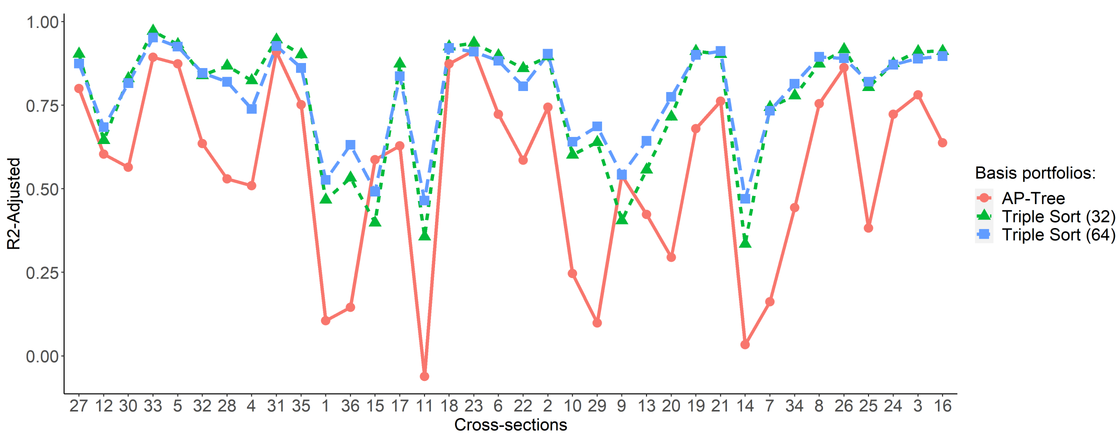
[17]The only cross-sections where triple-sorting achieves performance similar to that of AP-Trees are those that include short-term reversal. This is due to a high time variation in this pricing signal and its relationship with other characteristics, in which case it would be advisable to rely on time-varying portfolio weights, e.g., estimated via a rolling-window approach. We investigate this further in Section IV.6.



(a): Sharpe ratio of the robust SDF spanned by AP-Trees and conventional sorts.



(b) Alpha of the robust SDF spanned by AP-Trees and conventional sorts w.r.t. *FF5*



(c) Cross-sectional R^2 with AP-Trees and triple sorts as basis assets w.r.t. *FF5*

Figure 5: Cross-sections based on three characteristics

Panel (a) displays monthly out-of-sample Sharpe ratios of the robust MVE (mean-variance efficient) portfolios spanned by pruned AP-Trees (40 portfolios), triple sorts (32/64 portfolios), and cross-section-specific factors (XSF). SDFs based on triple sorts are constructed using mean and variance shrinkage. Panel (b) displays the t-statistics of the out-of-sample SDFs alpha relative the Fama-French five-factor model. Panel (c) depicts XS-R^2 within each of the cross-section of basis assets, relative to the the Fama-French five-factor model. All the cross-sections are sorted by the SR achieved with AP-Trees.

factors, and this is indeed largely the case for triple sorts and Fama-French five-factor model. In contrast, one hardly gets the average level fit of over 50% on AP-Trees, with some of the cross-sections having $XS\text{-}R^2$ below 25%.

Our findings are robust to the choice of the benchmark model: They are almost identical to those obtained with cross-section-specific factors, as well as an 11-factor model (see Figure C.2 in the Appendix). This suggests that AP-Tree portfolios are *overall* harder to price, and our results on the Sharpe ratio and SDF alpha were not driven by a small number of really challenging standalone portfolios. Instead, it seems to be a general feature of the data, once again suggesting that unconditional sorting does not provide a reliable benchmark for a cross-section of portfolios based on characteristics. In the spirit of Barillas and Shanken (2016), we investigate this further by performing a simple spanning test.

Lettau and Pelger (2020) recently proposed a generalization of the conventional Principal Component Analysis, RP-PCA, which extracts latent factors designed to price a particular cross-section of asset returns. Leading RP-PCA factors, therefore, provide a reliable small-dimensional representation of a large set of portfolios, and thus are ideal for spanning tests. We extract five leading latent RP-PCA factors from 40 AP-Tree and 32/64 triple-sorted portfolios in each of the 36 cross-sections. Note that traditional long-short portfolios used in cross-section-specific factors can also be viewed as a special case of these latent factors (extracted from the corresponding single sorts). Our out-of-sample analysis extends to spanning tests as well; that is, we test whether latent factors from different cross-sections price their optimal combination feasible to investors.

Figure C.3 in the Appendix summarizes our findings. First, we confirm that RP-PCA provides a reliable summary of the original cross-sections: SR, achievable out-of-sample by combining leading RP-PCs, is close to the one delivered with our robust SDF recovery in the whole cross-section (see Figure C.3a). In line with the previous findings, out-of-sample SR achievable with RP-PCs from triple sorts is higher than the one based on only cross-section-specific long-short factors, yet substantially inferior to the RP-PCs based on AP-Trees.

Importantly, Figure C.3b confirms that five latent factors based on AP-Trees can price most of the AP-Tree portfolios, with only a few cross-sections with significant pricing errors, indicating the need for more latent factors to capture the complex pricing patterns. RP-PCs based on triple sorts, however, face a significant challenge in all the cross-sections, that is, RP-PCs based on triple-sorted portfolios do not span AP-Tree portfolios. Conversely, Figure C.3c shows that the latent factors based on AP-Trees span the SDF based on triple sorts. In other words, AP-Trees span triple sorts out-of-sample, but not the other way around. Our results are robust to the number of RP-PCA factors, and additional robustness checks are in the Internet Appendix in Section IA.3.3.

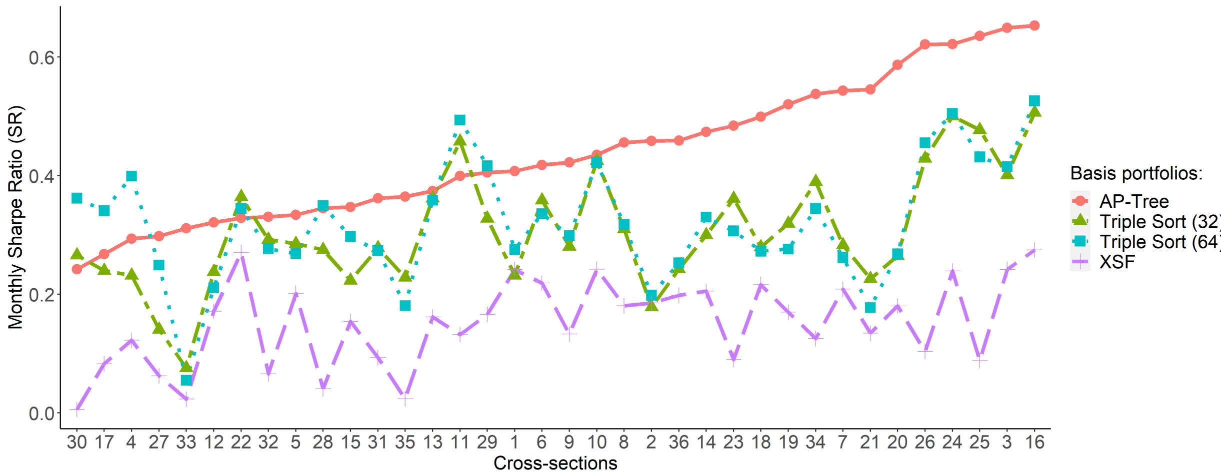


Figure 6: AP-Trees (10 portfolios) vs. triple sorts

The figure displays monthly out-of-sample Sharpe ratios of the robust MVE portfolios spanned by AP-Trees, pruned to 10 portfolios, relative to those spanned by 32/64 triple sorts and cross-section-specific factors (XSF). All the cross-sections are sorted by the SR achieved with AP-Trees pruned to 10 portfolios.

2. How Many Portfolios?

In the previous section we chose to prune AP-Trees to 40 portfolios in order to make results comparable across different cross-sections, without the additional contamination caused by the degrees of freedom. However, in practice, increasing the number of portfolios may not always be beneficial. One one hand, a larger number of portfolios could generally yield better spanning of the SDF. On the other hand, mechanical addition of the test assets could also lead to unnecessary repackaging of the same data without providing any new information. Indeed, a simple and rather parsimonious structure of the portfolios could often be enough to adequately capture the SDF projected on them. How many portfolios are enough to span the SDF? In this subsection, we investigate the role of sparsity in the cross-section of portfolios. We show that the pruning stage of AP-Trees leads to a small number of portfolios that are more diversified and liquid than conventional sorts, yet provide a much better span of the SDF. In other words, the question should be not of “how many?” but “which ones?”

We use lasso penalty as a way of building cross-sections of different size, and find that almost all of our empirical results carry through using only a quarter of the original basis assets. Figure 6 depicts monthly out-of-sample Sharpe ratios and SDF pricing error spanned by only 10 selected AP-Tree portfolios, relative to those of 32/64 triple sorts and the corresponding long-short factors. For most of the cases, using just a quarter of the original cross-section is enough to retain roughly 90% of the original Sharpe ratio and its alpha relative to the Fama-French five-factor model. The optimal number of portfolios depends on the complexity of the conditional SDF, projected on these characteristics, and could be chosen optimally based on the validation sample or using a full cross-validation. The general conclusion is clear, however: The sheer number of portfolios used to build a cross-section is often a poor reflection of the underlying information captured by those portfolios. Conventional measures of fit often used in empirical work, like R^2, are not robust to recombining assets into larger, denser portfolios, and, as a result, are prone to a substantial bias.

The key driver behind the ability of AP-Trees to condense a large amount of information in a small number of assets lies in the pruning methodology outlined in Section I.3. Since our method is designed to select portfolios not only by the type and value of the characteristics used for splitting but also the *depth* of the split, it effectively merges smaller assets together whenever it is optimal for the overall mean-variance-efficient (MVE) optimization. As as a result, our approach creates endogenously optimal portfolios that contain a larger number of securities (both in count and market cap) designed to span the SDF. Indeed, we find that intermediate nodes of the trees (as well as the market itself) constitute a substantial fraction of the chosen basis assets and substantially contribute to the overall projection of the pricing kernel. We revisit this issue further in Section IV.5, where we focus on a particular cross-section and examine the SDF structure identified with AP-Trees.

3. Nonlinearities and Characteristic Interactions

We now set to identify the sources of the AP-Trees performance. One of the advantages of using AP-Trees lies in their natural ability to impose various constraints on the set of potential basis assets, which can be achieved by simply removing particular nodes of the trees before applying the pruning procedure. Nonlinearity of characteristic impact has been the focus of many recent papers in asset pricing (see, e.g., Freyberger, Neuhierl, and Weber (2020)). However, this nonlinearity can come from two major sources: a nonlinear impact of a particular characteristic on asset returns/volatility and complicated general interaction effects among characteristics. To test these competing channels, we remove all the interaction nodes with different characteristics of the trees. Resulting AP-Trees essentially reflect the universe of single-sorted basis assets. While the robust mean-variance approach of Section II still allows us to track the nonlinear impact of stand-alone characteristics, it shuts down the channel for interactions. Importantly, our approach does not require exact parametrization of these effects; that is, combining conditional sorts with pruning allows us to capture general forms of nonlinearities without having to model them explicitly.

Figure 7 reveals that even faced with only three characteristics at a time, out-of-sample Sharpe ratios without interactions are on average half as large as those of our general AP-Trees that include interaction nodes. This further highlights the importance of modeling de-

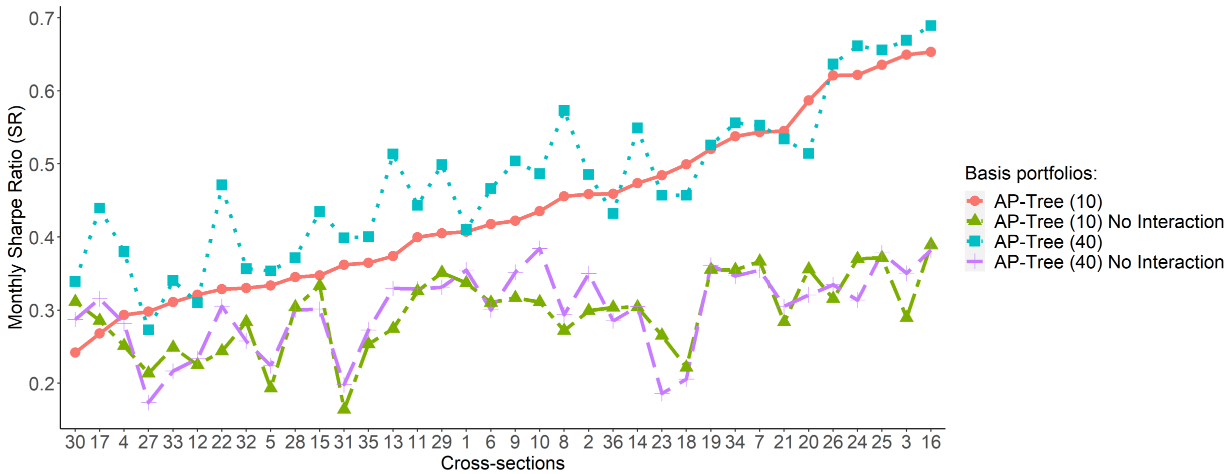


Figure 7: AP-Trees: the impact of interaction nodes

The figure displays monthly out-of-sample Sharpe ratios of the MVE portfolios spanned by AP-Trees with and without interaction nodes. All the cross-sections are sorted by the SR achieved with AP-Trees pruned to 10 portfolios.

pendencies of the conditioning information in the SDF (and resulting cross-sections). Based on Figure 5a, Sharpe ratios of the MVE portfolios, that are based on simple cross-section-specific long-short factors *XSF*, are lower than those of AP-Trees even when the latter do not include interaction nodes. This suggests that the SDF places asymmetric weights on the extreme quantiles of single-sorted portfolios, which form the long and short legs of conventional factors. In other words, nonlinearities matter even in the absence of characteristics interactions.

In summary, we find that both types of nonlinearities are important in the cross-section of asset returns. Characteristic interactions, however, seem to play a crucial role in the recovery of test asset spanning the SDF.

4. Importance of Economic Objective Function

The second driving force behind the AP-Tree performance is its economic objective function. A flexible non-parametric mapping between characteristics and portfolios alone is not sufficient to capture pricing information and represent profitable investment opportunities. There are two conceptual differences separating AP-Trees from the standard off-the-shelf data-reduction techniques and cross-section building. First, our approach implicitly takes into account the conditional impact of firm characteristics on *both* expected returns and variances of portfolios. As explained in Section I.1, recognizing this is a crucial step in creating an optimal cross-section. In contrast, most of the traditional approaches leading to portfolio formation focus on either purely return prediction (e.g., cross-sectional regressions of returns on characteristics, random forest, or conventional deep-learning prediction approaches) or pure sources of co-movement in asset returns (e.g., principal components). Second, economic theory suggests that finding an optimal set of portfolios combines the elements of both prediction and classification, making traditional ML techniques focused on pure return prediction invalid. We now show that both of these features are crucial drivers of the AP-Tree performance.

First, we consider a particular case of AP-Tree portfolios, with the mean shrinkage parameter, λ_0, set to infinity, thus using the same average return for all the portfolios in the portfolio selection step. Since this is equivalent to making a choice based only on portfolio covariance, we call such test assets *V-trees*. V-Trees start from the same pool of portfolios as AP-Trees, but differ in the way they select a low-dimensional cross-section: Similar to conventional PCA, they focus primarily on the variance reduction and neglect information in the mean. However, the mean of the selected low-dimensional basis assets is used to form a robust SDF.

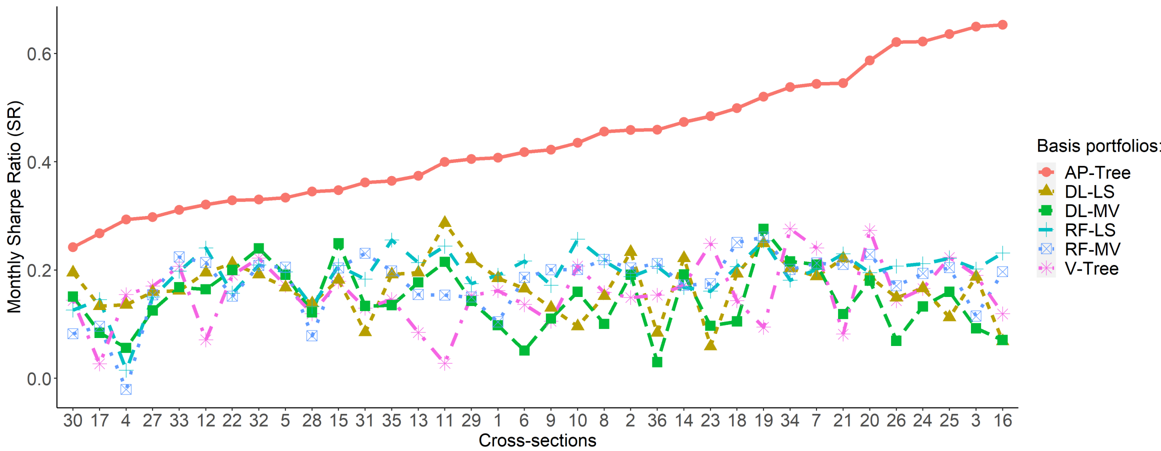
Second, we use decile-sorted portfolios based on the conditional expected returns, predicted with leading ML techniques. Popular in many recent papers, these portfolios are usually used to form an overall long-short ML-based trading strategy. Many off-the-shelf prediction-based approaches are intuitively appealing, as they easily take advantage of the flexible estimation techniques.[18] Note, however, that by construction these portfolios focus only on the return prediction step, and are not designed to span the SDF. Furthermore, return-sorted portfolios bundle different pricing signals together, often making it very difficult to disentangle individual contributions and lacking interpretability. In short, while prediction-based portfolios constructed from multiple characteristics can have high Sharpe ratios and alphas relative to benchmark models, there is nothing per se in their construction that delivers basis assets spanning the SDF, projected on the space of characteristics.

We use the best performing deep neural network of Gu, Kelly, and Xiu (2020b) to predict next period returns based on the current period characteristics and sort the stocks into decile portfolios based on the prediction.[19] The first decile portfolio is a value-weighted average of the 10% of stocks with the highest predicted returns. We also consider portfolios based on return prediction via random forest, another popular ML technique exploiting a collection of characteristic-based decision trees for return prediction.[20] Deep neural network and random forest are the two best empirically performing methods for return prediction (see Gu, Kelly,

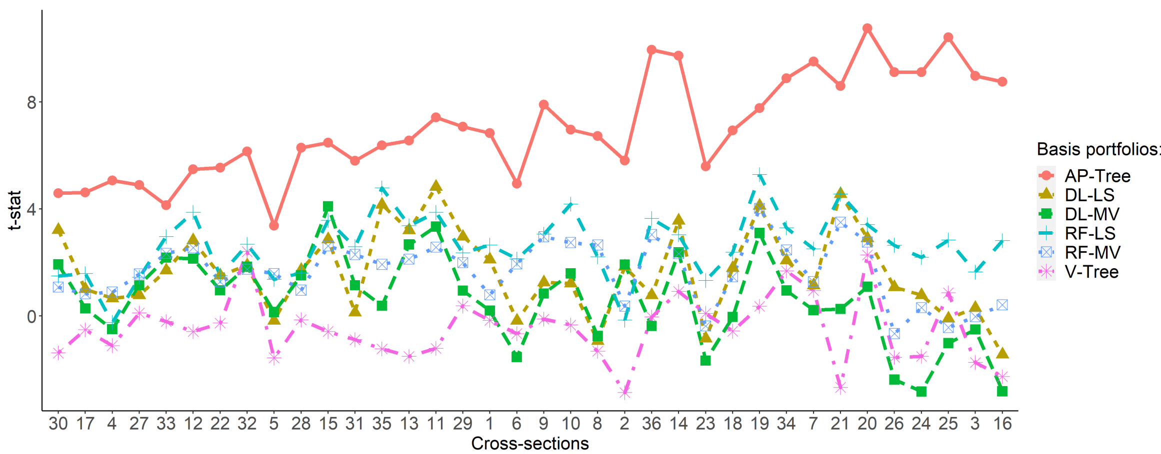
[18]Recent examples include the non-parametric adaptive lasso of Freyberger, Neuhierl, and Weber (2020) and a battery of machine-learning tools employed by Gu, Kelly, and Xiu (2020b).

[19]The best performing model is a three-layer neural network, with all the tuning parameters set to the same values as in Gu, Kelly, and Xiu (2020b). The tuning parameters of the random forest are optimally selected on the validation data.

[20]We set the random forest depth to four splits, making them comparable in granularity to triple-sorted portfolios and AP-Trees.



(a) Sharpe ratios of robust SDFs spanned by AP-Trees and standard ML-based portfolios



(b) Alpha of robust SDFs spanned by AP-Trees and standard ML-based portfolios w.r.t. *FF5*

Figure 8: AP-Trees vs. cross-sections built with off-the-shelf ML tools

Panel (a) presents monthly out-of-sample Sharpe ratios of the robust MVE portfolio spanned by AP-Trees, V-Trees, and portfolios based on return forecasts with deep learning (DL-MV) and random forest (RF-MV). RF-LS and DL-LS denote long-short portfolios based on the corresponding prediction deciles, formed with deep learning and random forest. Penal (b) reports out-of-sample t-statistics of the SDF alpha relative to the Fama-French five-factor model.

and Xiu (2020b)) and, hence, provide an appropriate benchmark for the off-the-shelf ML tools. We build robust MVE portfolios from 10 prediction deciles formed with both methods (labeled *DL-MV* and *RF-MV* for deep learning and random forest, correspondingly). In addition, we also consider a popular value-weighted long-short strategy, labeled *DL-LS* and *RF-LS*, of buying stocks in the highest prediction decile and selling those in the lowest one.

Figure 8 presents out-of-sample Sharpe ratios and pricing errors of the SDFs spanned by AP-Trees, V-Trees, and ML-based portfolios based on return prediction. First, AP-Trees clearly stand out in terms of Sharpe ratios, which are always two to three times larger than those of alternative cross-sections. The Fama-French five-factor model is successful in spanning at least half of the cross-sections based on leading ML methods, yet it is challenged by the AP-Trees. Second, Figure 8 also clearly indicates that information in the mean is crucial in selecting basis assets: V-Trees provide the least informative and easiest-to-price cross-sections, although they start from the same conditional sorting and stability-inducing ridge penalty in the pruning as AP-Trees. Their performance is poor, since targeting just the variance in selecting basis assets is an obviously wrong criterion. Finally, leading return-prediction portfolios, DL-MV, DL-LS, RF-MV, and RF-LS, display a very similar subpar performance. This clearly shows that while flexible off-the-shelf machine-learning methods do a great job at predicting returns per se, these are not the right tools for building a cross-section of asset returns, and they require significant adaptation rooted in economic theory. In other words, the signal extraction for an asset pricing problem requires an asset pricing objective.

5. Zooming into the Cross-Sections

We now turn to a particular example of the cross-section to further investigate the structure of the SDF spanned by AP-Trees: Portfolios based on size, investment, and operating profitability. This is a fairly representative cross-section, with all the results presented here naturally extending to other characteristics as well. We report the results for all other cross-sections in the Internet Appendix in Section IA.6.

Table 1 presents the summary statistics for cross-sections, built with AP-Trees and triple sorts. SDFs spanned by AP-Trees have higher SR and generally larger pricing errors α (or, equivalently, lower cross-sectional $XS\text{-}R^2$) than the conventional portfolios, based on the unconditional quantiles. Interestingly, performance of the SDF based on only 10 portfolios, is already very close to the full SDF based on 40 test assets. This is reassuring, since smaller cross-sections are also intuitively very appealing, and allow for easier interpretations of the spanning tests, sources of risk premia, or factors based on them. Furthermore, this decision is directly guided by the data-driven selection from the validation sample. Figure C.8a in the Appendix shows that a small number of original test assets is indeed enough to span most of the SDF, without introducing the danger of overfitting.

Figure C.8b in the Appendix presents the heatmap of tuning parameters and their optimally chosen values based on the validation data-set. In this particular case, the impact of the L2 penalty (ridge) on the composition of the cross-section was rather small; however, the model performance depends on the choice of the shrinkage to the mean, λ_0. In this particular case, the chosen value of λ_0 was 0.15, corresponding to an important, but not excessive, shrinkage toward the minimum variance portfolio induced by the estimation

Table 1: Cross-sections based on size, operating profitability, and investment

		Type of cross-section			
		AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)
SDF SR		0.65	0.69	0.51	0.53
α	FF3	0.94	0.90	0.75	0.84
		[10.11]	[11.03]	[7.40]	[8.13]
	FF5	0.81	0.76	0.47	0.61
		[8.76]	[9.60]	[5.57]	[6.73]
	XSF	0.81	0.76	0.46	0.61
		[8.77]	[9.46]	[5.39]	[6.69]
	FF11	0.89	0.80	0.37	0.65
		[9.12]	[9.60]	[4.29]	[6.91]
$XS\text{-}R^2$	FF3	18.0%	51.0%	82.0%	82.0%
	FF5	11.0%	64.0%	91.0%	90.0%
	XSF	28.0%	65.0%	91.0%	90.0%
	FF11	–	42.0%	92.0%	87.0%

The table presents aggregate properties of the cross-sections 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 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 the Fama-French three- and five-factor models, cross-section-specific factors (market, size, investment, and operating profitability), and the composite FF11 model.

error in expected returns.[21] Optimal amount and type of shrinkage can be different for each cross-section: Shrinkage to the mean here, for example, indicates that there is a substantial estimation uncertainty in the estimated expected returns relative to their actual spread. Note, however, that even without mean and variance shrinkage, AP-Trees would still outperform conventional sorting, since most of the empirical performance is derived from relying on conditional tree-based splits (and their selection based on the explicit asset pricing objective).

What are these 10 portfolios left after pruning the tree? Table 2, Panel A, provides the general description of these basis assets and their main characteristics: The relative number of stocks and its value-weighted counterpart, based on the market cap of the securities that go into a corresponding portfolio, as well as pricing errors, relative to leading reduced-form asset pricing models. Out of 10 selected portfolios, only five correspond to the final nodes of the trees, containing just over 6% of all the stocks; other assets include trading strategies that can be constructed from only one or two characteristic splits, and even the market itself. For example, portfolio 6 is created by taking the bottom 50% of the stocks based on their size (LME) and within them, the bottom 25% of the firms sorted by investment, as a result, containing 12.5% of the stocks at all the time periods.[22] This portfolio presents a challenge

[21]The mean shrinkage is scaled by the cross-sectional average, that is $\hat{\mu} + \lambda_0 \bar{\mu} \mathbb{1}$.

[22]The portfolio ID of AP-Trees has two parts: the characteristic used for the split (LME/OP/INV) and

Table 2: Portfolios in the cross-sections

Portfolio ID	Portfolio description	% of stocks	WV size quantile	α_{FF3}	α_{FF5}	α_{XSF}	$\alpha_{FF_{11}}$
Panel A: 10 AP-Tree portfolios							
1 (1111.1)	market	100%	0.95	–	–	–	–
2 (1111.121)	medium caps	25%	0.71	-0.03	-0.37	-0.05	-0.03
				[-0.34]	[-0.36]	[-0.40]	[-0.26]
3 (3111.11)	low profitability	50%	0.94	0.03	-0.11**	-0.11**	-0.11**
				[-0.54]	[-2.33]	[-2.20]	[-2.46]
4 (3111.11111)	low investment, small caps	6.25%	0.07	1.49***	1.32***	1.32***	1.63***
				[6.14]	[5.33]	[5.33]	[6.47]
5 (3211.1121)	low investment, high profit., small caps	12.5%	0.51	0.23***	0.09	0.07	0.06
				[2.95]	[1.19]	[0.66]	[0.83]
6 (1221.1111)	small caps, low profitability	12.5%	0.37	-0.37***	-0.34***	-0.34***	-0.29**
				[-2.81]	[-3.00]	[-3.01]	[-2.41]
7 (2221.11111)	low profitability, small caps	6.25%	0.19	-0.16	-0.22	-0.21	-0.05
				[-0.717]	[-0.95]	[-0.89]	[0.21]
8 (3331.12221)	high investment, small caps	6.25%	0.45	-0.82***	-0.62***	-0.62***	-0.46***
				[-5.86]	[-4.81]	[-4.78]	[-3.75]
9 (2122.11111)	low profitability, small caps	6.25%	0.27	-0.37*	-0.43**	-0.42**	-0.20
				[-1.95]	[-2.21]	[-2.13]	[-1.00]
10 (3133.12122)	high investment, small caps	6.25%	0.48	-0.84***	-0.64***	-0.64***	-0.50***
				[-5.92]	[-4.84]	[-4.81]	[-3.92]
Panel B: Top 10 most "challenging" assets from 64 triple-sorted portfolios							
122	LME(0-0.25) × OP(0.25-0.5) × Inv(0.25-0.5)	1.91 %	0.17	0.97***	0.84***	0.84***	0.95***
				[6.61]	[5.60]	[5.54]	[6.01]
214	LME(0.25-0.5) × OP(0-0.25) × Inv(0.75-1)	1.55 %	0.40	-0.93***	-0.61***	-0.61***	-0.66***
				[-5.27]	[-3.60]	[-3.61]	[-3.77]
132	LME(0-0.25) × OP(0.5-0.75) × Inv(0.25-0.5)	1.03 %	0.17	0.39*	0.40*	0.38*	0.77***
				[1.93]	[1.90]	[1.74]	[3.65]
121	LME(0-0.25) × OP(0.25-0.5) × Inv(0-0.25)	1.9 %	0.17	0.59***	0.45**	0.44**	0.66***
				[3.38]	[2.48]	[2.45]	[3.56]
112	LME(0-0.25) × OP(0-0.25) × Inv(0.25-0.5)	1.66 %	0.17	0.24	0.25	0.24	0.49***
				[1.45]	[1.45]	[1.38]	[2.79]
123	LME(0-0.25) × OP(0.25-0.5) × Inv(0.5-0.75)	1.47 %	0.17	0.38**	0.32**	0.32**	0.43***
				[2.56]	[2.10]	[2.08]	[2.71]
222	LME(0.25-0.5) × OP(0.25-0.5) × Inv(0.25-0.5)	1.95 %	0.40	0.46***	0.35***	0.34***	0.28**
				[4.69]	[3.52]	[2.98]	[2.69]
244	LME(0.25-0.5) × OP(0.75-1) × Inv(0.75-1)	1.18 %	0.41	-0.57***	-0.47***	-0.49**	-0.34**
				[-3.44]	[-3.18]	[-2.65]	[-2.28]
334	LME(0.5-0.75) × OP(0.5-0.75) × Inv(0.75-1)	1.08 %	0.66	0.35**	0.34**	0.32*	0.37**
				[2.33]	[2.18]	[1.77]	[2.25]
111	LME(0-0.25) × OP(0-0.25) × Inv(0-0.25)	5.04 %	0.16	0.42*	0.21	0.22	0.48**
				[1.89]	[0.95]	[0.97]	[2.19]

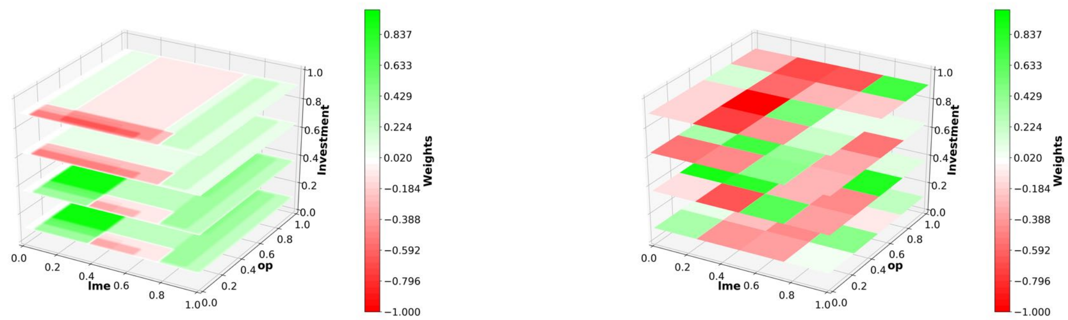
The table presents the properties of the portfolios, spanning the impact of size, investment, and operating profitability on asset returns. Panel A presents the set of 10 portfolios, created from trees of depth 4 (which exclude the extreme portfolios, representing 1/16 of the stocks sorted by the same 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 three- and five-factor 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, respectively. Panel B 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 portfolio ID for AP-Trees denotes the sequence of characteristic (LME/OP/INV) in the first part and the direction (low/high) of the split in the second part. Figure C.10 in the Appendix illustrates the composition for all 10 AP-Tree portfolios. The portfolio ID for triple sorts denotes the quantile for (LME/OP/INV).

the direction of the split (low/high). The first number of the second component denotes the parent node (market) and not a split; i.e., the split direction starts with the second number. For example, portfolio 6 with the id of 1221.1111 has its first split with characteristic 1 (size) to the left denoted by 1 (bottom 50%). Within this portfolio, the next split is with characteristic 2 (investment) to the left (bottom 50%). The third and final split is with characteristic 2 (investment) to the left (bottom 50%). Portfolio 8 with id 3331.12221 denotes characteristic 3 (operating profitability) to the right three times (top 12.5%) and characteristic 1 (size) to the left (bottom 50%).

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, portfolio 8 is constructed by taking small cap stocks among those highest in investment and has an alpha of 41-86 bp, depending on the underlying model. Notably, the average (value-weighted) market cap of the stocks forming this portfolio, is close to being 50% of the market, so this pricing performance cannot be driven by microcaps. Overall, six of the chosen 10 tree-based portfolios have consistently significant stand-alone alphas. These portfolios are not due to extreme quantiles of the underlying characteristics but are in fact characterized by a complex interaction structure. This is illustrated by Figure C.10 in the Appendix, which shows the composition for all 10 AP-Tree portfolios in the characteristic space. 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 persists when basis assets are priced with the Fama-French five-factors, or even 11 cross-sectional factors (see Table 2, Panel A).

Table 2, Panel B, lists the top 10 portfolios from the cross-section of 64 triple-sorted assets that are most difficult to price for the most comprehensive FF11 model. While there are obviously substantial alphas associated with some of these portfolios, the sheer number of "significant alphas" is somewhat misleading: Most of these portfolios consist of a small number of stocks (usually about 1.5% of the market) and often reflect assets that are very similar to each other in terms of characteristics: Small in size and profitability and high in investment. Since our pruning algorithm selects the basis assets in both types of characteristics used for splits and their depth, these portfolios could actually be grouped together by AP-Trees.

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 that 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 XS-R^2, based on the simple OLS estimates. This is precisely why the R^2 of a GLS estimation is often empirically lower than its OLS counterpart (see, e.g., Lewellen, Nagel, and Shanken (2010)): Many portfolios do not provide a source of incremental Sharpe ratio, which is reflected in the GRS statistic and other quantities that target investment opportunities, rather than a linear measure of fit. The stark difference between the number of correctly priced test assets and its spanning is illustrated in Figure C.9 in the Appendix. It shows the pricing errors for each of the individual basis assets with respect to XS-specific factors, with the candlestick denoting 5%



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

(b) SDF spanned by TS (64)

Figure 9: SDF weights in AP-Trees and triple sorts

This figures shows the structure of the conditional SDF spanned by 10 AP-Trees portfolios and 64 triple-sorted basis assets for size, investment, and operating profitability.

confidence intervals. The large number of triple-sorted portfolios with insignificant pricing errors represent largely redundant assets in the "center" of the sorting cube.

The need for portfolio depth is particularly visible when we compare the structure of the SDFs in Figure 9. Stocks generally high in investment, are grouped together in AP-Trees and present the same exposure to the SDF, compared to six separate portfolios spanning the same characteristic space in triple sorts (the top layer of the three-dimensional graph in Figure 9)). While some of the general patterns of the loadings are shared across cross-sections, it is immediately clear that AP-Trees reflect a more sophisticated data-generating process, which triple sorts aim to capture 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 new long-short patterns in the data, and could present a new challenge for both reduced-form and structural models.

Our main empirical results are unlikely to be driven by microcaps. First, in constructing AP-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 a single characteristic throughout the tree (e.g., 16 portfolios sorted only by size). Second, not only tree-based portfolios generally include a larger number of stocks, efficiently diversifying the idiosyncratic noise, and endogenously grouping similar securities together: They are either similar to or more liquid than the most challenging test assets among triple sorts. For the case of size, profitability, and investment (see Table 2), 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. Interestingly, superior performance of the AP-Tree-based SDF cannot be attributed to turnover, as the

Table 3: Size, operating profitability, and investment: Cross-sections without small caps

		Panel A: Portfolios with size quantile above 0.4				**Panel B:** Stocks with market cap above 0.01%			
		AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)	AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)
SDF SR		0.48	0.45	0.25	0.40	0.30	0.22	0.17	0.17
α	FF3	0.68	0.73	0.47	0.76	0.70	0.89	0.77	0.82
		[7.51]	[7.64]	[2.30]	[6.00]	[5.04]	[4.27]	[2.74]	[2.78]
	FF5	0.45	0.47	-0.11	0.54	0.32	0.37	0.16	0.25
		[5.70]	[6.11]	[-0.72]	[4.44]	[3.87]	[2.57]	[0.72]	[1.00]
	XSF	0.44	0.47	-0.11	0.54	0.31	0.36	0.14	0.22
		[4.76]	[5.04]	[-0.74]	[4.21]	[3.75]	[2.49]	[0.65]	[0.88]
	FF11	0.43	0.43	-0.25	0.34	0.29	0.36	0.18	0.38
		[5.37]	[5.27]	[-1.62]	[2.76]	[3.71]	[2.45]	[0.83]	[1.55]
XS-R^2	FF3	62.0%	73.0%	76.0%	86.0%	42.0%	96.0%	91.0%	90.0%
	FF5	70.0%	85.0%	91.0%	93.0%	70.0%	98.0%	95.0%	93.0%
	XSF	76.0%	85.0%	92.0%	93.0%	71.0%	98.0%	95.0%	93.0%
	FF11	–	86.0%	84.0%	94.0%	–	98.0%	95.0%	93.0%

The table presents the aggregate properties of the cross-sections based on size, operating profitability, and investment, created from AP-Trees (pruned to 10 and 40 portfolios, correspondingly) and triple sorts (32 and 64 portfolios). Panel A is based only on the basis portfolios that have a value-weighted size quantile greater than 0.4, while Panel B describes cross-sections created only from stocks with market capitalization greater than 0.01% of the total market capitalization, i.e., currently around top 600 stocks in the U.S. For each 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 the Fama-French three- and five-factor models, cross-section-specific factors (market and long-short portfolios, reflecting size, operating profitability, and investment), and the composite FF11 model which includes the market portfolio, along with the 10 long-short portfolios, based on the cross-sectional characteristics.

long and short legs of the MVE portfolio have the same turnover of 24-28% as the MVE portfolio based on cross-section specific factors.

Finally, the flexible nature of AP-Trees allows us to directly investigate the role of small caps in the SDFs spanned by different test assets. Table 3, Panel A, presents results after removal of all the nodes that have an average size quantile below 40% of the market-wide distribution, thus implicitly not allowing for splits that lead to portfolios with too many small-cap stocks. We find that AP-Tree portfolios are still harder to price and remain more informative about the SDF, even after removing portfolios consisting of small cap stocks. Following Kozak, Nagel, and Santosh (2020), we also start with the universe of liquid stocks having a capitalization above 0.01% of the total market, which currently corresponds to the universe of roughly 600 largest stocks in the United States (Panel B). This removes not only micro- but also many small- to medium-cap stocks. As expected, Sharpe ratios and pricing errors become significantly lower than using the whole data-set, but the qualitative difference between AP-Trees and conventional sorting remains. For example, the out-of-sample Sharpe ratio of the SDF, spanned by 10 AP-Tree portfolios, is almost *double* that of triple sorts. Furthermore, all AP-Tree SDFs alphas remain significant even at 1%, while triple sorts are routinely spanned by conventional factors.

6. Time Variation

All of our results so far assumed a constant projection of the SDF on the set of basis assets. To study time variation in these weights, we fix the portfolio selection based on the training and validating samples and estimate the optimal robust combination of the basis assets on a rolling window of 20 years.

Figure 10 summarizes our findings. First, time variation in the SDF weights seems to be quite important empirically for both triple sorts and AP-Trees. It definitely improves the out-of-sample performance of the robust SDF, as it allows us to achieve a higher Sharpe ratio with almost all of the cross-sections. The difference is large not only statistically but economically as well. Second, using AP-Trees with flexible weights clearly uniformly dominates triple sorts across the whole set of characteristics, including short-term reversal, which presented an earlier challenge (cross-sections 30, 4, and 11).

Second, we measure the sensitivity of empirical results with respect to a particular type of shrinkage used. Our main results are reported in Figure C.6 in the Appendix, comparing mean and variance shrinkage effect on the Sharpe ratio, achievable with time-varying SDF weights. Both AP-Trees and triple sorts benefit from shrinkage to the mean, which substantially stabilizes empirical performance. These effects, however, are not homogeneous across different cross-sections: For many sets of basis assets, the effect is fairly small (with a magnitude of 1–5%), but for other cross-sections, characterized by a large time variation in risk premia, mean shrinkage leads to large improvements in performance (up to 50%). Interestingly, there is clear synergy in combining mean and variance shrinkage, as it leads to the largest overall performance gains.[23] Section IA.3.1 in the Internet Appendix reports the same results using a fixed degree of mean and variance shrinkage, that is, if we don't choose the penalty parameters optimally for each cross-section on the validation data set, but instead pre-specify them to a fixed value. Using a combination of the mean and variance shrinkage leads to better results than no shrinkage or only variance shrinkage, even if tuning parameters are chosen non-optimally. Hence, our SDF construction is very robust, as its performance is driven by *some degree of combined shrinkage* rather than choosing a particular optimal level. In short, empirical performance of the AP-Trees seems to be mainly driven by the construction of the new basis assets, rather than a particular choice of tuning parameters that leads to their contribution to the SDF.

Finally, we also repeat the empirical exercise by using a cross-validation approach similar to Kozak, Nagel, and Santosh (2020), which allows us to also evaluate our model on the initial subsample of the data. We use three-fold cross-validation by splitting the whole

[23]Note that to isolate partial effects of shrinkage, we keep the optimal mean and/or variance shrinkage tuning parameters selected on the validation data.

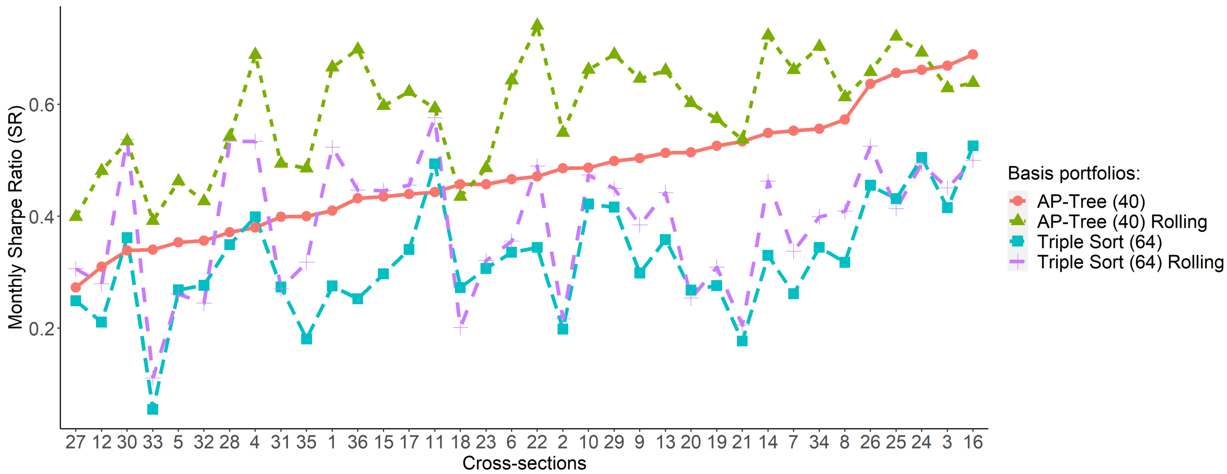


Figure 10: Sharpe ratios of the SDF estimated on a rolling window

The figure displays the monthly out-of-sample Sharpe ratio of the robust mean-variance efficient portfolios with time-varying weights estimated on a rolling window of 20 years and constant weights. The basis portfolios are the pruned AP-Trees (40 and 10 portfolios) and triple sorts. The AP-Tree portfolios are selected on the training data and kept the same. The variance and mean shrinkage is selected optimally on the validation data. All the cross-sections are sorted by the SR achieved with AP-Trees (40).

sample of data into three equal time-series blocks. Each block is used as a training period, and the model is evaluated on the remaining two. Tuning parameters are then selected to optimize the overall cross-validation results. Figure C.7 in the Appendix reports resulting Sharpe ratios and SDF alphas, confirming that the relative performance of AP-Trees and triple sorts is robust to the choice of evaluation time period. We further show in the Internet Appendix that these findings extend to all the other main results in our paper and that the estimated SDF structure remains the same under cross-validation and in the forward-looking out-of-sample approach of Section III.2.

V. AP-Trees in Large Dimension

Historically, stacking small-dimensional cross-sections together has been the only way to build a set of test assets reflecting more than two or three characteristics at the same time. Usually this involves combining long-short portfolios, decile sorts, or double-sorted portfolios together as, for example, in Feng, Giglio, and Xiu (2020), Giglio and Xiu (2021) or Hou, Mo, Xue, and Zhang (2021), among many others. With an ever-increasing list of anomalies, the size of such a cross-section explodes, especially after accounting for potential interactions. Furthermore, many composite cross-sections repackage the same underlying securities into differently labeled portfolios without providing evidence on whether they reflect different economic risks. AP-Trees and their pruning stage provide a way to identify optimal splits in

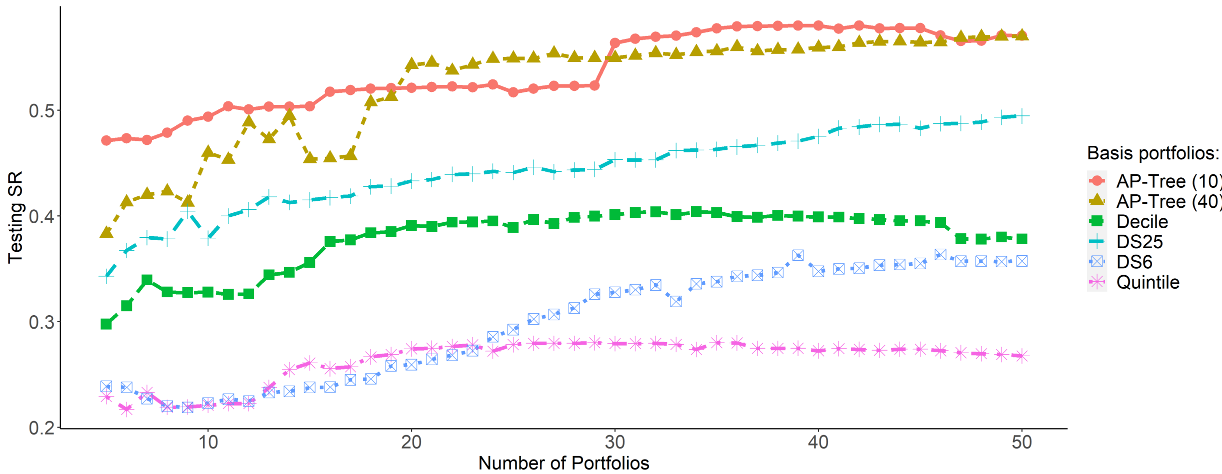


Figure 11: Sharpe ratios of the SDFs spanned by AP-Trees and traditional sorts.
Out-of-sample Sharpe ratios of SDFs based on 10 characteristics as a function of the number of basis assets constructed with AP-Trees (based on 36 AP-Trees with 10, respectively, and 40 portfolios in each), 10 quintile sorts, 10 decile sorts, combination of double sorts based on size and the other characteristic (either 6 or 25 double sorted assets per specific portfolio). Optimal mean and variance shrinkage is applied to *all* the basis assets. The number of selected portfolios is controlled by the lasso shrinkage.

the characteristic space that directly target SDF spanning and eliminate redundant assets and/or combine them together.

There is no universal benchmark for building an interpretable model-independent cross-section of portfolios reflecting more than 2–3 characteristics at a time. Therefore, we consider several popular approaches to generate the set of basis assets reflecting all 10 characteristics: a) Sets of 10 quintile portfolios, uniformly sorted by characteristics (50 assets altogether); b) sets of 10 decile-sorted portfolios (100 assets); c) a combination of six double-sorted portfolios, with each based on size and some other characteristic (54 assets), and d) a combination of 25 double-sorted portfolios, with each based on size and some other characteristic (225 assets). Tree-based portfolios are constructed based on combining selected assets (10/40 portfolios) from each of the 36 cross-section we described in Section IV.1 (starting from 360 and 1,440 assets, correspondingly).[24] Empirically, most papers focus on pricing the cross-sections of no more than 50–60 portfolios; hence, the combination of all the available decile or double sorts, if anything, overrepresents cross-sections used in the literature.

Figure 11 presents the SDF Sharpe ratio for the optimal mean-variance strategies, spanned by different basis assets. As before, to make results comparable, we rely on shrinkage optimal in each of the cross-sections, and study their performance out-of sample. Therefore, the only

[24]We have also pruned original AP-Trees based on 10 characteristics. We found that more than four-way interactions do not provide a significant contribution to the overall SDF or portfolio selection, and that starting from assets reflecting the information in three characteristics achieves the same result, both quantitatively and qualitatively. Detailed empirical results are available upon request.

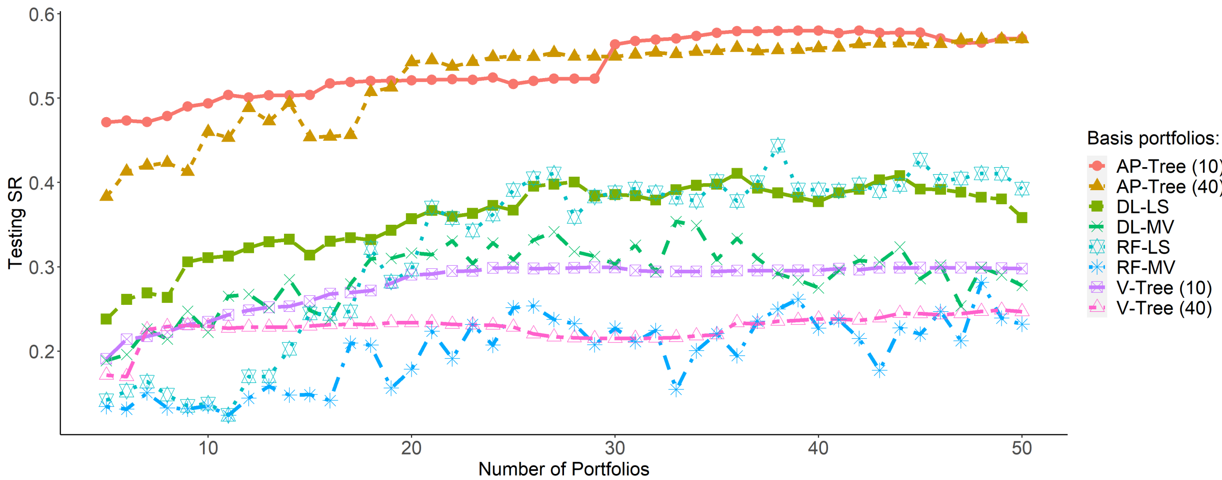


Figure 12: SR of the SDFs spanned by AP-Trees, V-Trees, and prediction portfolios

Out-of-sample Sharpe ratios of SDFs based on 10 characteristics as a function of the number of basis assets constructed with AP-Trees (based on 36 AP-Trees with 10, respectively, and 40 portfolios in each), V-Trees constructed as in Section IV.4, forecasted sorted portfolios based on deep learning DL-MV and random forest RF-MV. We apply robust shrinkage with lasso to all basis assets and choose the optimal validation mean and variance shrinkage. The number of selected portfolios is controlled by the lasso shrinkage. We also include long-short portfolios denoted as DL-LS and RF-LS based on a highest and lowest prediction quantile.

conceptual source of difference in the empirical performance, is the original set of portfolios (e.g., standard unconditional sorting vs. AP-Trees).

Several observations are in order. First, quintile-sorted portfolios represent probably the least informative cross-section one could in good faith build based on the 10 characteristics we consider. While univariate sorts are generally considered to be the staple of the empirical literature of anomalies, doubling the depth of the analysis (e.g., moving from five to potentially 10 portfolios based on the same characteristics) easily doubles the out-of-sample Sharpe ratio and presents a more realistic picture of the informational content of the same characteristics. Using only six double-sorted portfolios for each of the non-size anomalies (i.e., a potential cross-section of 54 assets) does not reflect most of the investment opportunities either.

Using the best (sparse and robust) combination of the 25×9 double-sorted portfolios is probably the best one could do within the standard framework, allowing one to achieve monthly out-of-sample Sharpe ratios of about 0.4–45, depending on the number of portfolios. AP-Trees, however, raise that bar even higher: On average, they provide a 0.1 increase in the monthly Sharpe ratio relative to the 25×9 double-sorted portfolios (in relative terms, this is roughly a 20% increase over 25×9 double sorts and a 80%–100% increase over anomaly-based deciles). Interestingly, most of the empirical performance of the SDF is achieved with about 20-30 portfolios, indicating large redundancies and repackaging in the cross-section sorted by 10 characteristics.

The increase in the composite Sharpe ratio, delivered by AP-Trees, does not come from conventional risk exposure: Figure C.5 in the Appendix presents out-of-sample alphas of the SDFs spanned by different basis assets. Regardless of the cross-sectional dimension, tree-based portfolios are considerably harder to price, and they have roughly a 20%–30% gain over the optimal subset of the double-sorted portfolios, built from the same characteristics (25×9). Decile-sorted portfolios are even easier to price.

Our results indicate that contrary to the conventional wisdom, one needs a fairly lean cross-section of assets to efficiently reflect an SDF projected on the span of a relatively large number of characteristics. Interactions, captured by conditional splits, account for almost half of the total SDF Sharpe ratio. This implies that even if the candidate model successfully explains a wide set of long-short anomalies or even decile sorts, it could still miss half of the SDF, projected on the space of returns via underlying characteristics.

To further understand the sources of AP-Tree performance, we follow the approach presented in Section IV.4. Figure 12 presents out-of-sample Sharpe ratio results for alternative test assets based on V-Trees and prediction-based portfolios from Section IV.4. V-Trees, basis assets selected from conditional sorts ignoring the mean information, have consistently two to three times lower Sharpe ratios than the robust SDF spanned by AP-Trees. Most of the prediction-based portfolios have a relatively unstable and equally subpar performance, yielding at best half the AP-Trees Sharpe ratio. In summary, our large-dimensional cross-section highlights the same features of the return space that underlie AP-Trees' advantage over simple triple sorts: flexible modeling of conditional and unconditional characteristic impact, and an objective function that directly targets SDF spanning by selecting portfolios optimal in depth and coverage.

We now turn to illustrating the effect of using different test assets in traditional horse races between asset pricing models.

1. Evaluating Asset Pricing Models

Horse races are popular among reduced-form asset pricing models, whether it is to judge the performance of a new SDF, or to compare existing benchmarks. In this subsection we illustrate two conceptual points. First, as expected, popular empirical metrics for model evaluation and comparison depend on the choice of tests assets. Optimal cross-sections, however, should span the SDF projected on stocks via characteristics, and allow for an easy diagnostic of model performance. Traditional sorts, especially deciles and simple long-short anomalies, do not span the underlying SDF (see Section V) and, therefore, even a model that successfully captures most of the alpha among hundreds of decile sorts, could be severely misspecified. As a consequence, uninformative simple sorts can be of limited use for the absolute or relative comparison of models.

Second, we show that popular measures of model performance used in larger cross-sections suffer from redundancy caused by asset repackaging. Since most of the empirically used cross-sections are naturally built by stacking portfolios built from the same universe of stocks, the sheer number of significant alphas, or OLS cross-sectional $XS\text{-}R^2$, do not reveal whether they identify independent sources of alpha. Instead, we propose to focus directly on the SDF alpha, since it allows us to aggregate all the information from the test assets and measures its out-of-sample feasible to investors performance. Finally, we show that even in stand-alone portfolio pricing, AP-Trees provide a more informative set of test assets, as the sources of its "alpha" are more independent and less prone to repackaging.

Our main findings are summarized in Table 4. We consider several high-dimensional cross-sections from Section V and evaluate leading reduced-form asset pricing models based on the empirically popular set of criteria: average absolute alpha, share of significant alphas in the cross-section, $XS\text{-}R^2$, and the alpha of the feasible out-of-sample SDF spanned by the test assets. It is important to emphasize that our goal is not to find the "best" model but to highlight that the empirically popular measures of model performance in a large cross-section crucially depend on the test assets and metrics. As before, all of our results are out-of-sample.

First, Table 4 confirms our results from Section V, since cross-sections based on AP-Trees are significantly harder to price, almost independent of the metric used in the evaluation. The relative ranking of model performance, of course, is different, and depends on the particular choice of the test assets. For example, the majority of the models seem to be very successful at pricing 100 decile-sorted portfolios, which would often be interpreted as empirical success (with top-performing models having R^2 of 91–92%). As the differences in explaining the decile-sorted portfolios is only minor among most factor models, they are of limited use to distinguish the relative performance of models. Double sorts are priced well too, albeit to a lesser extent. The same models, however, achieve a very modest level of fit, when facing cross-sections based on AP-Trees (-2–20% on a set of 40 portfolios), despite the latter consisting of, on average, more diversified, larger cap portfolios. This is not surprising, given the battery of empirical results presented earlier. Importantly, the relative difference between model performances is also more pronounced, allowing us to identify the weaker models.

Second, conventional metrics, like the absolute pricing error or the number of significant alphas (often reported in empirical literature), are misleading, as they tend to ignore portfolio repackaging, and, hence, the risk of focusing on redundant and/or duplicated test assets. Indeed, 100 deciles and 225 double-sorted portfolios have many close-to-identical portfolios.

Table 4: Horse races for reduced-form asset pricing models

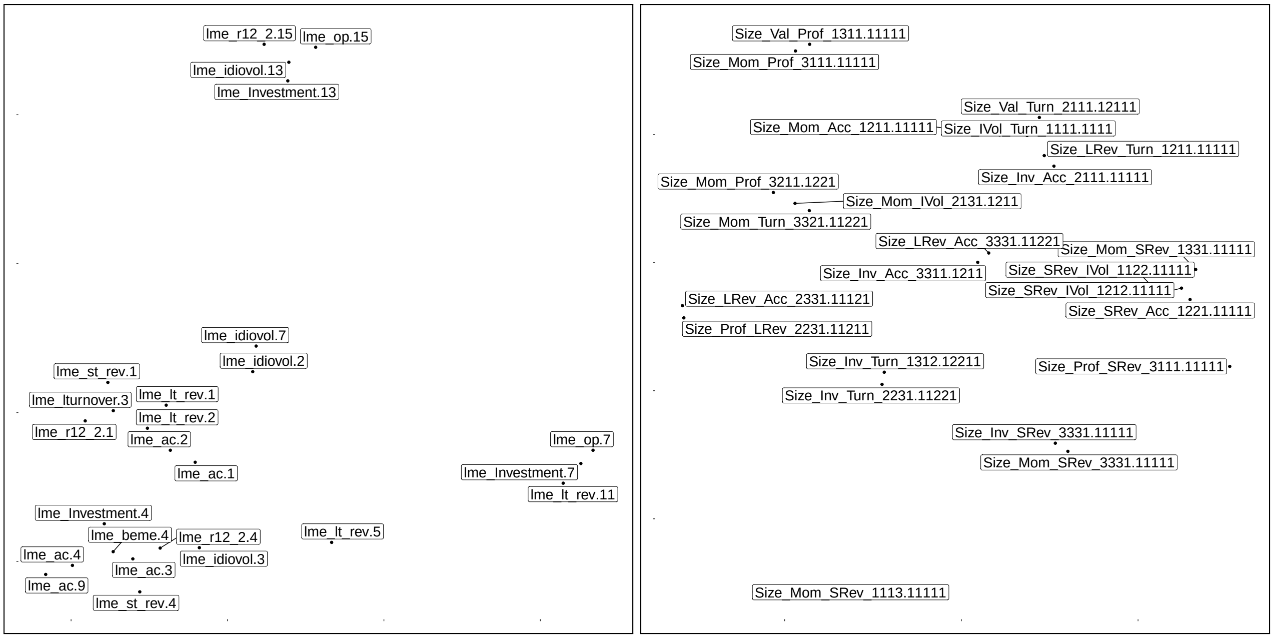
Statistics	Porfolios	Carhart	FF-3	FF-5	FF-6	Hirshleifer, Jiang (2010)	Q5	Stambaugh, Yuan (2017)
Avg Abs Alpha	Decile 100	0.0014	0.0018	0.0013	**0.0011**	0.0012	0.0012	0.0013
	DS 225	0.0025	0.0023	**0.0021**	0.0026	0.0023	0.0031	0.0032
	DS 40	0.0034	0.0039	0.0032	0.0031	**0.0029**	0.0037	0.0040
	DS 70	0.0032	0.0034	**0.0029**	0.0031	0.0028	0.0035	0.0039
	AP-Trees 40	0.0071	0.0081	0.0067	**0.0063**	0.0065	0.0064	0.0067
	AP-Trees 70	0.0058	0.0065	0.0053	**0.0052**	0.0053	0.0054	0.0057
# Signif. Alphas (%)	Decile 100	10 (10%)	18 (18%)	10 (10%)	8 (8%)	**5 (5%)**	6 (6%)	9 (9%)
	DS 225	77 (34%)	58 (26%)	**39 (17%)**	68 (30%)	47 (21%)	75 (33%)	74 (33%)
	DS 40	14 (35%)	16 (40%)	8 (20%)	12 (30%)	**7 (18%)**	11 (28%)	13 (33%)
	DS 70	28 (40%)	25 (36%)	**16 (23%)**	25 (36%)	17 (24%)	25 (36%)	27 (39%)
	AP-Trees 40	23 (58%)	29 (73%)	20 (50%)	18 (45%)	18 (45%)	**14 (35%)**	**14 (35%)**
	AP-Trees 70	37 (53%)	39 (56%)	27 (39%)	28 (40%)	28 (40%)	**24 (34%)**	**24 (34%)**
XS-R^2	Decile 100	0.88	0.78	0.89	**0.92**	0.91	0.91	0.89
	DS 225	0.82	0.83	**0.84**	0.77	**0.84**	0.70	0.67
	DS 40	0.65	0.54	0.66	0.56	**0.69**	0.44	0.37
	DS 70	0.72	0.69	0.74	0.65	**0.75**	0.57	0.52
	AP-Trees 40	0.17	0.12	**0.23**	0.03	0.20	-0.02	-0.12
	AP-Trees 70	0.36	0.33	**0.40**	0.24	0.38	0.18	0.09
SDF α (t-stat)	Decile 100	6.47	7.03	5.93	5.64	5.60	5.03	**4.95**
	DS 40	8.64	9.05	7.95	7.82	7.52	6.64	**6.62**
	DS 70	9.04	9.48	8.36	8.19	7.95	**6.81**	6.96
	AP-Trees 40	9.78	10.08	8.95	8.98	8.35	7.86	**7.33**
	AP-Trees 70	10.92	11.32	10.46	10.34	9.57	8.92	**8.50**
SR out-of-sample		0.20	0.16	0.21	0.23	0.32	**0.43**	0.30

The table shows a comparison between different linear factor models using a variety of empirical metrics. The models are: Fama-French-Carhart four-factors, Fama-French three- and five-factors, Fama-French five + momentum, Hirshleifer and Jiang (2010) four-factors, Q5 of Hou, Mo, Xue, and Zhang (2021), and a four-factor model of Stambaugh and Yuan (2017). The test assets are based on 10 characteristics and include 100 decile-sorted portfolios, 225 double-sorted portfolios (DS) based, robust mean-variance selection of the 225 DS portfolios to 40 and 70 basis assets, and AP-Trees with 40 and 70 portfolios. The performance metrics are the average absolute pricing errors of all the test assets with respect to the factor models, the number and fraction of pricing errors significant at 95%, the cross-sectional $XS\text{-}R^2$, and the t-statistic of the robust SDF alpha. Finally, we report the out-of-sample Sharpe ratios of the MVE portfolios constructed with the factors. All results are reported monthly out-of-sample, from January 1994 to December 2016. The best model according to a specific cross-section and metric is indicated in bold.

Hence, the average pricing errors or proportion of significant alphas can be fairly small if the dominating duplicated assets are explained well (Lewellen, Nagel, and Shanken (2010)). It should not come as a surprise, therefore, that Fama-French factors appear to perform well in this regard. Note that an appropriate metric in this context takes the form of an out-of-sample GRS-style test, which takes into account dependencies in the test assets. However, a conventional test is not applicable in our large-dimensional setup. Building on this intuition, our alphas of a feasible SDF are a spanning test for a large-dimensional cross-section relative to a reduced-form model. As a result, while all the models appear to be significantly misspecified, the mispricing factors of Stambaugh and Yuan (2017) explain these 10 characteristics the best.[25]

Figure 13 further illustrates the issue of data repackaging. We use a well-known machine-

[25]The factor models of Hirshleifer and Jiang (2010), Hou, Mo, Xue, and Zhang (2021), and Stambaugh and Yuan (2017) are based on information in addition to the 10 characteristics (in contrast to the versions of Fama-French factors), which explains the variation in Sharpe ratios. Hence, the relative performance could be different when we condition on other characteristics.



(a) Double Sorted (70) for Q5 factors

(b) AP-Tree (70) for Q5 factors

Figure 13: Similarity clusters of alpha-portfolios.

The figures show the t-Distributed Stochastic Neighbor Embedding (t-SNE) plots for the portfolios with significant pricing errors relative to the Q5-factor model of Hou, Mo, Xue, and Zhang (2021). We use the robust SDF recovery to select 70 double-sorted portfolios (from the set of 225) and 70 AP-Tree portfolios.

learning visualization tool for high-dimensional data, t-Distributed Stochastic Neighbor Embedding (t-SNE), to capture similarities of portfolios that have significant alphas.[26] Each portfolio is characterized by a 10-dimensional vector of value-weighted characteristic ranks of the stocks that it is composed of. The visualization with t-SNE displays portfolios with similar characteristic composition closer to each other. The 70 portfolios of both double-sorted and AP-Trees have a similar number of the assets with significant alphas (24–25 portfolios) relative to Q5 benchmark model. The alpha portfolios based on double sorts appear to form five clusters, corresponding to five groups of *economically similar* portfolios. On the other hand, AP-Trees' alpha portfolios seem to require more than twice as many clusters, which indicates that the revealed mispriced patterns of the SDF projection are more independent and less subject to the risk of simple repackaging. This illustrates that AP-Trees benefit substantially more from the sparse SDF selection because of their special overlapping basis function structure.

[26]The Internet Appendix provides further details on the t-SNE visualizations in Section IA.7.

VI. Conclusion

We propose a novel way to build characteristic-based cross-sections of returns that capture the complex information of a large number of cross-sectional stock return predictors. AP-Trees deliver a small number of long-only portfolios that (1) reflect information in many stock-specific characteristics allowing for conditional interactions and nonlinearities, (2) provide test assets for asset pricing models that are closer to spanning the SDF and are considerably harder to price than conventional cross-sections, and (3) act as the building blocks for the 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 among many characteristics and selects a sparse set of portfolios with the most relevant and non-redundant information. As a result, it reflects the importance of both the type of characteristics (and their interactions) and the granularity of the underlying splits. We show that conventional cross-sections (including hundreds of portfolios stacked against each other) 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 expected returns. Our approach is flexible and allows for many additional restrictions and modifications to the underlying procedure.

We show that popular double and triple sorts, as well as their typical combinations, are a poor reflection of the underlying stock returns. Yet, the majority of candidate tradable risk factors are constructed from suboptimal portfolios and have been tested on similar suboptimal cross-sections. Our findings, therefore, have important implications for both popular test assets and empirical reduced-form models in asset pricing.

REFERENCES

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61(1), 259–299.

Banz, R. (1981): "The Relationship between Return and Market Value of Common Stocks," *Journal of Financial Economics*, 9, 3–18.

Barillas, F., and J. Shanken (2016): "Which Alpha?," *The Review of Financial Studies*, 30(4), 1316–1338.

Basu, S. (1983): "The Relationship between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence," *Journal of Financial Economics*, 12(1), 129–156.

Bondt, W. F. M. D., and R. Thaler (1985): "Does the Stock Market Overreact?," *Journal of Finance*, 40(3), 793–805.

Chen, L., M. Pelger, and J. Zhu (2019): "Deep Learning in Asset Pricing," available at SSRN: https://ssrn.com/abstract=3350138.

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

Daniel, K., and S. Titman (2012): "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, (1), 103–139.

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.

Fabozzi, F. J., D. Huang, and G. Zhou (2010): "Robust Portfolios: Contributions from Operations Research and Finance," *Annals of Operations Research*, 176, 191–220.

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

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33(1), 3–56.

Fama, E. F., and K. R. French (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 (2020): "Taming the Factor Zoo: A Test of New Factors," *Journal of Finance*, forthcoming.

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting Characteristics Nonparametrically," *Review of Financial Studies*, forthcoming.

Garlappi, R., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *Review of Financial Studies*, 20(1), 41–81.

Giglio, S., and D. Xiu (2021): "Asset Pricing with Omitted Factors," *Journal of Political Economy*, 129(7), 1947–1990.

Gu, S., B. T. Kelly, and D. Xiu (2020a): "Autoencoder Asset Pricing Models," *Journal of Econometrics*, forthcoming.

——— (2020b): "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies*, forthcoming.

Hirshleifer, D., and D. Jiang (2010): "A Financing-Based Misvaluation Factor and the Cross Section of Expected Returns," *Review of Financial Studies*, 23(9), 3401–3436.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021): "An Augmented q-Factor Model with Expected Growth," *Review of Finance*, 25(1), 1–41.

Jegadeesh, N. (1990): "Evidence of Predictable Behavior of Security Returns," *Journal of Finance*, 45, 881–898.

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

Kelly, B. T., S. Pruitt, and Y. Su (2019): "Characteristics Are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 134(3), 501–524.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the Cross-Section," *Journal of Financial Economics*, 135(2), 271–292.

Lettau, M., and M. Pelger (2020): "Factors That Fit the Time-Series and Cross-Section of Stock Returns," *The Review of Financial Studies*, forthcoming.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A Skeptical Appraisal of Asset Pricing Tests," *Journal of Financial Economics*, 96(2), 175–194.

Martin, I., and S. Nagel (2019): "Market Efficiency in the Age of Big Data," available at SSRN: https://ssrn.com/abstract=3511296.

Moritz, B., and T. Zimmerman (2016): "Tree-Based Conditional Portfolio Sorts: The Relation Between Past and Future Stock Returns," available at SSRN: https://ssrn.com/abstract=2740751.

Nagel, S., and K. J. Singleton (2011): "Estimation and Evaluation of Conditional Asset Pricing Models," *Journal of Finance*, 66(3), 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.

Stambaugh, R. F., and Y. Yuan (2017): "Mispricing Factors," *Review of Financial Studies*, 30(4), 1270–1315.

Appendix A. List of the Firm-Specific Characteristics

Table A.1: Characteristic variables as listed in the Kenneth French Data Library

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 shareholders 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 per value (item PSTK) for PS. The market value of equity (PRC*SHROUT) is as of December t-1.	Basu (1983), Fama and French (1992)
IdioVol	Idiosyncratic volatility	Standard deviation of the residuals from a regression of excess returns on the Fama and French three-factor model	Ang, Hodrick, Xing, and Zhang (2006)
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	Fama and French (2015)
LME	Size	Total market capitalization at the end of the previous month defined as price times shares outstanding	Banz (1981), Fama and French (1992)
LT_Rev	Long-term reversal	Cumulative return from 60 months before the return prediction to 13 months before	Bondt and Thaler (1985)
Lturnover	Turnover	Previous 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.	Jegadeesh and Titman (1993)
ST_Rev	Short-term reversal	Prior month return	Jegadeesh (1990)

This table presents the characteristic variables as listed on Kenneth French's website with acronym, definition, and academic reference.

Appendix B. Additional Tables

Table B.1: Summary statistics of the cross-sections, built from AP-Trees

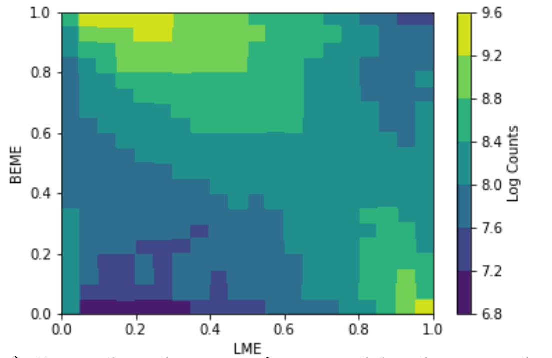
Id	Char 1	Char 2	Char 3	SR	α_{FF3}	α_{FF5}	α_{XSF}	α_{FF11}	λ_0	λ_2
30	Size	SRev	Turn	0.24	1.03	0.98	0.7	0.64	0.40	3.16E-06
					[7.36]	[6.83]	[6.65]	[5.78]		
17	Size	Prof	SRev	0.27	0.83	0.72	0.76	0.81	0.55	1.00E-08
					[6.8]	[5.81]	[6.31]	[6.53]		
27	Size	SRev	LRev	0.30	0.83	0.73	0.76	0.79	0.90	1.00E-08
					[10.2]	[8.97]	[9.55]	[9.26]		
4	Size	Val	SRev	0.30	0.98	1.03	0.92	1.04	0.90	1.00E-08
					[4.88]	[5.0]	[6.82]	[7.55]		
33	Size	LRev	Turn	0.31	0.62	0.51	0.61	0.66	0.35	1.00E-08
					[4.19]	[3.37]	[4.18]	[4.15]		
12	Size	Mom	LRev	0.32	0.84	0.71	0.83	0.88	0.80	1.00E-08
					[5.89]	[4.94]	[5.82]	[5.89]		
22	Size	Inv	SRev	0.33	0.95	0.91	0.82	0.89	0.80	1.00E-07
					[10.22]	[9.5]	[9.17]	[9.43]		
32	Size	LRev	IVol	0.33	0.79	0.74	0.79	0.83	0.05	1.00E-08
					[7.46]	[6.72]	[7.47]	[7.3]		
5	Size	Val	LRev	0.33	1.1	1.05	0.81	0.74	0.70	1.00E-08
					[8.61]	[7.9]	[7.91]	[6.97]		
28	Size	SRev	Acc	0.35	1.05	0.96	0.68	0.69	0.65	1.00E-08
					[7.75]	[6.97]	[6.56]	[6.22]		
15	Size	Mom	Turn	0.35	1.12	1.17	0.82	0.83	0.15	1.00E-08
					[7.38]	[7.42]	[8.8]	[8.53]		
31	Size	LRev	Acc	0.36	1.09	1.03	0.65	0.58	0.10	1.00E-08
					[5.97]	[5.42]	[5.39]	[4.64]		
35	Size	Acc	Turn	0.36	1.11	1.09	0.73	0.71	0.75	1.00E-08
					[6.84]	[6.56]	[5.98]	[5.4]		
13	Size	Mom	Acc	0.37	1.13	1.08	0.77	0.74	0.50	1.00E-08
					[10.59]	[9.73]	[9.75]	[8.92]		
11	Size	Mom	SRev	0.40	1.12	1.08	0.67	0.64	0.85	1.00E-07
					[6.95]	[6.47]	[6.79]	[5.96]		
29	Size	SRev	IVol	0.41	0.94	0.81	0.81	0.89	0.35	1.00E-08
					[10.11]	[8.76]	[8.77]	[9.12]		
1	Size	Val	Mom	0.41	1.03	1.06	0.83	1.06	0.00	1.00E-08
					[4.51]	[4.61]	[5.71]	[7.31]		
6	Size	Val	Acc	0.42	0.93	0.82	0.82	0.93	0.30	3.16E-07
					[7.74]	[6.93]	[7.11]	[7.55]		
9	Size	Mom	Prof	0.42	0.99	0.94	0.93	1.04	0.05	1.00E-08
					[8.35]	[7.76]	[7.89]	[8.43]		
10	Size	Mom	Inv	0.44	0.97	0.94	0.82	0.9	0.45	1.00E-08
					[11.22]	[10.75]	[9.93]	[10.67]		
8	Size	Val	Turn	0.46	1.01	0.97	0.95	1.03	0.15	1.00E-08
					[9.12]	[8.59]	[8.76]	[8.95]		
2	Size	Val	Prof	0.46	0.98	1.04	0.91	0.98	0.85	1.00E-08
					[5.43]	[5.54]	[]7.55	[8.03]		
36	Size	IVol	Turn	0.46	0.78	0.6	0.65	0.61	0.00	1.00E-08
					[6.87]	[5.59]	[6.13]	[5.45]		
14	Size	Mom	IVol	0.47	0.96	0.84	0.85	0.91	0.00	1.00E-08
					[10.17]	[9.1]	[9.28]	[9.37]		
23	Size	Inv	LRev	0.48	0.99	0.93	0.84	0.89	0.90	1.00E-08

					[11.33]	[10.41]	[10.09]	[10.22]		
18	Size	Prof	LRev	0.50	0.93	0.8	0.79	0.83	0.20	1.00E-08
					[9.53]	[9.11]	[9.56]	[9.39]		
19	Size	Prof	Acc	0.52	0.91	0.99	0.85	0.94	0.35	1.00E-08
					[4.63]	[4.89]	[6.83]	[7.09]		
34	Size	Acc	IVol	0.54	1.06	1.12	0.93	1.07	0.05	3.16E-06
					[6.11]	[6.29]	[7.42]	[8.23]		
7	Size	Val	IVol	0.54	1.05	1.1	0.92	1.04	0.00	3.16E-07
					[6.86]	[7.07]	[8.86]	[10.07]		
21	Size	Prof	Turn	0.55	1.02	1.14	0.89	1.01	0.00	1.00E-08
					[4.18]	[4.58]	[6.22]	[6.62]		
20	Size	Prof	IVol	0.59	1.01	0.93	0.95	1.02	0.05	3.16E-07
					[6.54]	[5.8]	[6.11]	[6.1]		
26	Size	Inv	Turn	0.62	1.03	0.99	0.79	0.98	0.00	1.00E-08
					[6.6]	[6.14]	[5.38]	[6.51]		
24	Size	Inv	Acc	0.62	0.81	0.69	0.75	0.7	0.20	1.00E-08
					[4.85]	[4.14]	[4.84]	[4.15]		
25	Size	Inv	IVol	0.64	1.01	0.98	0.86	0.92	0.05	1.00E-08
					[9.46]	[8.88]	[8.37]	[8.62]		
3	Size	Val	Inv	0.65	1.04	1.02	0.99	1.26	0.50	1.00E-08
					[6.78]	[6.37]	[6.4]	[7.98]		
16	Size	Prof	Inv	0.65	1.06	1.02	0.82	0.85	0.15	1.00E-08
					[10.59]	[9.94]	[9.95]	[9.94]		

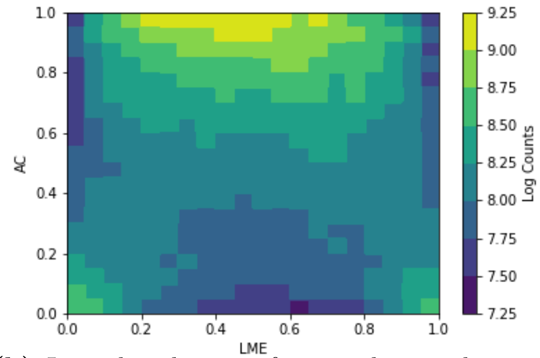
The table summarizes 36 cross-sections, based on AP-Tree portfolios. For each combination of three characteristics we built the trees of depth four, and excluded the nodes that result from using the same characteristics for splits four times. The set of final and intermediate nodes is then pruned to 10 portfolios, with the corresponding SDF spanning them, following the approach outlined in Sections II and I.3. The table presents the values of the optimally selected tuning parameters (λ_0 for shrinkage to the mean and λ_2 for ridge, while the lasso term is set to select 10 assets). We also report the out-of-sample properties of the SDFs spanning these portfolios: their Sharpe ratio and alphas with respect to Fama-French three- and five-factor models, cross-section-specific factors (and market), and the ultimate FF11 model, which includes the market and 10 long-short portfolios, based on characteristics. The t-statistics are in brackets.

Appendix C. Additional Figures

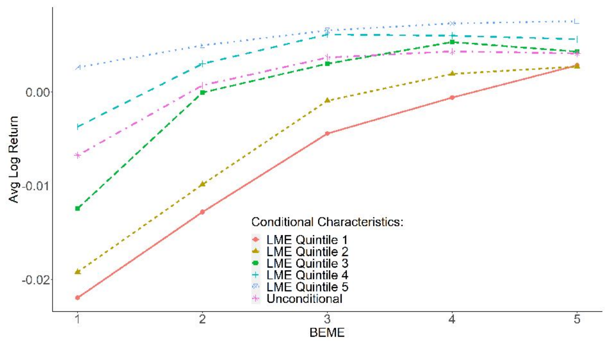
Appendix C.1. Stylized Empirical Facts



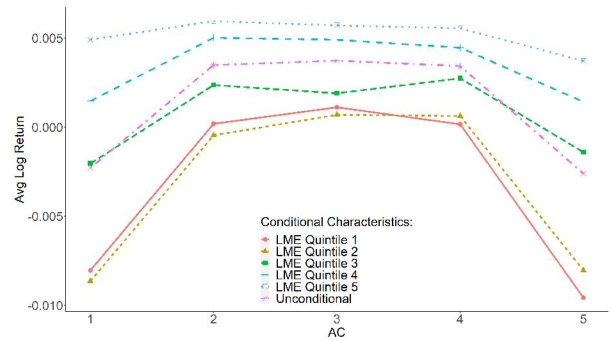
(a) Joint distribution of size and book-to-market



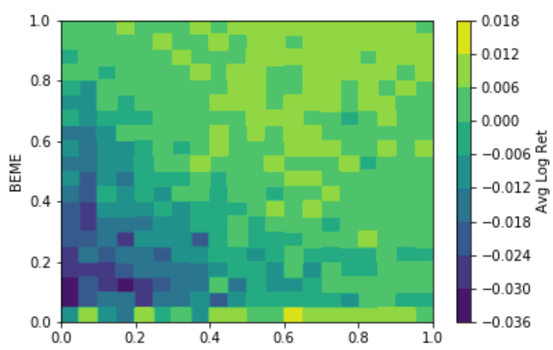
(b) Joint distribution of size and accruals



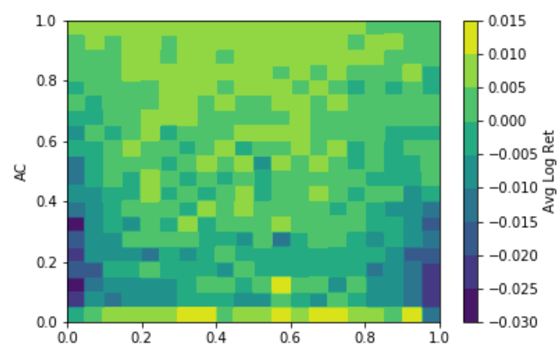
(c) Conditional mean return of size and book-to-market



(d) Conditional mean return of size and accruals



(e) Conditional mean return of size and book-to-market

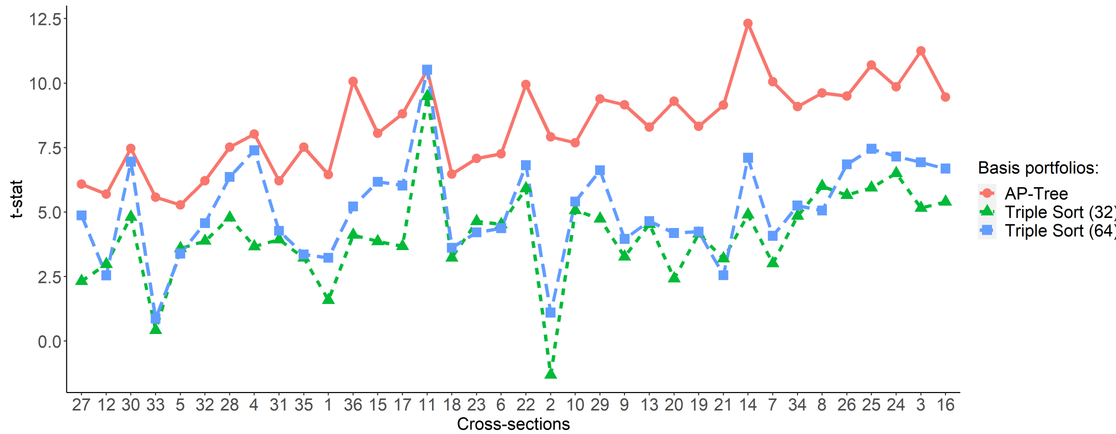


(f) Conditional mean return of size and accruals

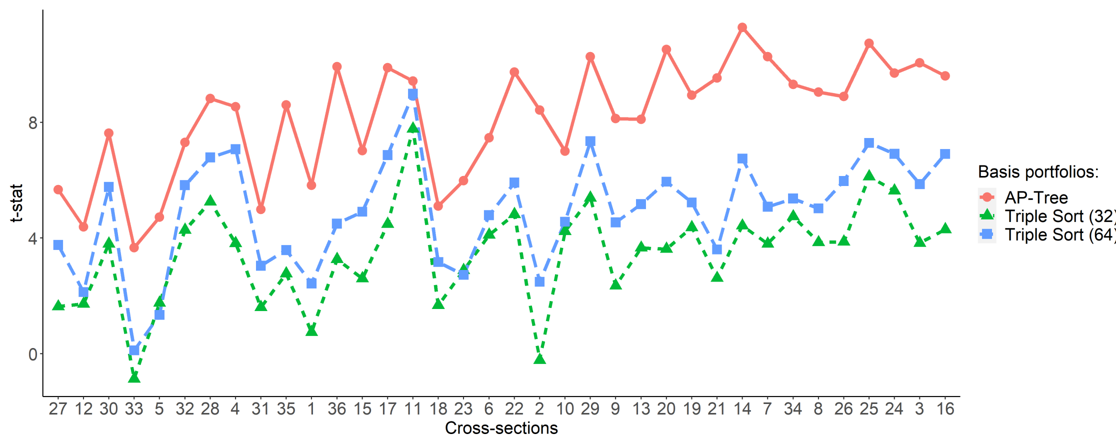
Figure C.1: Joint distribution of characteristics and conditional expected returns

These figures show the joint distribution of characteristics in the cross-section of stocks, their conditional and unconditional impact, and the expected returns for the combination of size and book-to-market and size and accrual. Subfigures (a) and (b) represent the pairwise empirical cross-sectional distribution of the characteristic quantiles across the stocks. The frequency is computed on a quantile 20×20 grid. Subfigures (c) and (d) represent the impact of a characteristic quintile on expected returns unconditionally and conditionally on the second characteristic. For example, Subfigure (c) describes the average returns on stocks sorted by their book-to-market ratio (quintiles 1–5), both unconditionally and conditionally on belonging to quintiles 1–5 based on size. Subfigures (e) and (f) represent the 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 × 20 unconditional quantiles by averaging historical returns on the stocks belonging to that portfolio. Section IA.5 in the Internet Appendix includes the results for other characteristics.

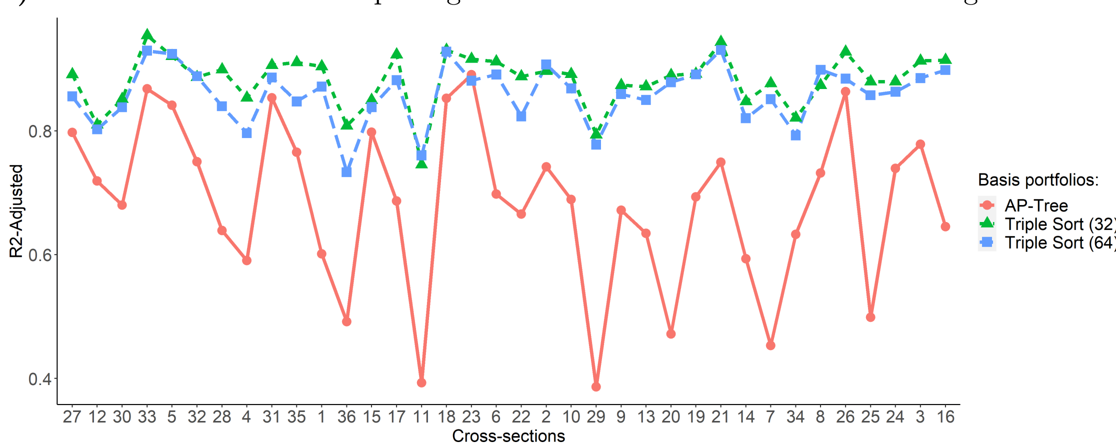
Appendix C.2. Additional Empirical Results for All Cross-Sections



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



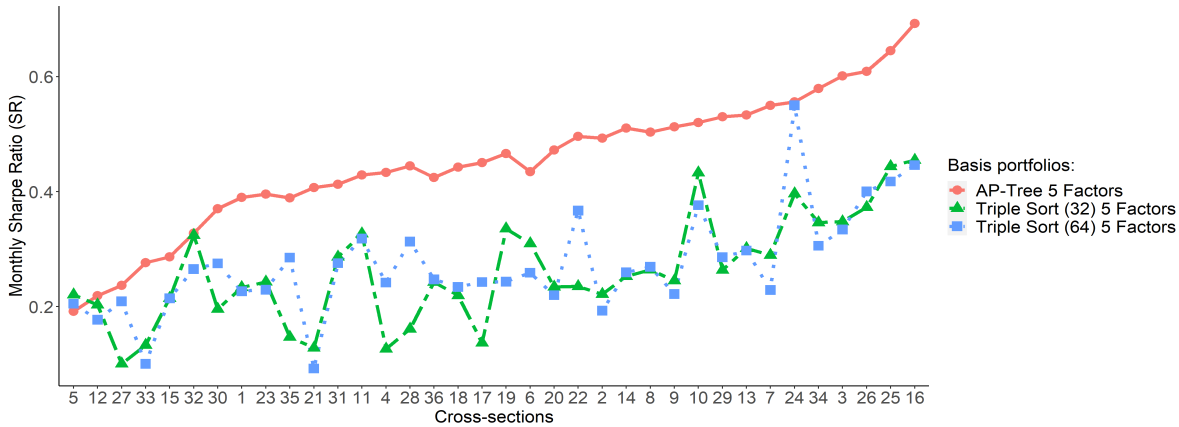
(b) SDF α: t-statistics of the pricing error relative to the market and 10 long-short factors



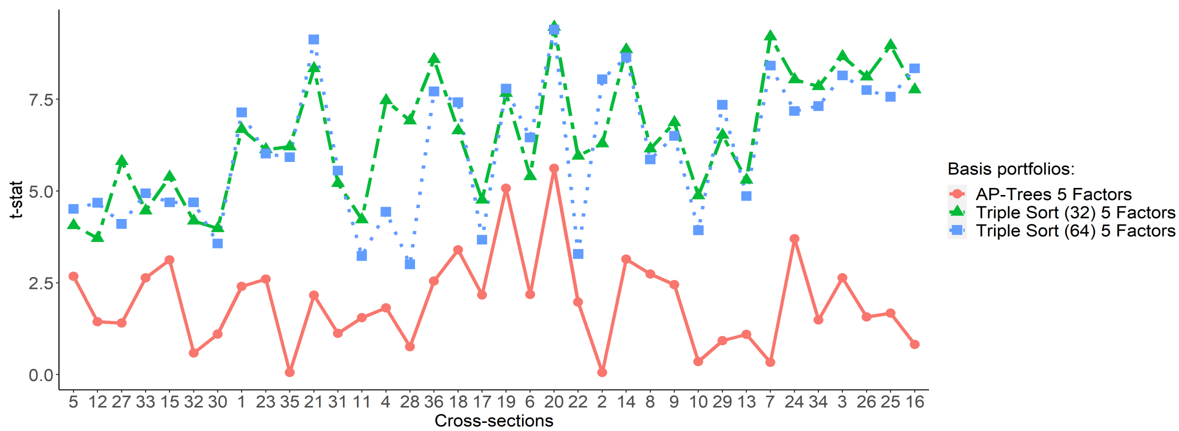
(c) XS-R^2: pricing portfolios with XS-specific factors

Figure C.2: Pricing errors and XS-R^2 for AP-Trees and triple sorts

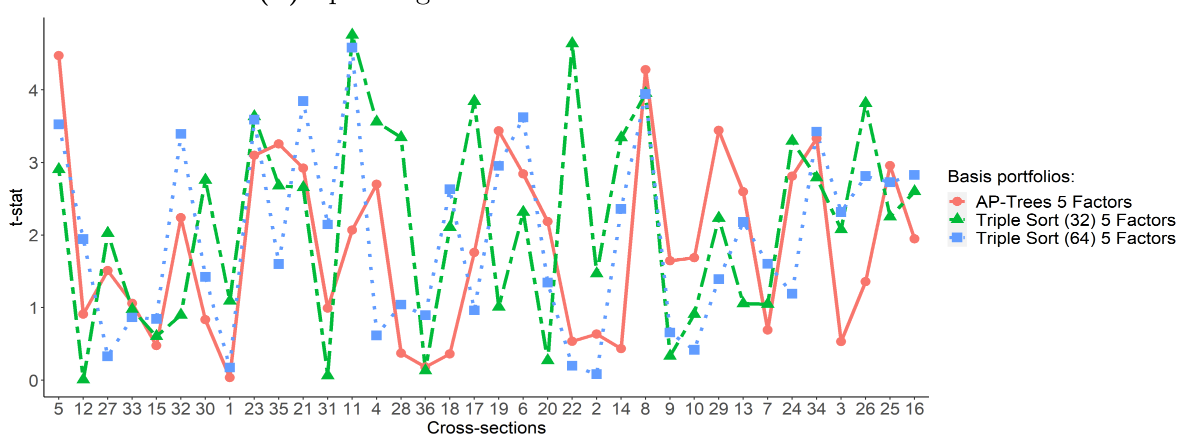
Pricing errors of the SDFs spanned by AP-Trees (40 portfolios) and triple sorts and cross-sectional fit measured by XS-R^2. The t-statistics of the pricing errors are relative to the cross-section-specific factors (market + 3 long-short factors) and all 11 factors (market + 10 long-short factors). All the cross-sections are sorted by the SR achieved with AP-Trees pruned to 40 portfolios.



(a): Sharpe ratios of the SDFs spanned by 5 RP-PCA of AP-Trees and triple sorts



(b) Spanning AP-Tree SDFs with 5 RP-PCA factors



(c) Spanning triple-sort SDFs with 5 RP-PCA factors

Figure C.3: Cross-sectional spanning tests via RP-PCA

Panel (a) displays monthly out-of-sample Sharpe ratios of the MVE portfolios spanned by 5 RP-PCA factors extracted from 40 AP-Tree portfolios and 32/64 triple-sorts. Following Lettau and Pelger (2020), the risk-premium weight is set to 10. All the cross-sections are sorted by the SR, achieved with RP-PCA factors based on AP-Trees. Panel (b) depicts t-statistics of the out-of-sample AP-Tree-SDF alphas with respect to 5 RP-PCA factors obtained from different cross-sections. Panel (c) shows spanning tests of the SDF based on 32 triple sorts.

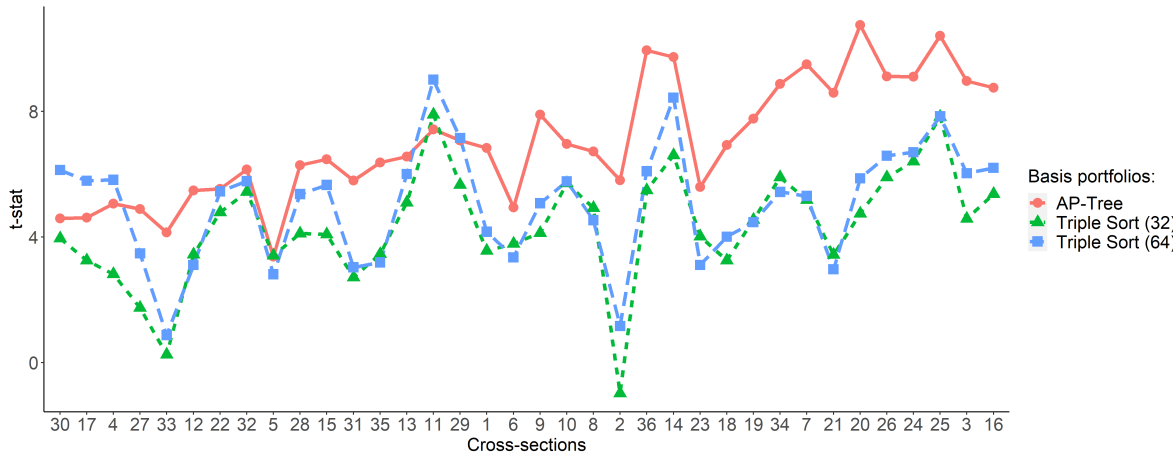


Figure C.4: SDF α: t-statistics of the pricing errors of SDF spanned by AP-Trees (10 portfolios) with respect to 5 Fama-French factors

The figure displays the out-of-sample t-statistics of the pricing errors of the robust mean variance efficient portfolios spanned by pruned AP-Trees (10 portfolios), triple sorts, and XSF (market and the three cross-section-specific characteristics) with respect to the five Fama-French factors. The SDF based on triple sorts is relying on either 32 or 64 assets and considers mean and variance shrinkage. Cross-sections are sorted by the SR achieved with AP-Trees (10).

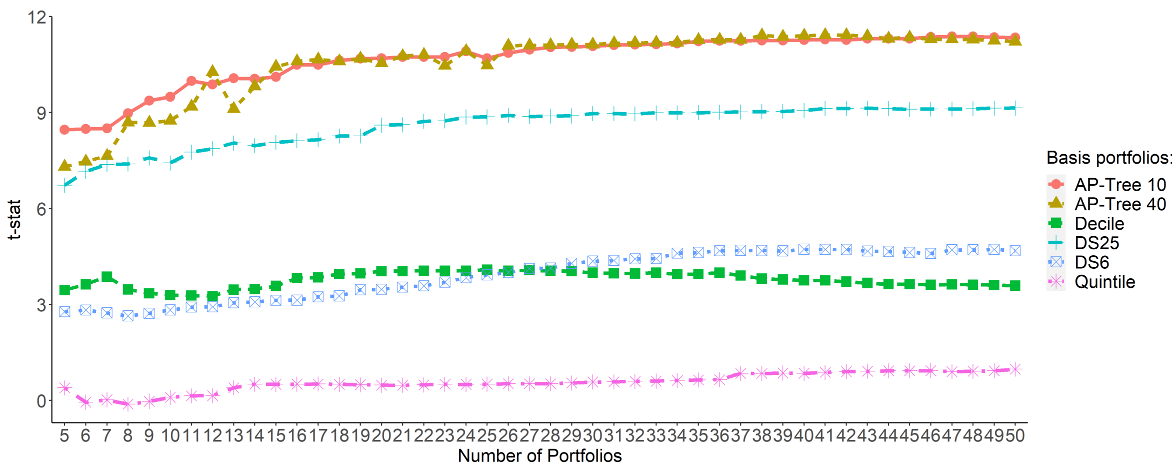
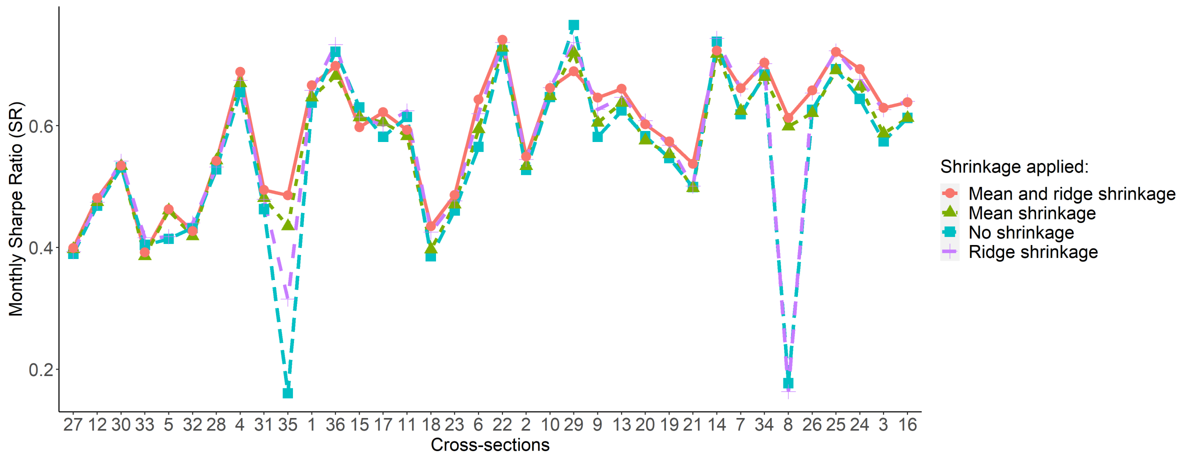
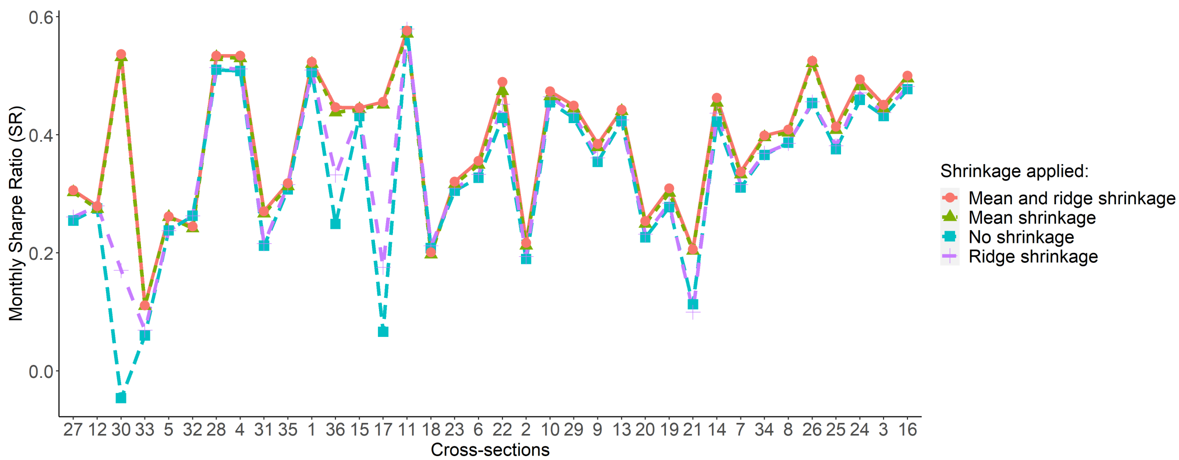


Figure C.5: Out-of-sample SDF alpha t-statistics, based on the FF11-factor model (market + 10 long-short factors) for 10 characteristics

SDF alpha t-statistics calculated on the out-of-sample robust mean-variance efficient portfolios spanned by different basis assets. The SDFs are constructed as in Figure 11. The pricing errors are calculated with respect to all 11 factors.



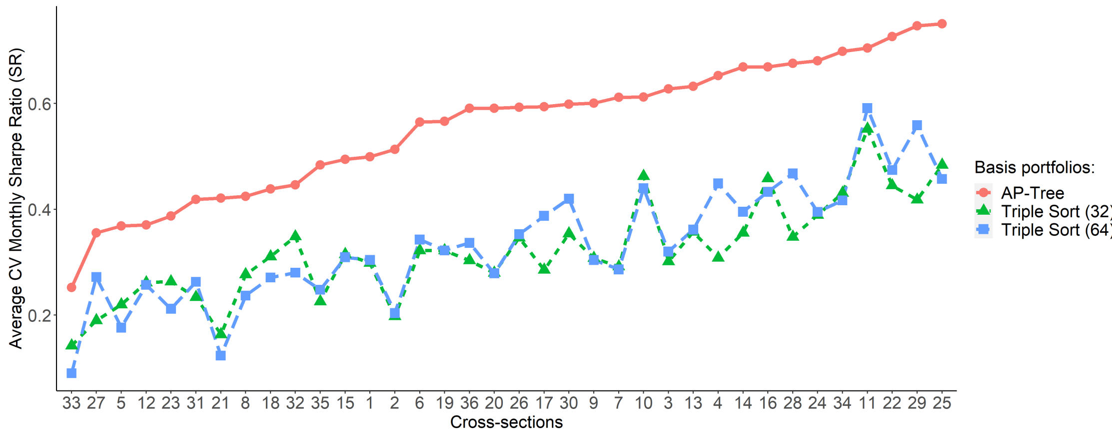
(a) Sharpe ratios of AP-Trees (40) with time-varying SDF weights



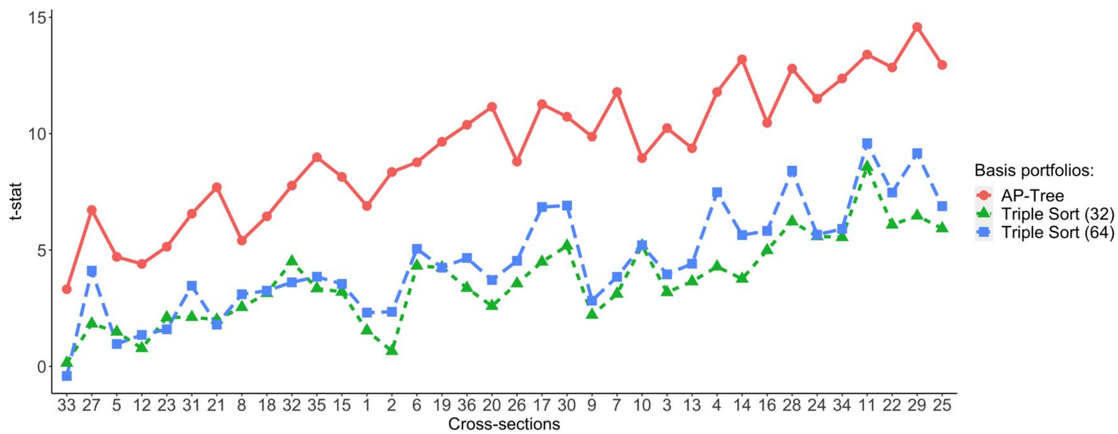
(b) Sharpe ratios of triple sorts (64 portfolios) with time-varying weights SDF weights

Figure C.6: Sharpe ratios of AP-Tree (40) and TS (64) SDFs with rolling window estimates of portfolio weights

The top figure displays the monthly out-of-sample Sharpe ratio of the SDFs of pruned AP-Trees (40 portfolios) with and without different forms of shrinkage, while the bottom figures show the results for 64 triple-sorted portfolios. The AP-Tree portfolios are selected on the training data and kept the same. The SDF portfolio weights for both figures are estimated on a rolling window of 20 years; that is, we allow for time-variation in the SDF portfolio weights. The tuning parameters (mean and/or variance shrinkage) are selected optimally on the rolling validation data. Cross-sections are sorted by the SR achieved with AP-Trees (40). Section IA.3.1 in the Internet Appendix includes additional results.



(a) Sharpe ratios of SDFs of AP-Trees (40 portfolios) and triple sorts under cross-validation

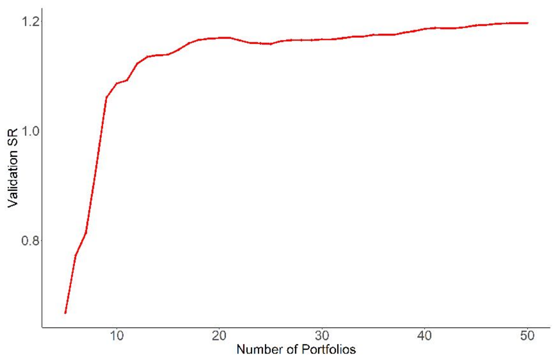


(b) t-stat of robust SDF alpha (w.r.t. to 11 factors) under cross-validation

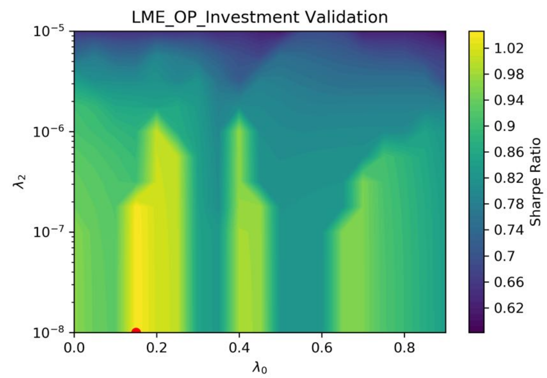
Figure C.7: Sharpe ratios and pricing errors of SDFs spanned by pruned AP-Trees (40 portfolios) and triple sorts under cross-validation

The figure displays the monthly out-of-sample Sharpe ratio and pricing errors of the mean-variance efficient portfolios spanned by AP-Trees (40 portfolios) and triple sorts under three-fold cross-validation. We split the time-series into three equal blocks and use each block as a training period and evaluate the model on the remaining two blocks. The cross-validation results report the average over these three estimates, where the tuning parameters are selected to optimize the cross-validation results. Panel A shows the Sharpe ratios and panel B the pricing errors relative to the 11-factor model. All the cross-sections are sorted by the SR achieved with AP-Trees.

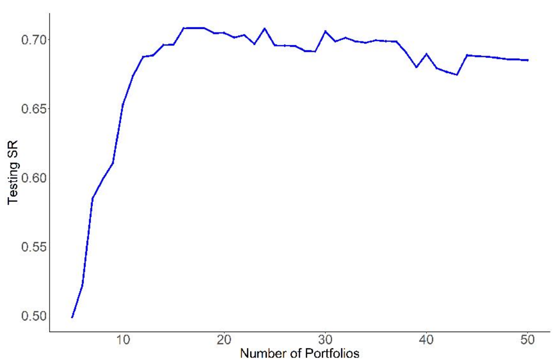
Appendix C.3. Size, Profitability, and Investment



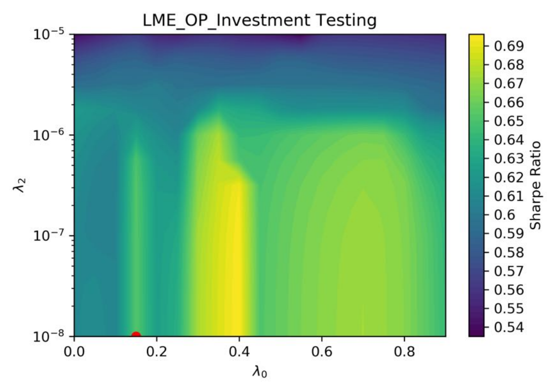
(a) Validation: Portfolio number impact



(b) Validation: Mean and variance shrinkage



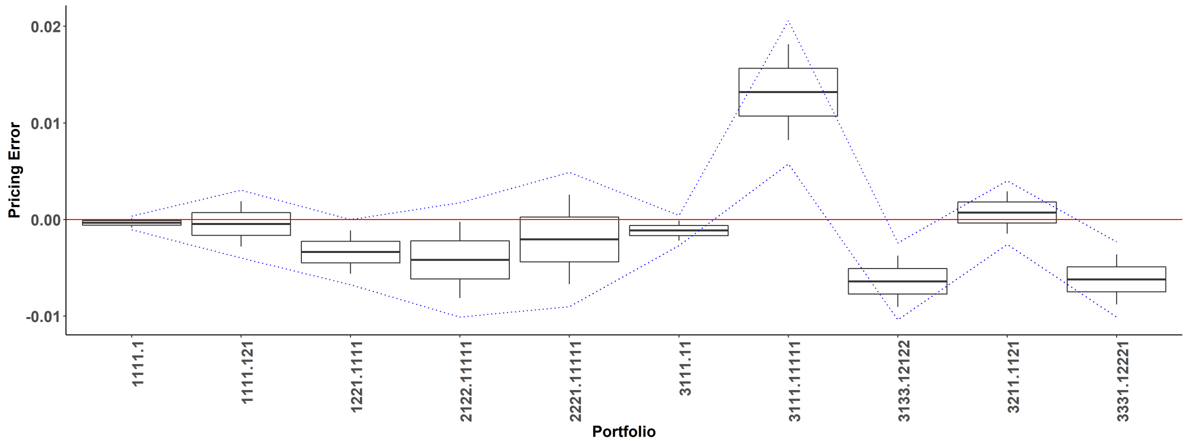
(c) Testing: Portfolio number impact



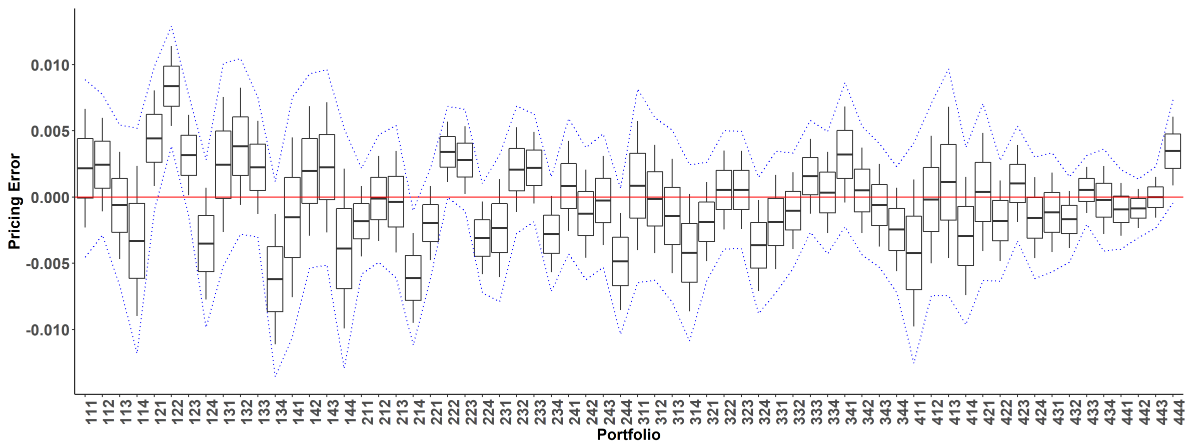
(d) Testing: Mean and variance shrinkage

Figure C.8: SR as a function of tuning parameter during validation and testing samples

The figure shows Sharpe ratio as a function of tuning parameters for AP-Trees built on size, investment, and operating profitability. Subplots (a) and (c) present the SR as function of the number of portfolios using the optimal mean and variance shrinkage. Subplots (b) and (d) depict the SR as a function of mean (λ_0) and variance (λ_2) shrinkage for 10 portfolios. Yellow regions correspond to higher values. Optimal validation-based tuning parameters are indicated by red dots.



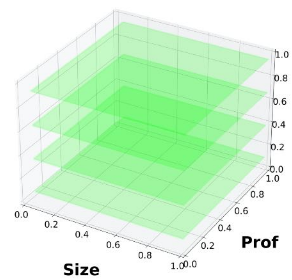
(a) Alphas of AP-Trees (10) portfolios, controlling for cross-section specific factors (*XSF*)



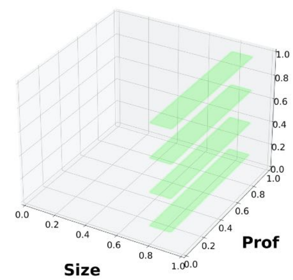
(b) Alphas of 64 triple-sorted portfolios, controlling for cross-section specific factors (*XSF*)

Figure C.9: Pricing errors for individual AP-Trees and triple-sorted portfolios

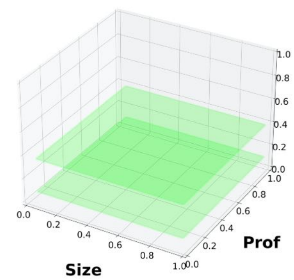
This figure shows the pricing errors and their significance level of 10 AP-Trees portfolios and 64 triple-sorted basis assets based on size, profitability, and investment. Subplot (a) presents the pricing errors for AP-Trees and subplot (b) for the triple sorts relative to the cross-section-specific factors (XSF) with significance levels (candlestick=5%, dashed line=1%, box=10%). The portfolio ID for triple sorts denotes the quantile for (LME/OP/INV).



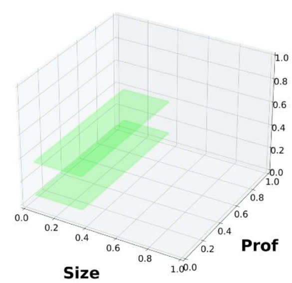
(a) Portfolio 1: The Market (id 1111.1)



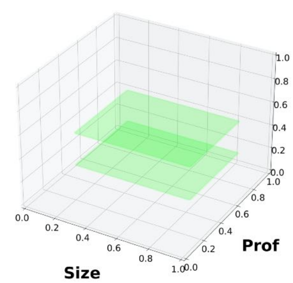
(b) Portfolio 2: Medium Size (50–75%) (id 1111.121)



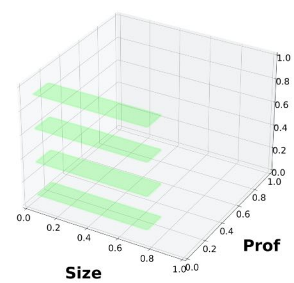
(c) Portfolio 3: Bottom 50% on Operating Profitability (id 1111.1)



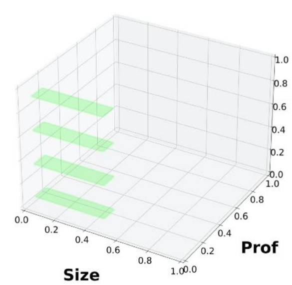
(d) Portfolio 4: Bottom 50% on Investment, bottom 25% on Size (id 3111.11111)



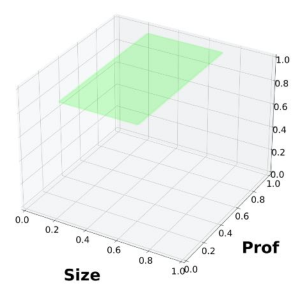
(e) Portfolio 5: Bottom 50% on Investment, top 50% on Operating Profitability, bottom 50% on Size (id 3211.1121)



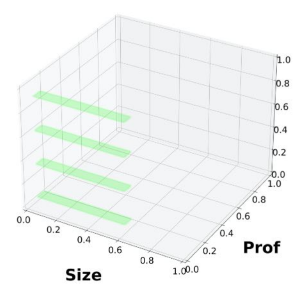
(f) Portfolio 6: Bottom 50% on Size, bottom 25% on Operating Profitability (id 1221.1111)



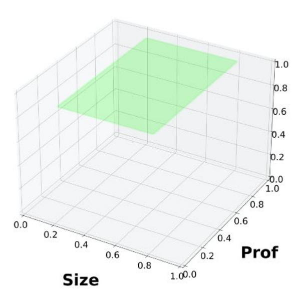
(g) Portfolio 7: Bottom 25% on Operating Profitability, bottom 50% on Size (id 2221.11111)



(h) Portfolio 8: Top 12.5% on Investment, bottom 50% on Size(id 3331.12221)



(i) Portfolio 9: Bottom 50% on Operating Profitability, bottom 50% on Size, bottom 25% on Operating Profitability (id 2122.11111)



(j) Portfolio 10: Top 50% on Investment, bottom 50% on Size, Top 25% on Investment (id 3133.12122)

Figure C.10: AP-Tree (10) portfolios for Size, Investment, and Operating Profitability

The graphs highlight the composition of the 10 AP-Tree portfolios in the characteristic space. The tree node id reflects the order, direction, and number of splits (first set of numbers is the characteristic order for the nodes, and the second set the direction and number of splits).