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

Svetlana Bryzgalova[†] Markus Pelger[‡] Jason Zhu[§]

First version: August 2019
This version: May 2020

ABSTRACT

We show how to build a cross-section of asset returns, that is, a small set of basis assets that capture complex information contained in a given set of stock characteristics. We use decision trees to generalize the concept of conventional sorting and introduce a new approach to the robust recovery of a low-dimensional set of portfolios that span the stochastic discount factor (SDF). Constructed from the same pricing signals as conventional double- or triple-sorted portfolios, our cross-sections have on average 30% higher Sharpe ratios and pricing errors relative to the leading reduced-form asset pricing models. They include long-only investment strategies that are well diversified, easily interpretable, and that could be built to reflect many characteristics at the same time. Empirically, we show that traditionally used cross-sections of portfolios and their combinations often present too low a hurdle for candidate asset pricing models, as they miss a lot of the underlying information from the original returns.

Keywords: Asset pricing, sorting, portfolios, cross-section of expected returns, decision trees, elastic net, stock characteristics, machine learning.
JEL classification: G11, G12, C55, C58.

---

*We thank Caio Almeida (discussant), Victor DeMiguel, Kay Giesecke, Serhiy Kozak, Andreas Neuhierl (discussant), Serge Nyawa (discussant), Juan Passadore, Andrew Patton, Ilaria Piatti (discussant), Ken Singleton, Raman Uppal, Michael Weber, Dacheng Xiu (discussant), Paolo Zaffaroni, Goufu Zhou, and seminar and conference participants at Stanford, Duke, London Business School, Imperial College London, Chicago Asset Pricing Conference, Fourth International Workshop on Financial Econometrics, Toulouse Financial Econometrics Conference, SIAM Financial Mathematics, Adam Smith Junior Workshop, Yale Junior Finance Conference, London CFE, 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

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

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

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

Inspired by the empirical success of machine learning techniques, we develop a new approach to *pruning* a large set of potential portfolios and end up with a cross-section that is sparse, interpretable, well-diversified, and informative about the impact of the underlying characteristics. This means we obtain a small number of tree portfolios that can span the SDF projected on these characteristics. Our paper generalizes the robust approach of Kozak, Nagel, and Santosh (2020) to the SDF recovery and relies on dual shrinkage in the variance *and* mean to find an optimal solution.

In a large scale empirical application, we build a set of 36 cross-sections for different combinations of firm-specific characteristics, and compare their performance and informational content to that achieved by conventional triple-sorted portfolios in each of the cases. While AP-Trees could be used to reflect a larger set of characteristics, our main results are focused on the combinations of three variables at a time, since this is empirically a limit for conventional sorting-based procedures and provides a natural benchmark against which to judge our cross-sections.

We find that for every single cross-section, relative to triple sorts, the test assets constructed with AP-Trees have dramatically higher Sharpe ratios, sometimes up to a factor of three. Importantly, this does not come from higher loadings on conventional sources of risk: The difference in these returns is not spanned by leading asset pricing models, and remains significant – both statistically and economically, even when the cross-section is pitted against 11 candidate risk factors. The resulting portfolios are well-diversified and do not load on the extreme deciles of the cross-sectional sorts. In fact, their construction with decision trees enforces a stable and balanced composition by default, making them fully comparable and often more diversified than triple sorts of the same *depth*. Finally, we create small-dimensional (e.g., $10 - 50$ assets) cross-sections of portfolios that reflect the information contained in *all 10 characteristics*. Compared to an optimal use of decile-sorted portfolios (100 assets) it doubles both the Sharpe ratio achieved out-of-sample and the alphas associated with the SDF spanned by these portfolios. Relative to the best combinations of the size including double-sorted portfolios,[1] it offers roughly a 30% increase in both Sharpe ratios and the SDF alpha. Most importantly, our method provides a general flexible way of building the most informative and still interpretable cross-sections subject to various constraints which can include the number of desired portfolios, the degree of interactions among characteristics, and restrictions on the minimum number of shares or their total market cap, among others.

All our results are obtained out-of-sample. Martin and Nagel (2019) highlight a potentially stark difference between the in-sample predictability reflected by traditional statistical tests and its feasible, out-of-sample counterpart. In spirit with their work, we focus on the out-of-sample Sharpe ratio and the alphas of the SDF, not spanned by the traditional risk factors. We show that the main driving forces behind the superior performance of AP-Trees is a) their ability to efficiently capture the interactions among characteristics and b) creating a set of optimally pruned overlapping *basis functions* to better reflect the patterns in expected returns. We find that triple sorts, while largely reflecting some of the true patterns in the data, present a very coarse structure of the cross-section, and they often fail to reflect the underlying stocks.

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

The first key observation lies in using conditional sorts to span the universe of managed portfolios and the diversity of “building blocks” that it naturally provides. This is particularly true for characteristics that are cross-sectionally correlated, as changing the order of the variables used for splits will lead to potentially different sets of basis assets. Conditional portfolio sorts, based on a selected set of characteristics, are intuitive to build and interpret within the setting of decision trees. First, we categorize firms into two equally sized groups according to the high/low value of a

[1] This is the combination of 25 size-value portfolios + 25 size-momentum portfolios + etc.

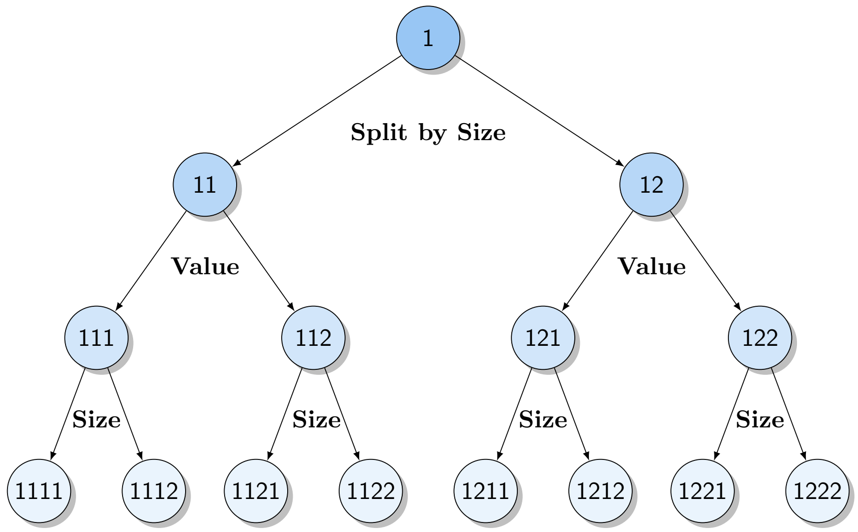


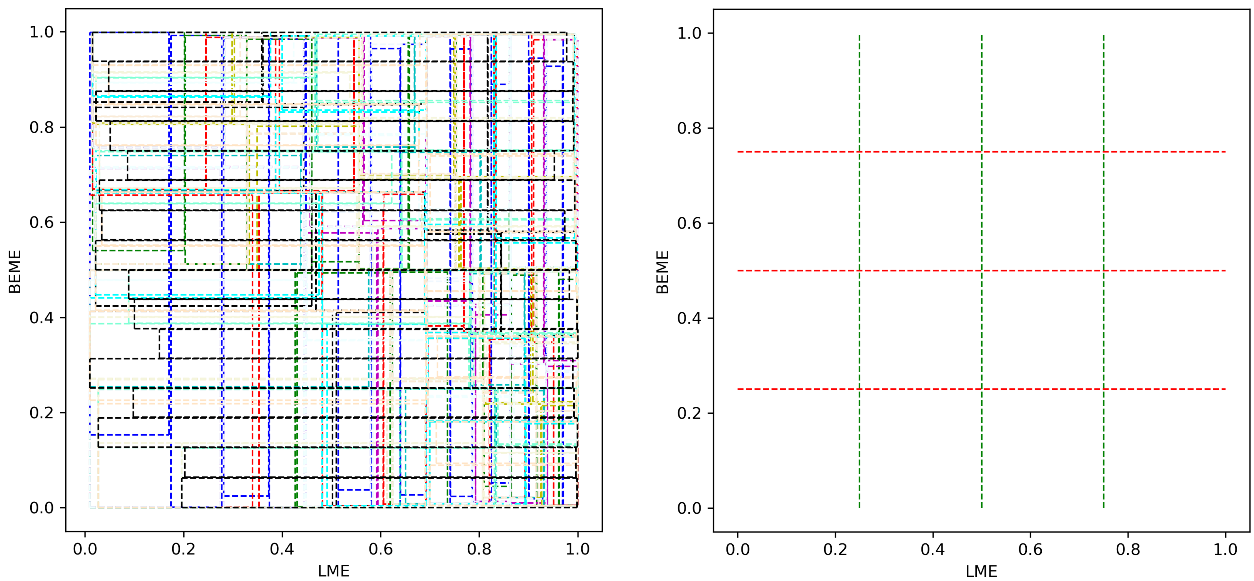
Figure 1. Example of a conditional tree based on size and value

The figure presents an example of an AP-Tree of depth 3 based on size and book-to market. 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.

characteristic variable, for example, size. Then within each group, stocks are further split into two smaller groups of the same dimension based on the value of some other variable, for example, value, creating four groups in total. This step can be repeated until the desired depth of the tree d is reached, obtaining 2^d groups of stocks in total, each of the size $\frac{N}{2^d}$, where N is the total number of stocks. The stocks in each group can then be combined into an equally or value-weighted portfolio. Figure 1 gives an example of a conditional tree of depth 3, where stocks are first sorted by size, then value, and then size again.

Figure 2 illustrates the difference between the kind of portfolios one could get with conditional trees and double sorting based on the information contained in probably the most well-known pair of characteristics: size and value. We build simple conditional trees of the same depth as the double-sorted portfolios (i.e., each node constructed with AP-Trees contains at least 1/16 of all the stocks at any given point of time). The outcome is staggering: The type of patterns and strategies one could span with conditional splits, using the same level of granularity as double sorts, is unquestionably richer. Furthermore, it directly incorporates the joint cross-sectional distribution of characteristics: Since AP-Trees rely on sequential splits, the cutoffs are internally chosen to give the balanced, equally populated set of assets and could never end up with empty portfolios. Building the whole universe of potential basis assets is important, but so is reducing their dimension to eliminate redundant strategies and identify the informative ones.

We introduce a novel way to extract a small set of basis assets capturing the key pricing information from a large set of portfolios by carefully selecting the basis functions that span the SDF. Finding the weights of these securities is equivalent to finding the tangency portfolio in the mean-variance space, that is the portfolio with the highest Sharpe ratio. We find a sparse representation of the SDF weights by including shrinkage in the mean-variance optimization problem. As our



(a) Conditional trees

(b) Double-sorting

Figure 2. Cross-sectional quantiles of the portfolio splits based on conditional trees and double sorting with size and value.

The figure shows the unconditional quantiles in the space of characteristics (size and book-to-market) that define the portfolio construction in AP-Trees (left) and the double-sorted cross-section (right). AP-Trees are constructed up to depth 4, that is allowing at most four consecutive splits, based on (repeating two) characteristics, i.e., all the portfolios have at least 1/16 of all the stocks. The figure reflects the quantiles of both final and intermediate nodes. The number of double-sorted portfolios (16) is chosen to match the depth of the trees based on the granularity of the splits.

optimal mean-variance portfolio with shrinkage could potentially include any final or intermediate nodes of the trees, we prune a large set of assets both in the choice of splits and their depth. Intuitively, applying our estimator to tree portfolios corresponds to choosing an optimal bandwidth in a nonparametric regression, with the decision to use a finer set of nodes in spanning the SDF being equivalent to choosing a smaller bandwidth in the characteristic space. Instead of the conventional bias-variance trade-off for parameter estimation, however, we set up a global objective motivated by the fundamental nature of the problem.

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

The rest of the paper is organized as follows. Section I highlights the role of sorting as a simple nonparametric conditional estimator of expected returns and introduces conditional trees as the collection of alternative basis assets. In Section II we describe the methodology behind Asset Pricing

Pruning based on the shrunk mean-variance optimization. We illustrate the basic properties of our approach with a simple simulation setup in Section III, and we provide empirical results in Section V. Finally, Section VI concludes. Additional empirical results, as well as alternative simulation setups, are delegated to the Appendix. The extensive Internet Appendix shows that all the main findings generalize to other cross-sections.

Related Literature

Our paper contributes to the growing literature in asset pricing that tackles the "multidimensional challenge," as formulated by Cochrane (2011) in his AFA Presidential Address. On a fundamental level this literature extracts basis assets with statistical and economic models that span the SDF. The asset pricing equation has a left-hand side with the test assets that need to be explained and a right-hand side with the asset pricing model spanned by basis assets. In the correct formulation, both sides of the equation should include all basis assets that can span the SDF or at least the projection of the SDF on the information set under consideration. So far this literature has almost exclusively focused on the right-hand side of the asset pricing equation with factor models as the most important class of asset pricing models. However, the test assets on the left-hand side of the asset pricing equation impose a higher hurdle, as it is not sufficient to only span the SDF, but the test assets should ideally be interpretable and long-only assets. Hence, even if a set of statistical factors, for example based on PCA, captures the asset pricing information in certain characteristics, it might not provide a good set of test assets, as it lacks the interpretability of our AP-Trees. In this sense, our approach is *complementary* to the effort of estimating asset pricing factors, as it pursues a different objective.

Our paper contributes to the literature of extracting the SDF. Most of these models estimate factors by applying a version of PCA to characteristic-sorted or characteristic-projected data, which does not offer the same interpretability as AP-Trees.[2] Our shrunk mean-variance optimization generalizes the approach suggested by Kozak, Nagel, and Santosh (2020), who estimate a small number of SDF basis assets by solving an efficient frontier optimization problem with a lasso penalty that selects their subset and a ridge component that shrinks the contribution of lower order principal component assets. Our approach generalizes their estimator by applying an additional shrinkage in the mean, the degree of which we choose optimally. As a result, we can show empirically that the special case of Kozak, Nagel, and Santosh (2020) can be improved.[3] The second differentiating

[2] 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 the SDF as a function of its characteristics, and as a result apply PCA to managed portfolios that represent linear (and nonlinear) projections of asset returns on characteristics.

[3] 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). We build on their insights and demonstrate that the methods are tightly linked to each other within a global portfolio selection problem and can efficiently enhance each other's impact when used jointly.

factor is the use of trees instead of simple sorted portfolios, linear long-short factors, or their principal components. Shrunk tree basis functions offer an interpretable alternative to PCA. For example, consider building a cross-section based on size and value using PCA on double-sorted portfolios. The resulting factors are typically a market portfolio, a long-short size factor, a long-short value factor, and some higher-order PCs capturing convexity effects in the two characteristics. Instead, we will select long-only portfolios based on different tree cuts. One basis asset can be the market portfolio, which is just the original node of the tree. One cut in the size dimension results in small cap and large cap portfolios, and similar with value and growth. However, if the relevant pricing information requires the interaction between size and value, an additional cut in the interaction dimension would result in small value stock portfolios and other combinations (see Figure 2). As a result, one can always trace a particular asset back to stock fundamentals. Hence, our AP-Trees combine the advantages of PCA methods of grouping similar stocks together with our requirement of obtaining interpretable, long-only basis assets.

A lot of pathbreaking contributions have recently been made to studying the impact of characteristics on returns directly, without imposing an underlying risk model or a no-arbitrage condition. Freyberger, Neuhierl, and Weber (2020) estimate conditional expected returns as a function of characteristics with adaptive group lasso, allowing for a high-dimensional structure with nonlinearities, but rule out the crucial interaction effects between characteristics. DeMiguel, Martin-Utrera, Nogales, and Uppal (2019) and Feng, Giglio, and Xiu (2020) focus on characteristics–based factor selection with a lasso-type penalty, while Gu, Kelly, and Xiu (2020b) use machine learning techniques, like neural networks, to estimate asset pricing models that account for general functional forms including interactions. Chen, Pelger, and Zhu (2019) and Gu, Kelly, and Xiu (2020a) combine economic model restrictions with the flexibility of neural networks to estimate an SDF. None of these papers, however, offers the same portfolio interpretability as our AP-Trees.

To the best of our knowledge, Moritz and Zimmerman (2016), Gu, Kelly, and Xiu (2020b), and Rossi (2018) are the only papers relying on decision trees in estimating conditional moments of stock returns. Moritz and Zimmerman (2016) apply tree-based models to studying momentum, while Gu, Kelly, and Xiu (2020b) use random forest to model expected returns on stocks as a function of characteristics. Rossi (2018) uses Boosted Regression Trees to form conditional mean-variance efficient portfolios based on the market portfolio and the risk-free asset. Since we use decision trees not for a direct prediction of returns but for constructing a set of basis assets that span the efficient frontier, none of the standard pruning algorithms available in the literature is applicable in our setting because of its global optimization nature. Section II highlights this difference further and introduces an alternative criterion we develop for pruning.

Naturally, our paper expands the literature on constructing optimal test assets. Lewellen, Nagel, and Shanken (2010) argue that conventional double-sorted portfolios, exposed to a small number of characteristics, often present a low hurdle for asset pricing models due to their strong embedded factor structure, and they recommend mixing them with other cross-sections. For a given set of characteristics we offer an alternative set of test assets that is harder to price, as it extracts additional information due to inherent interactions and nonlinearities in both the impact of characteristics on asset returns, and the distribution of the assets in the cross-section. Ahn, Conrad, and Dittmar (2009) use return correlation to group stocks into portfolios and find that it substantially decreases the estimation error of constructing the efficient portfolio frontier and other objects of interest. Nagel and Singleton (2011) introduce optimal managed portfolios based on a General Method of Moment (GMM) argument that essentially builds a set of optimal instruments for a particular pair of the null and alternative models. Our test assets are long-only portfolios chosen as a robust span of the SDF, projected on a given space of characteristics. As a result, while incorporating the information about both expected returns and their comovement, they are not designed to improve the power of the test when making inference for *a particular* parameter and/or factor, but rather try to answer the question of whether a given set of basis assets is representative of the information contained in the original stocks and their characteristics.

Finally, our pruning approach also contributes to the econometrics literature of shrunk mean-variance estimation and nonparametric mean estimation. We offer a new perspective of mean-variance optimization by solving the Markowitz optimization problem, that is minimizing the variance for a specific target mean return, with an elastic net penalty, and treating the target mean as a tuning parameter. We show that this corresponds to tracing out a robust efficient frontier and finding the portfolio with the highest Sharpe ratio under parameter uncertainty. As a result, our estimator has three different statistical interpretations. First, it can be interpreted as robust mean-variance optimization under uncertainty bounds on the estimated mean and variance. Second, it has the flavor of a regression with a ridge–type shrinkage in the variance and mean, and a lasso penalty to obtain a sparse set of coefficients. In particular, a conventional elastic net regression (Zou and Hastie (2005)) is a special case of our approach with a different weighting in the ridge penalty and without the mean shrinkage. Last but not least, by extending the findings of Kozak, Nagel, and Santosh (2020), our estimator can be viewed from the Bayesian perspective as imposing a specific prior on the mean.

I. Conditional Sorting and Trees

A conventional view of the impact of characteristics on asset returns is that they proxy for the underlying exposure to the systematic sources of risk.[4] While the characteristics of individual companies and hence, their risk exposure can rapidly change over time, sorting-based portfolios provide a time-varying rotation from individual securities to their baskets, not only diversifying the idiosyncratic risk but also providing (hopefully) a time-invariant exposure to the underlying

[4] Fan, Liao, and Wang (2016) and Kelly, Pruitt, and Su (2019) model the conditional exposure to common risk factors in the following way:

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

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

risk factors. While it certainly restricts the functional dependency of expected returns on characteristics, the ultimate goal is to move away from the individual assets and focus on the stable underlying factor structure in the characteristic space. In this setting sorting can also be viewed as a nonparametric estimator of expected returns on individual stocks, with a particular choice of a kernel that corresponds to cross-sectional quantiles (see Cochrane (2011)). Cattaneo, Crump, Farrell, and Schaumburg (2019) formalize this intuition further and derive the optimal portfolio splits for estimating expected returns in the case of a single characteristic.

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

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

Conditional portfolio sorts, based on a selected set of characteristics, are intuitive to build and interpret within the setting of decision trees. Figure 1 gives an example of a conditional tree of depth 3, where stocks are first sorted by size, then value, and then size again. Conditional splits are built using the information at time t, and hence, similar to double sorting, can easily handle an unbalanced panel of stocks. At time t, all the stocks that have valid size and value information from $t-1$ (or previous periods), are first sorted into 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. Lastly, the four groups 111, 112, 121, and 122 are each further split into two portfolios each, to form eight level-three portfolios. The notation of each node is therefore reflecting a chosen path along the tree with a specified list and *order* of the split criteria.

If stock-specific characteristics are independent, the order of the variables used for splits does not matter, and we end up with the same quantiles for splits as double sorting. Unconditionally, the cumulative density function of each characteristic has a uniform distribution. However, it is well known empirically that characteristics have a complicated joint relationship that question the validity of coarse double sorting as an appropriate tool to reflect expected returns. For example, Figure E.1, Panel A, shows the sample distribution of characteristics in the cross-section of stocks and their conditional and unconditional impact on expected returns for two examples: size/value and size/accruals. On average, there seems to be a negative cross-sectional correlation between size and book-to-market with a clear clustering around the north west and south east corners.[5] As a result, double-sorted portfolios are heavily unbalanced across the characteristic spectrum. Depending on the pair of characteristics, their joint distribution could exhibit very different shapes and relationships that are far from a simple linear correlation structure.

However, it is not only the dependence structure of the characteristics that substantially affects sorting-based portfolios but also their price impact, which is often highly nonlinear and full of interactions.[6] 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, medium and small securities reveal a striking inverted U-shape pattern. Figures E.1 and E.2 in the Appendix further demonstrate different cross-sectional distributions of characteristics and their conditional and unconditional effect on expected returns, which sometimes are parallel to each other, speaking to the additive nature of the data-generating process, but often are not.

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

To sum up, in case of independent characteristics, conditional trees lead to creating portfolios in line with the double sorting methodology. However, if the underlying distribution is in fact

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

[6] 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, Gu, Kelly, and Xiu (2020b) and Chen, Pelger, and Zhu (2019) for the machine learning approach.

dependent, the collection of all possible conditional splits delivers a multitude of basis assets that are fundamentally different from those created with double sorting, that, as a result, could span a different type of SDF.

II. Asset Pricing Pruning

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

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

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

A. Robust SDF Estimation

We find SDF weights by solving a mean-variance optimization problem with elastic net shrinkage applied to all final and intermediate nodes of AP-Trees. This approach combines three crucial features[7]:

[7] Including shrinkage is also necessary to deal with the challenge that the number of portfolios N is large relative to the number of time periods T. In this case, it is well-known that the naively estimated tangency portfolio can deviate significantly from the true population weights. One fundamental problem is the estimation of a large dimensional covariance matrix where the number of parameter is of a similar order as the convergence rate based on T. Hence, the

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

We summarize the properties of AP-Pruning in Propositions 1 to 3. The formal mathematical statements, detailed derivations and description of the implementation are in Appendix A.

The search of the tangency portfolio can effectively be decomposed into two separate steps. First, we construct a robust mean-variance efficient frontier using the standard optimization with shrinkage terms. Then, we select the optimal portfolio located on the robust frontier: we find the tangency portfolio on the validation data set which has not been used in the first step.

Consider the whole cross-section of excess returns on the portfolios built with AP-Trees, and denote their sample estimates of mean and variance-covariance matrix by $\hat{\mu}$ and $\hat{\Sigma}$. For each target expected return μ_0, we find the minimum variance portfolio weights $\hat{\omega}_{\text{robust}}$ with an elastic net penalty. This is the classical Markowitz problem with additional shrinkage. All the tuning parameters, that is the target mean μ_0, the lasso weight of λ_1, and ridge with λ_2 are treated as fixed at this step, and chosen separately on the validation data set. In other words, the estimation proceeds as follows[8]:

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

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

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

estimation errors in each entry become non negligible when inverting the sample covariance matrix. The literature has proposed several ways to deal with this issue, including a low rank factors structure, e.g., in Fan, Liao, and Mincheva (2011), shrinkage of the covariance matrix to a diagonal matrix, e.g., Ledoit and Wolf (2004), or using a lasso-type penalty in the mean-variance estimation to obtain sparse portfolio weights, e.g., Ao, Yingying, and Zheng (2018). Fundamentally, all these approaches reduce the dimensionality of the parameter space. In our setup the variance shrinkage to a diagonal matrix by the ridge penalty and the sparse SDF weights due to lasso take care of the large N and T issues. Note that the sparse structure in our SDF weights puts us implicitly into a small N and large T environment that circumvents the dimensionality issues.

[8] Note that the two-step approach for finding a robust tangency portfolio is a method that is not specific to using tree-based basis assets, but can be applied to any potential cross-section. However, we argue that it is particularly appealing when using AP-Trees, as it obtains a small number of interpretable basis assets, which are selected from a larger set of potentially highly correlated portfolios. While there are other techniques that could efficiently handle a case of highly correlated securities, like PCA, they often lack interpretability, and hence, investigating a failure of a given candidate model to price them could be challenging at best.

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

Without imposing any shrinkage on the portfolio weights for the SDF, the problem has an explicit solution, $\hat{\omega}_{naive} = \hat{\Sigma}^{-1}\hat{\mu}$. Our estimator is a shrinkage version as specified in Proposition 1.

PROPOSITION 1 (Robust Mean-Variance Optimization): *Tracing out the efficient frontier (without an elastic net penalty, $\lambda_1 = \lambda_2 = 0$) out-of-sample to select the tangency problem is equivalent to applying conventional in-sample mean-variance optimization but with a sample mean shrunk toward the cross-sectional average. It results in the weights*

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

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

Tracing out the robust efficient frontier out-of-sample that it includes a ridge penalty (i.e., no lasso penalty, $\lambda_1 = 0$, but general λ_2) is equivalent to conventional in-sample mean-variance optimization but with a shrunk sample mean and a sample covariance matrix shrunk toward a diagonal matrix. It has the weights

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

Our approach generalizes the SDF estimation approach of Kozak, Nagel, and Santosh (2020) by including a mean shrinkage to the variance shrinkage and sparsity. We show that tracing out the whole efficient frontier is generally equivalent to different levels of shrinkage on the mean return, and generally does not have to be zero, which is imposed in Kozak, Nagel, and Santosh (2020). In fact, using cross-validation to find the optimal value of this shrinkage for a set of 36 different cross-sections we build in the empirical application, we find that it in most cases it is not equal to 0. Intuitively, since the estimation of expected returns is severely contaminated with measurement error, 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.[9]

PROPOSITION 2 (Robust SDF Recovery): *The robust mean-variance optimization in Definition 1 generalizes the robust SDF recovery of Kozak, Nagel, and Santosh (2020). It is equivalent to applying the robust SDF recovery to the sample covariance matrix and the sample mean vector shrunk toward the cross-sectional average mean. It is identical to the SDF recovery of Kozak,*

[9]The same reasoning underlines the use of the so-called adjusted stock betas by Bloomberg that shrink their sample estimates toward 1, the average in the overall cross-section:

$$\hat{\beta}_{adj} = 0.67 \times \hat{\beta}_{sample} + 0.33 \times 1.$$

Nagel, and Santosh (2020) for the mean shrinkage of $\lambda_0 = 0$ or if the sample covariance matrix is a diagonal matrix. In the second case, the robust SDF weights equal[10]

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

In the general case of a non-diagonal sample covariance matrix the solution on the active set, that is for the non-zero values of ω_i are characterized by the following:

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

For the special case of $\lambda_0 = 0$ this is identical to the corresponding first order condition of Kozak, Nagel, and Santosh (2020). The robust SDF of Kozak, Nagel, and Santosh (2020) can be interpreted as a portfolio on the robust efficient frontier, which is not necessarily the tangency portfolio.

Our estimator can also be interpreted as a robust approach to the mean-variance optimization problem, when there is estimation uncertainty about the mean and variance-covariance matrix of returns. Each type of shrinkage has a one-to-one correspondence to a specific type of uncertainty in the estimation.[11]

PROPOSITION 3 (Robust Estimation Perspective): *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*

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

Each shrinkage has a direct correspondence to an uncertainty set: Mean shrinkage provides robustness against Sharpe ratio estimation uncertainty, variance shrinkage governs robustness against variance estimation uncertainty, and lasso controls robustness against mean estimation uncertainty. A higher mean shrinkage can also be interpreted as a higher degree of risk aversion of a mean-variance optimizer.

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.

[10] We define $(x)_+ = \max(x, 0)$, and $sign(x)$ is the sign of x.

[11] 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, and have led to empirical gains in a number of applications (see, e.g., Garlappi, Uppal, and Wang (2007) or Fabozzi, Huang, and Zhou (2010) for a review).

B. A Toy Simulation

We are illustrating the effect of shrinkage with a simple simulation example that considers only mean shrinkage. Recall that shrinking the cross-section of expected returns towards their mean is equivalent to shrinking the in-sample tangency portfolio toward the minimum variance one, which is easiest to see without any additional types of shrinkage (i.e., $\lambda_2 = \lambda_1 = 0$): $\hat{\omega}_{\text{robust}} = \alpha_{\mu_0}\hat{\omega}_{\text{naive}} + (1-\alpha_{\mu_0})\hat{\omega}_{\text{var}}$. The weights $\hat{\omega}_{\text{naive}}$ correspond to the naive tangency portfolio and $\hat{\omega}_{\text{var}}$ to the minimum variance portfolio. Note that there is a one-to-one mapping between α_{μ_0}, the mean shrinkage λ_0 and the target return μ_0. As with other shrinkage estimators, the main idea underlying the use of this weighting is a trade-off between the mean and the variance: Whenever the estimation error in expected returns is large enough, using naive optimal sample weights will lead to a substantial instability, especially out-of-sample. It could therefore be optimal to shrink the SDF weights toward a more feasible portfolio that does not rely on the poorly estimated returns – the minimum variance portfolio. As long as the original measurement error is large, relative to the actual spread in returns, decreasing the variance through shrinkage to the mean will lead to a gain in the overall Sharpe ratio, moving the frontier closer to the actual optimum.

To illustrate the impact of shrinkage explicitly, we shut down all the additional sources of uncertainty, and focus on the case where there is a cross-section of n independent portfolio returns with known variance, however, different (and unknown to the investor) expected returns as follows:

$$R \sim \text{i.i.d. } N(\mu_i, n\sigma^2), \quad i = 1, ..., n,$$

where $\mu_i \sim U[0,1]$ and $\sigma = 1$. The variance of the returns is scaled by n, the number of stocks, to make sure that it does not disappear for a well-diversified portfolio of stocks.

After observing the sample of T_{train} return realizations on the training dataset, an investor uses the estimated sample mean to make his portfolio decision, which we evaluate out-of-sample.[12] Note, that while on average the sample mean estimates are unbiased, in a given cross-section of n stocks some of them will be biased upward, while others downwards, which in case of the naive asset allocation will overweight and underweight the assets accordingly. We consider the following set of potential asset allocations:

$$\hat{\omega}_{\text{naive}} = \frac{1}{\mathbb{1}^\top \hat{\mu}}\hat{\mu}, \quad \hat{\omega}_{true} = \frac{1}{\mathbb{1}^\top \mu}\mu, \quad \hat{\omega}_{\text{robust}} = (1-\alpha)\hat{\omega}_{\text{naive}} + \alpha\hat{\omega}_{\text{var}} = (1-\alpha)\frac{1}{\mathbb{1}^\top \hat{\mu}}\hat{\mu} + \alpha\frac{1}{\mathbb{1}^\top \mathbb{1}}\mathbb{1}$$

for a range of $\alpha \in \{0.2, 0.5, 1\}$. Once the allocations are fixed, we observe a different realization of returns over the out-of-sample period T_{test}.

Figure 3 shows the distribution of the out-of-sample Sharpe ratios, achieved by different allocation strategies simulating each time-series for 1,000 times. The risk-return trade-off, achieved by

[12] Note that in this simple simulation setup the estimates of the expected returns can be simulated as $\hat{\mu}_i \sim \mu_i + \frac{1}{\sqrt{T_{train}}}N(0, n\sigma^2)$. Similarly, the average returns over the out-of-sample period T_{test} can be simulated as $\hat{\mu}_i \sim \mu_i + \frac{1}{\sqrt{T_{test}}}N(0, n\sigma^2)$.

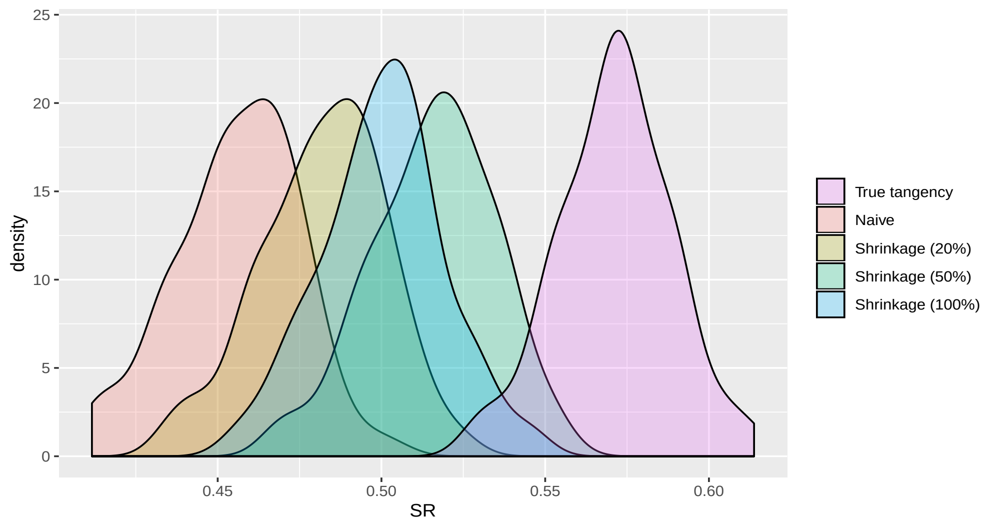


Figure 3. Distribution of out-of-sample Sharpe ratios, achieved with shrinkage-based allocations. The figure presents the out-of-sample distribution of Sharpe ratios for investment strategies, designed to reflect the tangency portfolio. The empirical densities characterize the oracle investment strategy (true tangency, denoted by purple), the naive one (based on the in-sample estimates of average returns, denoted by orange), and the set of shrinkage based-strategies that rely on shrinking the in-sample average returns toward their cross-sectional average (20%, 50%, and 100%, correspondingly). 100% shrinking corresponds to the minimum variance portfolio. $T_{train} = 100$, $T_{test=1000}$, $n = 50$, with the densities computed for a sample path.

naive, in-sample weights for the tangency portfolio, is generally subpar to both the shrinkage-based strategies, and obviously, the true tangency portfolio. Note, that while a modest amount of shrinkage (20-50%) helps to get closer to the unbiased estimates of the Sharpe ratio, a strong degree of it is harmful, since such an asset allocation disregards all the information in expected returns.

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. As a result, it essentially searches across the out-of-sample scenarios and picks the optimal values that get closer to the actual, infeasible, fully unbiased estimation.

C. Pruning Depth

Our portfolio selection approach prunes the AP-Tree in both depth and width, since we include all the intermediate and final nodes in the robust mean-variance portfolio selection. Hence, our selected tree portfolios can be higher-level nodes without further splits. As an example consider a univariate sorting with only one characteristic and depth 2. If only high values of the characteristics have an effect on mean returns, the low level of characteristics can be grouped together in one large portfolio. In this example, a possible selection could be a subtree of depth 1 for lower values that has 50% of the observations and two subtrees of level four for high values that each have 25% of the stocks.

Figures 4 and E.6a further illustrate the importance of pruning in depth of the tree, which amounts to selecting larger, denser portfolios. In the case of a single characteristic, simple univariate sort with the same population density as the tree of depth 3, would yield eight portfolios that are equally spaced from each other in the characteristic space, similar to the standard decile-sorted portfolios. In contrast, depending on the data-generating process, pruning the tree could yield a simple cross-section of a smaller set of assets, some of which are retained from original octiles, while others come from the intermediate nodes of the tree, essentially merging higher level nodes, as long as there is not enough informational gain from doing such a split.[13] In the particular example of the selection in Figure 4, 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. Why does it happen?

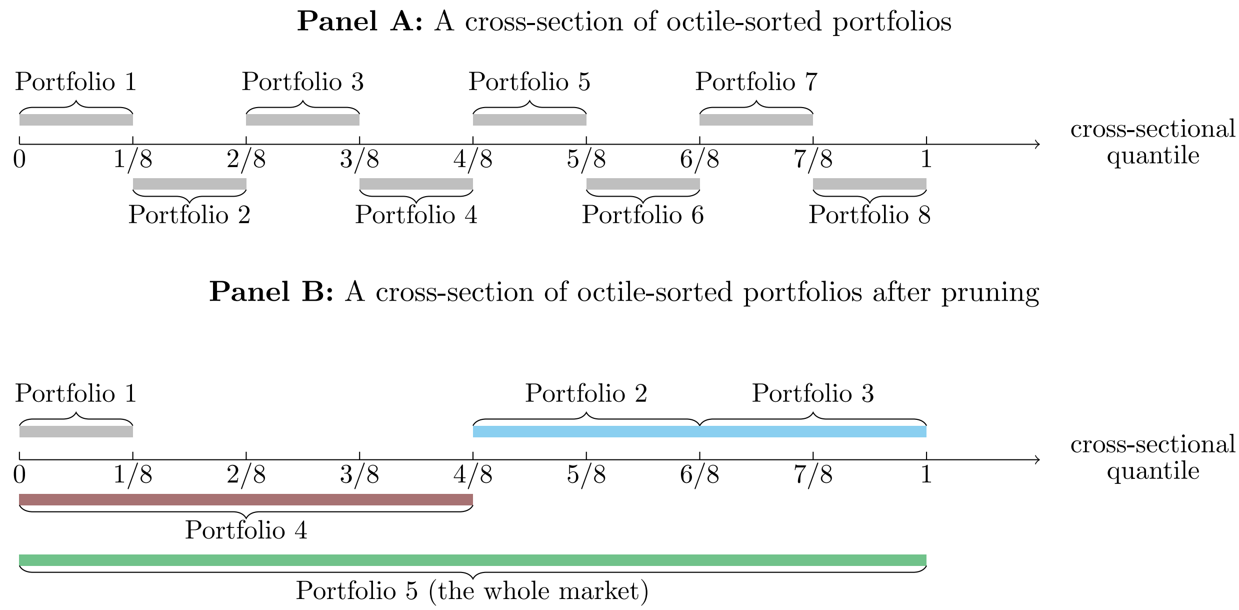


Figure 4. Selection of basis portfolios through pruning

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.

The problem of split selection fundamentally reflects the trade-off between the estimation error and bias. Tree portfolios at higher (intermediate) nodes are more diversified, naturally leading to a smaller variance of their mean estimation, and so forth, while more splits allow to capture a more complex structure in the returns at the cost of using investment strategies with higher variance. To mitigate this trade-off, we use the weighting scheme inspired by the properties of the GLS estimator. The idiosyncratic noise in each of the tree 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 $\sqrt{N_i}$. An equivalent approach would be to simply weight each portfolio by $\frac{1}{\sqrt{2^{d_i}}}$, where d_i is the depth of portfolio i in the tree.[14]

The selection of depth in an AP-Tree can be interpreted as the choice of bandwidth in a non-

[13]An illustration of potential pruning output for the case of two characteristics can be found in Figure E.6.

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

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.

AP-Pruning 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 similar arguments as 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 at achieving 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.[15]

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

III. Illustrative Simulation Example

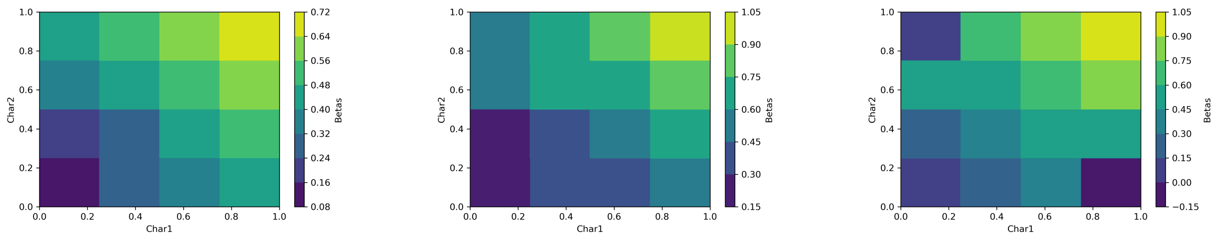
We illustrate the benefits of using tree-based portfolios in uncovering the patterns of expected returns in the characteristic space and the efficiency of our pruning approach with a simple simulation that is designed to capture some of the stylized features of the data.

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

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

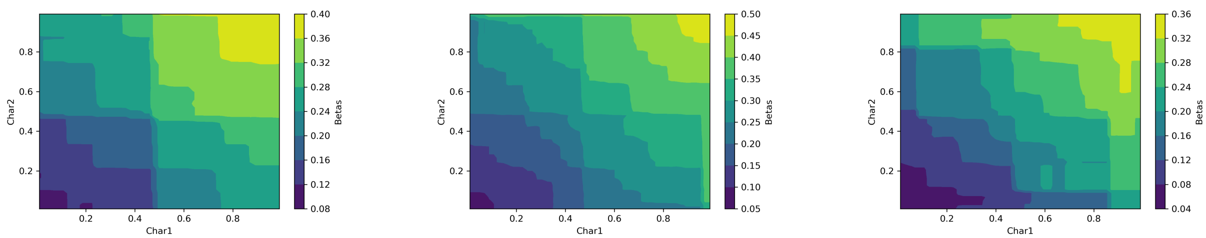
In our simple model the factor follows $F_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_F, \sigma_F^2)$ and the idiosyncratic component $\epsilon_{t,i} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. In the main text, we model the risk loadings as a linear function of characteristics $\beta_{t,i} = C^{(1)}_{t,i} + C^{(2)}_{t,i}$. Motivated by a wide range of empirical patterns in the joint empirical distribution of characteristics and their impact on expected returns, we consider nonlinear formulations for the

[15] Alternatively, one could compare our pruning in depth approach to switching between lasso and group lasso, where, depending on the underlying characteristics, the algorithm automatically switches between selection among the underlying variables and their groups.



Correlation = 0 Correlation = 0.5 Correlation = 0.9

Panel A: Beta patterns, spanned by double-sorted portfolios



Correlation = 0 Correlation = 0.5 Correlation = 0.9

Panel B: Beta patterns, spanned by AP-Trees

Figure 5. SDF loadings (betas and expected returns) as a function of characteristics in a linear model.

Loading function β as a function of the two characteristics in the linear model for different levels of correlation in the characteristics. The top plots show the loadings estimated with double-sorted portfolios and the bottom plots the corresponding results with pruned AP-Trees.

risk-loadings in Appendix B, which lead to the same insight as the simple linear model.

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

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

where Corr-Uniform$[0, 1, \rho]$ denotes a pair of uniformly distributed random variables that have correlation ρ and marginal densities $U[0, 1]$. To model the impact of characteristic dependence on the portfolio structure, we consider three cases: $\rho \in \{0, 0.5, 0.9\}$. The details for other parameters are in Appendix B.

We build double-sorted portfolios (DS), based on two characteristics distributed on a 4×4 grid, which creates a cross-section of 16 portfolios. Similarly, we build AP-Trees with a depth of 4, so that each portfolio consists of at least 1/16th of all the stocks and has a similar granularity as the double-sorted cross-section. We set cross-sectional dimensions to $N = 800, \quad T = 600$. In line with the strategy used for our empirical applications, we rely on the first $T_{train} = 240$ time series observations for training, the next $T_{val} = 120$ for parameter validation, and the last $T_{test} = 240$ observations to form the actual dataset for studying model performance. Once all the portfolios have been constructed (and in the case of AP-Trees, pruned to the set of 20 potentially overlapping basis assets) we estimate stock betas spanned by these basis assets.

Our focus is the estimation of the stock loadings $\beta_{t,i}$ on the SDF, because they incorporate all the relevant information for asset pricing. Figure 5 presents the scaled version of the estimated SDF betas (and hence, conditional expected returns) for different basis assets. Similar to the way portfolio buckets are used to estimate expected returns, we compute security betas by averaging the stock betas that belong to the same portfolios (and in case of the trees, averaging across the overlapping portfolios). The pattern in expected returns is quite striking. Ideally, it should be close to the one used to generate the data (Figure E.4, left two plots). However, in practice there is a substantial difference between the type of shapes and figures one could get with double-sorted portfolios relative to those reflected by conditional trees. The difference remains substantial even in the simplest case of independent characteristics, where both tree-based portfolios and double-sorted ones are most similar to each other, having 1/16 of all the stocks. Clearly, averaging betas across conditional basis assets allows tracking the underlying SDF loadings in a the characteristic space.

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

For the sake of brevity, we have discussed here results for the case of linear factor loadings and present additional findings in the Appendix. To summarize, even for the case of factor loadings, linear in characteristics we find that: (1) Overall, conditional sorts build portfolios that provide a substantially more accurate reflection of expected returns in the characteristic space and remain fully interpretable. (2) The higher is the correlation among characteristics, the larger is the benefit of using conditional sorts.

IV. Data

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

To build the sorting variables for the decision trees, we have constructed 10 firm-specific characteristics as defined on the Kenneth French Data Library.[16] All these variables are constructed

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

from either accounting variables from the CRSP/Compustat database or past returns from CRSP. Monthly updated variables are updated at the end of each month for use in the next month. Yearly updated variables are updated at the end of each June following the Fama-French convention. The full details on the construction of these variables are in the Appendix.

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

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

V. Empirical Results

A. Estimation and Hyperparameter Tuning

To minimize the possibility of overfitting, we divide all the data into three samples: training, validating, and testing datasets. By fixing the portfolio structures estimates from the training sample, and by optimally choosing the tuning parameters on the validating dataset, we focus on the out-of-sample behavior of triple-sorted portfolios vs. those produced by pruned AP-Trees, and the SDFs, spanned by different basis assets. We consider a cross-section of portfolios, constructed with AP-Trees with depth 4, and its closest analogue in the standard methodology: 32 and 64 triple-sorted portfolios.[17] All the portfolios are value-weighted. For additional robustness, we also exclude all the level 4 nodes that are sorted only on one characteristic to avoid going into extreme tails of the distribution without interaction with other variables. As an example, from the set of tree portfolios generated for size, value, and turnover, we exclude all the nodes corresponding to a single 1/16th quantile of any of these characteristics.

Training sample. We use the first 20 years of data (1964 – 1983, or 240 monthly observations) to form AP-Trees of depth 4, and estimate the vector of average returns and the covariance matrix for both final and intermediate nodes that are then used to construct an efficient frontier with

[17]32 triple-sorted portfolios consider only a 50/50 split based 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.

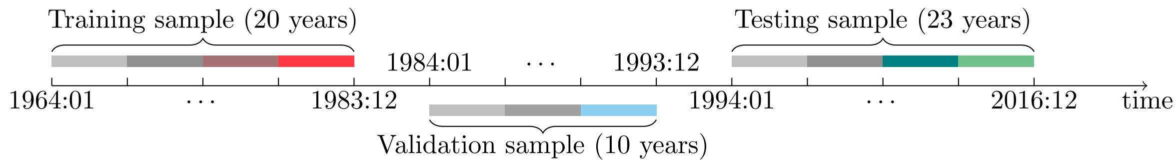


Figure 6. Timeline of the empirical strategy

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).

elastic net. For different levels of shrinkage (lasso and ridge) and target expected return, we find the optimal test assets and the SDF spanned by them with the corresponding weights. The same is done for triple-sorted portfolios, where we also use the elastic net to construct robust portfolio weights to mitigate the impact of estimation noise and sample variation. In other words, we allow triple-sorted portfolios to also benefit from the stability of shrinkage. This makes the overall comparison more fair and allows us to focus on the direct impact of using different basis assets to span the SDF.

Validating sample. The middle 10 years (1984 – 1993, 120 monthly observations) of returns serve as the validation set for hyperparameter tuning: We pick the model based on Sharpe ratio of the tangency portfolio on the validation dataset, fixing the SDF weights at their training values. Table **??** reports the hyperparameters we used for the SDF construction. For each combination of $(\lambda_0, K, \lambda_2)$ the lasso penalty λ_1 is chosen such that the number of non-zero weights reaches the target number K. In particular, we tune the value of λ_1 for AP-Trees to select 40 portfolios, which makes the dimension of pruned trees comparable to Fama-French triple-sorted 32 and 64 portfolios.

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

B. Evaluation Metrics

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

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

We evaluate the pricing information of these basis assets in two different dimensions: First, how much of the investments opportunities are spanned by different basis assets?[18] Second, how well can these basis assets be priced by conventional factor models, that is whether achievable Sharpe ratio really comes from loading on the conventional sources of risk, or something that is not spanned by them. We run the 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 a market, size and value factor;
- *FF5*: Fama-French five-factor model, which adds an investment and profitability factor to *FF3*;
- *XSF*: a cross-section specific model that includes the market factor, and three long-short portfolios, corresponding to the three characteristics used to build a cross-section. In order to build the factors, we use the same approach as used in constructing HML, momentum, or other conventional long-short portfolios.
- *FF11*: an 11-factor model, consisting of the market factor and all 10 long-short portfolios, based on the full list of characteristics.

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

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

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

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

[18] Alternatively, we could also measure the information content of different cross-sections by the Hansen-Jagannathan distance (or entropy), relative to some of the basic models, e.g., Consumption-CAPM.

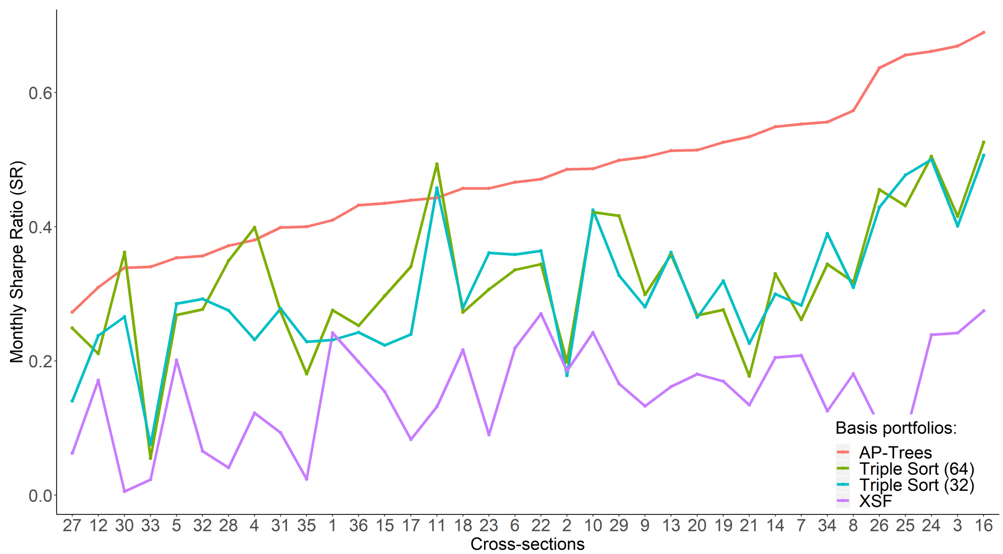


Figure 7. Sharpe ratios of SDFs spanned by pruned AP-Trees (40 portfolios), triple sorts, and XSF. Cross-sections are sorted by the SR achieved with AP-Trees.
The figure displays the monthly out-of-sample Sharpe ratio of the robust mean-variance efficient portfolios spanned by pruned AP-Trees (40 portfolios), triple-sorts, and XSF (market and the three cross-section specific characteristics). The SDF based on triple sorts is based on either 32 or 64 assets and considers mean and variance shrinkage.

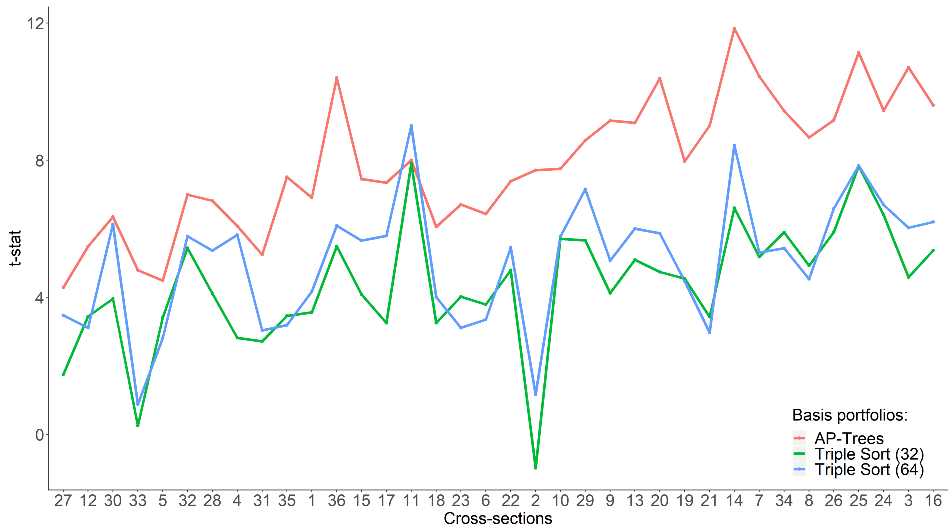
C. 36 Cross-Sections of Expected Returns

We start by comparing the SDFs spanned by cross-sections with different basis assets. Figure 7 summarizes their Sharpe ratios, when using *AP-Trees*, *TS(32)*, or *TS(64)*, as the "building blocks" of the SDF, along with the set of cross-section-specific long-short portfolios, accompanied by the market factor.

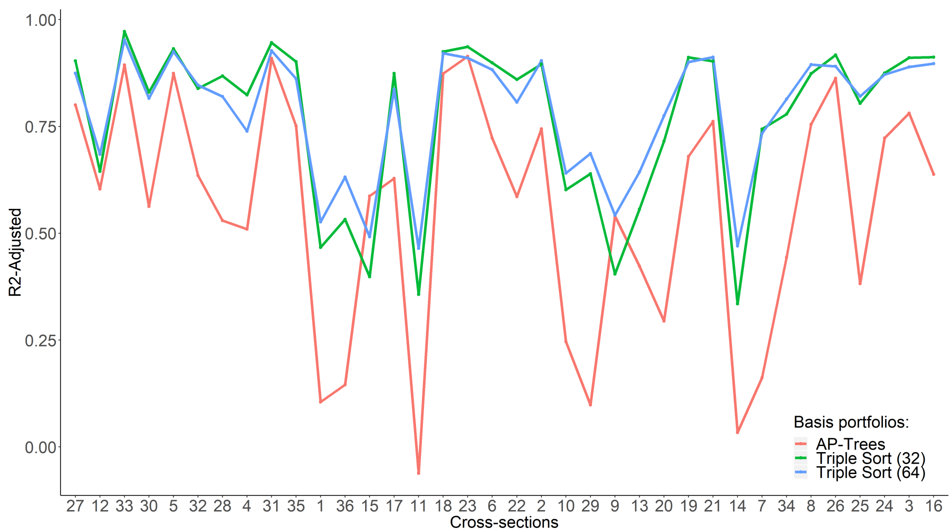
AP-Trees obtain considerably higher out-of-sample Sharpe ratios compared to the triple-sorted portfolios or conventional long-short factors. These 36 cross-sections are arranged according to the out-of-sample Sharpe ratio achieved with *AP-Trees*, with their labels and corresponding values reported in Table D.1. The differences in Sharpe ratios are striking. Compared to the case of simple long-short factors, our basis assets are able to deliver SR that are up to three times higher. As the Sharpe ratio measures the mean-variance efficiency of the SDF constructed with different basis assets, it implies that our AP-Trees extract more pricing information compared to the conventional benchmark basis assets. These results are particularly strong for such characteristics as investment, idiosyncratic volatility, and profitability.[19]

Where does this superior performance come from? First, note that cross-section-specific factors have around half of the Sharpe ratios spanned by triple sorts, that is the linear factors already miss information. This is expected, since even by construction the long-short factors cannot efficiently account for the interactions between the characteristics, while triple sorting can reflect their impact

[19]The only cross-sections where triple-sorting achieves a similar performance to AP-Trees are those related to short-term reversal. In Section V.F we show that this is due to instability in the portfolio weights, and even for those cases, AP-Trees outperform triple-sorted SDF weights when they are estimated on a rolling window.



(a) t-stat of the robust SDF alpha (w.r.t the Fama-French 5 factor model)



(b) R^2 within cross-sections (w.r.t the Fama-French 5 factor model)

Figure 8. Pricing errors of the SDFs spanned by AP-Trees (40 portfolios) and triple sorts, and R^2 within cross-sections.

The top figure displays the out-of-sample t-statistics of the pricing errors of the robust mean-variance efficient portfolios spanned by pruned AP-Trees (40 portfolios), triple sorts, and XSF (market and the three cross-section-specific characteristics). The bottom figure reports the adjusted cross-sectional R^2 for the 40 AP-Tree basis assets, respectively, 32, and 64, triple-sorted basis assets with respect to the five Fama-French factors. The SDF based on triple sorts is based on either 32 or 64 assets and considers mean and variance shrinkage. All the cross-sections are sorted by the SR achieved with AP-Trees.

at least to a partial extent. Having one or two splits in the size dimension, that is, 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: Tree-based portfolios are substantially harder to price using any standard factor model. Indeed, Figure 8 confirms that 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 AP-trees. Consider, for example, the case of size, value, and profitability (cross-section 2). The SDF, spanned by triple-sorted portfolios, does not have a significant alpha, when pitted against Fama-French five factors, while the one built from AP-trees has a t-statistics of 8. In other words, pricing conventional triple-sorted cross-sections, or spanning the factors that explain them, may be too low of a hurdle and has a real chance of missing important information contained in the original universe of stocks.

These findings are robust to the choice of benchmark factors. They are almost identical to the t-statistics obtained with cross-section-specific factors (see Figure E.7a in the Appendix), and survive even when faced with a whole set of 10 long-short portfolios built from all the characteristics used in our application, as well as the market (Figure E.7b). Interestingly, even when using all 11 cross-sectional factors, our AP-Trees have uniformly significant pricing errors. In contrast, around one-third of the triple sorts can be explained by the large set of factors. This again provides a strong support for using AP-trees as more informative test assets.

Cross-sectional R^2 is another benchmark often used to check whether the candidate model spans asset returns, and they confirm our alpha-based findings, as conventional factor models fail to explain the cross-section of expected returns constructed from AP-Trees. Figure 8b presents the pricing ability of the Fama-French five-factor model to span different sets of basis assets. Typically, cross-sections that obtain an R^2 of more than 80% would be considered as being well-explained by the set of factors, and this is largely the case for triple sorts. In contrast, one hardly gets the fit of over 50% on AP-Trees, if at all positive. This again illustrates that conventional sorting does not provide a sufficiently high hurdle for asset pricing models. This finding is robust to the choice of risk factors, with additional results reported in the Appendix.

D. How Many Portfolios?

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

We find that for most of the cross-sections, all the empirical results largely carry through even

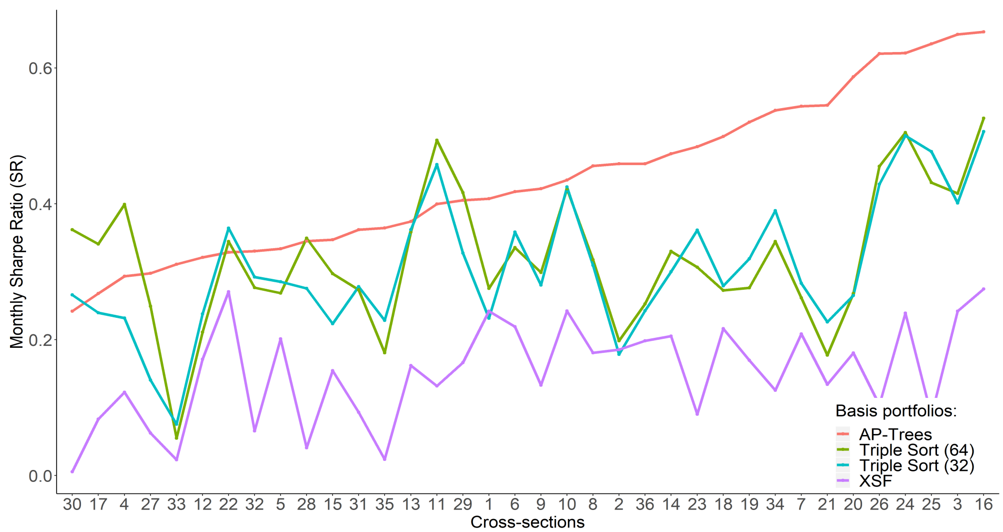


Figure 9. Sharpe ratios of the SDFs spanned by pruned AP-Trees (10 portfolios), triple sorts, and XSF. Cross-sections are sorted by the SR achieved with AP-Trees.
The figure displays the monthly out-of-sample Sharpe ratio of the robust mean variance efficient portfolios spanned by pruned AP-Trees (10 portfolios), triple sorts, and XSF (market and the three cross-section specific characteristics). The SDF based on triple-sorts is based on either 32 or 64 assets and considers mean and variance shrinkage.

for a relatively small number of basis assets. Figure 9 demonstrates monthly out-of-sample Sharpe ratios and pricing errors of the SDFs spanned by only 10 portfolios, based on trees, relative to those of triple sorts (32/64 portfolios) and the corresponding long-short factors. For most of the cross-sections, using only 10 portfolios is enough to retain roughly 90% of the original Sharpe ratio and its alpha relative to standard Fama-French five-factor model.[20] Of course, some characteristics are more heterogenous in their impact than others, and generally the optimal number of portfolios will differ, depending on the complexity of the conditional SDF, projected on these characteristics, and could be chosen optimally based on the validation sample or other data-driven techniques. The general point we would like to highlight is that the number of portfolios used to build a cross-section is often a poor reflection of the underlying information captured by that set of assets, with *effective* degrees of freedom quite often different from the sheer number of such portfolios. Naturally, simple measures of fit are not robust to recombining assets into larger, denser portfolios, and, as a result, are prone to a substantial bias.

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

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

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

		Type of the cross-section			
		AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)
SDF SR		0.65	0.69	0.51	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]
$XS\ R^2_{adj}$	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 the 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 of the cross-section, the table reports its monthly Sharpe ratio on the test sample, along with the alpha of the SDF spanned by the corresponding basis portfolios. Alphas are computed relative to the Fama-French three- and five-factor models, cross-section-specific factors (market and long-short portfolios, reflecting size, investment, and operating profitability), and the composite FF11 model that includes the market portfolio, along with the 10 long-short portfolios, based on the cross-sectional characteristics.

and the pricing error of the tangency portfolio spanned by these combinations of stocks.

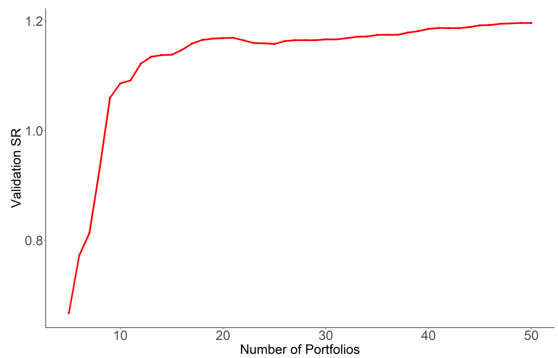
We now turn to a particular example of the cross-section to investigate further the structure of the SDF, spanned by AP-Trees, the characteristics of the optimal basis assets, and the patterns they reveal in expected returns.

E. Zooming into the Cross-Section

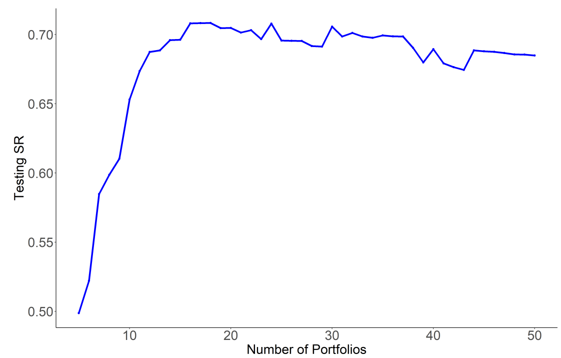
We focus on representative cross-sections to better understand the source of the pricing performance of AP-Trees. In the main text we consider the set of portfolios that could be built to reflect size, investment, and operating profitability. The results generalize to the other cross-sections. The Appendix illustrates that the results generalize to the cross-sections based on size, value, and idiosyncratic volatility, while the Internet Appendix collects the detailed results for all 36 cross-sections.

Table I presents the summary statistics for cross-sections, built with AP-Trees and triple sorts. As we have already observed, the AP-Trees have higher SR and generally larger pricing errors α (or, equivalently, lower cross-sectional R^2_{adj}) than the conventional portfolios, based on the unconditional quantiles. Interestingly, almost all of the information, relevant for asset pricing, is already contained in the set of 10 portfolios. A small number of assets is intuitively appealing and allows easily analyzing the source of the model performance. Table D.3 in the Appendix confirms that the results extend to the other set of characteristics.

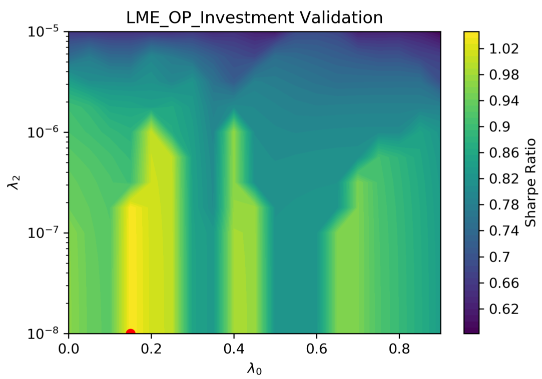
The selection of 10 portfolios, while seemingly surprising, is supported by the data, with both the validation and testing sample suggesting that having 10 portfolios is enough to capture most of



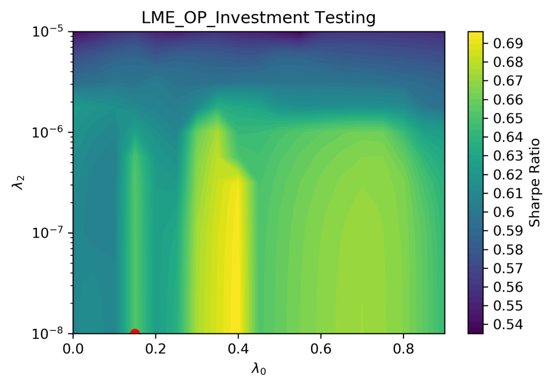
(a) Number of portfolios (validation sample)



(b) Number of portfolios (testing sample)



(c) Mean and variance shrinkage (validation)



(d) Mean and variance shrinkage (testing)

Figure 10. SR as a function of tuning parameters, ex ante (validation) and ex post (oracle).

The figure shows the validation and out-of-sample Sharpe ratio as a function of tuning parameters for AP-Trees built on size, investment, and operating profitability. Subplots (a) and (b) present the SR as function of the number of portfolios using the optimal mean and variance shrinkage. Subplots (c) 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. The optimal validation shrinkage is indicated by the red dot. We use the mean $\hat{\mu} + \lambda_0 \bar{\mu} \mathbb{1}$.

the variation in asset returns, without introducing the danger of overfitting (see Figure 10a). This decision is therefore in line with the traditional approach of choosing the optimal tuning parameter with one standard deviation of the one maximizing the cross-validation of some criteria. Inherently, both methods aim at selecting a sparse set of parameters/portfolios that give a performance very similar to that of the first best.

The heat-map of the tuning parameters indicates the range of shrinkage that was chosen as optimal on the validation dataset (Figure 10c). In this particular case, the impact of L2 penalty (ridge) on the composition of the cross-section was rather small; however, the model performance crucially depends on the choice of the shrinkage to the mean, λ_0. This value, which out-of-sample, could have been even higher, corresponds to the shrinkage of extreme sample returns towards the average in the cross-section, reflecting the intuition that returns both too high and too low relative to the benchmark, are probably due to a statistical error and will revert back at some point in the future. In this particular case, the chosen value of λ_0 was 0.15, corresponding to an important, but not excessive, shrinkage toward the minimum variance portfolio induced by the estimation error in expected returns.

Optimal amount and type of shrinkage can be different for each cross-section. Consider, for example, portfolios based on size, value, and idiosyncratic risk (see Figure E.15 in the Appendix). As before, 10 basis assets summarize most of the pricing information. However, for these portfolios, variance shrinkage seems to be more relevant than mean shrinkage. Note that on the test dataset which is not available for the choice of tuning parameters, some degree of mean shrinkage would in general have been beneficial. For both sets of cross-sections optimal mean and variance shrinkage matters, as it can increase the Sharpe ratio on the test data between 5% to 10%, but it is not the driver of the large differences between AP-Trees and triple-sorted cross-sections. Even without mean and variance shrinkage, AP-Trees clearly outperform conventional sorting.

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

Figure 11 displays the structure of the SDF, conditional on characteristics and the pricing errors for each of the individual basis assets with respect to XS-specific factors, with the candlestick denoting 5% confidence intervals. Overall six of the chosen 10 tree-based portfolios have consistently significant alphas. These portfolios are not just reflecting extreme quantiles of the underlying characteristics but are in fact characterized by a complex interaction structure. The prevalence of pricing errors, and the diversity of the stocks that go into such portfolios, are robust to the choice of risk factors, as almost the same pattern persists when basis assets are priced with the Fama-French 5, or 11 cross-sectional factors (see Table II, Panel A). Examining conventional triple sorts, in turn, reveals a completely different pattern (see Figure 11, Panel (d)) with generally substantially smaller pricing errors that are largely subsumed by standard risk factors. Exactly the same pattern holds for the other cross-sections, as illustrated by Figure E.16 in the Appendix.

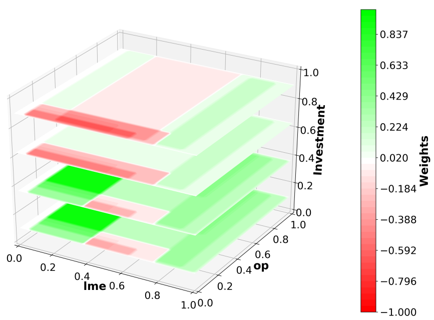
Table II, Panel B, lists the top 10 portfolios from the cross-section of 64 triple-sorted assets that are most difficult to price, according to the most comprehensive FF11 model. While there are obviously substantial alphas associated with some of these portfolios, the sheer number of them is somewhat misleading: Most of these portfolios consist of a small number of stocks (usually about 1.5%) and often reflect securities that are very similar to each other in terms of characteristics:

Table II Portfolios in the cross-sections

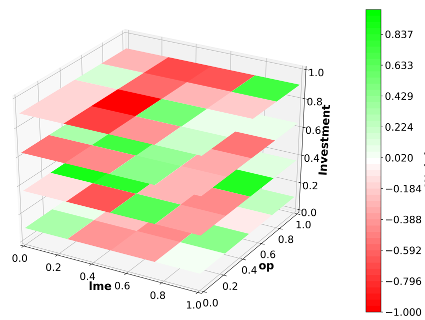
Portfolio ID	Portfolio construction	% of stocks	WV q_{LME}	α_{FF3}	α_{FF5}	α_{XSF}	$\alpha_{FF_{11}}$
Panel A: 10 AP-Tree portfolios							
1111.1	Market	100%	0.95	–	–	–	–
1111.121	LME(0.5-0.75)	25%	0.71	-0.03	-0.37	-0.05	-0.03
				[-0.34]	[-0.36]	[-0.40]	[-0.26]
3111.11	OP(0-0.5)	50%	0.94	0.03	-0.11**	-0.11**	-0.11**
				[-0.54]	[-2.33]	[-2.20]	[-2.46]
3111.11111	Inv(0-0.5) → LME(0-0.125)	6.25%	0.07	1.49***	1.32***	1.32***	1.63***
				[6.14]	[5.33]	[5.33]	[6.47]
3211.1121	Inv(0-0.5) → OP(0.5-1) → LME(0-0.5)	12.5%	0.51	0.23***	0.09	0.07	0.06
				[2.95]	[1.19]	[0.66]	[0.83]
1221.1111	LME(0-0.5) → OP(0-0.25)	12.5%	0.37	-0.37***	-0.34***	-0.34***	-0.29**
				[-2.81]	[-3.00]	[-3.01]	[-2.41]
2221.11111	OP(0-0.25) → LME(0-0.5)	6.25%	0.19	-0.16	-0.22	-0.21	-0.05
				[-0.717]	[-0.95]	[-0.89]	[0.21]
3331.12221	Inv(0.875-1) → LME(0-0.5)	6.25%	0.45	-0.82***	-0.62***	-0.62***	-0.46***
				[-5.86]	[-4.81]	[-4.78]	[-3.75]
2122.11111	OP(0-0.5) → LME(0-0.5) → OP(0-0.25)	6.25%	0.27	-0.37*	-0.43**	-0.42**	-0.20
				[-1.95]	[-2.21]	[-2.13]	[-1.00]
3133.12122	Inv(0.5-1) → LME(0-0.5) → Inv(0.75-1)	6.25%	0.48	-0.84***	-0.64***	-0.64***	-0.50***
				[-5.92]	[-4.84]	[-4.81]	[-3.92]
Panel B: Top 10 portfolios from TS (64)							
122	LME(0-0.25) × OP(0.25-0.5) × Inv(0.25-0.5)	1.91 %	0.17	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 (that 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.

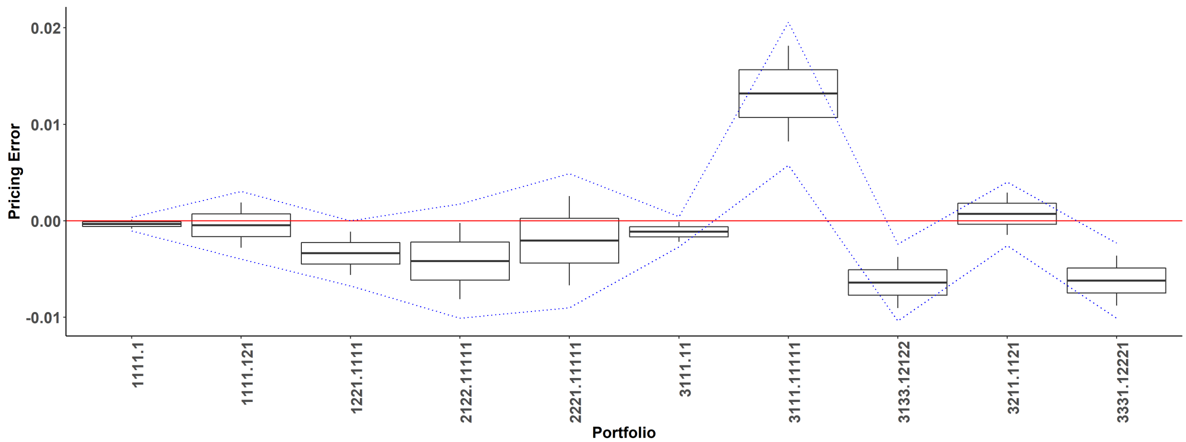
Small in size and profitability, high in investment. Since our pruning algorithm selects the basis assets in both characteristics and depth, these portfolios are actually often grouped together by AP-Trees, bundling these securities together. This pattern is particularly visible if you compare the structure of the SDF, spanned by different portfolios (see Figure 11, Panels a) and b)). For example,



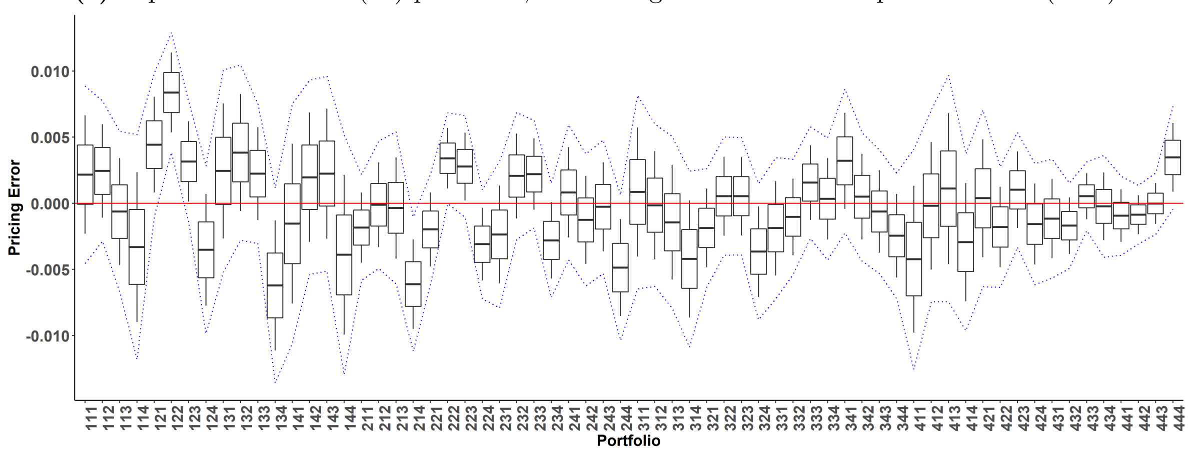
(a) SDF spanned by AP-Trees (10)



(b) SDF spanned by TS (64)



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



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

Figure 11. SDF weights and pricing errors for basis assets with 10 AP-Tree and 64 TS.
Composition of conditional SDF with 10 AP-Trees and 64 TS basis assets and corresponding pricing errors for each basis asset. Subplots (a) and (b) show the SDF weight as a function of the size, profitability, and investment quantile. Subplots (c) and (d) present the pricing errors relative to the cross-section-specific factors (XSF) with significance levels (candlestick=5%, dashed line=1%, box=10%).

the stocks that are generally high in investment (the top layer of the three-dimensional graph in Panel a)) are roughly grouped together and present the same exposure to the SDF, compared to six separate portfolios spanning the same characteristic space in triple sorting. Naturally, since the tree-based assets are constructed with both conditional and unconditional quantiles, the actual space of returns could be somewhat different, but the general pattern remains: There is often not enough signal in the data to warrant such a granular split.

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

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 R^2, based on the simple OLS estimates. This is precisely why for many popular cross-sections using GLS to evaluate the model performance often leads to a substantially lower measure of fit (see, e.g., Lewellen, Nagel, and Shanken (2010)), since treating separately these assets does not provide an incremental Sharpe ratio, which is reflected in the GRS statistic and other quantities that target investment opportunities, rather than a linear measure of fit.

Our main empirical results are unlikely to be driven by micro-caps. First, in constructing the trees, we specifically eliminated extreme groups of stocks that are heavily loading on a single characteristic, for example, size. We excluded all the splits that use the same characteristic to make the splits (e.g., 16 portfolios sorted only by size). Second, not only are tree-based portfolios generally composed of a larger number of stocks, efficiently diversifying the idiosyncratic noise, and endogenously grouping similar securities together, they are often comparable, if not better, to the triple sorts in terms of the actual market cap of the stocks that drive most of the difference in performance. Table II presents the value-weighted (since the portfolios are value-weighted themselves) size quantile of the stocks, that comprise each of the basis assets. Compared to the triple sorts, there is only one portfolio that loads heavily on the small caps (bottom 50% on investment and bottom 12.5% on the size within), and the rest have the same market cap as those based on triple sorts or larger.

Table III summarizes additional robustness checks. First, taking advantage of the intuitive and

Table III 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]
$XS\ R^2_{adj}$	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 of the cross-section the table reports its monthly Sharpe ratio on the test sample, along with the alpha of the SDF spanned by the corresponding basis portfolios. Alphas are computed relative to 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 that includes the market portfolio, along with the 10 long-short portfolios, based on the cross-sectional characteristics.

adaptive structure of the trees, we removed all the nodes that have an average size quantile below 40%, thus implicitly not allowing for splits that lead to portfolios with too many small-cap stocks (Panel A).[21] 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. Second, we also built different types of cross-sections using only stocks with a market 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 dataset, but the qualitative difference between AP-Trees and conventional sorting remains. For example, an out-of-sample Sharpe ratio of the SDF, spanned by 10 AP-Tree portfolios, is almost double that of triple sorts. Furthermore, all SDF alpha of AP-Trees remains significant even at 1%, while triple sorts are routinely spanned by conventional factors. Similar results for another example, cross-sections based on size, value, and idiosyncratic risk, could be found in the Appendix (Table D.4).

Finally, we can check whether most of the superior performance of the SDFs based on AP-Trees is coming at the expense of an unrealistically high turnover, which could prove it impossible to implement the strategies empirically (Table IV). For the baseline case of monthly rebalancing, an SDF portfolio, spanned by AP-Trees, achieves a monthly turnover in the range of 20–35%,

[21] The results for other cutoff levels are similar and available upon request. Furthermore, the tree-based approach naturally allows to incorporate any other restriction on the portfolios.

Table IV Portfolio turnover

		AP-Trees (10)	XSF	Operating Profitability	Investment
SR	Monthly	0.65	0.30	0.10	0.08
	Yearly	0.27	0.29	-0.07	-0.10
Turnover+	Monthly	0.28	0.24	0.35	0.35
	Yearly	0.13	0.13	0.11	0.12
Turnover-	Monthly	0.24	0.24	0.28	0.28
	Yearly	0.09	0.11	0.09	0.09

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

depending on the leg, which is generally comparable to investing in the corresponding long-short factors, while providing a higher Sharpe ratio. With annual rebalancing Sharpe ratio naturally declines but remains substantial, contrary to many passive long-short strategies. Pricing errors, cross-sectional fit, and other empirical results remain similar to the baseline case and are available upon request.

F. Additional Robustness Checks

All of our results so far assumed a constant projection of the SDF on the set of basis assets. To investigate the role of time variation in these weights, we fix portfolio selection, based on the training and validating datasets, and estimate the optimal robust combination of the basis assets on a rolling window of 20 years. Figure 12 summarizes our main results across all 36 cross-sections.

First, time variation in the SDF seems to be quite important empirically for both triple sorts and AP-Trees. Allowing for the time variation in portfolio weights definitely improves the performance of the robust tangency portfolio, as for almost all of the cross-sections it achieves a higher Sharpe ratio. The difference is large not only statistically but economically as well: For some of the cross-sections, Sharpe ratio with AP-Trees is double that achieved with constant weights. Second, using AP-Trees with flexible weights clearly dominates either version of the triple sorts. With fixed weights, there have been only three cross-sections, where triple sorting yielded the same or slightly higher Sharpe ratio, with all of them including short-term reversal (Figure 12, cross-sections 30, 4, and 11). However, with portfolio weights allowed to be time-varying, there is a clear advantage to AP-Trees.

Second, we tried to identify the sensitivity of these empirical results with respect to particular types of shrinkage used. Our main results are reported in Figures E.10 and E.11 in the Appendix. To investigate this issue, we compare the effect of mean and variance shrinkage on the Sharpe ratio for the time-varying SDF weights. Both AP-Trees and triple sorts benefit from shrinkage to the mean, which substantially stabilizes the SDF weights. These effects are not homogeneous across different cross-sections: For many sets of basis assets, the effect is rather small (with a magnitude of 1–5%), but for some cross-sections, mean shrinkage leads to large improvements on a

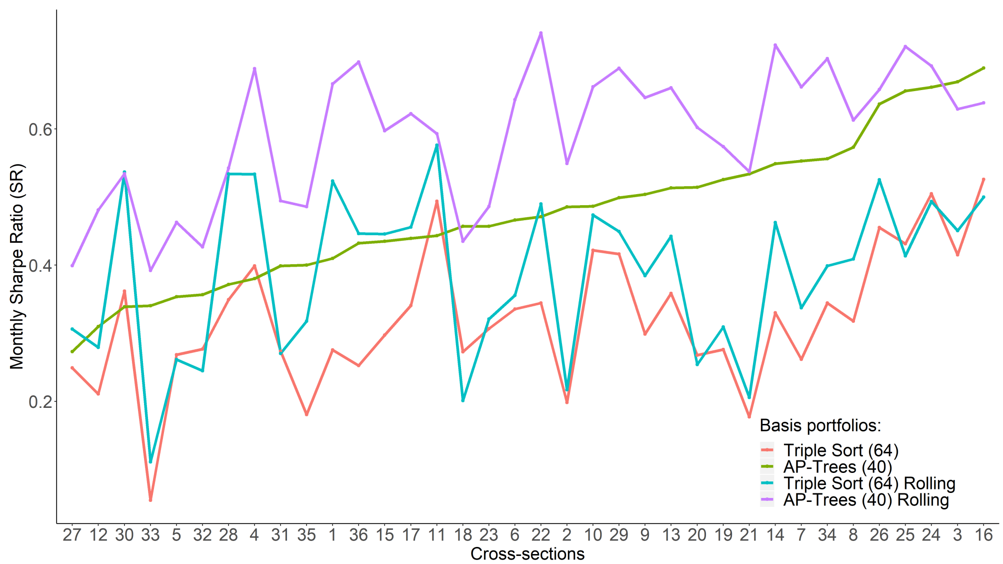


Figure 12. Sharpe ratios of the SDF estimated on a rolling window

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.

magnitude of up to 10–50%. A combination of mean and variance shrinkage results in the largest performance gains relative to the absence of any shrinkage at all.[22] In the Internet Appendix we also report the same results using a fixed 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 the tuning parameters are chosen non-optimally. In fact, the various shrinkage results are very close for optimally chosen tuning parameter or a non-optimal shrinkage. Hence, the robust SDF construction is relatively robust to the choice of tuning parameters, and it is often more important to include some degree of shrinkage than choosing the optimal one.

In short, better empirical performance of AP-Trees seems to be mainly driven by the construction of the new basis assets, not a particular choice of tuning parameters, as the large differences between conventional sorting and our novel AP-Trees are not explained by the shrinkage. However, both types of the portfolios benefit from constructing a robust SDF, using ridge and shrinkage to the mean.

G. A 10-Characteristic Cross-Section

Since our methodology relies on conditional splits, and the AP-Tree portfolios can generally reflect a large number of characteristics, it provides a general way to build cross-sections based

[22]Note that we estimate the mean and/or variance shrinkage tuning parameters optimally on the validation data, which can lead to different mean or ridge penalties if considered in isolation or combined.

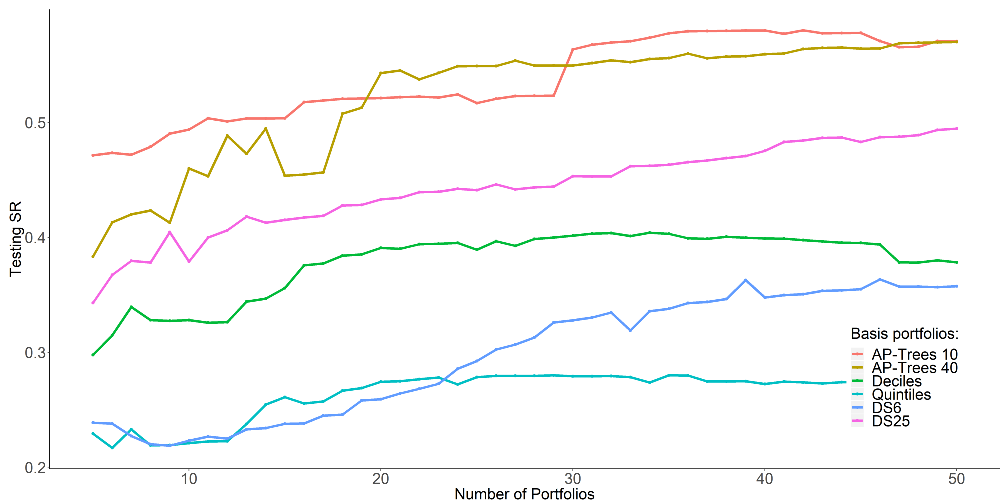


Figure 13. Sharpe ratios of the SDFs spanned by AP-Trees and combining double/univariate sorts based on 10 characteristics.

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). We apply robust shrinkage with lasso to all basis assets and choose the optimal validation mean and variance shrinkage.

on a relatively large (compared to what could naturally be handled by double- or triple-sorting) number of characteristics. In this section we describe the cross-section of portfolios, based on 10 characteristics at the same time.

Historically, there has been only one way to build a set of test assets that reflects more than two or three characteristics at the same time: Bundling several separate cross-sections together, usually either as a combination of several double or single sorts (e.g., 25 portfolios, sorted by size and value augmented by 10 portfolios sorted by momentum, five portfolios sorted by short-term reversal, five portfolios sorted by accrual). With an exploding dimension of such a cross-section, not only it is hard to make sure existing portfolios are fully reflecting the underlying characteristics and their interactions but also whether the portfolios challenging to price are actually economically different: If containing roughly the same stocks, treating them separately will not reflect the real economic advantage, which could also be captured out-of-sample within the SDF framework).

To provide a reasonable benchmark against which to compare the performance of tree-based portfolios, we consider several alternatives to generate the set of basis assets: a) Sets of 10 quintile portfolios, uniformly sorted by characteristics (50 assets altogether), b) Sets of 10 decile 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 V.C (360 and 1,440 assets, correspondingly).[23] Empirically, most papers focus on pricing

[23]We have also implemented tree construction based on 10 characteristics from first principles, optimizing the original set of portfolios. However, we have found that considering four-way interactions does not provide a significant

the cross-sections of no more than 60 portfolios; hence, the combination of all the available decile or double sorts, if anything, overrepresents the set of assets used in most papers.

Figure 13 presents the Sharpe ratio for the optimal mean-variance strategies, spanned by different basis assets. As before, to make the results comparable, we apply optimally selected shrinkage to each of the cross-sections, selected on the validating dataset, and study their performance out-of sample. Therefore, the only source of difference in the performance of the portfolios is the way they have originally been constructed (e.g., standard sorting vs. AP-Trees).

There are several important empirical observations. First, quintile-sorted portfolios that represent various anomalies, represent probably the least informative cross-section one could in good faith build based on the 10 characteristics we consider. While in general, univariate sorts are 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: Ultimately the Sharpe ratio one could get is either comparable to or lower than that of relying on the quintiles.

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, which is already at least twice higher than what standard empirical applications would deliver. The use of AP-Trees, however, raises that bar even higher: On average, they provide a 0.1 increase in the 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, Sharpe ratios of strategies that are built on tree-based portfolios (based on the prior selected 10 or 40) are very close to each other in terms of their efficiency, for almost all of the cross-sectional dimensions (e.g. 15–50 ultimately selected portfolios).

This increase in the composite Sharpe ratio is not coming from predominantly loading on the conventional sources of risk: Figure E.9 in the Appendix illustrates the out-of-sample alphas of the SDFs spanned by different basis assets. Almost regardless of the cross-sectional dimensions, 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). The relative order of the other types of the cross-sections often used empirically stays the same, with the stacks of quintiles based on the anomalies being the easiest one to be priced (out-of-sample).

Our results demonstrate that, first, for a large set of potential portfolios the actual information relevant for asset pricing (be it the composite spanning of the tangency portfolio, projected on the characteristic space, or the alphas relative to leading asset pricing models) can be reliably reflected

contribution to the overall SDF properties or portfolio selection, and that starting from assets reflecting the information in three characteristics achieves the same result, both quantitatively and qualitatively. Additional empirical results are available upon request.

in a smaller, optimally chosen subset of securities. Second, we show that using AP-Trees leads to at least a 20%–30% gain in both the total risk-return trade-off, and the ultimate mispricing reflected by the basis assets. Finally, we show how to construct a cross-section based on many (10, in our case) characteristics at once. Our approach is based on the symmetric treatment of all the information sets, the endogenous (adaptive) selection of the characteristics used for splits, and the depth of such selection. Finally, our method is flexible enough to incorporate all the additional constraints one could want to impose on the portfolio creation: Minimum market cap of the companies, number of stocks included, and many others.

VI. Conclusion

We propose a novel way to build basis assets for asset pricing that capture the complex information of a large number of cross-sectional stock return predictors. Our Asset-Pricing Tree portfolios are a small number of long-only portfolios that (1) reflect the information in many stock-specific characteristics allowing for conditional interactions and nonlinearities, (2) provide test assets for asset pricing that are considerably harder to price than conventional cross-sections, and (3) act as the building blocks for 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. We show that conventional cross-sections do not fully reflect the information contained in the characteristics of the underlying stocks and often present a rather crude, if not misguided, description of the expected returns.

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

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

REFERENCES

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

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.

Ao, M., L. Yingying, and X. Zheng (2018): "Approaching Mean-Variance Efficiency for Large Portfolios," *The Review of Financial Studies*, 32(7), 2890–2919.

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

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.

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

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.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998): "Liquidity and Stock Returns: An Alternative Test," *Journal of Financial Markets*, 1(2), 203 – 219.

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

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.

——— (2015): "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 116(1), 1 – 22.

Fan, J., Y. Liao, and M. Mincheva (2011): "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *Annals of Statistics*, 39(6), 3320–3356.

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.

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.

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

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.

Ledoit, O., and M. Wolf (2004): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 30(4), 110–119.

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.

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

Appendix A. Formal Statements and Derivations

Appendix A.1. Robust Portfolio Optimization (Proposition 1)

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

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

It is easy to notice that this problem is equivalent to the conventional Markowitz problem, where the sample covariance matrix $\hat{\Sigma}$ is replaced by $\left(\hat{\Sigma} + \lambda_2 I_N\right)$. Its solution is well-known in closed form as follows:

$$\hat{\omega}_{\text{robust}} = \alpha_{\mu_0}\hat{\omega}_{\text{tan},\lambda_2} + (1 - \alpha_{\mu_0})\hat{\omega}_{\text{var},\lambda_2},$$

where

$$\hat{\omega}_{\text{tan},\lambda_2} = c_{\text{tan}}\left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\hat{\mu}, \qquad \hat{\omega}_{\text{var},\lambda_2} = c_{\text{var}}\left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\mathbb{1}, \qquad \alpha_{\mu_0} = \frac{\mu_0 - \hat{\mu}^\top \hat{\omega}_{\text{var},\lambda_2}}{\hat{\mu}^\top \hat{\omega}_{\text{tan},\lambda_2} - \hat{\mu}^\top \hat{\omega}_{\text{var},\lambda_2}}$$

$$c_{\text{tan}} = \left(\mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\hat{\mu}\right)^{-1}, \qquad c_{\text{var}} = \left(\mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\mathbb{1}\right)^{-1}.$$

The optimal portfolio for a given target mean μ_0 is a weighted combination of the tangency portfolio $\hat{\omega}_{\text{tan},\lambda_2}$ and the minimum variance portfolio $\hat{\omega}_{\text{var},\lambda_2}$. The ridge penalty λ_2 affects the shape of the efficient frontier, while the target mean μ_0 pins down the location of the portfolio. The conventional efficient frontier without ridge-shrinkage sets $\lambda_2 = 0$, resulting in the weights $\hat{\omega}_{\text{naive}} = \hat{\omega}_{\text{tan},0}$ and $\hat{\omega}_{\text{var}} = \hat{\omega}_{\text{var},0}$. A higher target mean μ_0 increases the weight α_{μ_0} in the tangency portfolio. Note that if the target mean μ_0 is chosen on the training data to maximize the Sharpe ratio, it naturally yields the solution $\alpha_{\mu_0} = 1$.

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

$$\begin{aligned}\hat{\omega}_{\text{robust}} &= \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\hat{\mu} + \underbrace{\frac{\hat{\mu}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)\hat{\mu} - \mu_0 \mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)\mathbb{1}}{\mu_0 \mathbb{1}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)\mathbb{1} - \hat{\mu}^\top \left(\hat{\Sigma} + \lambda_2 I_N\right)\mathbb{1}}}_{\lambda_0}\left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}\mathbb{1} \\ &= \left(\hat{\Sigma} + \lambda_2 I_N\right)^{-1}(\hat{\mu} + \lambda_0 \mathbb{1}).\end{aligned}$$

Note that the shrinkage parameter λ_0 is a decreasing function in μ_0, that is, a lower target expected return implies more shrinkage toward the mean. Naturally, there is a direct mapping between these parameters, as $\mu_0 \in \left[\frac{\mu^\top(\hat{\Sigma}+\lambda_2 I_N)\mu}{\mathbb{1}^\top(\hat{\Sigma}+\lambda_2 I_N)\mathbb{1}}, \frac{\mu^\top(\hat{\Sigma}+\lambda_2 I_N)^{-1}\mathbb{1}}{\mathbb{1}^\top(\hat{\Sigma}+\lambda_2 I_N)\mathbb{1}}\right)$ corresponds to $\lambda_0 \in [0, +\infty)$. In particular, using $\lambda_0 = 0$ is equivalent to setting the target expected return to $\mu_0 = \frac{\mu^\top(\hat{\Sigma}+\lambda_2 I_N)^{-1}\mu}{\mathbb{1}^\top(\hat{\Sigma}+\lambda_2 I_N)^{-1}\mathbb{1}}$, which corresponds to the optimal portfolio weight of

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

Appendix A.2. Shrinkage Perspective (Proposition 2)

Our estimator generalizes the robust SDF recovery in Kozak, Nagel, and Santosh (2020). The general estimator of Kozak, Nagel, and Santosh (2020) solves the following problem:

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

If $\hat{\Sigma}$ is a diagonal matrix denoted by $\hat{D}$, this is conceptually equivalent to using our estimator in the PCA space, as advocated by Kozak, Nagel, and Santosh (2020). This has the following closed-form solution[24]:

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

with $(x)_+ = max(x, 0)$ element wise. As shown below, our estimator has the explicit solution

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

which coincides with that of Kozak, Nagel, and Santosh (2020) for an appropriate choice of λ_1. Hence, our estimator is equivalent to the one used by Kozak, Nagel, and Santosh (2020) in the case of uncorrelated assets. The analytical solution for a diagonal covariance matrix is based on the following argument. We solve the optimization problem with Lagrange multipliers:

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

The first order condition on the active set, that is, for the non-zero values of ω_i equals

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

which yields $\hat{\omega}_{\text{robust}} = \left(\hat{D} + \lambda_2 I_N\right)^{-1} \left(\tilde{\gamma}_1\hat{\mu} + \tilde{\gamma}_2\mathbb{1} - \lambda_1\mathbb{1}\right)_+$. As both Lagrange multipliers are functions of μ_0, we can reformulate the problem as $\hat{\omega}_{\text{robust}} = \left(\hat{D} + \lambda_2 I_N\right)^{-1}\left(\hat{\mu} + \lambda_0\mathbb{1} - \tilde{\lambda}_1\mathbb{1}\right)_+$. Here we have relaxed the constraint $\hat{\omega}_{\text{robust}}^{\top}\mathbb{1} = 1$, which can be enforced ex post and substituted $\lambda_0 = \frac{\tilde{\gamma}_2}{\tilde{\gamma}_1}$ and $\tilde{\lambda}_1 = \frac{\lambda_1}{\tilde{\gamma}_1}$.

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

[24] See Lettau and Pelger (2020) for a derivation.

multipliers as follows:

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

The first order condition for the general optimization problem with Lagrange multipliers on the active set, that is, for the non-zero values of ω_i, implies

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

which in turn can be formulated as follows:

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

Note that the corresponding first order condition for Kozak, Nagel, and Santosh (2020) equals

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

which shows that the solutions coincide if we use $\hat{\mu} + \lambda_0\mathbb{1}$ instead of $\hat{\mu}$ in their problem.

Appendix A.3. General Robust Estimation Perspective (Proposition 3)

Our estimator can also be interpreted as a robust approach to the mean-variance optimization problem, when there is uncertainty about the mean and variance-covariance matrix of returns. Each type of shrinkage has a one-to-one correspondence to a specific type of uncertainty in the estimation.

We start with constructing the robust mean-variance efficient frontier. Assume that the true mean and covariance matrix lie in an uncertainty set around their sample estimates:

$$\begin{aligned} S_\Sigma &= \left\{\Sigma : \Sigma_{i,j} = \widehat{\Sigma}_{i,j} + e^\sigma_{i,j}; \quad \|e^\sigma\|_2^2 \le \kappa_\sigma \quad \Sigma \text{ is positive definite and } \kappa_\sigma \ge 0\right\} \\ S_\mu &= \{\mu : \mu_i = \hat{\mu}_i + e^\mu_i; \quad |e^\mu_i| \le \kappa_\mu \quad \text{and } \kappa_\mu \ge 0\}. \end{aligned}$$

The parameter κ_μ captures the uncertainty in the mean estimate, while κ_σ reflects the covariance estimation uncertainty. The robust version of the classical mean-variance portfolio optimization is equivalent to finding the solution under the worst case scenario as follows:

$$\min_w \max_{\Sigma\in S_\Sigma,\, \mu\in S_\mu} \frac{1}{2}\omega^\top\Sigma\omega - \gamma_1\left(\omega^\top\mu - \mu_0\right) - \gamma_2\left(\omega^\top\mathbb{1} - 1\right).$$

Concentrating out μ yields

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

where sign$(\cdot)$ is the sign function. Concentrating out Σ yields

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

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

$$\min_{w}\ \frac{1}{2}\omega^{\top}\hat{\Sigma}\omega+\lambda_2\|\omega\|_2^2+\lambda_1\|\omega\|_1-\gamma_1\left(\omega^{\top}\hat{\mu}-\mu_0\right)-\gamma_2\left(\omega^{\top}\mathbb{1}-1\right).$$

Note that the lasso shrinkage λ_1 increases with more uncertainty in the mean, while the ridge shrinkage λ_2 has a direct correspondence to the variance uncertainty. Therefore, the first stage regularized optimization results in a robust estimation of the whole mean-variance efficient frontier, which we trace out in the second step to find the tangency portfolio.

The mean shrinkage can be linked to uncertainty in the Sharpe ratio of the tangency portfolio. In order to show this, we consider the equivalent problem of an investor with mean-variance utility function and risk aversion γ. It is instructive to start with the conventional problem without any lasso or ridge shrinkage as follows:

$$\max_{\omega}\quad \omega^{\top}\hat{\mu}-\frac{\gamma}{2}\omega^{\top}\hat{\Sigma}\omega-\gamma_1\left(1-\omega^{\top}\mathbb{1}\right).$$

The well-known solution is a combination of the tangency portfolio and the minimum variance portfolio $\omega^*=\alpha_{\gamma}\hat{\omega}_{\text{naive}}+(1-\alpha_{\gamma})\hat{\omega}_{\text{var}}$ with $\alpha_{\gamma}=\frac{1}{\gamma}\mathbb{1}^{\top}\hat{\Sigma}^{-1}\hat{\mu}$. Hence, each target mean μ_0 in the Markowitz problem corresponds to a risk aversion parameter γ. Consider now the problem of mean-variance optimization with the sample mean shrunk toward its cross-sectional average $\bar{\mu}$ as follows:

$$\max_{\omega}\quad \omega^{\top}\left(\delta\hat{\mu}+(1-\delta)\bar{\mu}\mathbb{1}\right)-\frac{\gamma}{2}\omega^{\top}\hat{\Sigma}\omega-\gamma_1\left(1-\omega^{\top}\mathbb{1}\right).$$

The solution is again a convex combination of the tangency and minimum variance portfolio $\omega^*=\alpha_{\delta}\hat{\omega}_{\text{naive}}+(1-\alpha_{\delta})\hat{\omega}_{\text{var}}$, with weights $\alpha_{\delta}=\frac{\delta}{\gamma}\mathbb{1}^{\top}\hat{\Sigma}^{-1}\hat{\mu}$. Hence, mean shrinkage can equivalently be formulated as a higher degree of risk aversion. Next, we consider a robust optimization problem with uncertainty in the Sharpe ratio. The uncertainty set[25]

$$S_{\text{SR}} = \left\{ \mu : (\mu - \hat{\mu})^\top \Sigma^{-1} (\mu - \hat{\mu}) \leq \kappa_{\text{SR}}; \;\; \kappa_{\text{SR}} \geq 0 \right\}$$

is used in the worst case optimization

$$\max_{\omega} \min_{\mu \in S_{\text{SR}}} \quad \omega^\top \mu - \frac{\gamma}{2} \omega^\top \Sigma \omega - \gamma_2 \left(\kappa_{\text{SR}} - (\mu_2 - \hat{\mu})^\top \Sigma^{-1} (\mu_2 - \hat{\mu}) \right) - \gamma_1 \left(1 - \omega^\top \mathbb{1} \right),$$

yielding the solution $\omega^* = \alpha_{\kappa_{\text{SR}}} \hat{\omega}_{\text{naive}} + (1 - \alpha_{\kappa_{\text{SR}}}) \hat{\omega}_{\text{var}}$ with $\alpha_{\kappa_{\text{SR}}} = \frac{1}{\frac{\kappa_{\text{SR}}}{\sigma_p} + \gamma} \mathbb{1}^\top \hat{\Sigma}^{-1} \hat{\mu}$. Note that $\sigma_p = \sigma_p(\omega^*) = \sigma_p(\kappa_{SR}, \gamma)$ is the volatility of the optimal portfolio and generally can be found as a unique solution to a higher order polynomial equation.[26] However, all these problems can easily be seen as tracing the same efficient frontier, with the one-to-one mapping between the coefficient of risk aversion, the degree of shrinkage to the mean, and the elliptical uncertainty on the expected returns, with the latter two equivalent to a higher value of the effective risk aversion in the standard Markowitz optimization. Hence, robust estimation with uncertainty in the Sharpe ratio can be motivated as a conventional mean-variance optimization with shrinkage in the mean or a higher degree of risk aversion.

Including the variance shrinkage λ_2 is straightforward. We define the uncertainty set as

$$S_{\text{SR},\Sigma} = \left\{ \mu, \Sigma : \quad \begin{array}{l} \Sigma = \hat{\Sigma} + dI_N; \;\; d \leq \kappa_\sigma; \;\; \Sigma \text{ is positive definite and } \kappa_\sigma \geq 0 \\ (\mu - \hat{\mu})^\top \Sigma^{-1} (\mu - \hat{\mu}) \leq \kappa_{\text{SR}}; \;\; \kappa_{\text{SR}} \geq 0 \end{array} \right\},$$

and the solution simply replaces $\hat{\Sigma}$ with $\hat{\Sigma} + \lambda_2 I_N$ (that is, we use $\hat{\omega}_{\tan,\lambda_2}$, $\hat{\omega}_{\text{var},\lambda_2}$ and the corresponding $\alpha_{\kappa_{\text{SR}}}$) and identifies $\kappa_\sigma = \lambda_2$ in the equivalent shrinkage formulation.[27] The general case includes uncertainty in the mean resulting in the uncertainty set

$$S_{\text{SR},\Sigma,\mu} = \left\{ \mu, \Sigma : \quad \begin{array}{l} \Sigma = \hat{\Sigma} + dI_N; \;\; d \leq \kappa_\sigma; \;\; \Sigma \text{ is positive definite and } \kappa_\sigma \geq 0 \\ \mu = \mu_1 + \mu_2; \;\; |\mu_1 - \hat{\mu}| \leq \kappa_\mu; \;\; (\mu_2 - \hat{\mu})^\top \Sigma^{-1} (\mu_2 - \hat{\mu}) \leq \kappa_{\text{SR}}; \;\; \kappa_{\text{SR}}, \kappa_\mu \geq 0 \end{array} \right\}$$

Adding the lasso term, then, is a simple generalization of the previous results, which, however, do not have a closed-form solution anymore. Note that the objective function becomes additive in μ_1 and μ_2 and after concentrating out the parameters, it is equivalent to the mean-variance utility maximization with shrunk mean, lasso, and ridge penalty. In particular, κ_{SR} can be directly linked to the mean shrinkage λ_0, κ_μ to the lasso penalty λ_2 and κ_σ to the ridge penalty λ_2.

We now provide the detailed derivation for the link between mean shrinkage and the uncertainty

[25]The uncertainty set puts constraints on the mean return, which can be mapped into Sharpe ratio constraints. Assume $\mu^\top \hat{\Sigma}^{-1}(\hat{\mu} - \mu) \geq 0$, that is, the sample mean overestimates the mean. In this case the constraint is equivalent to $\hat{\mu}^\top \hat{\Sigma}^{-1} \hat{\mu} - \mu^\top \Sigma^{-1} \mu \leq \kappa_{\text{SR}}$, which bounds the estimation error of the sample Sharpe ratio from above. In the case of $\hat{\mu}^\top \hat{\Sigma}^{-1}(\mu - \hat{\mu}) > 0$, that is, the sample mean underestimates the mean, the bound is equivalent to $\mu^\top \Sigma^{-1} \mu - \hat{\mu}^\top \hat{\Sigma}^{-1} \hat{\mu} \leq \kappa_{\text{SR}}$. This is a lower bound on how much the Sharpe ratio is underestimated.

[26]See Garlappi, Uppal, and Wang (2007).

[27]Note that in this formulation we use a simpler uncertainty set than in the robust efficient frontier construction in order to obtain the closed-form solution.

in the Sharpe ratio of the tangency portfolio. It is instructive first to consider the case without the restrictions that would lead to a lasso penalty, to demonstrate the argument with a closed-form solution. The robust version of the classical mean-variance portfolio optimization is equivalent to finding the solution under the worst case scenario as follows:

$$\max_{\omega} \min_{\Sigma,\mu \in S_{\mathrm{SR},\Sigma}} \quad \omega^\top \mu - \frac{\gamma}{2}\omega^\top \Sigma \omega - \gamma_2 \left(\kappa_{\mathrm{SR}} - (\mu_2 - \hat{\mu})^\top \Sigma^{-1} (\mu_2 - \hat{\mu})\right) - \gamma_1 \left(1 - \omega^\top \mathbb{1}\right).$$

First, note that the objective function is monotonic in κ_σ; hence $\tilde{\Sigma}^* = \hat{\Sigma} + \kappa_\sigma I_N$.[28] The FOC w.r.t. $\tilde{\mu}$ implies $\tilde{\mu}^* = \hat{\mu} - \frac{1}{2\tilde{\gamma}_1}\tilde{\Sigma}^* w$, which, after substituting into the loss function and solving for $\tilde{\gamma}_1$, yields

$$\frac{\gamma}{2\tilde{\gamma}_1^2}\omega^\top \tilde{\Sigma}^* \omega - \kappa = 0 \implies \tilde{\gamma}_1 = \sqrt{\frac{\gamma \omega^\top \tilde{\Sigma}^* \omega}{2\kappa_{\mathrm{SR}}}},$$

and the whole problem becomes

$$\max_{w} \quad \omega^\top \hat{\mu} - \frac{\gamma}{2}\omega^\top \tilde{\Sigma}^* \omega \left[1 - \frac{2\sqrt{\kappa_{\mathrm{SR}}}}{\gamma\sqrt{\omega^\top \tilde{\Sigma} \omega}}\right] - \tilde{\gamma}_2 \left(1 - \omega^\top \mathbb{1}\right). \tag{A1}$$

The solution to this optimization satisfies the following equations:

$$\omega^* = \left(\frac{\sigma_p}{\sqrt{\kappa_{\mathrm{SR}}} + \gamma \sigma_p}\right) \tilde{\Sigma}^{*-1} \left(\hat{\mu} - \tilde{\gamma}_2 \mathbb{1}\right),$$
$$w^{*\top}\mathbb{1} = 1 \implies \tilde{\gamma}_2^* = \frac{1}{\mathbb{1}^\top \Sigma^{*-1} \mathbb{1}} \left(\hat{\mu}^\top \Sigma^{*-1} \mathbb{1} - \frac{\sqrt{\kappa_{\mathrm{SR}}} + \gamma \sigma_p}{\sigma_p}\right),$$

where $\sigma_p \equiv \sqrt{\omega^{*\top} \tilde{\Sigma}^* \omega^*}$ is the volatility of the optimal portfolio. This implies that the optimal asset allocation is described by the following equation:

$$\omega^* = \frac{\sigma_p}{\sqrt{\kappa_{\mathrm{SR}}} + \gamma \sigma_p} \tilde{\Sigma}^{*-1} \left(\hat{\mu} - \frac{1}{\mathbb{1}^\top \Sigma^{*-1} \mathbb{1}} \left(\hat{\mu}^\top \Sigma^{*-1} \mathbb{1} - \frac{\sqrt{\kappa_{\mathrm{SR}}} + \gamma \sigma_p}{\sigma_p}\right)\right). \tag{A2}$$

Note that σ_p can be found as a unique positive solution to a higher order polynomial equation (see, e.g., Garlappi, Uppal, and Wang (2007)).

It is interesting to note that there is a one-to-one mapping between the optimal asset allocation, solving problem (A3), and a standard mean-variance optimization problem, where expected returns are shrunk toward their cross-sectional mean, that is, there exists $\alpha \in [0, 1]$ s.t. the portfolios choice

[28] To see this, note that for a positive definite quadratic form $x^\top (\hat{\Sigma} + dI_N)^{-1} x$,

$$\frac{\partial (x^\top (\hat{\Sigma} + dI_N)^{-1} x)}{\partial d} = -Tr\left((\hat{\Sigma} + dI_N) x x^\top (\hat{\Sigma} + dI_N) \frac{\partial ((\hat{\Sigma} + dI_N)^{-1})}{\partial d}\right)$$

$$= Tr\left((\hat{\Sigma} + dI_N) x x^\top (\hat{\Sigma} + dI_N) \left[-(\hat{\Sigma} + dI_N)^{-1} I_N (\hat{\Sigma} + dI_N)^{-1}\right]\right) = Tr\left((\hat{\Sigma} + dI_N) x x^\top (\hat{\Sigma} + dI_N)^{-1}\right) = Tr\left(x x^\top\right) = \sum_{i=1}^{n} x_i^2 \geq 0.$$

Hence, the optimal solution is to set $d = \kappa_{\mathrm{SR}}$.

in eq. (A2) coincides with the solution to the following problem:

$$\max_{w} \quad \omega^\top ((1-\alpha)\hat{\mu} + \alpha\bar{\mu}) - \frac{\gamma}{2}\omega^\top \tilde{\Sigma}^* \omega - \tilde{\gamma}_2 \left(1 - \omega^\top \mathbb{1}\right). \tag{A3}$$

This is a standard portfolio optimization problem, with the optimal solution satisfying the following first order conditions:

$$\omega^*_\alpha = \frac{1-\alpha}{\gamma}\tilde{\Sigma}^{*-1}\left(\hat{\mu} - \frac{\tilde{\gamma}_2 - \alpha\bar{\mu}}{1-\alpha}\mathbb{1}\right)$$

$$w^{*\top}_\alpha \mathbb{1} = 1 \implies \frac{\tilde{\gamma}_2 - \alpha\bar{\mu}}{1-\alpha} = \frac{1}{\mathbb{1}^\top \Sigma^{*-1}\mathbb{1}}\left(\hat{\mu}^\top \Sigma^{*-1}\mathbb{1} - \frac{\gamma}{1-\alpha}\right).$$

Hence,

$$w^*_\alpha = \frac{1-\alpha}{\gamma}\left(\hat{\mu} - \frac{1}{\mathbb{1}^\top \Sigma^{*-1}\mathbb{1}}\left(\hat{\mu}^\top \Sigma^{*-1}\mathbb{1} - \frac{\gamma}{1-\alpha}\right)\right). \tag{A4}$$

Finally, note that for each optimal portfolio w^* in eq. (A2), with the known volatility $\sigma_p(w^*)$, there exists a value of $\alpha^* \in [0,1]$:

$$\alpha^* = \frac{\sqrt{\kappa_{\text{SR}}}}{\sqrt{\kappa_{\text{SR}}} + \gamma\sigma_p}$$

such that $w^* = w^*_\alpha(\alpha = \alpha^*)$. This equivalence between two optimization problems persists in the corner solutions as well: as $\kappa_{\text{SR}} \to 0$, $\tilde{\mu} \to \hat{\mu}$, $\alpha \to 0$, and we're back to the standard mean-variance optimization that relies on the sample estimates of the mean. On the other hand, as $\kappa_{\text{SR}} \to \infty$, $\tilde{\mu} \to \bar{\mu}$, and $\alpha \to 1$, which is equivalent to the minimum variance portfolio, and full shrinkage of expected returns towards their cross-sectional average.

Adding the lasso term then is a simple generalization of the previous results, which, however, do not have a closed-form solution anymore. Consider the uncertainty sets for the means and variances of returns $S_{\text{SR},\Sigma,\mu}$. Note that the objective function becomes additive in μ_1 and μ_2, and after concentrating the parameters out, is equivalent to the following optimization:

$$\max_{w} \quad \omega^\top (\hat{\mu} - \bar{\psi}\text{sign}(\omega)) - \frac{\gamma}{2}\omega^\top \tilde{\Sigma}^* \omega \left[1 - \frac{2\sqrt{\kappa_{\text{SR}}}}{\gamma\sqrt{\omega^\top \tilde{\Sigma}^* \omega}}\right] - \tilde{\gamma}_2\left(1 - \omega^\top \mathbb{1}\right)$$

$$= \max_{w} \quad \omega^\top \hat{\mu} - \frac{\gamma}{2}\omega^\top \hat{\Sigma}\omega \left[1 - \frac{2\sqrt{\kappa_{\text{SR}}}}{\gamma\sqrt{\omega^\top \hat{\Sigma}\omega}}\right] - \tilde{\gamma}_2\left(1 - \omega^\top \mathbb{1}\right) - \lambda_1 \sum_{i=1}^n |\omega_i| - \lambda_2 \sum_{i=1}^n \omega_i^2$$

for certain values of $\lambda_1 > 0$ and $\lambda_2 > 0$. The equivalence between the solution to this optimization and the standard mean-variance approach with elastic net (that is, both lasso and ridge penalties) is similarly established by examining the first order conditions for the active set of $\{w_i : w_i \neq 0\}$.

Appendix A.4. AP-Pruning Implementation

The return of each tree portfolio is multiplied by $\frac{1}{\sqrt{2^d_i}}$ before running the mean-variance optimization, where d_i denotes the depth of a tree portfolio. We calculate the covariance matrix $\hat{\Sigma}$ and mean vector $\hat{\mu}$ of the reweighted assets. The cross-sectional mean is denoted by $\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mu}_i$. We obtain the SDF weights by running an elastic net regression of $\tilde{\Sigma}$ on $\tilde{\mu}$, where

$$\tilde{\Sigma} = \hat{\Sigma}^{1/2} \qquad \tilde{\mu} = \hat{\Sigma}^{-1/2}\left(\hat{\mu} + \lambda_0 \bar{\mu}\mathbb{1}\right)$$

and

$$\hat{\omega}_{\text{robust}} = \arg\min_{\omega} \left(\tilde{\mu} - \tilde{\Sigma}\omega\right)^{\top}\left(\tilde{\mu} - \tilde{\Sigma}\omega\right) + \lambda_2\|\omega\|_2^2 + \lambda_1\|\omega\|_1.$$

The elastic net implementation actually simplifies to a lasso regression. Define

$$X = \begin{pmatrix} \tilde{\Sigma} \\ \sqrt{\lambda_2} I_N \end{pmatrix} \qquad y = \begin{pmatrix} \tilde{\mu} \\ 0_N \end{pmatrix}.$$

Then the problem is equivalent to the lasso regression

$$(y - X\omega)^{\top}(y - X\omega) + \lambda_1\|\omega\|_1.$$

The algorithm requires the square root and inverse of a large dimensional sample covariance matrix. We suggest the following steps. First, we obtain an eigenvalue-eigenvector decomposition of the sample covariance matrix

$$\hat{\Sigma} = V^{\top} D V,$$

where V are the eigenvectors and D is a diagonal matrix of sorted eigenvalues. We keep only the $\delta = \min(T, M)$ non-zero eigenvalues where M is the number of tree portfolios and T is the length of the training sample. Then

$$\begin{aligned} \tilde{V} &= V(:, 1:\delta) & \tilde{D} &= D(1:\delta, 1:\delta) \\ \tilde{\Sigma} &= \tilde{V}^{\top}\tilde{D}^{1/2}\tilde{V} & \tilde{\mu} &= \tilde{V}^{\top}\tilde{D}^{-1/2}\tilde{V}\left(\hat{\mu} + \lambda_0\bar{\mu}\mathbb{1}\right). \end{aligned}$$

We implement the lasso regression with the LARS (Least Angle Regression) algorithm.

We have also considered to correct for over-shrinkage induced by double-shrinkage in the mean through lasso and the mean shrinkage. Our modification is a convex combination of the sample mean and cross-sectional average, that is, we use

$$\tilde{\mu} = \hat{\Sigma}^{-1/2}\left(\alpha_0\hat{\mu} + (1 - \alpha_0)\bar{\mu}\mathbb{1}\right)$$

in the lasso regression. The results are very close to $\tilde{\mu} = \hat{\Sigma}^{-1/2}\left(\hat{\mu} + \lambda_0\bar{\mu}\mathbb{1}\right)$, which is our benchmark case.

Appendix B. Simulation

This section includes the details for Section III and additional simulation results. Suppose there is a single factor that drives expected returns on a cross-section of stocks, with loadings being a function of two stock-specific characteristics as follows:

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

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

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

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

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

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

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

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

where Corr-Uniform$[0, 1, \rho]$ denotes a pair of uniformly distributed random variables that have correlation ρ and marginal densities $U[0, 1]$. To model the impact of characteristic dependence on the portfolio structure, we consider three cases: $\rho \in \{0, 0.5, 0.9\}$. The risk factor has $\mu_F = 1, \sigma_F = 2$, implying a Sharpe ratio of 0.5, and the idiosyncratic noise has a standard deviation of $\sigma_e = 8$ to mimic a noisy and volatile process for stock returns. Since $\mathbb{E}(f_t) = 1$, the stock's beta is exactly equal to its expected return. Finally, we set cross-sectional dimensions to $N = 800, \quad T = 600$. Figure E.4 illustrates the resulting empirical distribution of stock returns in the characteristic space and their true loadings on the SDF, which is equal to the expected returns up to a proportionality constant.

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

Figure E.5 presents the scaled version of the estimated SDF betas (and hence, conditional expected returns) for different basis assets for the nonlinear model.[29] Clearly, averaging betas across conditional basis asset allows tracking the underlying SDF loadings in the characteristic space. As in the linear case, the correlation between characteristics matters, as it leads to an unbalanced composition of basis assets for conventional sorting. Table D.6 reports the Sharpe ratios of the SDFs spanned by different assets. The AP-Tree SDF has clearly higher out-of-sample Sharpe ratios compared to the DS approach. Note that for this particular simulation setup, obtaining just a high SR is too low a bar. Indeed, as we mentioned earlier, in this particular setting the data is coming from a very simple one-factor model, which drives both the time-series of returns and their cross-sectional spread. As a result, almost any well-diversified stock portfolio (equally weighted, or long-short in a given characteristic, or a simply first principal component) would also be loading on the same latent factor, and, as a result, asymptotically recover both the highest SR and an accurate SDF projection on returns. The true test of the basis assets lies in their ability to span expected returns in the characteristic space and match its empirical counterpart, and this is where AP-Trees really shine.

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

Appendix C. List of the Firm-Specific Characteristics

Table C.1 Characteristic variables as listed on 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	Last month's volume (VOL) over shares outstanding (SHROUT)	Datar, Naik, and Radcliffe (1998)
OP	Operating profitability	Annual revenues (REVT) minus cost of goods sold (COGS), interest expense (TIE), and selling, general, and administrative expenses (XSGA) divided by book equity (defined in BEME)	Fama and French (2015)
r12_2	Momentum	To be included in a portfolio for month t (formed at the end of month t-1), a stock must have a price for the end of month t-13 and a good return for t-2. In addition, any missing returns from t-12 to t-3 must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of month t-1.	Jegadeesh and Titman (1993)
ST_Rev	Short-term reversal	Prior month return	Jegadeesh (1990)

Characteristic variables as listed on Kenneth French's website with acronym, definition, and academic reference.

Appendix D. Additional Tables

Appendix D.1. Additional Empirical Results for All Cross-Sections

Table D.1 Summary statistics of the cross-sections, built from AP-Tree

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 II.C. 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 that includes the market and 10 long-short portfolios, based on characteristics. The t-statistics are in brackets.

Table D.2 Selection of hyperparameters

Notation	Hyperparameters	Candidates
λ_0	Mean shrinkage	$\{0, 0.05, 0.1, ..., 0.9\}$
K	Number of portfolios	$\{10, 11, ..., 40\}$
λ_2	Variance shrinkage	$\{0.1^5, 0.1^{5.25}, ..., 0.1^8\}$

The table presents the list of hyperparameters for pruning AP-Trees, and the actual range of their values used for simulation and empirics. The mean shrinkage is scaled by the cross-sectional average, $\hat{\mu} + \lambda_0 \bar{\mu} \mathbb{1}$.

Appendix D.2. Results for Size, Value, and Idiosyncratic Volatility

Table D.3 Cross-sections based on size, value, and idiosyncratic volatility

		Type of the cross-section			
		AP-Trees (10)	AP-Trees (40)	TS (32)	TS(64)
SDF SR		0.54	0.55	0.28	0.26
α	FF3	0.95	0.94	0.73	0.94
		[10.22]	[11.07]	[4.95]	[5.50]
	FF5	0.91	0.91	0.75	0.96
		[9.50]	[10.45]	[5.01]	[5.50]
	XSF	0.82	0.82	0.33	0.65
		[9.17]	[10.05]	[3.01]	[4.08]
	FF11	0.89	0.88	0.42	0.83
		[9.43]	[10.27]	[3.80]	[5.08]
$XS\ R^2_{adj}$	FF3	-41.0%	10.0%	76.0%	74.0%
	FF5	-93.0%	16.0%	74.0%	73.0%
	XSF	-22.0%	45.0%	88.0%	85.0%
	FF11	–	24.0%	83.0%	79.0%

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

Table D.4 Size, value, and idiosyncratic volatility: 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		0.30	0.32	0.18	0.19	0.19	0.14	0.10	0.06
α	FF3	0.70	0.74	0.32	0.70	0.58	0.82	0.59	0.66
		[5.65]	[6.34]	[1.22]	[2.79]	[3.24]	[3.15]	[1.32]	[0.80]
	FF5	0.64	0.70	0.18	0.62	0.30	0.35	-0.02	-0.42
		[5.20]	[6.16]	[0.68]	[2.44]	[1.81]	[1.55]	[-0.05]	[-0.54]
	XSF	0.32	0.41	-0.24	0.04	0.17	0.26	-0.14	-0.56
		[4.06]	[4.94]	[-1.07]	[0.20]	[1.70]	[1.57]	[-0.39]	[-0.78]
	FF11	0.38	0.48	-0.26	0.21	0.12	0.08	-0.38	-1.08
		[4.85]	[5.83]	[-1.08]	[1.03]	[1.17]	[0.54]	[-1.04]	[-1.50]
$XS\ R^2_{adj}$	FF3	-50.0%	31.0%	64.0%	78.0%	79.0%	88.0%	87.0%	85.0%
	FF5	-85.0%	43.0%	63.0%	79.0%	94.0%	96.0%	94.0%	91.0%
	XSF	55.0%	83.0%	78.0%	90.0%	98.0%	98.0%	95.0%	92.0%
	FF11	–	77.0%	45.0%	86.0%	–	97.0%	93.0%	91.0%

The table presents the aggregate properties of the cross-sections based on size, value and idiosyncratic volatility, 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 United States. For each of the cross-sections, the table reports its monthly Sharpe ratio on the test sample, along with the alpha of the SDF spanned by the corresponding basis portfolios and cross-sectional R^2_{adj} obtained with conventional factor models for each of the portfolio sets. Alphas and R^2_{adj} 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 that includes the market portfolio, along with the 10 long-short portfolios, based on the cross-sectional characteristics.

Table D.5 Portfolio turnover for size, investment and idiosyncratic volatility

		AP-Trees	XSF	Invest	IdioVol
SR	Monthly	0.64	0.09	0.08	0.03
	Yearly	0.28	0.06	-0.10	-0.01
Turnover+	Monthly	0.82	0.36	0.35	0.66
	Yearly	0.15	0.11	0.12	0.12
Turnover-	Monthly	1.00	0.31	0.28	0.65
	Yearly	0.11	0.09	0.09	0.11

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

Appendix D.3. Additional Simulation Results

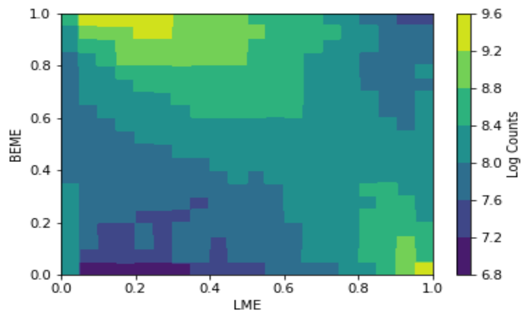
Table D.6 SR and SDF recovery in a simulated environment

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

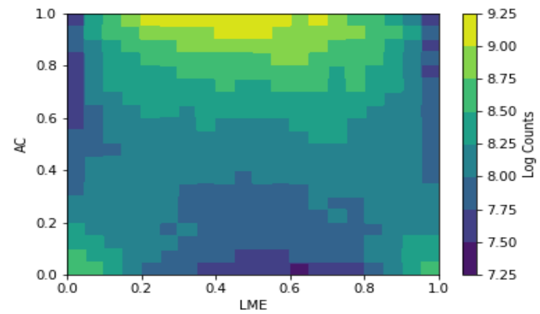
Sharpe ratio (SR) of the SDF factor for the tree portfolios and double-sorted 16 portfolios under two different loading functions and three different correlations of the two characteristic variables. The data is generated with an SDF factor with Sharpe ratio $SR = 0.5$. $N = 800, T = 600$, $T_{tr} =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

Appendix E. Additional Figures

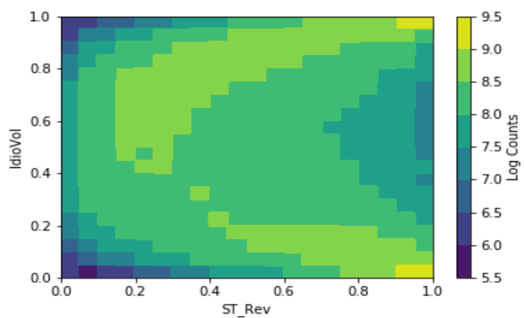
Appendix E.1. Additional stylized empirical facts



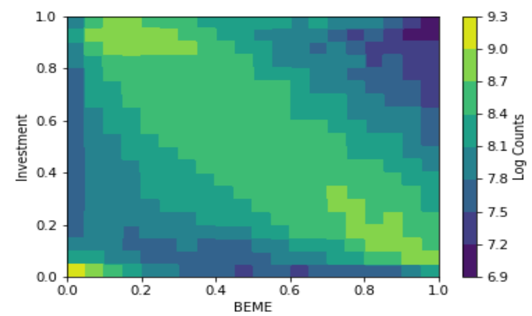
(a) Size and book-to-market



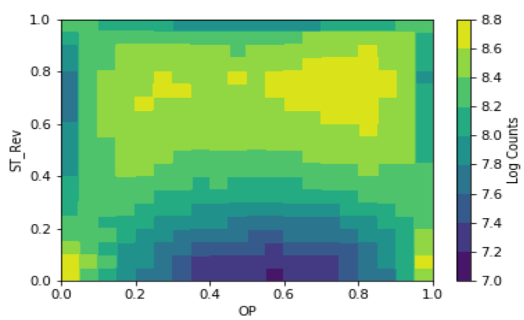
(b) Size and accruals



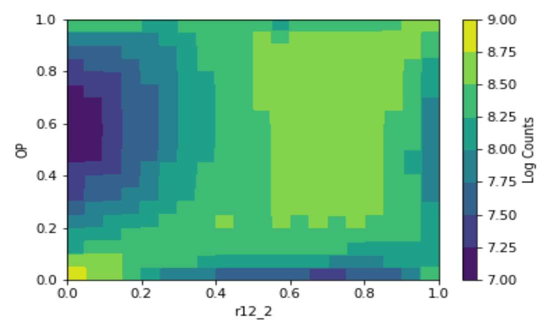
(c) Short-term reversal and idiosyncratic vol.



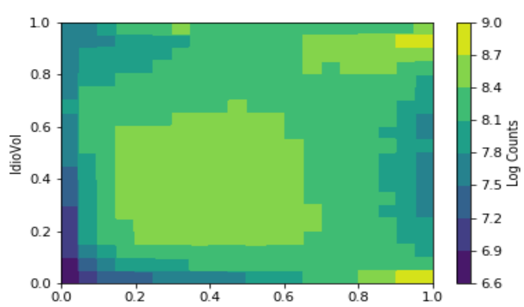
(d) Book-to-market and investment



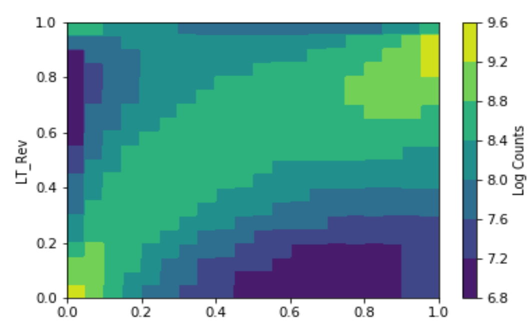
(e) Operating profitability and short-term reversal



(f) Momentum and operating profitability



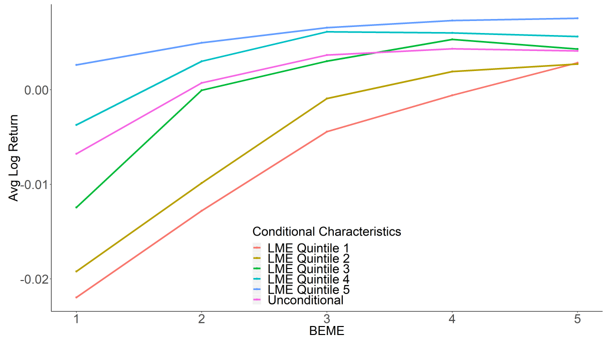
(g) Book-to-market and idiosyncratic volatility



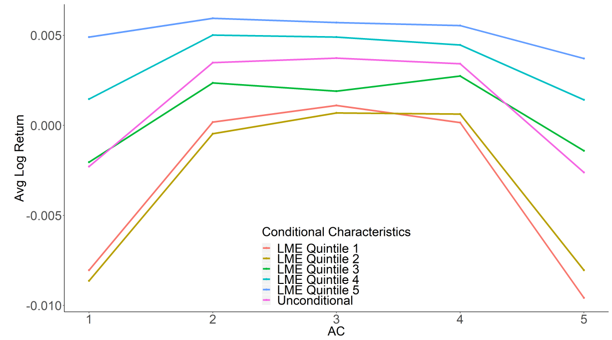
(h) Operating profitability and long-term reversal

Figure E.1. Joint empirical distribution of characteristics in the cross-section of stocks

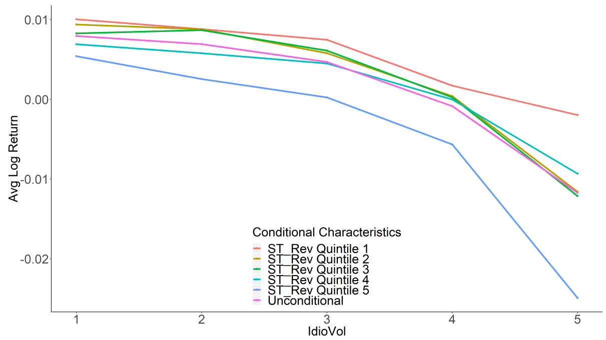
The graphs represent the pairwise empirical cross-sectional distribution of the characteristic quantiles across the stocks. The frequency is computed on a quantile 20×20 grid.



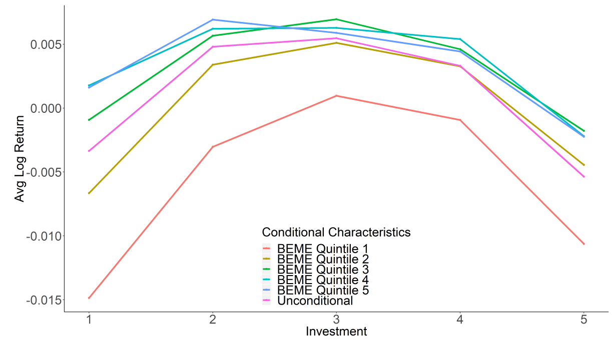
(a) Size and book-to-market



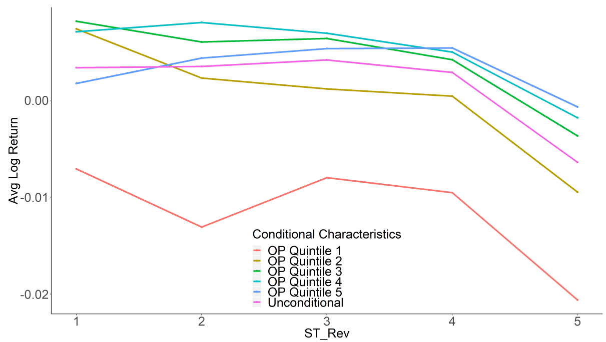
(b) Size and accruals



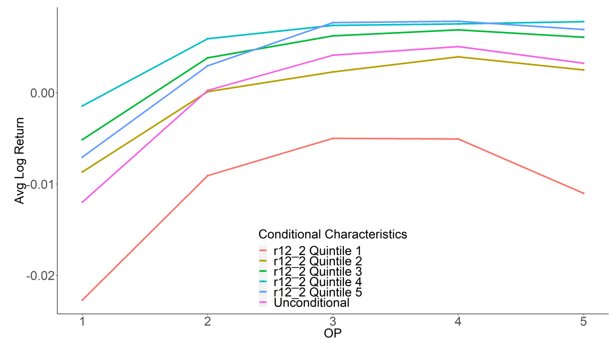
(c) Short-term reversal and idiosyncratic vol.



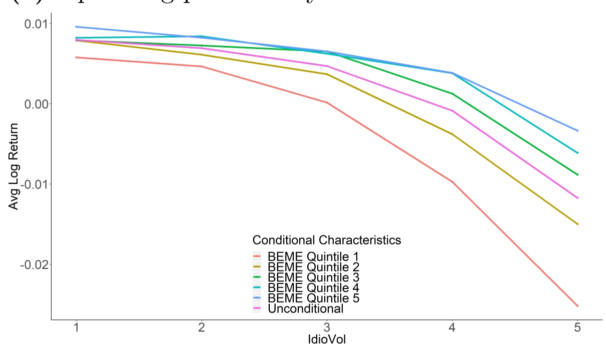
(d) Book-to-market and investment



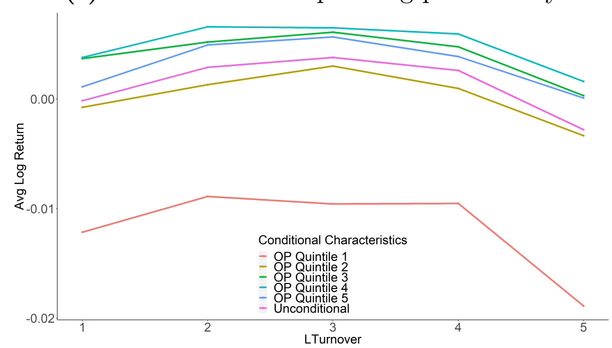
(e) Operating profitability and short-term reversal



(f) Momentum and operating profitability



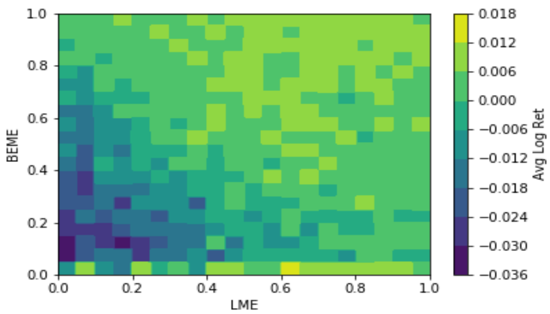
(g) Book-to-market and idyosyncractic volatility



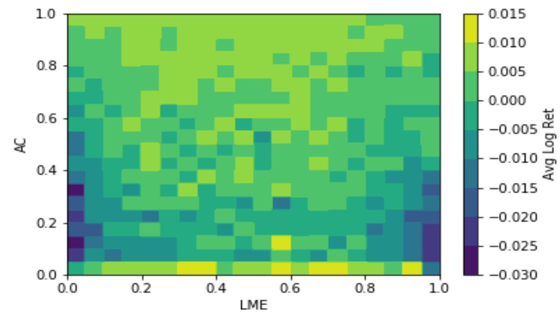
(h) Operating profitability and long-term reversal

Figure E.2. Conditional and unconditional characteristic impact

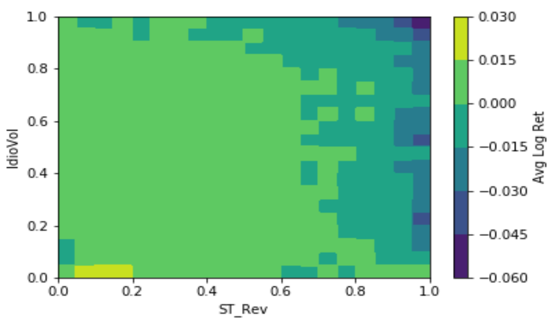
The graphs represent the impact of a characteristic quintile on expected returns unconditionally and conditionally on the second characteristic. For example, Panel (a) describes the average returns on stocks sorted by their idyosyncractic volatility (quintiles 1–5), both unconditionally and conditionally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



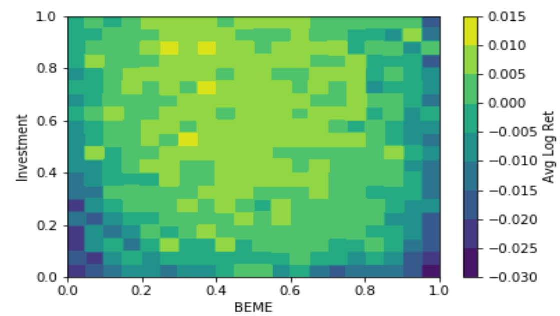
(a) Size and book-to-market



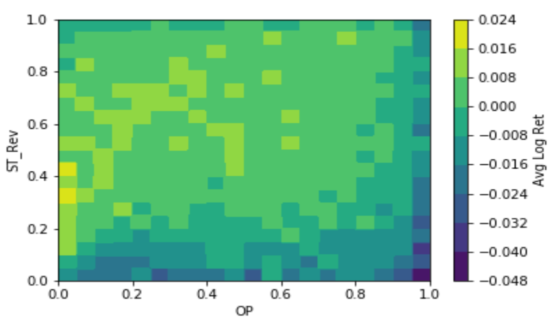
(b) Size and accruals



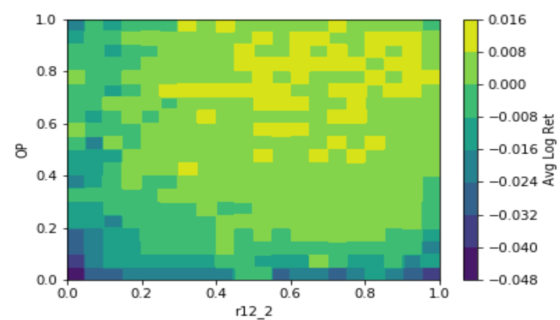
(c) Short-term reversal and idiosyncratic vol.



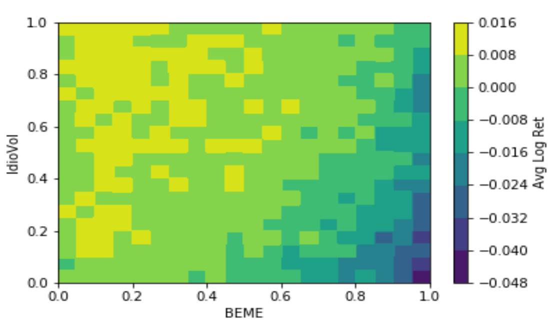
(d) Book-to-market and investment



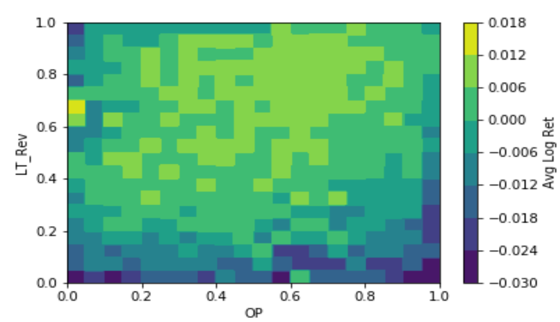
(e) Operating profitability and short-term reversal



(f) Momentum and operating profitability



(g) Book-to-market and idiosyncratic volatility

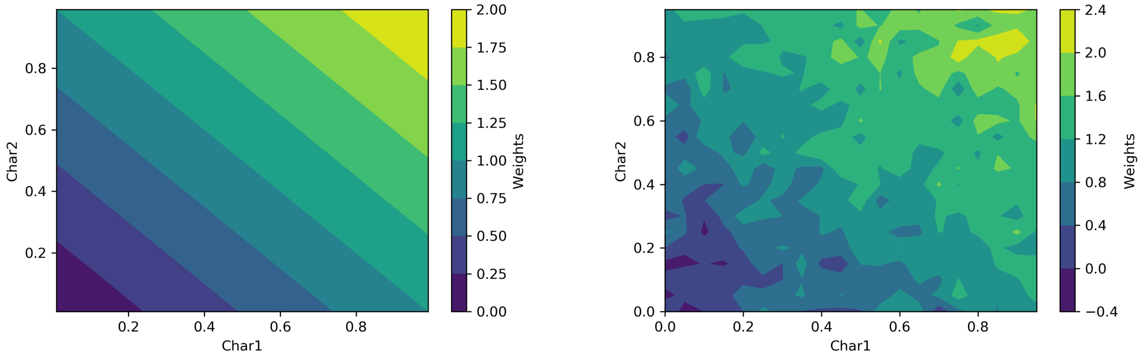


(h) Operating profitability and long-term reversal

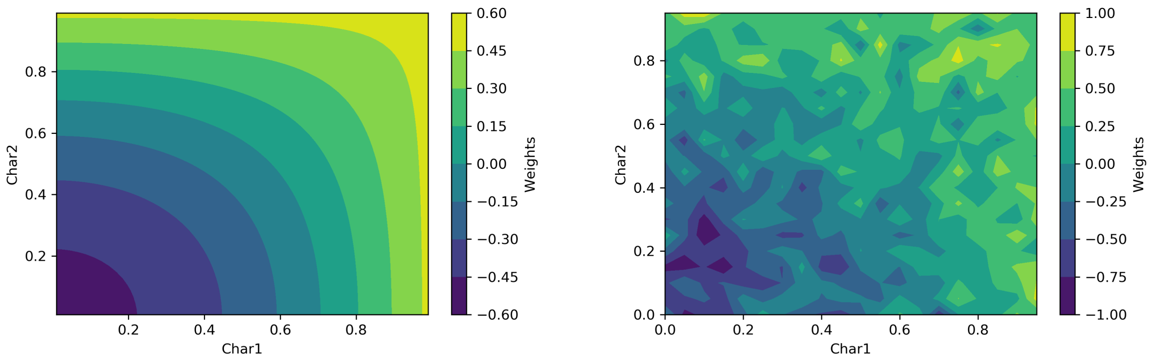
Figure E.3. Expected return as a function of stock characteristics

The graphs represents 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.

Appendix E.2. Additional Simulation Results



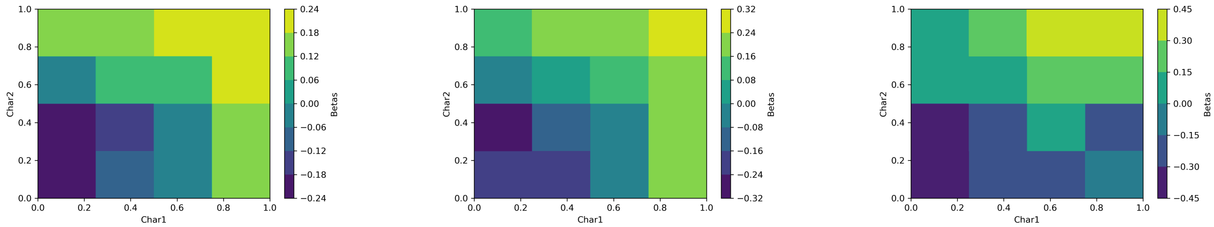
(a) An additively linear model



(b) A nonlinear model

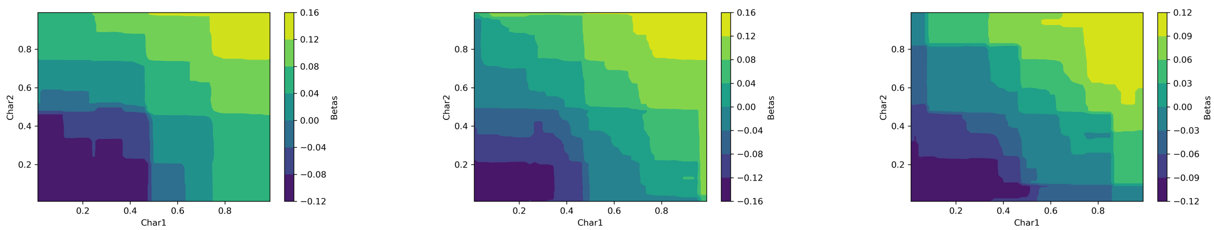
Figure E.4. True factor loadings and sample average returns as a function of characteristics

Two simulation setups for the loadings β as a function of two characteristics. The figure shows the factor loadings β and the average returns for a cross-section of stocks as a function of the characteristics. The left plots are based on the linear model $\beta_{t,i} = C_{t,i}^{(1)} + C_{t,i}^{(2)}$, and the right ones on the nonlinear model $\beta_{t,i} = \frac{1}{2} - [1 - (C_{t,i}^{(1)})^2] \cdot [1 - (C_{t,i}^{(2)})^2]$.



Correlation = 0 | Correlation = 0.5 | Correlation = 0.9

Panel A: Double-sorted portfolios



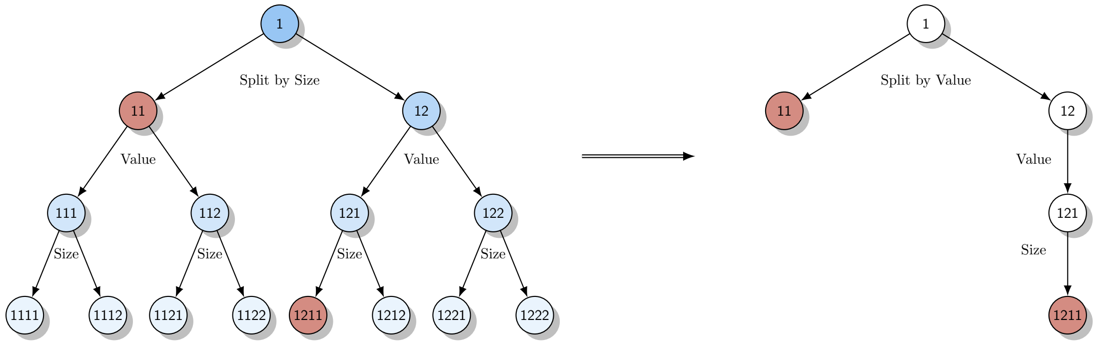
(a) Correlation = 0 | **(b)** Correlation = 0.5 | **(c)** Correlation = 0.9

Panel B: AP-Trees

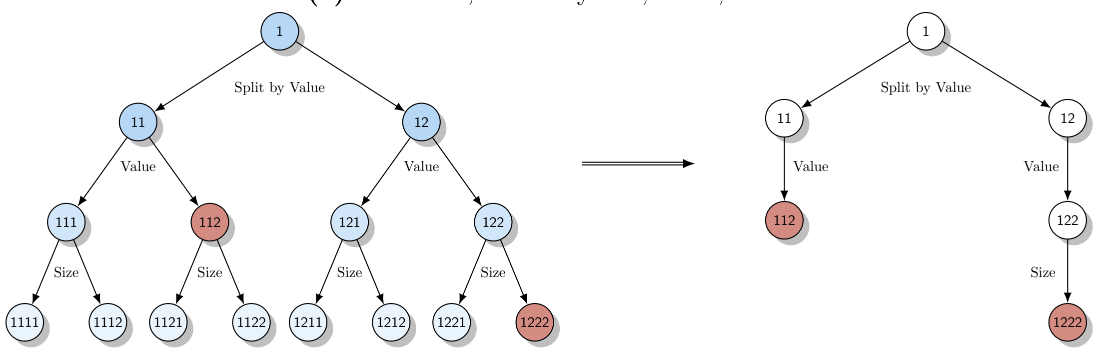
Figure E.5. Estimated SDF loadings as a function of characteristics in the nonlinear model

Estimated loading function β as a function of the two characteristics in the nonlinear model for different levels of correlation in the characteristics. The top plots show the loadings estimated with double-sorted portfolios and the bottom plots the corresponding results with pruned AP-Trees.

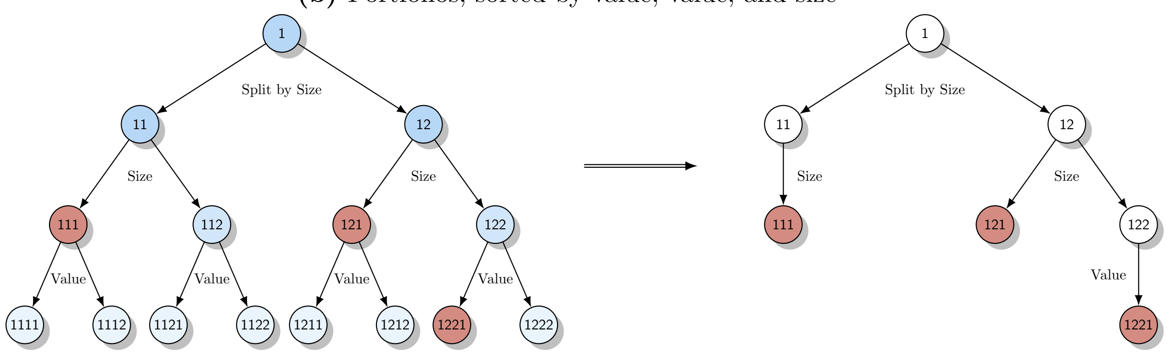
Appendix E.3. Pruning Illustration



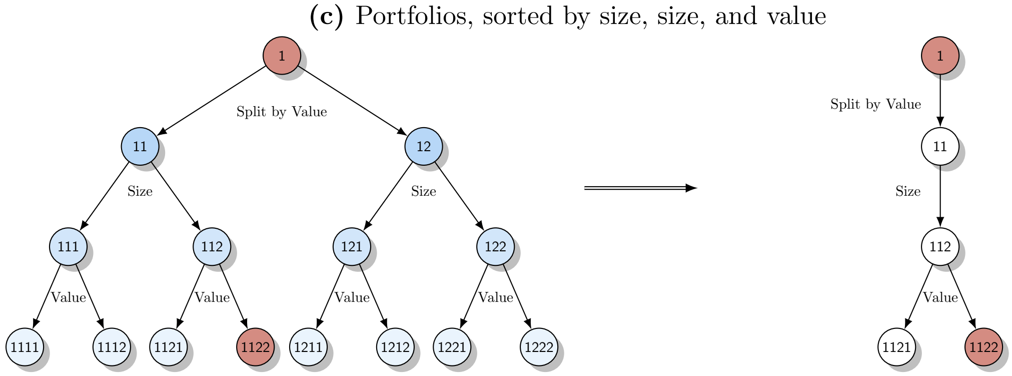
(a) Portfolios, sorted by size, value, and size



(b) Portfolios, sorted by value, value, and size



(c) Portfolios, sorted by size, size, and value

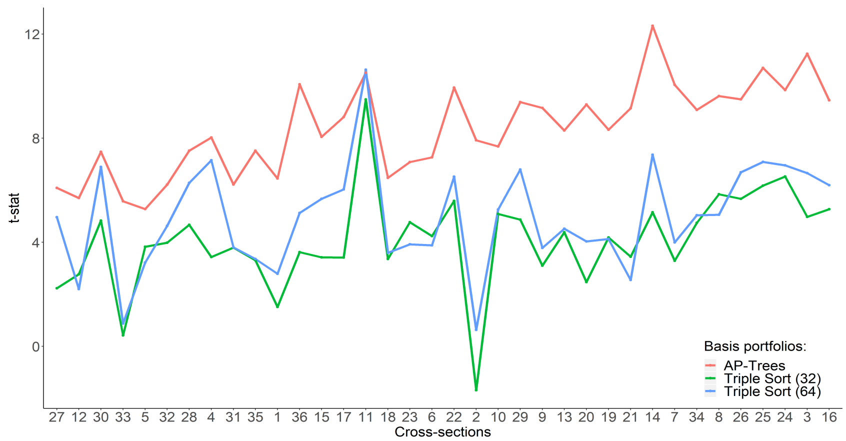


(d) Portfolios, sorted by value, size, and value

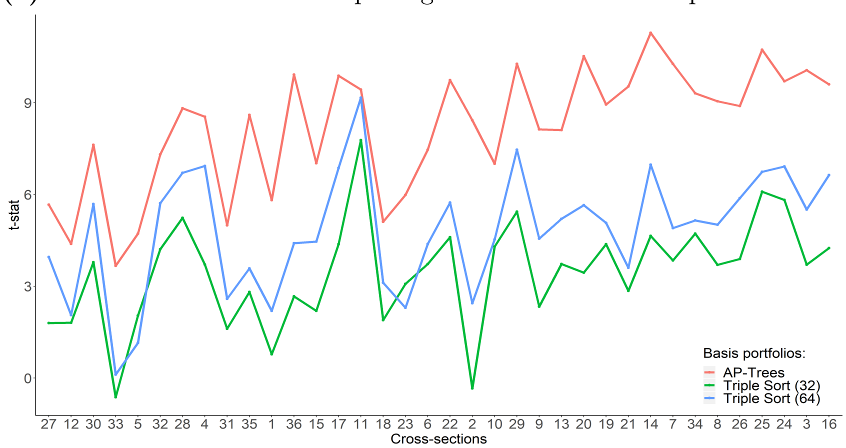
Figure E.6. Portfolio selection: Original and pruned trees

The figure shows sample trees, original and pruned for portfolios of depth 3 constructed based on size and book-to-market as the only characteristics. The fully pruned set of portfolios is based on eight trees, with Panels a) – d) illustrating potential outcomes.

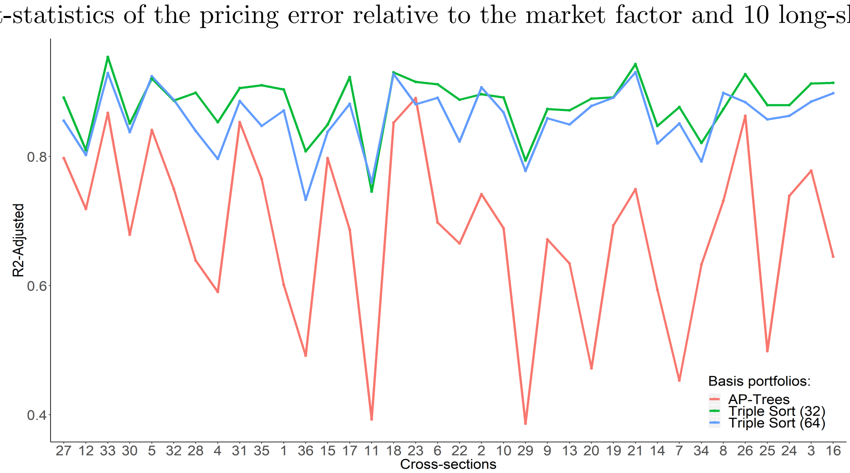
Appendix E.4. Additional Empirical Results for All Cross-Sections



(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 factor and 10 long-short portfolios



(c) $XS\text{-}R^2$: pricing portfolios with XS-specific factors

Figure E.7. Pricing errors and $XS\text{-}R^2$ 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 $XS\text{-}R^2$. 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).

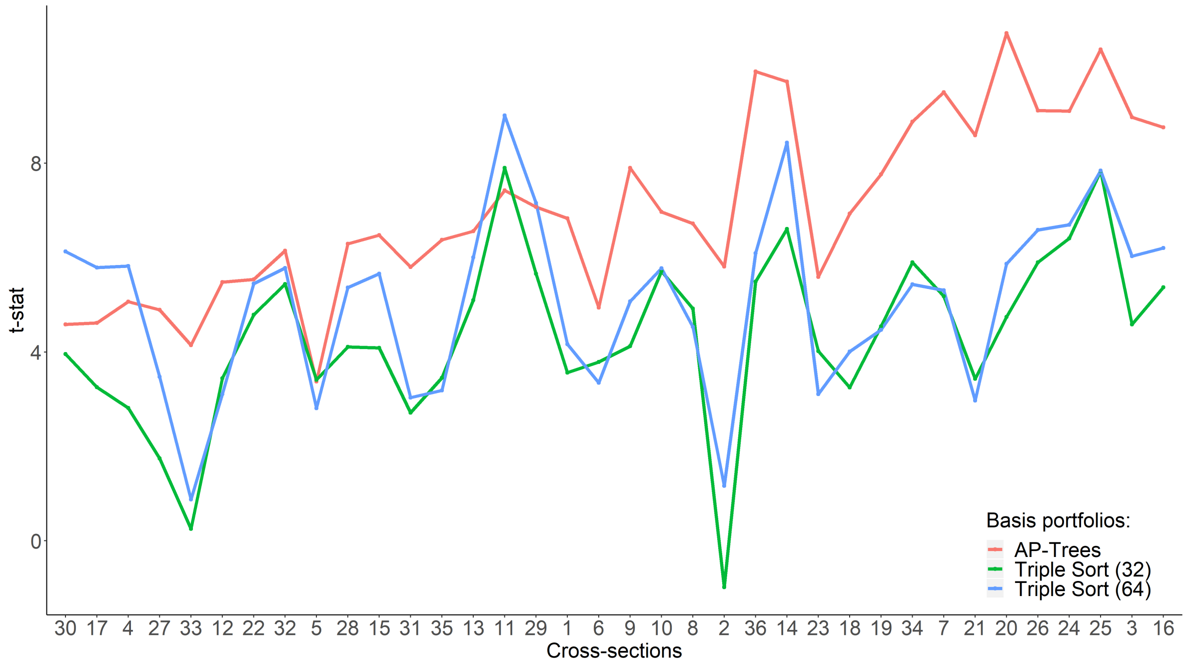


Figure E.8. SDF α: t-statistics of the pricing errors of SDF pruned AP-Trees (10 portfolios) with respect to 5 Fama-French factors

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 (40).

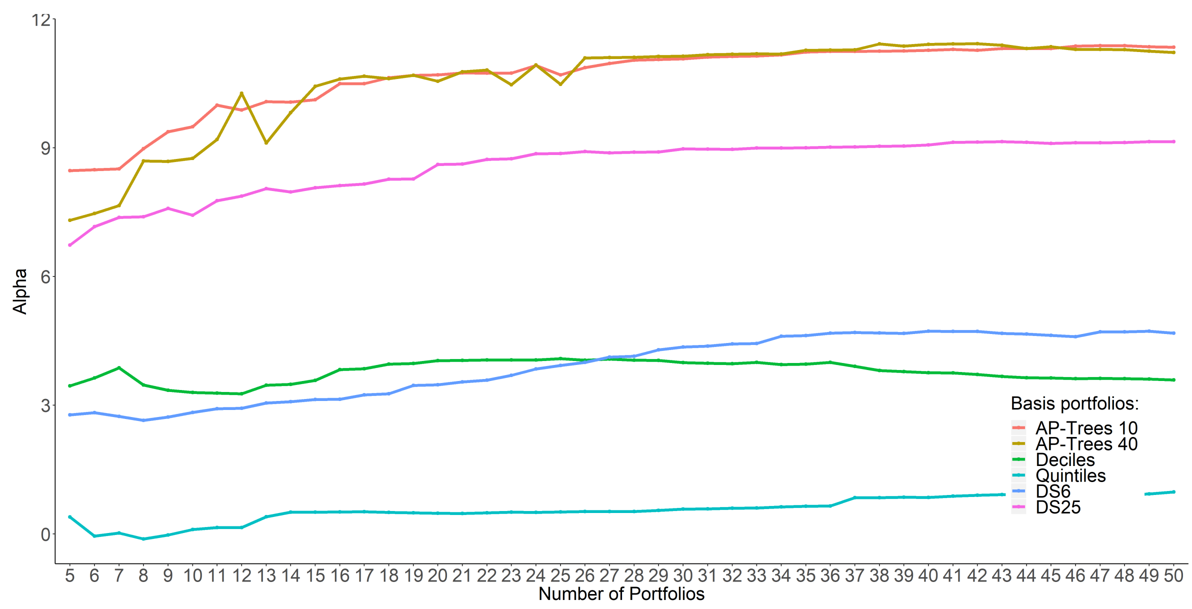
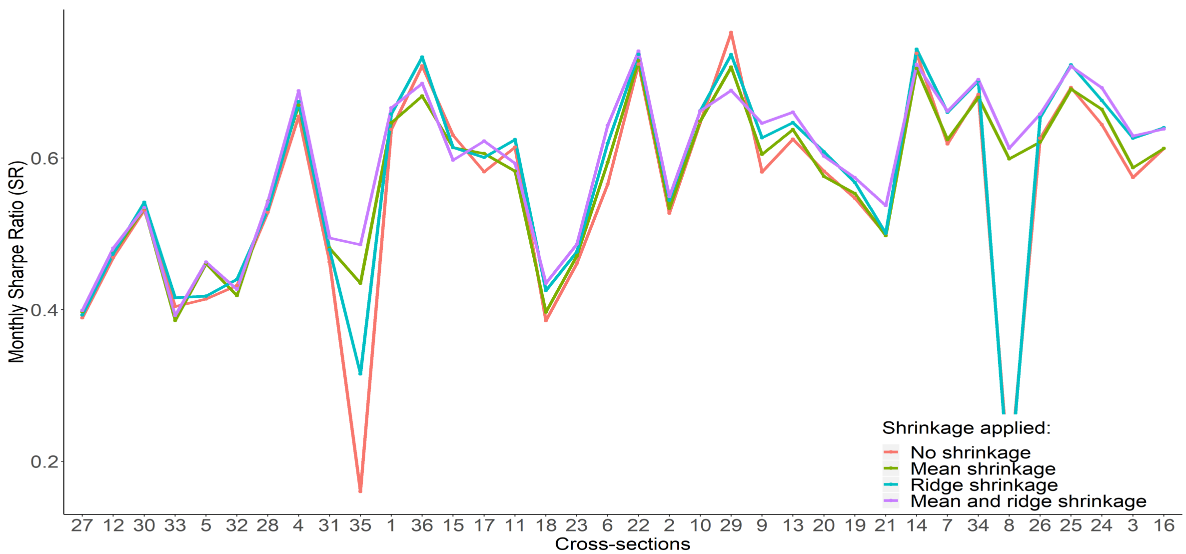


Figure E.9. Out-of-sample SDF alpha t-statistics, based on the FF11 factor model (market portfolio + 10 long-short portfolios) for 10 characteristics

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 13. The pricing errors are calculated with respect to all 11 factors.

Panel A: Sharpe ratios of AP-Trees (40) with time-varying SDF portfolio weights



Panel B: Sharpe ratios of AP-Trees (10) with time-varying SDF portfolio weights

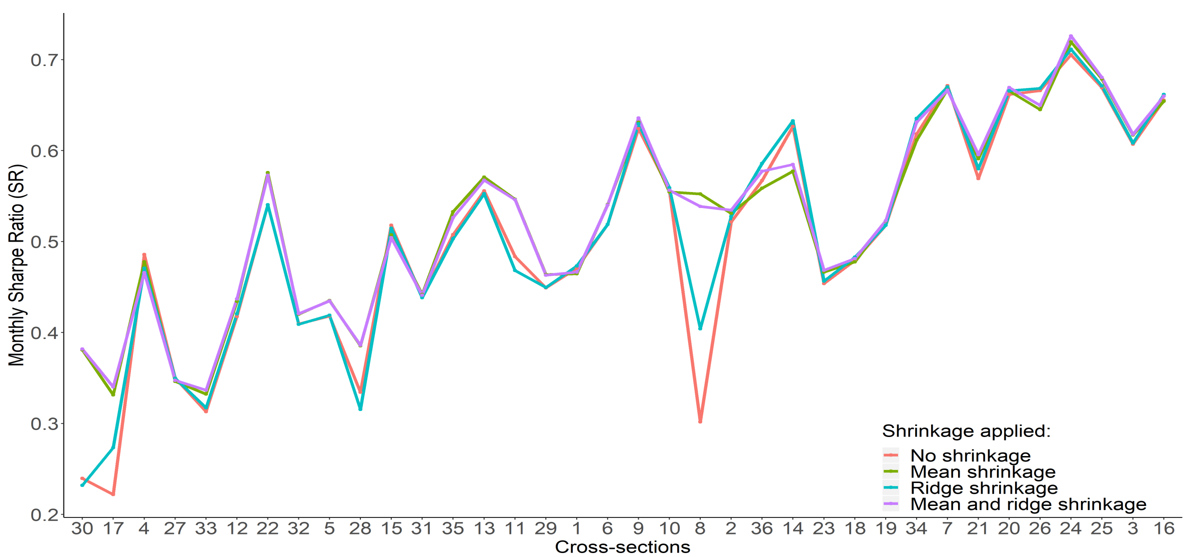
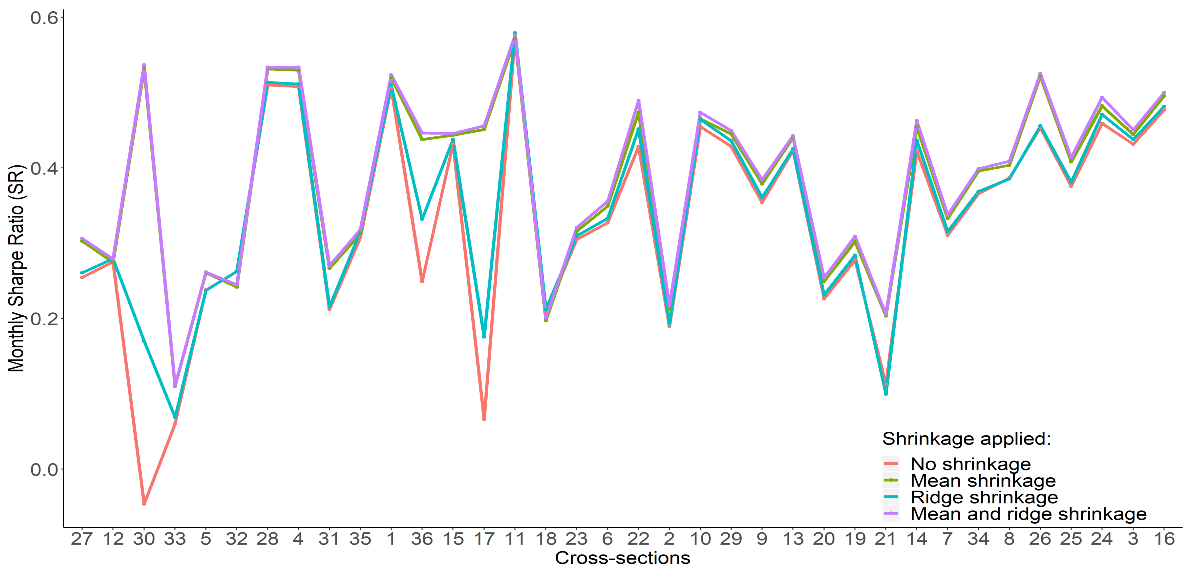
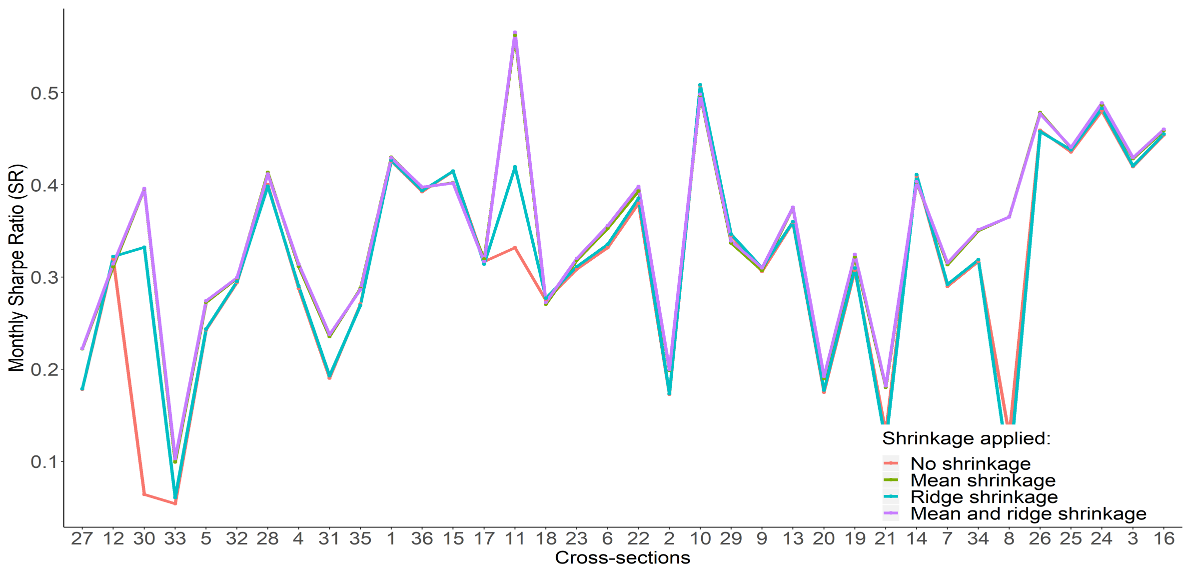


Figure E.10. Sharpe ratios of AP-Tree's SDFs with rolling window estimates of portfolio weights

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 AP-Trees (10 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 validation data.



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

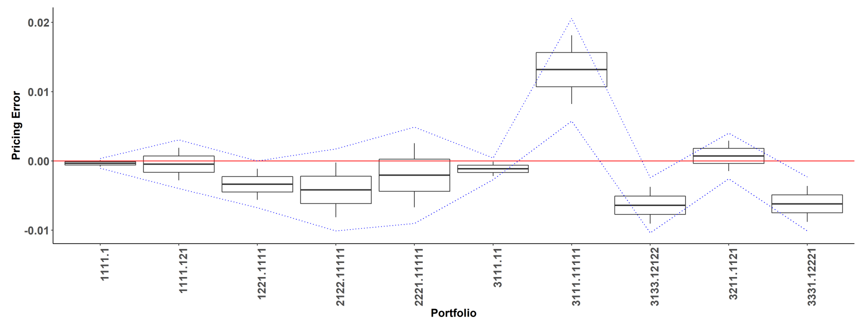


(b) Sharpe ratios of triple sorts (32 portfolios) with time-varying weights SDF portfolios weights

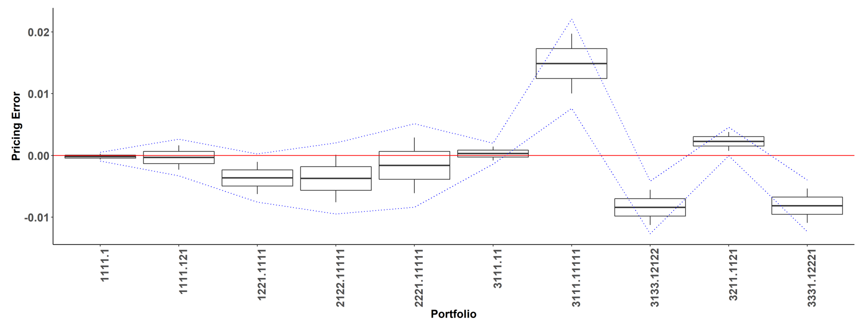
Figure E.11. Sharpe ratios of SDFs based on triple sorts with rolling window estimates of weights

The top figure displays the monthly out-of-sample Sharpe ratio of the SDFs of 64 triple-sorted portfolios with and without different forms of shrinkage, while the bottom figures show the results for 32 triple-sorted portfolios. 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 validation data.

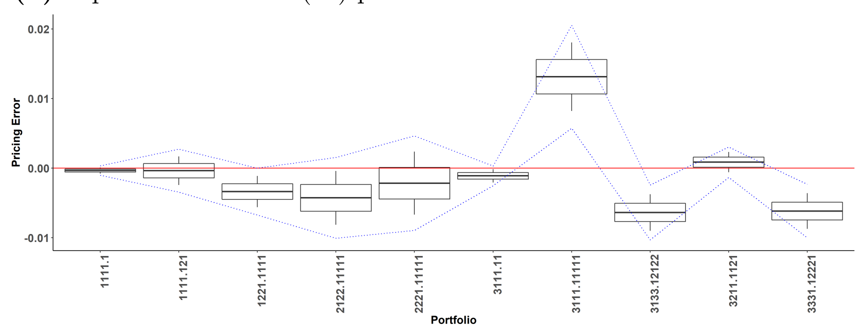
Appendix E.5. Additional Results for Size, Profitability, and Investment



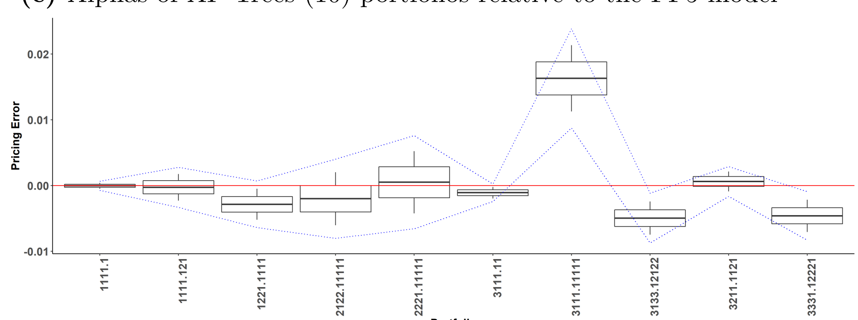
(a) Alphas of AP-Trees (10) portfolios relative to the XSF model



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

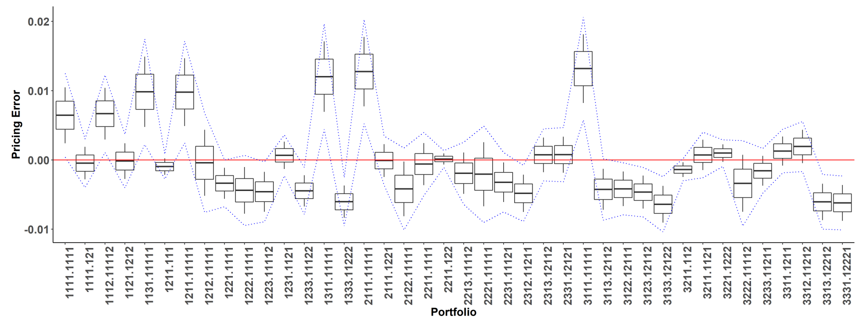


(c) Alphas of AP-Trees (10) portfolios relative to the FF5 model

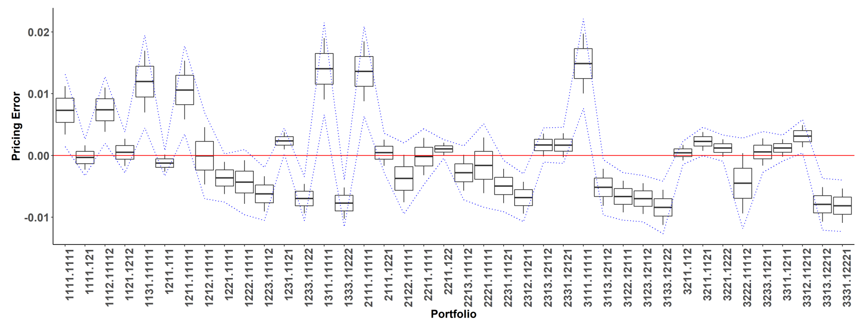


(d) Alphas of AP-Trees (10) portfolios relative to the FF11 model

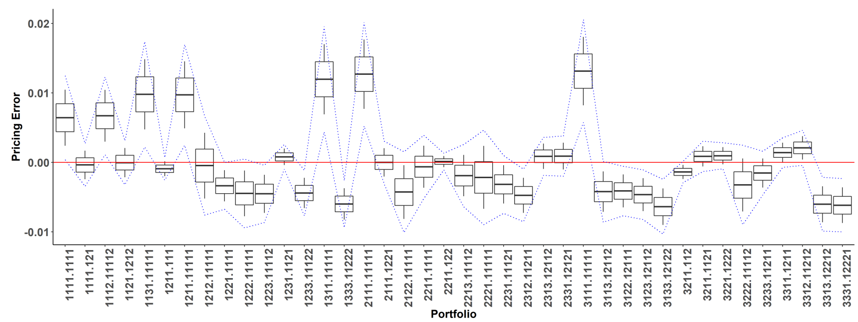
Figure E.12. Pricing errors α_i for AP-Trees (10) for size, operating profitability, and investment. The plots present the pricing errors relative to the FF3, FF5, FF11, and XSF factors with significance levels (candlestick=5%, dashed line=1%, box=10%).



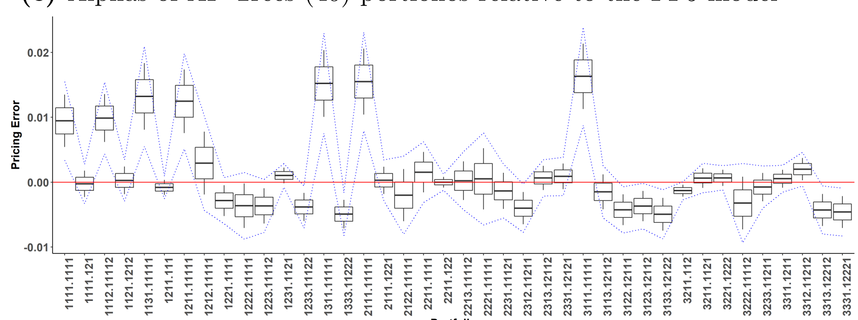
(a) Alphas of AP-Trees (40) portfolios relative to the XSF model



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



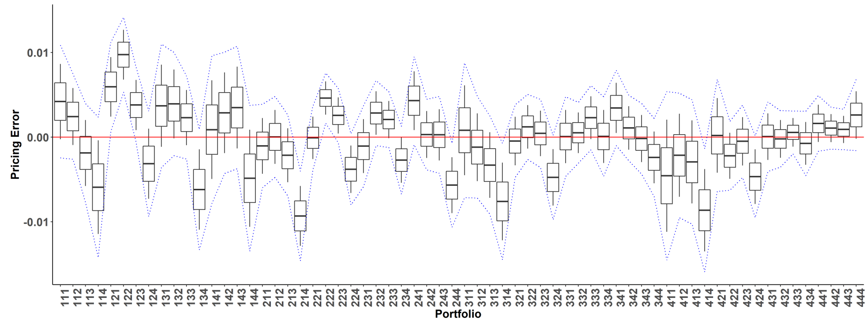
(c) Alphas of AP-Trees (40) portfolios relative to the FF5 model



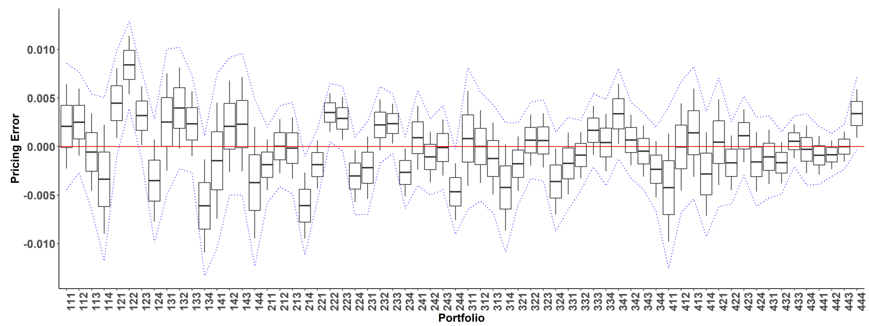
(d) Alphas of AP-Trees (40) portfolios relative to the FF11 model

Figure E.13. Pricing errors α_i for AP-Trees (40) for size, operating profitability, and investment

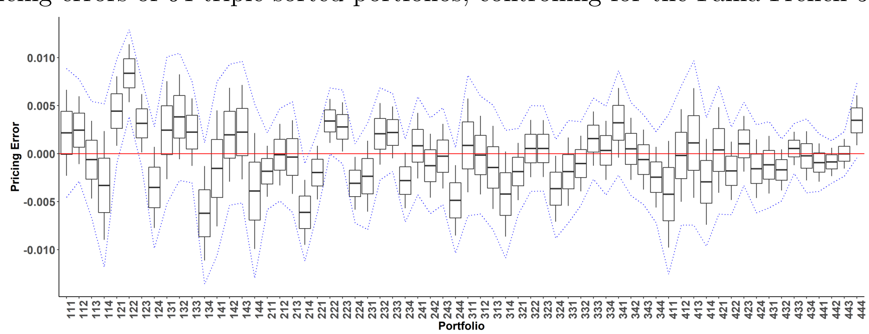
The plots present the pricing errors relative to the FF3, FF5, FF11, and XSF factors with significance levels (candlestick=5%, dashed line=1%, box=10%).



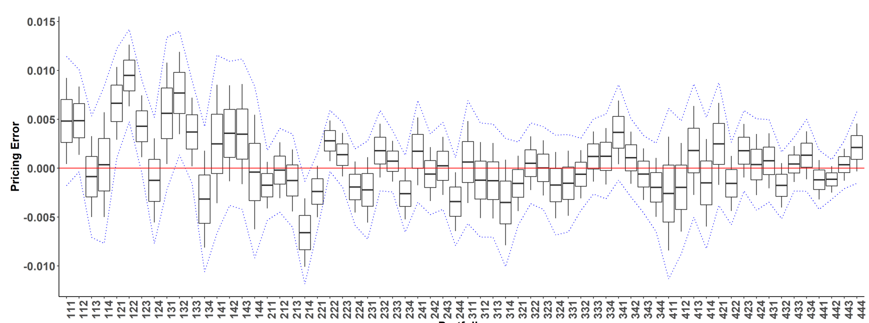
(a) Pricing errors of 64 triple-sorted portfolios, controlling for the Fama-French 3 factors



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



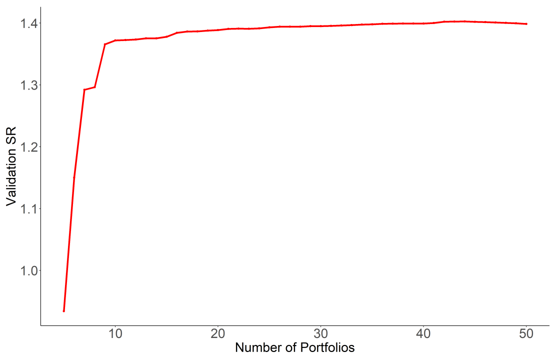
(c) Pricing errors of 64 triple-sorted portfolios, controlling for the cross-section-specific factors (market + size, profitability, investment)



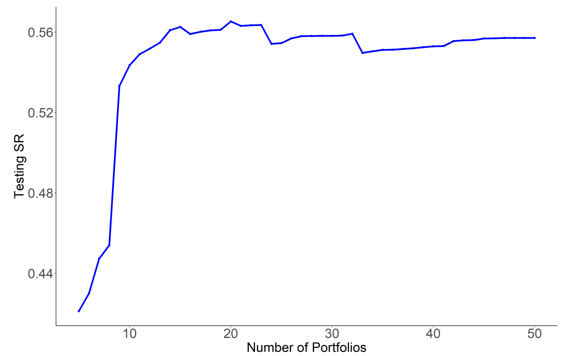
(d) Pricing errors of 64 triple-sorted portfolios, controlling for 11 cross-sectional factors

Figure E.14. Pricing errors α_i for TS (64 assets) for size, operating profitability, and investment. The plots present the pricing errors for triple-sorted portfolios relative to the FF3, FF5, FF11, and XSF factors with significance levels (candlestick=5%, dashed line=1%, box=10%).

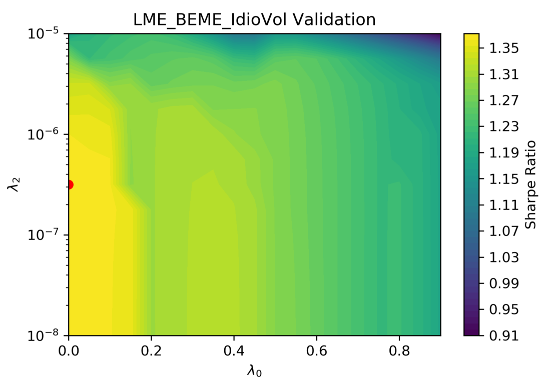
Appendix E.6. Results for Size, Value and Idiosyncratic Volatility



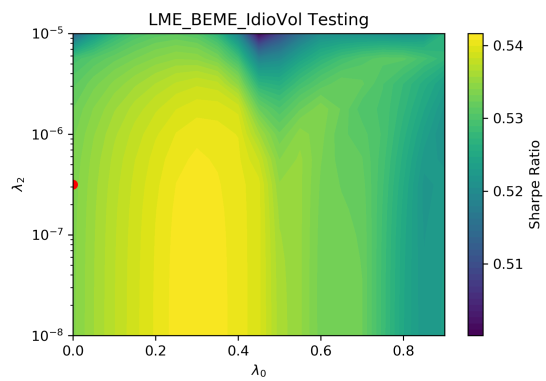
(a) Number of portfolios (validation sample)



(b) Number of portfolios (testing sample)



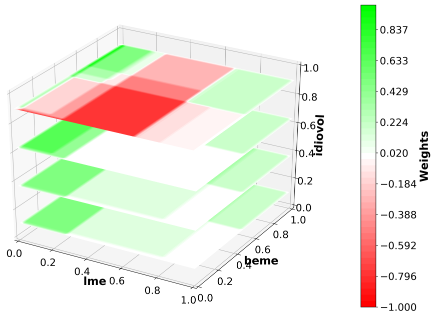
(c) Mean and variance shrinkage (validation)



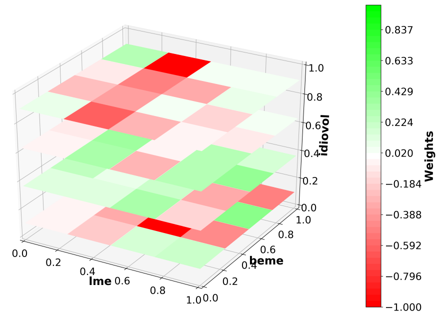
(d) Mean and variance shrinkage (testing)

Figure E.15. SR as a function of tuning parameters for size, value, and idiosyncratic volatility

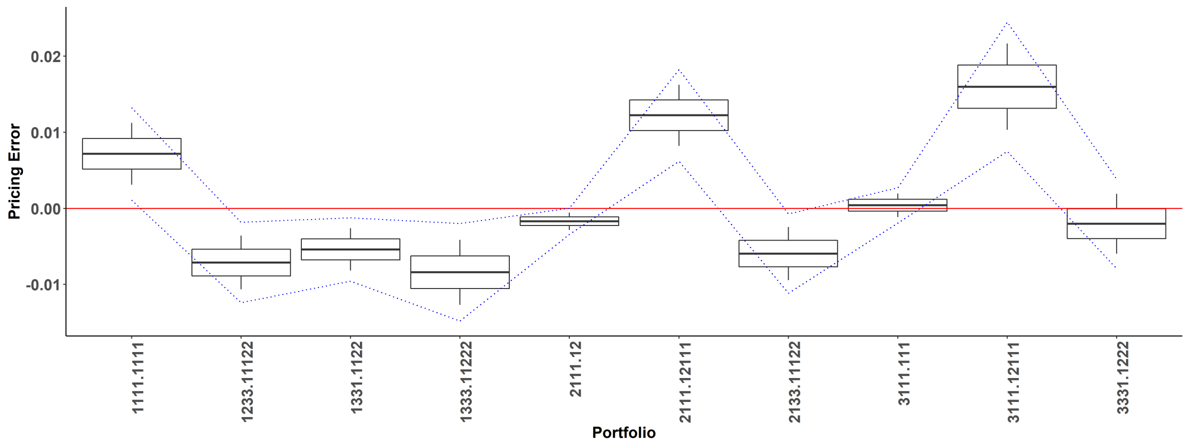
The figure shows the validation and out-of-sample Sharpe ratio as a function of tuning parameters for AP-Trees on size, value, and idiosyncratic volatility. Subplot (a) and (b) present the SR as function of the number of portfolios using the optimal mean and variance shrinkage. Subplot (c) and (d) depict the SR as a function of mean (λ_0) and variance (λ_2) shrinkage for 10 portfolios. Yellow regions correspond to higher values. The optimal validation shrinkage is indicated by the red dot. We use the mean $\hat{\mu} + \lambda_0 \bar{\mu} \mathbb{1}$.



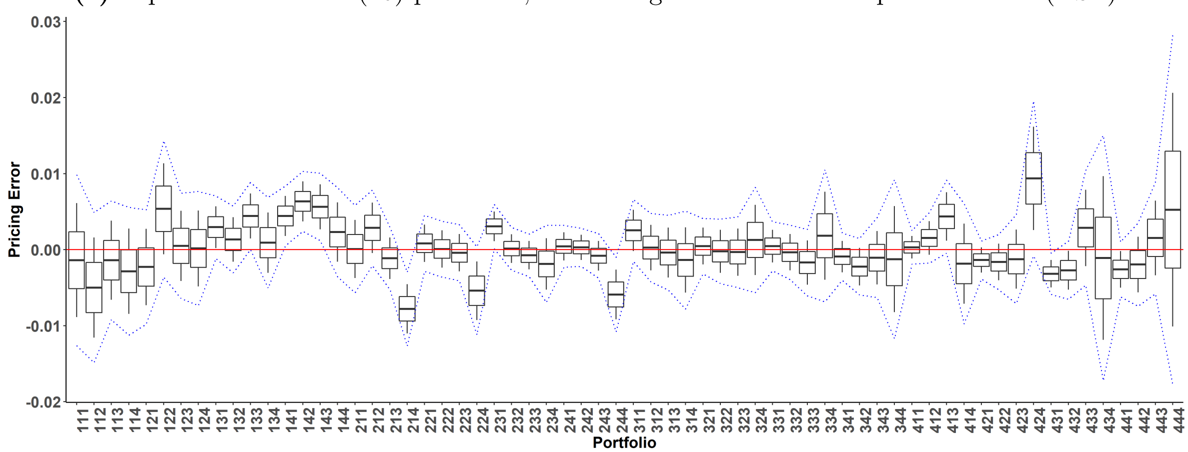
(a) SDF spanned by AP-Trees (10)



(b) SDF spanned by TS (64)



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



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

Figure E.16. Size, value, and idiosyncratic volatility: SDF weights and pricing errors for basis assets with 10 AP-Tree and 64 TS

Composition of conditional SDF with 10 AP-Tree and 64 TS basis assets and corresponding pricing errors for each basis asset. Subplots (a) and (b) show the SDF weight as a function of the size, value, and idiosyncratic volatility quantile. Subplots (c) and (d) present the pricing errors relative to the cross-section specific factors (XSF) with significance levels (candlestick=5%, dashed line=1%, box=10%).