Abstract

This thesis adds to the literature of option return predictability by leveraging supervised classification methods that have been left untouched until now. Using 766,524 option-month observations of S&P500 call options between 1996 and 2021, I show that nonlinear machine learning algorithms also outperform linear ones in classifying option returns. Further, I find that the class that captures the highest positive option returns is hard to predict, which is why long-short portfolios that avoid this class generate the statistically significant and economically highest risk-adjusted returns. Expected volatility and the uncertainty about it seem to play a role in explaining the long-short returns of the portfolio that additionally avoids the negative tail. Across all machine learning models, option features seem to be by far the most important, but stock features should still be included. Finally, I show that it can be beneficial to include a transformer in the model stack as it belongs to the models that perform the best.