Discussion of:

"Forest through the Trees: Building Cross-Sections of Stocks Returns"

by

Svetlana Brysgalova, Markus Pelger, and Jason Zhu

Caio Almeida

Princeton University

Fourth International Workshop in Financial Econometrics,
October 8, 2019
The Goal

- There is an active literature discussing how to characterize the SDF as a function of firms’ characteristics.

- There is also a huge literature on tests of asset pricing models.

- However, not much is said on **how to construct a set of basis assets** representative of the information contained in the large XS of US stocks returns.

  - There is only one paper, “Basis assets”, by Ahn, Conrad and Dittmar, RFS 2009, that proposes a new way to build basis assets based on sorts by correlations.

- This paper aims at contributing to these two topics.
The Goal

▶ Given a XS of stocks and corresponding characteristics, the goal is twofold.

▶ First, to build a small set of portfolios representative of that XS, capturing the complex dependence of expected returns on the conditioning characteristics.

▶ Second, find projections of a (linear) SDF into chosen sub-spaces of these characteristics that could be useful for diagnosing AP models.

▶ Instead of relying on sorts based on unconditional quantiles of each individual characteristic, they propose using conditional decision trees.
Double Sorting versus Conditional Trees

Diagram showing double sorting and conditional tree structure.
The Goal

- The combinatorial structure of those trees makes the number of generated assets large, even when dealing with a set of characteristics as small as three.

- To reduce this number, they use a regularized mean-variance approach.

- They reduce from hundreds of tree assets to a small XS (40) with size comparable to those obtained with triple sorts.

- **Main Conclusions:**
  
  i. Trees assets are much more informative (higher Sharpe Ratios / larger pricing errors) than the triple sort assets (TS-assets).

  ii. The recovered SDF projections are highly non-homogeneous functions of the characteristics.
Contribution

- Really innovative paper on the construction of new basis assets.

- The whole process of construction has many layers and by products that enrich the paper but also make it difficult to grasp.

- I will try to offer alternative perspectives on the metrics adopted (Sharpe Ratios and pricing errors) and on the objects of interest (basis assets; linear conditional SDF).
Discussion Outline

- Three versions of SDFs.

- Pricing errors in different layers (different subsets of assets).

- Informational content: Hansen & Jagannathan and Kullback-Leibler bounds.

- Informational content: skewness, kurtosis, and Generalized Sharpe Ratios.

- Disentangling nonlinearities from conditional interactions.

- A more appropriate benchmark than triple sorts.
Three Datasets

1. First, there is the whole XS of CRSP stocks. About 15000 stocks available daily.

2. 10 characteristics available: Fixing size and two other characteristics gives $C_2^9 = 36$ XS’s with $3^3 \times (2 + 4 + 8) = 378$ portfolios in each. A total of 13608 overlapping portfolios.

3. After a XS is pruned we reduce from 378 to 40 assets. Therefore the total of assets after pruning is $36 \times 40 = 1440$.

- I will define $R_{tree,378}^i (R_{tree,40}^i)$ to be the vector with the 378 (40) assets of XS $i = 1, \ldots, 36$, before pruning (after pruning).

- I will also define $R_{tree,378}$, $R_{tree,40}$ to be the corresponding vectors stacking all 36 XS’s together.

- For the purposes of this discussion we will be interested in datasets 2. and 3.
Three versions of SDFs: “Full” SDF

- In my view the three versions of SDFs implicit in the paper are useful but not yet clearly addressed.

1. First, there is the “full” SDF that should be a joint function of all the characteristics.

- Not estimated yet. In theory, easy to estimate applying their robust MV technique to all 36 XS’s jointly, i.e. to $R_{tree,378}$:

$$
min_{\omega} \frac{\omega^\top \hat{\Sigma} \omega}{2} + \lambda_1 \|\omega\|_1 + \frac{\lambda_2 \|\omega\|_2^2}{2}, \text{s.t.} \omega^\top \mathbf{1} = 1, \omega^\top \hat{\mu} \geq \mu_0. \quad (1)
$$

- The “full” SDF is then: $m(R_{tree,378}) = 1 - \hat{\omega}^\top R_{tree,378}$.

- Comparable to Santosh, Nagel and Kozak’s SDF implied from characteristics in the case without using PCs.

- Useful for the discussion: “which characteristics matter to build an SDF?”. 

2. For each of the 36 XS’s based on three characteristics, a projected SDF is obtained as a byproduct of their robust OOS MV method. It will have 40 tree assets after pruning.

For each XS $i$, it solves: $\min_\omega \frac{\omega^\top \hat{\Sigma}^i \omega}{2} + \lambda_1 \|\omega\|_1 + \frac{\lambda_2 \|\omega\|_2^2}{2}$, with $\omega^\top 1 = 1$, and $\omega^\top \hat{\mu}^i \geq \mu_0$. The SDF is $m^{i,OOS}(R_{tree,378}^i) = 1 - \hat{\omega}^\top R_{tree,378}^i$.

Useful to understand the (conditional) dependence of the projected SDF on interactions of characteristics.

Could be used to test AP models OOS.

For consistency: Projections of the “full” SDF $m$ into each of the 36 subspaces of characteristics would have to coincide with the corresponding $m^{i,OOS}, i = 1, \ldots, 36$. 
Third version SDF: Hansen Jagannathan with pruned assets

3. Finally, there is the SDF that, for each XS, is a linear projection on the subspace of the 40 tree assets remaining after pruning:

\[ m^{i,HJ} = m(R^{i}_{tree,40}) = 1 - \tilde{\omega}^\top R^{i}_{tree,40}, i = 1, \ldots, 36. \]

- Both 2. and 3. use the same tree assets as basis assets. But in 3., the weights \( \tilde{\omega} \) come from an in-sample projection while in 2. \( \hat{\omega} \) maximizes the OOS Sharpe Ratio.

- Interesting object to test asset pricing models based on a narrow subset of characteristics.

- We should expect the SDFs in 2. \( m^{i,OOS} \) to be much smoother than the corresponding ones in 3., \( m^{i,HJ} \). I am curious to see how much smoother they will be.

- It would help to make explicit the difference between \( m^{i,OOS} \) and \( m^{i,HJ} \) in the paper.

- It could also be valuable to have a discussion about each SDF version and its usefulness in empirical applications.
How to measure informational content?

- **Goal 1**: Obtain, for each XS, 40 basis assets having similar informational content to the original 378 portfolios.

- **Goal 2**: Show, for each XS on a relative basis, that the 40 assets capture more complex information than traditional triple sorts (TS’s).

- **Metrics for comparisons**: Maximum Sharpe ratio (SR), and Pricing Errors (PEs) given by traditional linear factor models (FF-type).

- **Can we complement these metrics?**
Pricing Errors

► Their robust MV estimator eliminates a large part of the trees assets (due to the Lasso term). It naturally introduces pricing errors.

► What is the behavior and magnitude of the pricing errors?

► Behavior of P.E.’s: Using convex duality, we note that regularizing with a Elastic net metric $\lambda_2 \|\omega\|_2^2 + \lambda_1 \|\omega\|_1$ on the Mean-Variance problem induces pricing errors that are (quadratically) penalized only if they are larger than the Lasso regularization parameter $\lambda_1$. Do you see that in your data?

► Magnitude of P.E.’s: Could show the Mean Absolute Pricing Errors of the eliminated assets to give us an idea of how much smoothing of the SDF is obtained with the regularization procedure.
Could test three layers of PE’s:

1. **PE’s for the whole XS of stocks:**
   Obtained with the full linear SDF extracted from all tree portfolios of the 36 different XS’s (total of 36* 378 assets).

2. **PE’s for the trees portfolios:**
   For each XS \(i\), given the tree assets before pruning (378), PE’s obtained with the linear SDF based on that specific XS, \(m_{i,OOS}^{\text{tree}}\) with 40 assets.

3. **PE’s between TS and trees assets:**
   For each XS \(i\), given the 40 pruned trees assets and corresponding 64 TS, use their implied linear SDFs \(m_{i,OOS,\text{tree}}\) and \(m_{i,OOS,\text{TS}}\) to obtain the PE for the other set. That is, price \(R_{\text{tree,40}}^i\) using \(m_{i,OOS,\text{TS}}\) and \(R_{\text{TS,64}}^i\) using \(m_{i,OOS,\text{tree}}\).
Sharpe Ratios and HJ bounds

- By a simple duality argument, an equivalent metric to compare trees and TS assets is given by the minimum variance bounds of Hansen and Jagannathan.

\[
\min_{m \text{ that price the basis assets}} \frac{\sigma(m)}{E(m)} = \max_{\text{all excess returns on basis assets}} \frac{|E(R^e)|}{\sigma(R^e)}
\]

- I see two advantages in including this metric in the paper.

1. Given that the new assets should be useful to diagnose AP models, observing the two HJ bounds together (AP and TS bounds) provides information on the relative difficult for a model to pass each bound.

2. Obtain non-negative minimum variance (sharper) bounds that should be satisfied by non-negative admissible SDFs. More appropriate to test general / partial equilibrium models.

- Below I plot these bounds for TS and trees assets.
These bounds are based on 40 portfolios conditioned on the following characteristics: size, book-to-market and accruals.

As a simple illustration I calibrate $\beta$ and $\alpha$ for the CCAPM model $m(CG_t) = \beta(CG_t)^{-\alpha}$, to find the minimum risk aversion coefficient that would make it pass each bound. It was $\alpha = 44$ for TS assets and $\alpha = 48$ for trees assets.
Informational content: Higher moments matter!

- With all the conditional interactions, I expect trees assets to be highly skewed and/or leptokurtic. The same could be true for TS assets since they come from nonlinear estimators.

- By looking at a fixed XS, higher moments are important for both trees and TS assets, with many being highly skewed and leptokurtic.

- How much information beyond variance the trees assets bring w.r.t TS assets?
To take into account higher moments, could adopt either Generalized Sharpe Ratios (Cerny, 2003) or information theoretical bounds (Almeida and Garcia, 2017; Ghosh, Julliard and Taylor, 2017).

Below I plot as an example the KLIC information bounds (Stutzer, 1995) generated by the two sets of assets.

Note that the informational difference between the two sets of assets is smaller once we account for skewness and kurtosis.
Expected returns and characteristics: Conditional interactions and nonlinearities.

- Characteristics are important drivers of expected returns. Big debate on how to better specify this dependence.

- The authors characterize this dependence with a functional form $g(\cdot)$ to define the conditional expectation and a joint multivariate uniform in the space of characteristics.

$$
E_t(R_i) = g(C_{t,i}^1, \ldots C_{t,i}^k), \text{ with } f(C_{t,i}^1, \ldots, C_{K,t}) \approx U([0, 1], \rho) \quad (2)
$$

- The patterns of observed expected return are then jointly determined by $g(\cdot)$ and $f(\cdot)$.

- $f(\cdot)$ controls for sparsity/homogeneity (number of firms) in regions of the space of characteristics while $g(\cdot)$ and $f(\cdot)$ control together sparsity/concentration in the space of expected returns.
Conditional splits of trees are adequate to capture information about the joint conditional distribution of the characteristics.

1. Is it possible to verify how effective the method is with respect to nonlinearities versus conditional interactions in capturing the XS of expected returns?

2. Moreover, is it possible to use your method to disentangle the role of nonlinearities from that of conditional interactions?

A possible way to answer 1., could be to shut down each channel in simulations and see how well the method does in each case. The metric could be a comparison of the SR of the whole XS of simulated assets versus the SR of the optimal portfolio with pruned assets.

To give a start to solving 2. could use projected-PCA (Fan, Liao, Wang, 2016) to roughly estimate betas as nonlinear functions of the characteristics with interactions.
KNS propose to normalize each characteristic (to be in $[-1, 1]$) to create long-short portfolios (factors) based on these characteristics, their interactions (products $C_i C_j$, $i, j = 1, \ldots K$) and nonlinearities (quadratic ($C_i^2$, $i = 1, \ldots K$) and cubic ($C_i^3$, $i = 1, \ldots K$)).

Fixed a set of characteristics, the benchmark in terms of out of sample SR and P.E.'s should be a more general version of the SDF of KNS based on charac (not PCs) but pruned with your new robust MV.

Their method doesn’t capture conditional information as yours (the order of their interactions / products doesn’t matter) but it allows for general nonlinear forms as products of polynomials $C_i^l C_j^u$, $i, j = 1, \ldots K$, $l, u \in \mathbb{N}$.

Maybe comparing their SDF to yours may reveal something about the ability to capture/separate nonlinearities from conditional interactions.
Conclusion

▶ Extremely interesting paper!

▶ Steps in the direction of answering important open questions:

▶ How to build relevant XS’s? How do the SDF and expected returns depend on a given set of characteristics? How can we optimally use all the XS’s obtained?

▶ I am looking forward to a complete version and to using their basis assets in empirical applications!