

Authors: Svetlana Bruzgalova (London Business School), Markus Pelger (Stanford University), Jason Zhu (Stanford University)

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

Abstract. Sorting-based strategy of building portfolios has been a default empirical approach in asset pricing for creating both test assets and factor-mimicking returns. One of the natural limitations of this technique, however, is its inability to adequately reflect the information contained in more than 2 characteristics and their interaction. Yet recent advances in empirical asset pricing have repeatedly highlighted the importance of the latter, e.g. Freyberger et al (2017), Kozak et al (2018). We propose to analyze the effect of a large number of characteristics on expected stock returns with the machine learning technique known as random forest. As an ensemble learning method for classification, the new approach is particularly well-suited for building composite cross-sections of portfolios that reflect the rich conditional information contained in a large number of characteristics simultaneously, and can be viewed as a natural generalization of the conventional sorting-based strategies. We build decision trees for various sets of stock-specific characteristics, and demonstrate that the new approach is able to create cross-sections that a) reflect the information in a joint conditional distribution of characteristics, b) are challenging to price based on the conventional models, even when pitted against the tradable factors based on the underlying characteristics, and c) imply investment strategies that achieve yearly out-of-sample Sharpe ratios above 2.