Forest Through the Trees
Building Cross-Sections of Stock Returns

Svetlana Bryzgalova ¹  Markus Pelger ²  Jason Zhu ²

¹London Business School   ²Stanford University

Chicago Booth Asset Pricing Conference
October 25, 2019
Introduction
Motivation: What explains the cross-section of expected returns?

- Asset pricing has two essential elements:
  1. A cross-section of test or basis assets (e.g. 25 size – value sorted portfolios)
  2. A model for their pricing (e.g. the Fama-French 3 factor model)

- Empirical results crucially depend on the choice of the test assets

- This paper: building cross-sections of interpretable test assets that capture the asset pricing information contained in stock characteristics

- Equivalently, we construct a stochastic discount factor (SDF) conditional on a given set of characteristics
Conventional approach: **Sorting**

- Portfolios diversify the idiosyncratic noise, and are easy to use
- A balanced and more stationary (constant beta?) cross-section

**Figure 1:** Unconditional quantiles, used to build 16 double-sorted portfolios
State of the literature: Building cross-sections

State of the literature:

- Sorting is routinely used to create cross-sections based on 1-2 characteristics
- More than 3 characteristics: stack several cross-sections together
  - Example: 25 size and value + 10 momentum + 10 accruals

Challenges and problems:

- Single sorting ignores any interactions between characteristics
- Multiple sorting is limited to coarse double or triple interactions
- Not considering firm distribution ⇒ empty and unbalanced portfolios

⇒ Sorting captures only limited interactions and ignores conditional distributions between characteristics.
⇒ Success of simple long-short factors due to overly simplistic test assets.
This paper

Asset Pricing Trees (AP-Trees) = Cross-section of basis assets that reflect all the information including interactions and non-linearities, relying on

1. Tree-sorted portfolios: Creating managed portfolios with arbitrary interactions
2. Pruning the trees: Selecting portfolios that are the most relevant for asset pricing

Advantages:

• Generalization of sorting (sorting is a special and suboptimal case of AP-Trees)
• Capture arbitrary interactions between characteristics
• Retains interpretation and alleviates the curse of dimensionality
• Combines interactions between many characteristics
• New (and not overstudied) test assets to test asset pricing theories

⇒ AP-Trees have higher OOS Sharpe ratios than sorted portfolios or conventional factors, and are harder to price
⇒ Triple sorting, etc are too low a bar for most models, and do not reflect well the underlying information
Methodological Contribution

Alleviating the curse of dimensionality in conditional asset pricing models

- Conventional sorting = Kernel projections (with univariate kernel)
- Multi-dimensional sorting = Curse of dimensionality of multivariate kernel projections
- Tree projections capture non-linearities not spanned by kernel projections

Introducing a novel global approach to pruning “asset pricing” trees

- Conventional machine-learning pruning is local and not appropriate for asset pricing
- Generalization of Kozak, Nagel and Santosh (2019)
- Novel and economically motivated approach to find robust SDF

Empirically

- Harder, more informative cross-sections
- Test assets that reflect multiple characteristics and their interactions (no stacking!)
- Fresh cross-sections (for p-hackers)
Sorting, trees, and forest
Conditional asset pricing: Sorting as nonparametric estimator

Figure 2: Cochrane (2011): BM-sorted portfolios as nonparametric estimator of conditional mean.

- Characteristics of the assets are **informative** for expected returns
- Sorting provides a nonparametric estimator of conditional mean returns
- Sorted portfolios that should have **time-invariant** exposure to factors
What is a tree?

The simplest tree of depth \( d \) based on \( k \) characteristics is the set of all possible conditional sorts yielding \( k^d \times 2^d \) overlapping portfolios.

![Conditional tree based on size and value (121), with depth 3](image)

**Figure 3:** Conditional tree based on size and value (121), with depth 3

- Conditional on the order of splits, each subtree yields \( 2^d \) non-intersecting portfolios.
- AP-Trees: \( k^d \times 2^d \) portfolios, with \( \frac{N}{2^d} \) stocks in each independent of \( k \).
- Sorting: \( 2^{k \times d} \) portfolios, with on average \( \frac{N}{2^{k \times d}} \) stocks in each.

Tree-based portfolio id is ABC.WXYZ:
- ABC indicates the characteristics used for each split
- WXYZ identifies actual portfolio at the end of this subtree.
Trees with two characteristics

- Simplest trees rely on the 50/50 split within each group of the assets
- Standard double sort disregards conditional distribution in the characteristics space
- Tree-based portfolios are conditional splits: Reflect joint distribution of characteristics without the need for parametric modeling

**Figure 4:** One of the conditional trees based on size and value, with depth 3
Size and Value: Trees

Figure 5: Cross-sectional quantiles of the portfolio splits based on conditional trees
Size and Value: Trees

**Figure 5:** Cross-sectional quantiles of the portfolio splits based on conditional trees

Piet Mondrian (1930): *Composition in red, blue and yellow*
Size and Value: Trees vs Double-Sorting

Figure 6: Cross-sectional quantiles of the portfolio splits based on conditional trees and double sorting with size and value.
Many characteristics have a complicated joint distribution:
⇒ Focusing on unconditional quantiles can produce either empty or too dense portfolios.

**Figure 7:** Left: size and value, Right: size and accruals
Conditional and Unconditional Characteristic Effects

The impact of characteristic on asset returns is often not monotonic, and interacts with other variables.

Figure 8: Left: size and value, Right: size and accruals
Dimensionality Reduction in AP-Trees

- With only 2 characteristics and depth 3: $2^3$ subtrees, each having $2^3$ portfolios, total 64 (intersecting) portfolios
- With 3 characteristics and depth 4: 1296 final nodes
- Baseline results with trees of depth 4 to have well-diversified portfolios

Usual techniques for portfolio reduction (e.g. pruning) is not applicable.

- Stocks with the same expected returns could have very different features
- Example: small stocks within value firms, large caps within growth
- Conventional machine learning tree pruning (bottom-up approach) is local in nature, and works only if the current split does not affect other nodes

Our approach: **Asset Pricing Pruning**

$\Rightarrow$ Optimal mean–variance portfolios + shrinkage applied to all the final and intermediate nodes of tree portfolios
Asset Pricing Pruning: Portfolios with Elastic Net

Optimal “basis” portfolios equivalent to basis assets of Stochastic Discount Factor (SDF):

- Tangency portfolio in the mean variance space
- Global decision problem: Comparing tree portfolios locally is not sufficient
- Our pruning reflects both risk $\hat{\Sigma}$ and mean return $\mu$.

Solution: tangency portfolio on a robust and sparse mean-variance efficient frontier:

1. Construct a portfolio frontier on training data with elastic net (under $\mu, \Sigma$ uncertainty):

   \[
   \begin{align*}
   &\text{minimize} \quad w^T \Sigma w + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \\
   &\text{subject to} \quad w^T \mathbb{1} = 1 \\
   &\quad w^T \mu \geq \mu_0
   \end{align*}
   \]

   Tuning parameters: target return $\mu_0$, sparsity lasso shrinkage $\lambda_1$, ridge shrinkage $\lambda_2$.

2. Choose robust tangency portfolio by selecting tuning parameters on the validation set.
Shrinkage of the SDF: Intuition (no lasso!)

Our approach generalizes the SDF shrinkage of Kozak, Nagel and Santosh (2019).

1. Unconstrained SDF = tangency portfolio of conventional mean-variance optimization

   \[ w_{\text{naive}}^* = \Sigma^{-1} \hat{\mu} \]

   Breaks down in large dimensions.

2. KNS (2019): Shrinkage in \( \Sigma \) applied directly on the unconstrained FOC:

   \[ w_{\text{ridge}}^* = \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} \hat{\mu} \]

   \( \Rightarrow \) gives a portfolio on the robust MVE frontier, but in general not the tangency one.

3. Our approach: Robust tangency portfolio and 2-fund separation theorem

   \[ w = \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} (\hat{\mu} + \lambda_0 1) \]

   \( \Rightarrow \) Shrinkage in the covariance matrix and sample mean towards the average!
   \( \Rightarrow \) Extreme sample means are likely wrong, so pulled towards the X-sectional average.
   \( \Rightarrow \) Analogue: Bloomberg’s adjusted beta is shrunk towards 1
Shrinkage of the SDF: Intuition (no lasso!)

Our approach generalizes the SDF shrinkage of Kozak, Nagel and Santosh (2019).

1. Unconstrained SDF = tangency portfolio of conventional mean-variance optimization

\[ w_{\text{naive}}^* = \Sigma^{-1} \hat{\mu} \]

Breaks down in large dimensions.

2. KNS (2019): Shrinkage in \( \Sigma \) applied directly on the unconstrained FOC:

\[ w_{\text{ridge}}^* = \left( \hat{\Sigma} + \lambda I_N \right)^{-1} \hat{\mu} \]

⇒ gives a portfolio on the robust MVE frontier, but in general not the tangency one.

3. Our approach: Robust tangency portfolio and 2-fund separation theorem

\[ w = \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} \left( \hat{\mu} + \lambda_0 \mathbb{I} \right) \sim \alpha w_{\text{min.var.}}^* + (1 - \alpha) w_{\text{ridge}}^* \]

⇒ Shrinkage in the covariance matrix and sample mean towards the average!

⇒ Extreme sample means are likely wrong, so pulled towards the X-sectional average.

⇒ Analogue: Bloomberg’s adjusted beta is shrunk towards 1
Shrinkage to the average: bias-variance trade-off

Example:
- iid stocks with the same known variance, but different (and estimated) expected returns

Figure 9: Out-of-sample Sharpe ratio for tangency portfolios, with/without shrinkage

⇒ Sacrifice some returns in exchange for much smaller variance
⇒ Optimal amount of shrinkage depends on the cross-section of assets
Pruning of the trees is done in **both depth and width**:

- Include all the intermediate and final nodes in robust mean-variance optimization
- Weight each node by $\sqrt{N_i}$ with $N_i =$ number of assets in portfolio $i$.
  - Intuition: GLS-like efficient weighting
  - Idiosyncratic risk diversified with a rate of $\frac{1}{\sqrt{N_i}}$ in portfolio $i$
  - Trade-off: Higher nodes have lower variance, but possibly more bias
- Building up on the PCA rotation of data as in KNS (2019)
  - Intuition: First PCA is market portfolio and scaled by rate $\sqrt{N_i}$.
  - First node of Asset Pricing Tree is market portfolio and scaled by $\sqrt{N_i}$.
  - Tree portfolios are still long-only strategies, and are easy to interpret.

⇒ A novel approach to optimal bandwidth selection based on the asset pricing criterion.
Empirical results
### Data

- Monthly stock returns from CRSP merged with Compustat
- Data sample: January 1964 to December 2016 (53 years)
- 46 firm-specific characteristics

<table>
<thead>
<tr>
<th>Past Returns</th>
<th>Investment</th>
<th>Profitability</th>
<th>Intangibles</th>
<th>Value</th>
<th>Trading Frictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum</td>
<td>Investment</td>
<td>Operating profitability</td>
<td>Accrual</td>
<td>Book to Market Ratio</td>
<td>Size</td>
</tr>
<tr>
<td>Short-term Reversal</td>
<td>Net operating assets</td>
<td>Profitability</td>
<td>Operating accruals</td>
<td>Assets to market cap</td>
<td>Turnover</td>
</tr>
<tr>
<td>Long-term Reversal</td>
<td>Change in prop. to assets</td>
<td>Sales over assets</td>
<td>Operating leverage</td>
<td>Cash to assets</td>
<td>Idiosyncratic Volatility</td>
</tr>
<tr>
<td>Return 2-1</td>
<td>Capital turnover</td>
<td>Fixed costs to sales</td>
<td>Price to cost margin</td>
<td>Cash flow to book value</td>
<td>CAPM Beta</td>
</tr>
<tr>
<td>Return 12-2</td>
<td>Profit margin</td>
<td>Return on net assets</td>
<td></td>
<td>Dividend to price</td>
<td>Residual Variance</td>
</tr>
<tr>
<td>Return 36-13</td>
<td>Return on assets</td>
<td>Return on equity</td>
<td></td>
<td>Earnings to price</td>
<td>Total assets</td>
</tr>
<tr>
<td></td>
<td>Expenses to sales</td>
<td>Expenses to sales</td>
<td></td>
<td>Tobin's Q</td>
<td>Market Beta</td>
</tr>
<tr>
<td></td>
<td>Capital intensity</td>
<td></td>
<td></td>
<td>Sales to price</td>
<td>Close to High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leverage</td>
<td>Spread</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unexplained Volume</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Variance</td>
</tr>
</tbody>
</table>

- For now a short list of 10 characteristics from K. French website: Size, Value, Operating Profitability, Investment, Momentum, Short-term Reversal, Long-term Reversal, Accrual, Asset Turnover, and Idiosyncratic Volatility.
- Basic set of characteristics: Size + 2 other: 36 cross-sections.
Empirical Implementation

- Split the data in training (20 years), validation (10 years), test (23 years)
- **Training**: Construct constant portfolio weights:
  - Estimate AP-Tree: 40 basis assets from penalized portfolio frontier applied to all tree-based portfolios with intermediate and final nodes
  - Construct weights of SDF based on AP-Tree and triple sorts (with elastic net)
- **Validation**: Select tuning parameters for elastic net for efficient frontier (for AP-Trees and SDF)
- **Test**: Out-of-sample time-series of AP-Trees and SDFs
  - Sharpe ratios, alphas for individual portfolios and SDF and test statistics
- All the portfolios are value-weighted
- AP-Trees with level 4 and 3 characteristics (but at most 3 splits for one characteristic to avoid extreme quantiles)
Asset Pricing Performance

Asset pricing test portfolios:
- AP-Tree with 40 basis assets
- 32 TS: triple sorted with 32 assets (only one size split)
- 64 TS: triple sorted with 64 assets (two size splits)

Factors:
- 11 Factors (FF11): market + all 10 cross-sectional specific factors
- 4 Factors (XSF): market + 3 cross-sectional specific factors
- Fama-French 3 and 5 factors (FF3 and FF5)

Benchmark numbers (all out-of-sample)
- \(SR\) (Sharpe ratio): Out-of-sample Sharpe ratio of conditional SDF based on different basis assets
- \(\alpha\) (Pricing error): t-statistic of out-of-sample time-series pricing error of SDF
- \(\alpha_i\) (Portfolio pricing error): time-series pricing error of basis asset \(i\)
- \(XS - R^2\) (Cross-sectional \(R^2\)): Adjusted cross-sectional pricing error

\[
1 - \frac{N}{N - K} \frac{\sum_{i=1}^{N} \alpha_i^2}{\sum_{i=1}^{N} E[R_i]^2}
\]
Sharpe ratios

**Figure 10**: Monthly out-of-sample Sharpe ratios of mean-variance efficient portfolios for AP-Trees and triple sorts and XSF. Cross-sections sorted by the SR of AP-Trees.
SDF unspanned by factors: FF5

Figure 11: SDF $\alpha$: t-statistics of $\alpha$ relative to Fama-French 5 factor model.
SDF unspanned by long-short XS-specific factors

Figure 12: SDF $\alpha$: t-statistics of $\alpha$ relative to cross-section specific factors.
SDF unspanned by factors: FF11

Figure 13: SDF $\alpha$: t-statistics of $\alpha$ relative to all 11 factors.
Zooming into the cross-sections
Structure of SDF and pricing errors for specific cross-sections

<table>
<thead>
<tr>
<th>Size-Investment-Profitability</th>
<th>AP-Tree 10</th>
<th>AP-Tree 40</th>
<th>32 TS</th>
<th>64 TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDF SR</td>
<td>0.65</td>
<td>0.69</td>
<td>0.50</td>
<td>0.52</td>
</tr>
<tr>
<td>FF11</td>
<td>9.12</td>
<td>9.60</td>
<td>4.46</td>
<td>7.09</td>
</tr>
<tr>
<td>XS-F</td>
<td>8.77</td>
<td>9.46</td>
<td>5.28</td>
<td>6.53</td>
</tr>
<tr>
<td>FF5</td>
<td>8.76</td>
<td>9.60</td>
<td>5.59</td>
<td>7.09</td>
</tr>
<tr>
<td>FF3</td>
<td>10.11</td>
<td>11.03</td>
<td>7.47</td>
<td>8.73</td>
</tr>
</tbody>
</table>

Representative Example: Size-Investment-Profitability:

- AP-Tree with 40 portfolios (as before) and with 10 portfolios
  $\Rightarrow$ 10 AP-Tree portfolios capture the same pricing information as 40
  $\Rightarrow$ $\alpha$ t-statistics twice as large for AP-Tree
Size-Investment-Profitability: How many portfolios?

**Figure 14:** SR as a function of portfolios, ex ante and ex post (oracle)

Variance shrinkage $\lambda_2$ and mean shrinkage $\lambda_0$ are chosen optimally

$\Rightarrow$ 10 AP-Tree portfolios capture the same pricing information as 40
Size-Investment-Profitability: Optimal Tuning Parameters

Figure 15: SR as a function of tuning parameters, ex ante and ex post (oracle)

- Choice of mean and variance shrinkage for \( N = 10 \) AP-Tree portfolios
- Optimal mean shrinkage \( \lambda_0 = 0.15 \)
- Optimal variance shrinkage \( \lambda_2 \) close to zero.

\( \Rightarrow \) KNS (2019) shrinkage is not optimal
Size-Investment-Profitability: How large are pricing errors?

Figure 16: Pricing errors $\alpha_i$ for AP-Trees with XS-specific factors

Significance: Candlestick = 5%, dashed line = 1%, box = 10%
Size-Investment-Profitability: How large are pricing errors?

Figure 17: Pricing errors $\alpha_i$ for AP-Trees with FF5 and FF11 for Size-Investment-Profitability
Size-Investment-Profitability: How large are pricing errors?

Figure 18: Pricing errors $\alpha_i$ for AP-Trees and 64 TS with XSF.
What does the SDF load on?

⇒ Conditional interactions matter!
⇒ Triple-sorting (≈ 3-dimensional kernel) too coarse.
⇒ Weights not only on small stocks but on all characteristics.

**Figure 19**: Composition of conditional SDF with 10 AP-Tree and 64 TS basis assets.
Challenges
**AP-Trees: a new benchmark for the conditional SDF and test assets**

**Fundamental questions/challenges:**

1. **What is the right way to recombine portfolios?**
   - A non-hierarchical problem for SDF or portfolio optimization
   - Shrinkage on the mean and covariance matrix
     (instead of just covariance)

2. **How to efficiently incorporate a large number of characteristics?**
   - AP-Trees is a natural generalization of sorting
   - Currently working on 10+interactions: Stay tuned!
     (already completely infeasible in traditional sorting)

3. **How to get back to structural modeling?**
   - Cross-products mask the true joint information set
   - Conditional asset pricing models with curse of dimensionality?

4. **Application-specific optimal cross-section:**
   - Generally, same factors do not have to span all the cross-sections
   - Maximizing the power against alternative: Nagel and Singleton (2010)
   - Existing double-sorted cross-sections are overstudied: Data-mined factors
Appendix
Interpretation with Robust Mean-Variance Optimization

Assume the true covariance and expected returns belong to the uncertainty sets:

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

\[ \mu \in S_{\mu} = \left\{ \tilde{\mu} : \tilde{\mu}_i = \hat{\mu}_i + \psi_i; \quad |\psi_i| \leq \bar{\psi} \quad \text{and } \bar{\psi} \geq 0 \right\} . \]

Optimal portfolios under the worst case scenario:

\[
\min_{\omega} \max_{\tilde{\Sigma} \in S_{\Sigma}, \tilde{\mu} \in S_{\mu}} \frac{1}{2} \omega^T \tilde{\Sigma} \omega - \tilde{\gamma}_1 (\omega^T \tilde{\mu} - \mu_0) - \tilde{\gamma}_2 (\omega^T \mathbb{1} - 1)
\]

Equivalent to our problem:

\[
\min_{\omega} \frac{1}{2} \omega^T \hat{\Sigma} \omega + \lambda_2 \|\omega\|_2^2 + \lambda_1 \|\omega\|_1 - \tilde{\gamma}_1 (\omega^T \hat{\mu} - \mu_0) - \tilde{\gamma}_2 (\omega^T \mathbb{1} - 1)
\]

⇒ In this sense robust mean variance efficient frontier.
Cross-sectional $R^2$ with XS-specific factors

Figure 20: $XS - R^2$: Cross-sectional pricing of portfolios XS-specific factors.
Cross-sectional $R^2$ with FF5

Figure 21: $XS - R^2$: Cross-sectional pricing of portfolios with 5 Fama-French factors.
How large are pricing errors? Example 1: Size-Value-Turnover.
How big are pricing errors? Example 2: Size-Investment-Profitability

Figure 23: Pricing errors $\alpha_i$ for AP-Trees with XS-specific factors.
How big are pricing errors? Example 2: Size-Investment-Profitability

Figure 24: Pricing errors for AP-Trees with FF5 and FF11.
Size-Investment-Profitability: Optimal Tuning Parameters

Figure 25: SR as a function of tuning parameters, ex ante and ex post (oracle)

- Choice of mean and variance shrinkage for $N = 40$ AP-Tree portfolios
- Optimal mean shrinkage $\lambda_0 = 0.15$
- Optimal variance shrinkage $\lambda_2$ close to zero.

$\Rightarrow$ Kozak, Nagel and Santosh (2019) shrinkage is not optimal
Size-Investment-Profitability: How large are pricing errors?

Figure 26: Pricing errors $\alpha_i$ for AP-Trees with XS-specific factors

Significance: Candlestick = 5\%, dashed line 1\%, box 10\%
Size-Investment-Profitability: How large are pricing errors?

Figure 27: Pricing errors $\alpha_i$ for AP-Trees with FF5 and FF11 for Size-Investment-Profitability.
Size-Investment-Profitability: How large are pricing errors?

Figure 28: Pricing errors $\alpha_i$ for AP-Trees and 64 TS with XSF.
What does the SDF load on?

(a) Tree-based portfolios

(b) 64 Triple-sorted portfolios

Figure 29: Composition of conditional SDF with 40 AP-Tree and 64 TS basis assets.

⇒ Conditional interactions matter!
⇒ Triple-sorting (= 3-dimensional kernel) too coarse.
⇒ Weights not only on small stocks but on all characteristics.
Why portfolios and not... individual stocks?

- CRSP panel is incredibly unbalanced: many firms exist only for a few years
- Structural models are often estimated by GMM and similar techniques
- AP models are not about fitting 'firm names', but composite economic objects
- Individual stocks have time-varying exposure to sources of risk (proxied by char.)

... reduced form SDFs directly?

- Models are not 'theories of everything', - they target only some data features
- Where exactly does the model fail? With PCA-like factors and SDFs it’s hard to say
- Does not reveal beta patterns in the characteristic space

... factors directly, e.g. like in Barillas and Shanken (2016)?

- Not suitable for nontradable factors or nonlinear models
- Flat priors may fail for model comparison and factor selection: BHJ (2019)

⇒ There are cases where you need portfolios
Asset Pricing Pruning: Portfolios with Elastic Net

The search of the optimal “basis” in the portfolio space is challenging:

• Comparing the mean and recombining is not enough
• SDF reflects information on both the risk and return
• Estimation risk of \( \hat{\Sigma} \)

Mean-variance optimization to reflect the efficient frontier:

\[
\begin{align*}
\text{minimize} & \quad w^T \Sigma w + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \\
\text{subject to} & \quad 1^T w = 1 \\
& \quad w^T \mu \geq \mu_0
\end{align*}
\]

Shrinkage necessary because of high-dimensionality of the problem. A combination of \( L_1 \) and \( L_2 \) norms yields individual sparsity and weight stability between similar portfolios.

Solution retains interpretability, and easily allows for:

• stable loadings
• a fixed number of portfolios (in and out-of-sample) for comparison
• additional restrictions on portfolio loadings (short sales, etc).
Interpretation with robust MV optimization

Assume the true covariance and expected returns belong to the uncertainty sets:

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

\[ \mu \in S_\mu = \left\{ \tilde{\mu} : \tilde{\mu}_i = \hat{\mu}_i + \psi_i; \quad |\psi_i| \leq \bar{\psi} \quad \text{and } \bar{\psi} \geq 0 \right\} . \]

Optimal portfolios under the worst case scenario:

\[ \min_w \max_{\tilde{\Sigma} \in S_\Sigma, \tilde{\mu} \in S_\mu} \frac{1}{2} \omega^T \tilde{\Sigma} \omega - \tilde{\gamma}_1 \left( \omega^T \tilde{\mu} - \mu_0 \right) - \tilde{\gamma}_2 \left( \omega^T \mathbb{1} - 1 \right) \]

Concentrating out \( \tilde{\mu} \) and \( \tilde{\Sigma} \):

\[ \min_w \frac{1}{2} \text{tr} \left( \omega^T \hat{\Sigma} \omega \right) + \bar{\Delta} \omega^T \omega - \bar{\gamma}_1 \left( \omega^T \hat{\mu} - \mu_0 \right) + \bar{\gamma}_1 \sum_{i=1}^{N} |\psi_i| \omega_i - \bar{\gamma}_2 \left( \omega^T \mathbb{1} - 1 \right) \]

Taking \( \lambda_1 = \bar{\gamma}_1 \bar{\psi} \) and \( \lambda_2 = \bar{\Delta} \), the problem is equivalent to the standard regularized optimization with the sample mean estimates of expected returns and the covariance matrix:

\[ \min_w \frac{1}{2} \omega^T \hat{\Sigma} \omega + \lambda_2 \|\omega\|_2^2 + \lambda_1 \|\omega\|_1 - \bar{\gamma}_1 \left( \omega^T \hat{\mu} - \mu_0 \right) - \bar{\gamma}_2 \left( \omega^T \mathbb{1} - 1 \right) \]
Pruning vs Kozak, Nagel and Santosh (2019)

This paper: mean-variance optimization of the efficient frontier:

\[
\begin{align*}
\text{minimize} & \quad w^T \Sigma w + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \\
\text{subject to} & \quad 1^T w = 1 \\
& \quad w^T \mu \geq \mu_0
\end{align*}
\]

KNS (2019): shrinkage directly on the unconstrained First Order Condition:

\[
\begin{align*}
\text{minimize} \quad & \frac{1}{2} (\hat{\mu} - \hat{\Sigma} w)^T \hat{\Sigma}^{-1} (\hat{\mu} - \hat{\Sigma} w) + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \\
\end{align*}
\]

Analogy with OLS and lasso:

\[
\begin{align*}
\text{minimize} \quad & (Y - X \beta)^T (Y - X \beta) + \lambda_1 \|\beta\|_1 \\
& \quad \text{vs} \\
\text{minimize} \quad & (X^T Y - X^T X \beta)^T (X^T Y - X^T X \beta) + \lambda_1 \|\beta\|_1
\end{align*}
\]
Optimal portfolios with L2 penalty

\[
\min_{\omega} \omega^T \hat{\Sigma} \omega + \lambda_2 \|\omega\|_2^2 \quad \text{subject to } \omega^T \hat{\mu} = \mu_0 \text{ and } \omega^T \mathbb{1} = 1
\]

The standard solution:

\[
\hat{\omega} = \delta \omega_{\mu} + (1 - \delta) \omega_1
\]

where

\[
\omega_{\mu} = \frac{1}{1^T (\hat{\Sigma} + \lambda_2 I_N)^{-1}} \hat{\mu}, \quad \omega_1 = \frac{1}{1^T (\hat{\Sigma} + \lambda_2 I_N)^{-1}} \mathbb{1}, \quad \delta = \frac{\mu_0 - \hat{\mu}^T \omega_1}{\hat{\mu}^T \omega_{\mu} - \hat{\mu}^T \omega_1}
\]

Relaxing the normalization of \( \omega^T \mathbb{1} = 1 \) yields

\[
\hat{\omega} = \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} \hat{\mu} + \frac{\hat{\mu}^T \left( \hat{\Sigma} + \lambda_2 I_N \right) \hat{\mu} - \mu_0 \mathbb{1}^T \left( \hat{\Sigma} + \lambda_2 I_N \right) \mathbb{1}}{\mu_0 \mathbb{1}^T \left( \hat{\Sigma} + \lambda_2 I_N \right) \mathbb{1} - \hat{\mu}^T \left( \hat{\Sigma} + \lambda_2 I_N \right) \mathbb{1}} \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} \mathbb{1}
\]

\[
= \left( \hat{\Sigma} + \lambda_2 I_N \right)^{-1} (\hat{\mu} + \gamma \mathbb{1})
\]
**Shrinkage perspective**

$\gamma$ is a decreasing function of $\mu_0$: shrinkage towards the average return.

Direct mapping: $\mu_0 \in \left[ \frac{\mu^T(\Sigma+\lambda_2 I_N)\mu}{1^T(\Sigma+\lambda_2 I_N)1}, \frac{\mu^T(\Sigma+\lambda_2 I_N)^{-1}1}{1^T(\Sigma+\lambda_2 I_N)^{-1}1} \right]$ corresponds to $\gamma \in [0, +\infty)$.

$\gamma = 0$ is equivalent to setting the target expected return to $\mu_0 = \frac{\mu^T(\Sigma+\lambda_2 I_N)^{-1}1}{1^T(\Sigma+\lambda_2 I_N)^{-1}1}$, which corresponds to the optimal portfolio weights of

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

This estimator is identical to Kozak et al. (2019) without the lasso term,

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

The same intuition works with adding a lasso penalty.
Design

Consider a single factor model:

\[ R_{t+1,i}^e = \beta_{t,i} F_{t+1} + \epsilon_{t+1,i} \]

where \( F_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_F, \sigma_F^2) \) and \( \epsilon_{t,i} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2) \).

Factor loadings (more complex models in paper):

- **Additively linear**: stock beta being simply the sum of two characteristics:
  \[
  \beta_{t,i} = C_{t,i}^{(1)} + C_{t,i}^{(2)}.
  \]

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

- Correlation between characteristics: 0, 0.5, 0.9.
Simulated cross-section of returns: betas and returns

(a) True factor loadings
(b) Sample average returns

Figure 30: True factor loadings and sample average returns as a function of characteristics.

⇒ Sample mean noisy measure of SDF beta.
SDF betas in the characteristic space

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

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

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

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

**Figure 31:** Tracking expected returns: SDF loadings as function of characteristics in linear model.
Size and Value revisited: 16 portfolios

Figure 32: SDF betas with different basis assets, based on Size and Value

(a) AP-Trees, OOS SR = 0.61
(b) Double-sorted portfolios, OOS SR = 0.29