How can machine learning advance quantitative asset management?*

David Blitz

Robeco, d.c.blitz@robeco.com

Tobias Hoogteijling

Robeco, t.hoogteijling@robeco.com

Harald Lohre

Robeco, Lancaster University Management School, h.lohre@robeco.com

Philip Messow

Robeco, p.messow@robeco.com

January 9, 2023

Journal of Portfolio Management, forthcoming.

Abstract

The emerging literature suggests that machine learning (ML) is beneficial in many asset pricing applications because of its ability to detect and exploit nonlinearities and interaction effects that tend to go unnoticed with simpler modelling approaches. In this paper, we discuss the promises and pitfalls of applying machine learning to asset management, by reviewing the existing ML literature from the perspective of a prudent practitioner. The focus is on the methodological design choices that can critically affect predictive outcomes and on an evaluation of the frequent claim that ML gives spectacular performance improvements. In light of the practical considerations, the apparent advantage of ML is reduced, but still likely to make a difference for investors who adhere to a sound research protocol to navigate the intrinsic pitfalls of ML.

Keywords: machine learning; asset management; portfolio management; factor investing

JEL Classification: G10, G12, G17

---

* The authors thank Guido Baltussen, Joris Blonk, Matthias Hanauer, Robbert-Jan ´t-Hoen, Fabio Martinetti, Frederik Muskens, and participants at the Robeco research seminars for valuable comments and suggestions. The views expressed in this article are not necessarily shared by Robeco or its subsidiaries.

1 Introduction

Machine learning (ML) has made inroads into many research areas to augment or improve upon human expertise. Activities that involve a lot of data and numbers are particularly amenable to benefit from the application of ML methods. Although finance and investment is ultimately a social science that revolves around human behavior, the use of quantitative methods and analysis has become firmly established. With the large amounts of data and numbers available, it goes without saying that many believe that ML has the potential to revolutionize the investment industry.

While ML theory has been around for many years, recent advances in (cloud) computing power have made it operationally feasible to assess how ML can contribute to investment management. The increasing popularity is reflected in the number of research papers published in recent years investigating the use of "artificial intelligence" and "machine learning" in quantitative asset management. Whereas just five years ago only a handful of studies on this topic were available, there is now a considerable body of literature that provides many interesting insights and keeps growing by the day.

Before the surge in computing power and available data, one would instead have resorted to an economic model to describe optimal decisions of rational individuals, rather than just letting the data speak in an ML model. In economics, theory prescribes the model, and data determines the associated estimates of model parameters. Such econometric modelling is looking to establish a relationship among features and target variables, but also whether these make sense; in other words, one is generally looking for causality and not just correlation. Machine Learning can be at an advantage over classical econometric approaches whenever the focus is on forecasting, especially in the absence of a theoretical model. It is well-equipped to deal with large and complex data sets, and particularly suitable for capturing nonlinearities and interaction effects.

In this article we describe the specific challenges and opportunities of ML methods in quantitative asset management and the road ahead from an institutional asset manager perspective. In the following section we explain the benefits and pitfalls of machine learning versus classical econometrics. Next, we discuss different practical design choices for applying ML models in quantitative asset management. The subsequent section focuses on the ability of ML to generate alpha in the stock market, but also discusses alternative applications, such as forecasting asset risk, constructing optimal portfolios, or optimally executing the resulting trades.

2 Machine Learning versus classical econometrics

The difference between machine learning and classical econometric models is often not well-defined. Herein, we define machine learning models as models that learn “by themselves” and then make predictions on unseen (out-of-sample) data. Classical econometric models are typically fit on the full set of data, without there being an explicit “learning” element. They tend to be relatively simple and are usually linear, such as an Ordinary Least Squares (OLS) regression. In this section, we contrast machine learning to econometrics approaches and discuss the resulting benefits and pitfalls of adapting a machine learning approach.

2.1 Benefits of Machine Learning

The advantages of ML models can roughly be grouped into three categories. First, ML methods are data-driven, which means that models are shaped depending on the nature of the data. ML is well-equipped to deal with large sets of features and "learns" to give the most weight to the most relevant variables. In contrast, simple linear regressions can quickly run into problems with large sets of explanatory variables, such as inflated standard errors and multicollinearity. Second, ML techniques are so-called "model-free" approaches that are not restricted to a specific functional form (Hastie, Tibshirani, and Friedman, 2009). In particular, nonlinearities or interaction effects can be identified in the data and modelled accordingly.

To illustrate, Exhibit 1 and 2 present stylized examples of nonlinearities and interaction effects that can be identified using machine learning models in the context of stock return prediction. Specifically, Exhibit 1 is looking into the prediction of returns based on dividend yield in an OLS regression. When fitting all observations, an OLS regression recovers a flat line, suggesting that there is no predictive relationship. However, there are a lot of companies that do not pay out dividends at all. Focusing on dividend-paying companies only reveals a positive relationship between dividend yield and return, which is a pattern that can be readily identified by ML models but goes unnoticed by OLS regression.

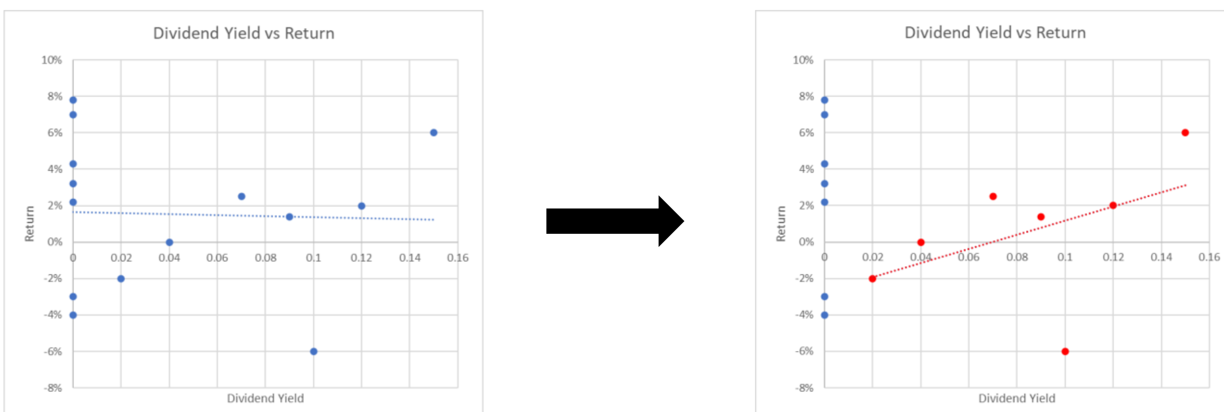

Exhibit 1: Stylized example of a nonlinear relationship between dividend yield and return that can be picked up by Machine Learning models.

Exhibit 2 illustrates a positive relationship between earnings-to-price and return. However, if one singles out observations for which earnings have been reported in the current month, we observe an interaction effect of earnings-to-price and earnings announcement month that can be uncovered with an ML model. In this example the positive relationship between earnings-to-price and return is stronger during earnings announcement months and much weaker in other months.

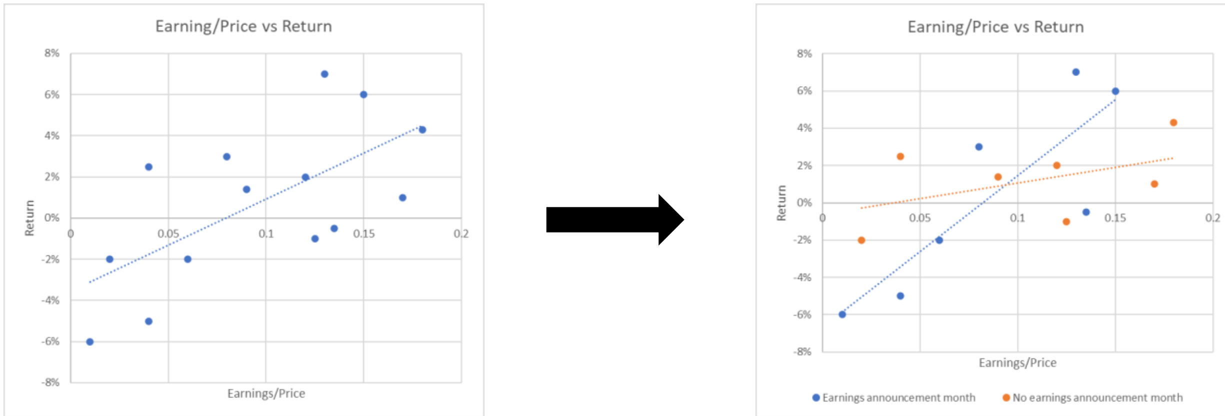

Exhibit 2: Stylized example of an interaction effect between earnings-to-price and earnings announcement month that can be picked up by Machine Learning models.

While nonlinearities and interaction effects could also be modelled in a regression setup, the researcher then needs to provide an exact specification, which is rarely straightforward and quickly inflates the pool of regressors. Conversely, ML methods enjoy greater flexibility. Neural networks, for example, can accommodate a very high degree of complexity, but they can also nest simple models such as linear regression.

Finally, ML algorithms are generally forward-looking, while traditional approaches select models based on in-sample fit. Using the same data set to both fit and evaluate models exacerbates the risk of p-hacking. The resulting models benefit from exploiting the full information set, but in-sample performance estimates tend to be biased upwards. In contrast, ML selects models based on out-of-sample predictive power and is thus less prone to such look-ahead biases. Although a properly calibrated ML signal will have a hard time competing in an in-sample backtest against a "simple" model that has been designed to navigate the in-sample period very well, the backtest of the ML model gives a more realistic assessment of out-of-sample predictive power.

2.2 Pitfalls of Machine Learning

Nevertheless, there are also pitfalls in the use of ML. For example, ML models are particularly useful for applications with a large amount of data and a high signal-to-noise ratio. In financial market research, however, the data sets are comparatively small and the signal-to-noise ratio tends to be low. Moreover, many data points are correlated, both in the time series and in the cross-section, which further reduces the effective number of observations.

Importantly, financial markets are constantly evolving, and we might see detected anomalies being arbitraged away over time. Israel, Kelly, and Moskowitz (2020) thus conceive an analogy to image recognition (a field of research that is particularly amenable for ML methods to shine): in financial markets, all cats might morph into dogs once the algorithm has learned how to determine a cat in an image, and the algorithm must start learning all over again. This analogy cautions that the relevance of past data points is not constant, since the data-generating process may change over time. It is therefore not surprising that financial markets are characterized by a very low signal-to-noise ratio.

One of the main concerns with machine learning is the risk of overfitting, which means models fit patterns that are present in a specific sample, but do not hold in general. On the one hand, machine learning models are more susceptible to overfitting because they come with more degrees of freedom than simple linear models. On the other hand, the focus on out-of-sample predictability reduces the risk of overfitting. Furthermore, there are many tools to combat overfitting, such as a train-validation split, regularization, early stopping and drop-out (Hastie, Tibshirani, and Friedman, 2009). In this vein, one could argue that basic ML algorithms such as the regularized Ridge and Lasso regressions are actually less prone to overfitting than traditional linear regression models.

It is worth remembering that traditional methods also suffer from overfitting bias when considering variables in a backtest before they were discovered in the literature. The aforementioned "model-free" approach can therefore be seen as both a blessing and a curse. Without a predefined structure, ML models are only interested in the best prediction, regardless of whether this makes economic sense. Heaton, Polson and Witte (2017) emphasize that, in finance, "*the focus is not on replicating tasks that humans already do well. Unlike recognizing a picture or responding appropriately to verbal prompts, humans do not have the innate ability to, for example, select a stock that is likely to perform well in a future period*." Nevertheless, ML methods have shown particular promise in replicating human behavior. Moreover, the increasing complexity of ML models leads to an increasing sensitivity to (erroneous) data, which underlines the importance of data quality.

3 Modeling choices in Machine Learning for asset management

Although ML models are data driven, the user still needs to make various important choices. These design choices can have a large impact on the quality of the model prediction. In general, many of these questions can be answered empirically, provided one has enough data. However, (independent) data is one of the biggest barriers in quantitative asset management, rendering these design choices even more important.

Furthermore, open source software is readily available for all common ML applications in free programming languages such as R or Python. Yet, this availability can be considered a blessing and a curse. While one can thus quickly come up with naïve ML-driven predictions, it takes domain knowledge and expertise in making meaningful design choices to guide a given ML algorithm. In this section, we discuss the most important of these design choices.

3.1 Methodological choices

ML studies typically apply a strict separation between training and validation data (to calibrate the model) as well as out-of-sample data (to evaluate its predictions), see Hastie, Tibshirani, and Friedman (2009). The training period can make use of an expanding window or a moving window. With an expanding window the model will converge towards selecting factors and factor interactions that were most effective overall. A moving window approach is more adaptive and better equipped to deal with variables that only become available later in the sample, but it reduces the number of observations and can also result in a more procyclical model.

The researcher can also choose whether to fit one model on the entire investment universe, or whether to let the ML algorithm fit different models for, e.g., different regions or sectors. Ultimately, this choice reflects a trade-off. On the one hand, training on a subsample will single out the patterns most relevant for the specific subset. On the other hand, patterns can hold more or less in general and focusing on a subset then leads to a large loss of informative training data, see Haley, Norvig, and Perreira (2019).

The choice can also be left to the ML algorithm, e.g., by making region or sector information an input variable. However, just including such information as a feature along with all other variables is not necessarily the best approach. For example, tree-based methods iteratively search to explain most variation in the target variable based on whether variables are higher or lower than a certain threshold. If the relationship between certain features and targets is different from one sector to the other, this may not be reflected in the decision trees if the target values are not very different across the two sectors.

Note that when information such as country or region is included in the feature set, and companies from a given country (say US) happen to have higher returns than those of other countries in the training data, the model might learn that simply being a US stock is a favorable characteristic on its own. If such outcomes are considered undesirable, it is better to train the model on stock returns relative to the returns of other stocks in the same country or region.

3.2 Choice of target

In line with the classic asset pricing literature, ML studies typically focus on predicting 1-month ahead stock returns. As a result, these models tend to prefer variables with strong short-term predictive power that involve high turnover, such as the well-known short-term reversal effect. This is not to say that ML models are less suitable for long-term return prediction per se, but a potential problem is that long-term return predictions require many more years of training data for the same number of independent observations. For example, 10 years of training data translate to 120 independent monthly observations, but only 10 independent yearly observations. For any estimation method, a time series of just 10 observations is likely to be too limited. While one could use overlapping observations to increase the sample size, these observations would not be independent.

Hence, given the limited available data history in finance, applying Machine Learning with sufficiently long training and testing samples is more challenging for long-term return prediction than for short-term return prediction. Switching from short-term to long-term return prediction is reflected in the importance of characteristics for the machine learning predictions. Long-term equity return predictions tend to relate more strongly to slow-moving factors, such as low volatility and beta; see for example Li, Simon, and Turkington (2022). Instead of considering forecast horizons longer than 1 month, one can also go in the opposite direction. Ait-Sahalia, Fan, and Xue (2022) study ultra-high frequency stock returns and find that ML models give rise to predictability that is large, systematic, and consistent over short horizons.

If one wishes to generate predictions that are less correlated with 'traditional' factors, one should adjust the target accordingly. One option is to cross-sectionally orthogonalize targets towards well-known factors. An example of this research direction is MSCI (2021), who find that an isolated, orthogonalized ML component capturing nonlinearities is among the strongest return predictors. Such cross-sectional orthogonalization is effective in achieving cross-sectional uncorrelatedness, but it does not necessarily translate into uncorrelatedness over time. Also, this approach is potentially subject to estimation error in the orthogonalization and ignores possible changes in factor dynamics over time.

Given the choice of target return, most ML studies focus on minimizing a loss function such as the mean squared error between predicted and realized returns. Wang et al. (2019) and Liu, Zhou, and Zhu (2021) propose alternative ML methods that attempt to directly maximize portfolio risk-adjusted returns and find that these give superior Sharpe ratios. Standard loss functions may result in top-minus-bottom quintile portfolios that are not necessarily optimal from a Sharpe ratio perspective. As a result, a variable which gives 3% alpha at a volatility of 6% (Sharpe ratio = 0.5) may be preferred over a variable with an alpha of 2% at a volatility of just 2% (Sharpe ratio = 1.0). However, predictions at the portfolio level suffer from a large loss in the number of training observations when moving from a whole cross-section of observations to just one observed portfolio return. Focusing on Sharpe Ratios or Information Ratios reduces the number of observations even more, as multiple time periods are translated into just one SR or IR observation.

3.3 Choice of input features

ML models tend to generate the best results with a broad and diverse set of input variables. In the next section we will discuss a large number of studies that follow this approach, and which typically claim striking improvements in predicting stock returns compared to traditional linear models.

However, there are also studies that use more narrow sets of input variables, for instance by focusing on past returns only. Moritz and Zimmermann (2016) use a tree-based approach to predict future stock returns from past stock returns, while Murray, Xiao, and Xia (2020) take inspiration from classic technical analysis and derive signals from past price plots. Both studies report strong results, but these are partly driven by picking up well-known return-based factors, such as short-term reversal, medium-term momentum, and seasonal effects. However, the authors claim that their ML methods offer added value on top of the implicit exposures to these classic return-based signals. In a related vein, Fisher and Krauss (2018) use deep learning on past prices for statistical arbitrage, but conclude that "*as of 2010, excess returns seem to have been arbitraged away with LSTM (Long Short-Term Memory) profitability fluctuating around zero after transaction costs*".

Instead of technical indicators one can also consider other input feature datasets. An example is Bloomberg (2020) who use more than 300 ESG variables to predict stock returns. They report a Sharpe ratio of 1.25 for a long-short ESG portfolio based on their model, albeit over a short period from 2014 to 2018.

In addition to cross-sectional firm characteristics, one can also feed ML models with macroeconomic information to further improve model predictions, see Gu, Kelly, and Xiu (2020) among others. However, given the limited number of business cycles in typical samples, this raises the same concern as with classical factor timing studies, namely that the additional use of macroeconomic information may well amplify overfitting.

4 Applying ML in asset management

In this section, we review and discuss the application of ML in asset management. Importantly, investors want to rationalize the resulting ML-driven investment portfolios and the first part of this section will therefore speak to the notion of interpretability in ML models. Next, we discuss various use-cases for ML in asset management. The first-order concern here is forecasting stock returns, hence it is not surprising to find many studies exploring ML in this context even though the associated low signal-to-noise ratio attenuates the expected benefits. Therefore, we also discuss use-cases that go beyond forecasting returns, where the caveat of low signal-to-noise ratios applies less. Lastly, we touch on the need for establishing and following a rigorous protocol to avoid the common pitfalls in applying ML for asset management.

4.1 The virtue of interpretability

The interpretability of the model and its outcome is of great relevance because it adds to the credibility of ML models in asset management applications. Indeed, the ability to rationalize a model's outcome and to shape a corresponding narrative sets humans apart from machines (Seiler, 2018), enabling asset managers to engage and build trust with clients. In that regard, Shapley values are an important tool for the interpretation of machine learning models (Shapley, 1953). Roughly speaking, they indicate for every observation how much each feature contributes to the prediction. Shapley values have a rigid mathematical foundation, and one can show that they are the only attribution method that satisfies the four axioms of fairness: efficiency, symmetry, dummy, and additivity (Molnar, 2022). Without getting into mathematical details, Shapley values therefore allow for making statements of the following kind: “We predict this stock to have an outperformance of 3.5%, of which 2% is due to value, 1% to low volatility and 0.5% to momentum.”

To illustrate the use and interpretation of Shapley values, we have created a boosted regression tree model that predicts one-month ahead relative returns. As described in more detail in the Appendix, we largely follow the setup of Leung et al. (2021), using the features book-to-price, market capitalization, profitability, investments, momentum, and short-term reversal. Exhibit 3 presents a Shapley force plot that attributes the difference between the average target value and a specific prediction to various features in the model. In this example, the average target has the standardized value of 0.194 and the specific prediction is slightly higher than average (0.20). The positive standardized momentum score (0.83) pushes the predicted value up, while the negative book-to-price score (-0.86) pushes the predicted value down.

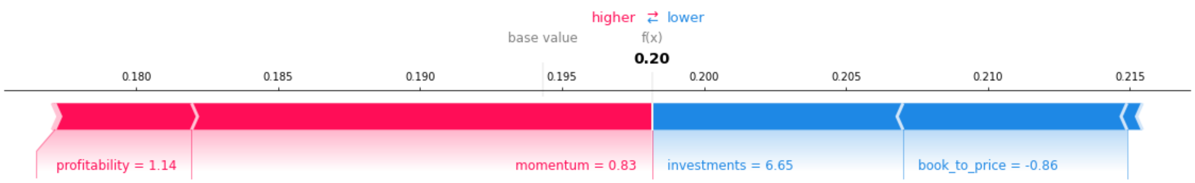

Exhibit 3: A Shapley force plot for a boosted regression tree model predicting 1 month ahead standardized returns. It illustrates how the difference between a prediction and the average target value can be decomposed in feature contributions.

In Exhibit 4, we illustrate how nonlinearities and interaction effects in Machine Learning return prediction models can be visualized using Shapley values. Shapley values are shown on the y-axis, with a Shapley value above 0 indicating that a feature has a positive impact on model predictions. The chart on the left indicates that there tends to be a positive relation between momentum and return predictions, consistent with well-known price momentum effects. However, this relationship is nonlinear, with the strongest effect being in the short leg. We also observe a reversal rather than a momentum effect for the stocks with the lowest momentum scores. The chart on the right shows an interaction effect between investments and momentum. Generally, the Shapley plot indicates that high investment tends to be followed by low future returns. However, this effect is stronger for stocks with poor momentum (blue dots) than for stocks with strong momentum (red dots).

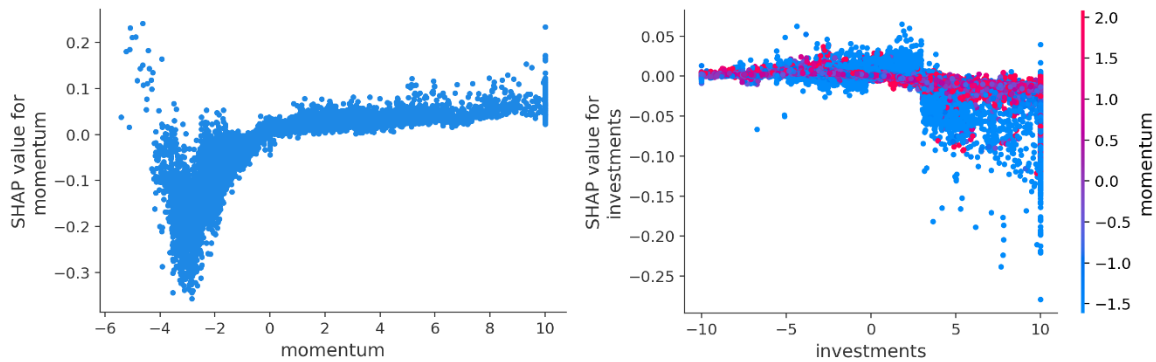

Exhibit 4: A Shapley scatter plot (left) and a Shapley dependence plot (right) for a boosted regression tree model predicting 1 month ahead standardized returns.

Another advantage of Shapley values is a fair attribution to all features in the data set (Molnar, 2022). Unfortunately though, Shapley values take rather long to compute, which leads to the need for approximations (Lundberg and Lee, 2017). Another important technical caveat that Shapley values share with other sampling-based attribution methods is the ignorance of feature correlations (Molnar, 2022) which can lead to either over- or underestimation of feature importance. This issue can be countered by conditional sampling, which, in turn, hinges on accurate estimation of correlations.

A different route is taken by Li, Turkington, and Yazdani (2022), who propose a technique called "model fingerprint" to decompose the results of complex models into subcomponents. This fingerprint method is independent of the employed model and aims to provide insights into the contribution of both global and local predictors, separating linear and non-linear effects for each feature and the interaction effects for each set of features. However, this approach is strongly related to partial dependence analysis and thus focuses on estimating the average marginal effect of a feature, keeping other features constant. Such an approach ignores correlation in features and heterogeneity in effects (if the effect of a variable is sometimes negative and sometimes positive, it might cancel out on average). Jensen et al. (2022) introduce a new measure for interpreting ML models called "economic feature importance" which is a refinement of regular feature importance based on economic rationales incorporated in the loss function.

Against this backdrop, there are some studies that promote to also integrate economic plausibility into the loss function as this helps to considerably increase the interpretability of models and results. For instance, Cong et al. (2022) use ML for directly determining portfolio weights based on stock characteristics, which reduces the importance of technical predictor variables. Adding economic structure has two key advantages. First, it reduces complexity. An economic structure can guide the ML model which therefore requires less data to forecast. Second, ML models often detect predictability of returns that is difficult to harvest in practice, i.e., predictability does not guarantee profitability. Explicitly accounting for turnover or using an asymmetric loss function that weighs observations differently to limit shorting are simple examples of integrating economic structure. Such economic insights prove similarly important as the technical specifications of the ML model.

Of course, complicated and computation-intensive methods are not a must to interpret machine learning predictions. Ultimately, any model produces alpha signals, which can be characterized using the classical toolkit. That is, one can resort to signal correlations with existing factors, spanning regressions, return correlations, or conditional performance analysis. Notwithstanding, it is clear that interpretability and performance attribution is one of the virtues of using simpler (linear) models. While advances are being made regarding the interpretability of ML models, it is a given that more complicated models are harder to understand. When investors want to move beyond simple factor-based models, they will have to accept that interpretation will become more involved. Of course, investors can gradually gain trust in machine learning models if they experience a good live track record.

Lastly, x-raying ML models furthers confidence in their workings and plausibility and can also improve our understanding of more complex market dynamics that are hard to detect in linear models. Hence, even if one is not going to solely rely on ML model decisions, one might learn about important diversifying features or meaningful interactions that can also help to enrich a simple traditional framework.

4.2 Use cases for Machine Learning in asset management

4.2.1 Forecasting stock returns

Various studies report promising results when predicting stock returns using ML with a large set of traditional predictor variables as input features, such as past return signals, valuation ratios, quality metrics, or risk measures. The number of input features included is typically over 50. Three of the most cited papers in this area are Gu, Kelly, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), and Light, Maslov, and Rytchkov (2017), who all find that ML strategies substantially outperform comparable linear strategies, generating Sharpe ratios of about 2 or higher. Notably, Baltussen, van Vliet, and van Vliet (2022) find that ML methods were effective in predicting stock returns in the "pre-CRSP" era, going back all the way to 1866.

Other recent ML studies that report strong stock selection results based on a broad set of input features are Rasekhschaffe and Jones (2019), Messmer (2017), Chen, Pelger, and Zhu (2022), and Bryzgalova, Pelger, and Zhu (2019). International evidence is provided by Tobek and Hronec (2021), European evidence by Drobetz and Otto (2021), emerging market evidence by Hanauer and Kalsbach (2022), and Chinese evidence by Leippold, Wang and Zhou (2022). Moreover, the Gu, Kelly, and Xiu (2020) model has already been extended in Kelly, Pruitt, and Su (2019) and Gu, Kelly, and Xiu (2021). Guijarro-Ordonez, Pelger, and Zanotti (2022) also present results in favour of ML methods, yet by pursuing an alternative route. They generalize different approaches to statistical arbitrage using deep learning, which combines global dependency patterns with local filters. This means that some patterns are found to exist in general, whereas other patterns and nuances only hold for a subset of the data. A trading strategy based on this model is found to significantly outperform the benchmark.

As explained earlier, adding economic structure to ML models has key advantages. For instance, Chen, Pelger, and Zhou (2022) use a no-arbitrage condition as part of a neural network to predict asset prices. They can thus extract the state of the economy by identifying the factors driving asset prices, which outperforms all competing out-of-sample benchmark approaches. Kozak, Nagel, and Santosh (2020) and Lettau and Pelger (2020) exploit the economic insight that the first and second moment of stock returns should be related. This economically and empirically motivated prior is used to construct a new stochastic discount factor that leads to better out-of-sample results. Jensen et al. (2022) combine several ideas in the literature on how to adapt ML methods to the economic realities of financial markets. They argue that portfolio optimization that integrates transaction costs works better than first fitting ML stock return predictions and then implementing these in a portfolio.

It is important to note that there are also studies that are more cautious regarding the added value of machine learning. Novy-Marx (2015) already expressed the general concern that "*strategies selected by combining multiple signals suffer severe overfitting biases, because underlying signals are typically signed such that each predicts positive in-sample returns*". He shows that even with randomly generated signals it is easy to generate highly significant performance in backtests. While this concern also applies to traditional multi-factor models, the additional degrees of freedom that come with ML approaches exacerbate the issue. Martin and Nagel (2021) make a similar argument and explicitly link the overfitting concern to the rise of big data.

ML studies try to mitigate such concerns by using training sets and cross-validation techniques, with strict in-sample versus out-of-sample testing, see for instance Arnott, Harvey, and Markowitz (2019). When applied correctly, cross-validation is strictly out-of-sample and there is no forward-looking bias. However, researchers can still cherry-pick from multiple tested configurations, play around with the set of hyperparameters, or even subconsciously steer the training process in a direction that they know would work well. Further, any experiment that includes a set of factors that are already known to perform well suffers from a forward-looking bias by design. For a more extensive discussion of why ML methods may fail in practice, see Lopez de Prado (2018).

Some papers specifically address the high Sharpe ratios reported in the empirical ML literature. Leung et al. (2021) conclude that "*ML models are superior to traditional linear models in predicting cross-sectional one-month-ahead stock returns using a set of well-documented stock characteristics*" but immediately downplay this conclusion by adding that "*however, the extent to which this statistical advantage can be translated into economic gains in portfolio backtests depends critically on the ability to take risk and implement trades efficiently*." Ultimately, they conclude that "*machine learning models have been somewhat more effective over the past decade at discerning valuable predictions from cross-sectional equity characteristics*".

Avramov, Cheng, and Metzker (2022) are even more critical, concluding that *"deep learning signals extract profitability from difficult-to-arbitrage stocks and during high limits-to-arbitrage market states. In particular, excluding microcaps, distressed stocks, or episodes of high market volatility considerably attenuates profitability. Machine learning-based performance further deteriorates in the presence of reasonable trading costs due to high turnover and extreme positions."*

4.2.2 Enhancing traditional factors

Many traditional factors hinge on simple fundamental firm characteristics. ML can be used to enhance the corresponding factors by providing better forecasts of the underlying fundamental metric. The key advantage of applying ML to forecasting fundamentals is that the signal-to-noise ratio tends to be higher. For instance, Bew et al. (2019) use ML to extract more value from analyst recommendations, solely relying on analyst data as input. Snow (2020) builds on ML to predict earnings surprises, not only leveraging earnings data but also past returns and other technical indicators such as trading volume. Vlan Binsbergen, Han, and Lopez-Lira (2022) create better earnings forecasts than analysts by utilizing an ML model based on fundamental variables, such as growth metrics, as well as macroeconomic variables. Anand et al. (2019) use ML to predict profitability using several dozen fundamental variables as input.

However, predictability in fundamentals does not necessarily translate to predictability in returns. Even if perfect foresight of future earnings is a source of alpha, and if future earnings are to some extent predictable, there is no guarantee that earnings predictions hold alpha. The reason is that the predictable earnings component might well be priced by the market, and the perfect foresight alpha basically derives from future surprises in earnings.

In a related vein, Hanauer, Kononova, and Rapp (2022) use ML for fair value estimation and thus improve on corresponding linear approaches as in Bartram and Grinblatt (2018). Geertsema and Lu (2022) also find strong results for a relative valuation signal based on a decision-tree ML approach. UBS (2021) use ML to construct a growth factor, for which they report a Sharpe ratio of 0.81 since 2010. However, it must be noted that growth factors did well over this period in general.

A general caveat of these studies is that factor improvements are to be expected when one adds variables that are known to predict future stock returns and future fundamental performance. For instance, a value signal that (implicitly) incorporates some notion of quality, momentum, or low-risk factors is likely to outperform a pure value signal by design. Evidence that a combination of factors outperforms individual factors is given in, e.g., Houweling and Van Zundert (2017) and Blitz, Hanauer, Honarvar, Huisman, and Van Vliet (2022).

4.2.3 Creating non-traditional variables and NLP

ML is frequently used to construct non-traditional variables. A lot of progress has been made especially in the field of Natural Language Processing (NLP), which can be considered both a separate field or a subfield of machine learning. Using deep neural networks, NLP techniques code words and sentences as numbers, and perform 'math with words'. NLP techniques enable converting text-based data into quantitative investment signals. Examples are the application of NLP to SEC filings data, earnings call transcripts, news articles, or patent data. Thus, ML can help turn unstructured data into structured data. An extensive discussion is beyond the scope of this article, but we refer to Zing, Cambria, and Welsch (2018) for a good overview of financial forecasting with NLP.

4.2.4 Predicting metrics other than return

Instead of predicting stock returns, ML can also be used to predict other relevant metrics, such as market betas (Jourovski et al., 2020) or stock volatilities (Filipovic and Khalilzadeh, 2021). However, the same caveat that applies to enhancing factors applies here as well, namely that simply incorporating some information from other variables such as value and momentum can already lead to better risk forecasts.

Furthermore, ML models can help predict corporate events, such as share buyback announcements, dividend announcements, or M&A announcements, see Ding et al. (2015) and Bhatia (2020). Another important use-case is the prediction of companies' sustainability scores, e.g., Chen et al. (2022) assess companies' alignment with the Sustainable Development Goals (SDGs) of the United Nations by applying NLP to corporate sustainability disclosures. The challenge with evaluating the quality of such models is that the 'true' scores are not unambiguously known.

4.2.5 Using ML methods in fixed income

Machine learning techniques have obviously also found their way into fixed income research and we highlight some notable angles. For instance, Cherief et al. (2022) use boosted regression trees and random forests to identify nonlinearities and interaction effects between risk factors in the EUR and USD credit space and compare the resulting performance relative to traditional linear methods when forecasting credit excess returns. They ultimately propose to combine linear and nonlinear methods, by first utilizing a linear Lasso approach and then estimating their tree-based model based on the error of the linear approach. According to the authors, combining linear and nonlinear methods increases the stability and robustness of their model.

Bali et al. (2022) confirm the efficacy of ML models in predicting corporate bond returns. They study random forests, feed-forward neural networks, and long short-term memory neural networks and document significant improvements in the out-of-sample performance of stock and bond features when predicting future bond returns. Furthermore, imposing economic structure based on the Merton model the ML models continue to significantly add value despite this restriction. Recently, Kaufmann, Messow, and Vogt (2021) apply boosted regression trees to improve existing stock momentum factors. The addition of liquidity and size effects speeds systematic improvements in performance, as the interactions of momentum factors with both size and liquidity can be explicitly accounted for.

One of the most widely cited studies on machine learning applied to fixed income is Bianchi, Büchner, and Tamoni (2021), who document strong out-of-sample predictive power of government bond returns using neural networks, boosted trees, random forests and extreme trees. However, a subsequent corrigendum (Bianchi, Büchner, Hoogteijling, and Tamoni, 2021) concludes that actually none of the tree-based methods have predictive power, highlighting the need to handle ML in finance with care.

4.2.6 Improving portfolio construction

With financial markets becoming more complex over time, traditional risk models may develop certain blind spots. Bartram, Branke, and Motahari (2020) argue that ML risk models can extract such information more efficiently from multiple sources of structured and unstructured data to produce more accurate forecasts of bankruptcy and credit risk, market volatility, macroeconomic trends, or financial crises. Indeed, there are strong parallels between weight optimization in machine learning and weight optimization in portfolio management, both choosing weights to minimize a loss function. Thus, ML models can be considered a natural extension of traditional portfolio construction.

Reflecting on modern portfolio construction, Snow (2020) reports a shift from Markowitz mean-variance optimization to reinforcement learning techniques that apply a more holistic approach. Specifically, reinforcement learning is an area of machine learning that deals with how intelligent agents should act in an environment to maximize the notion of cumulative reward. In the context of portfolio construction, reward is typically a performance metric such as Sharpe or Information Ratio. Snow (2020) also mentions other techniques that can be applied in portfolio construction, such as autoencoders in *deep* portfolios. An autoencoder is an unsupervised neural network that tries to model the full panel of inputs by only using the inputs themselves. Gu, Kelly, and Xiu (2021) describe autoencoders as the nonlinear counterpart to PCA. In particular, one thus avoids to explicitly model the variance-covariance structure. We refer to Heaton, Polson, and Witte (2017) for a thorough treatment of the use of autoencoders in constructing deep portfolios.

Another unsupervised learning technique deriving from graph theory has become popular in constructing hierarchical risk parity strategies, see Lopez de Prado (2016). It synthesizes the dependence structure among assets in terms of the so-called minimum spanning tree that is naturally linked to the hierarchical clustering of the underlying asset universe. Allocating capital along these clusters in a risk-based fashion is a robust way to construct diversified portfolios, see Lohre, Rother and Schaefer (2020) for an application in the context of multi-asset multi-factor investing.

4.2.7 Improving trading and execution

Given the increasing speed and complexity of trading, ML techniques are becoming an essential part of trading practices. In particular, they can reduce transaction costs by automatically analyzing the market and then determining the best time, size, and place to trade. Ritter (2017) shows that reinforcement learning, specifically Q-learning, can help solve the problem of discovering and implementing dynamic trading strategies in the presence of transaction costs. Q-learning is a model-free reinforcement learning algorithm that explicitly takes into account the current state and tries to find the best course of action conditional on this state.

Moallemi and Wang (2022) use a novel neural network architecture that combines two data-driven approaches—one based on supervised learning and one based on reinforced learning—to optimize execution timing, and thus obtain a significant cost reduction. Ning, Ho Ting Lin, and Jaimungal (2021) also address optimal trade execution using a double-deep Q-learning model. The input features are the current state of the limit order book, other trading signals, and available execution actions, while the output is the Q-value function that estimates future rewards for any given action. Applying their model to several stocks, the authors find it outperforming the standard benchmark approach. Similarly, ML techniques have also made inroads in corporate bond trading. Fedenia, Nam, and Ronen (2021) model the trade direction of US corporate bonds and find that random forests are best suited for this purpose.

4.3 Research governance and protocol

To successfully navigate the described promising landscape of ML modelling and to not fall for the dangers of data mining, a rigorous research governance and protocol is needed. The foundations of a healthy research protocol for quantitative investment managers have been laid out by Arnott, Harvey, and Markowitz (2019), who elaborate on seven pillars. First and foremost, they highlight that the motivation for research should always follow a sound rationale. Second, the more different strategies are tested, the higher the risk of false positives, so researchers should make sure to keep track of what has been tried and apply proper statistical methods, especially with respect to multiple testing biases. Third, data quality and sample choice are crucial ingredients to evaluating a given method. The latter also relates to the fourth pillar which advocates the use of cross-validation. Here, machine learning research is typically applying more rigor when compared to classic empirical finance work, as one is working with training and validation samples to calibrate a model that is then evaluated over an out-of-sample. Yet, the authors rightfully caution that, in investment management, the only true out of sample is live trading. Pillars 5 and 6 concern the dynamics and complexity that a given model should encompass. The authors' clear advice is to refrain from overfitting model dynamics and to pursue simplicity and regularization. Finally, an asset manager needs to establish a research culture that rewards research that is following the above quality criteria (rather than one that implicitly promotes the generation of overfitted investment strategies and signals that are a sure way to fail out of sample).

5 Conclusions

The extant evidence suggests that machine learning can boost quantitative investing by uncovering exploitable nonlinear patterns and interaction effects in the data. Being mindful of a positive publication bias, we caution that ML is not a panacea, as users need to make important methodological choices, the models can overfit the data, and they are based on the premise that past relations will continue to hold in the future.

Many studies that report strong results for ML models focus on predicting next 1-month returns based on a large number of traditional factor characteristics as input features. Although the models load on traditional short-term return predictors they are able to exploit additional nonlinear alpha opportunities. The challenge with these models is to turn the resulting fast alpha signals into a profitable investment strategy after costs and other real-life implementation frictions. The corresponding literature is scarce, and the few works naturally suggest that the opportunity set for ML models to outperform traditional ones is often reduced given the reliance of ML models on high-turnover signals. For this reason, there has recently been an effort to integrate economic structure into loss functions to have the ML model focus on better tradable stocks. Such efforts are expected to increase the likelihood of monetizing the predictive power of ML models.

Apart from forecasting returns, there are other promising use cases for ML, such as enhancing traditional factors, creating new variables from unstructured data, and predicting metrics other than return, such as risk or sustainability. So far, ML methods in asset management have therefore been more of an evolution than a revolution. Presumably, asset managers who will disregard advances in ML will see their performance wane relative to those who embrace ML. Naturally, mostly big institutional players can enjoy economies of scale in this competition given the high costs of running such operations in practice. The ability to automate tasks of traditional analysts, such as reading, seeing or hearing ultimately promises large gains in productivity—provided the asset manager possesses the necessary infrastructure and can investigate different big data sets and signals at scale. Yet, discarding economic theory altogether and turning to a fully data-driven approach can vice versa also set one up for failure. Investors can identify and evaluate the ability of an asset manager to succeed in advancing his investment process accordingly by scrutinizing what research protocol is in place. The latter is key to the success of ML in practice and to navigate the many pitfalls.

Altogether, researchers have just been scratching the surface of the endless possibilities offered by machine learning, and many exciting new discoveries can be expected in the years ahead. However, human domain knowledge is likely to remain important, because the signal-to-noise ratio in financial data is low, and the risk of overfitting is high.

Appendix

We construct a boosted regression tree model to predict one-month ahead relative returns, closely following the setup outlined in Leung et al. (2021). The predictive features are book-to-price, market capitalization, profitability (net profit margin), investments, 12-1 month momentum and 1-month reversal. Features with a skewness above 1 are transformed with a logarithm. Features and target are cross-sectionally standardized using robust z-scores. For missing values, the cross-sectional median is imputed. The model is trained on monthly data over the period 2001-01-01 till 2010-12-31, using all constituents of the MSCI World Index. The hyperparameters are set to a learning rate of 0.01, a maximum depth of 6 and a number of trees of 100. All other parameters are set to default value. Models have been created using the Python xgboost package v1.4.2 and the shap package v0.41.0. Return and fundamental data are sourced from Refinitiv.

References

Aít-Sahalia, Y., Fan, J., and Xue, L., 2022. "How and When are High-Frequency Stock Returns Predictable?." SSRN working paper no. 4095405.

Anand, V., Brunner, R., Ikegwu, K., and Sougiannis, T., 2019. "Predicting Profitability Using Machine Learning." SSRN working paper no. 3466478.

Arnott, R., Harvey, C., and Markowitz, H., 2019. "A Backtesting Protocol in the Era of Machine Learning." *Journal of Financial Data Science* 1(1), 64-74.

Avramov, D., Cheng, S., and Metzker, L., 2022. "Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability." *Management Science*, forthcoming.

Bali, T., Goyal, A., Huang, D., Jiang, F., and Wen, Q., 2022. "Predicting Corporate Bond Returns: Merton Meets Machine Learning." SSRN working paper no. 3686164.

Baltussen, G., van Vliet, B., and van Vliet, P., 2022. "The Cross-Section of Stock Returns before 1926 (and Beyond)." SSRN working paper no. 3969743.

Bartram, S., and Grinblatt, M., 2018. "Agnostic Fundamental Analysis Works." *Journal of Financial Economics* 128(1), 125-147.

Bartram, S., Branke, J., and Motahari, M., 2020. "Artificial Intelligence in Asset Management." CFA Institute Research Foundation, August 2020.

Bew, D., Harvey, C., Ledford, A., Radnor, S., and Sinclair, A., 2019. "Modelling Analysts' Recommendations via Bayesian Machine Learning." *Journal of Financial Data Science* 1(1), 75-98.

Bhatia, T., 2020. "Modeling Multifactor Event Driven Trading Strategy." SSRN working paper no. 3683454.

Bianchi, D., Büchner, M., Hoogteijling, T., and Tamoni, A. (2021). "Corrigendum: Bond Risk Premiums with Machine Learning." *Review of Financial Studies* 34(2):1090–1103.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). "Bond Risk Premiums with Machine Learning." *Review of Financial Studies* 34(2):1046–1089.

Blitz, D., Hanauer, M. X., Honarvar, I., Huisman, R., and van Vliet, P. (2022). Beyond Fama-French Factors: Alpha from Short-Term Signals. Available at SSRN 4115411.

Bloomberg, 2020. "Seeking Signals from ESG Data." Bloomberg Quant Research, April 2020.

Bryzgalova, S., Pelger, M., and Zhu, J., 2019. "Forest through the Trees: Building Cross-Sections of Stock Returns." SSRN working paper no. 3493458.

Chen, L., Pelger, M., and Zhu, J., 2022. "Deep Learning in Asset Pricing." *Management Science*, forthcoming.

Chen, M., Mussalli, G., Amel-Zadeh, A., and Weinberg, M., 2022. "NLP for SDGs: Measuring Corporate Alignment with the Sustainable Development Goals." *Journal of Impact and ESG Investing* (forthcoming).

Cherief, A., Ben Slimane, M., Dumas, J.-M., and Fredj, H., 2022. "Credit Factor Investing with Machine Learning techniques.", Amundi Asset Management, June 2022.

Cong, L., Tang, K., Wang, J., Zhang, Y., 2022. "AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI." SSRN working paper no. 3554486.

Ding, X., Zhang, Y., Liu, T., and Duan, J., 2015. "Deep Learning for Event-Driven Stock Prediction." *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (IICAI 2015), 2327-2333.

Drobetz, W., and Otto, T., 2021. "Empirical Asset Pricing via Machine Learning: Evidence from the European Stock Market." *Journal of Asset Management* 22, 507-538.

Fedenia, M., and Nam, S., and Ronen, T., 2021. "Machine Learning in the Corporate Bond Market and Beyond: A New Classifier." SSRN working paper no. 3848068.

Filipovic, D., and Khalilzadeh, A., 2021. "Machine Learning for Predicting Stock Return Volatility." Swiss Finance Institute Research Paper Series No. 21-95.

Fischer, T., and Krauss, C., 2018. "Deep Learning with Long Short-Term Memory Networks for Financial Market Prediction." *European Journal of Operational Research* 270(2), 654-669.

Freyberger, J., Neuhierl, A., and Weber, M., 2020. "Dissecting Characteristics Nonparametrically." *Review of Financial Studies* 33(5), 2326-2377.

Geertsema, P., and Lu, H., 2022. "Relative Valuation with Machine Learning." *Journal of Accounting Research*, forthcoming.

Gu, S., Kelly, B., and Xiu, D., 2020. "Empirical Asset Pricing via Machine Learning." *Review of Financial Studies* 33(5), 2223-2273.

Gu, S., Kelly, B., and Xiu, D., 2021. "Autoencoder Asset Pricing Models." *Journal of Econometrics* 222(1B), 429-450.

Guijarro-Ordonez, J., Pelger, M., and Zanotti, G., 2022. "Deep Learning Statistical Arbitrage." SSRN working paper no. 3862004.

Hanauer, M., and Kalsbach, T., 2022. "Machine Learning and The Cross-Section of Emerging Market Stock Returns." SSRN working paper no. 4287550.

Hanauer, M., Kononova, M., and Rapp, M., 2022. "Boosting Agnostic Fundamental Analysis: Using Machine Learning to Identify Mispricing in European Stock Markets." *Finance Research Letters* 48, 102856.

Halevy, A., Norvig, P., and Pereira, F.. "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems* 24, 8-12, 2009.

Hastie, T., Tibshirani, R., and Friedman, J. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Spring Series in Statistics, 2009

Heaton, J., Polson, N., and Witte, J., 2017. "Deep Learning for Finance: Deep Portfolios.", *Applied Stochastic Models in Business and Industry* 33(1), 3-12.

Houweling, P., and Van Zundert, J. (2017). "Factor Investing in the Corporate Bond Market." *Financial Analysts Journal* 73(2), 100-115.

Israel, R., Kelly, B., Moskowitz, T., 2020. "Can Machines "Learn" Finance?." *Journal of Investment Management* 18(2), 23-36.

Jensen, T., Kelly B., Malamud, S., Pedersen, L., 2022. ''Machine Learning and the Implementable Efficient Frontier." SSRN working paper no. 4187217.

Jourovski, A., Dubikovskyy, V., Adell, P., Ramakrishnan, R., and Kosowski, R., 2020. "Forecasting Beta Using Machine Learning and Equity Sentiment Variables." *Machine Learning for Asset Management: New Developments and Financial Applications*, First Edition, edited by E. Jurczenko.

Kaufmann, H., Messow, P., and Vogt, J., 2021. "Boosting the Equity Momentum Factor in Credit." *Financial Analysts Journal* 77(4), 83-103.

Kelly, B., Pruitt, S., and Su, Y., 2019. "Characteristics *Are* Covariances: A Unified Model of Risk and Return." *Journal of Financial Economics* 134(3), 501-524.

Kozak, S., Nagel, S., and Santosh, S., 2020. "Shrinking the cross-section." *Journal of Financial Economics* 135(2), 271-292.

Leippold, M., Wang, Q., and Zhou, W., 2022. "Machine Learning in the Chinese Stock Market." *Journal of Financial Economics* 145(2), 64-82.

Lettau, M., and Pelger, M., 2020. "Estimating Latent Asset-Pricing Factors." *Journal of Econometrics* 218(1), 1-31.

Leung, E., Lohre, H., Mischlich, D., Shea, Y., and Stroh, M., 2021. "The Promises and Pitfalls of Machine Learning for Predicting Stock Returns." *Journal of Financial Data Science* 3(2), 21-50.

Li, Y., Turkington, D., and Yazdani, A., 2022. "Beyond the Black Box: An Intuitive Approach to Investment Prediction with Machine Learning*", Journal of Financial Data Science* 2(1), 61-75.

Li, Y., Simon, Z., and Turkington, D., 2022. "Investable and Interpretable Machine Learning for Equities." *Journal of Financial Data Science* (forthcoming).

Light, N., Maslov, D., and Rytchkov, O., 2017. "Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach." *Review of Financial Studies* 30(4), 1339-1381.

Liu, Y., Zhou, G., and Zhu, Y., 2021. "Maximizing the Sharpe Ratio: A Genetic Programming Approach." SSRN working paper no. 3726609.

Lohre, H., Rother, C., and Schaefer, K., 2020. "Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations" *Machine Learning for Asset Management: New Developments and Financial Applications*, First Edition, edited by E. Jurczenko.

López del Prado, M., 2016. "Building Diversified Portfolios that Outperform Out of Sample.", *Journal of Portfolio Management* 42(4), 59-69.

López del Prado, M., 2018. "The 10 Reasons Most Machine Learning Funds Fail.", *Journal of Portfolio Management* 44(6), 120-133.

Lundberg, M. and Lee, S. (2017). "A Unified Approach to Interpreting Model Predictions". *Advances in Neural Information Processing Systems* 30, 4765–4774.

Martin, I., and Nagel. S., 2021. "Market Efficiency in the Age of Big Data." *Journal of Financial Economics* (forthcoming).

Messmer, M., 2017. "Deep Learning and the Cross-Section of Expected Return." SSRN working paper no. 3081555.

Moallemi, C., Wang, M., 2022. "A Reinforcement Learning Approach to Optimal Execution." *Quantitative Finance* 22(6), 1051-1069.

Moritz, B., and Zimmermann, T., 2016. "Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns." SSRN working paper no. 2740751.

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). Available at christophm.github.io/interpretable-ml-book/

MSCI, 2021. "Machine Learning Factors: Capturing Nonlinearities in Linear Factor Model." MSCI Research Insights, March 2021.

Murray, S., Xiao, H., and Xia, Y., 2020. "Charting by Machines." Working paper, available at http://www.apjfs.org/resource/global/cafm/2020-14-1.pdf

Ning, B., Ho Ting Lin, F., Jaimungal, S., 2021. "Double Deep Q-Learning for Optimal Execution." *Applied Mathematical Finance* 28(4), 361-380.

Novy-Marx, R., 2015. "Backtesting Strategies Based on Multiple Signals." SSRN working paper no. 2629935.

Rasekhschaffe, K., and Jones, R., 2019. "Machine Learning for Stock Selection." *Financial Analysts Journal* 75(3), 70-88.

Ritter, G., 2017. "Machine Learning for Trading." *Risk* 30(10), 84-89.

Seiler, D., 2018. "Artificial Intelligence in Investing." Vontobel Asset Management, May 2018.

Shapley, S. (1953). "A Value for n-person Games". In Kuhn, H. W.; Tucker, A. W. (eds.). *Contributions to the Theory of Games. Annals of Mathematical Studies* 28, 307–317. Princeton University Press.

Snow, D., 2020. "A Surprising Thing: The Application of Machine Learning Ensembles and Signal Theory to Predict Earnings Surprises." SSRN working paper no. 3420722.

Snow, D., 2020. "Machine Learning in Asset Management: Part 2: Portfolio Construction – Weight Optimization." *Journal of Financial Data Science* 2(2), 17-24.

Tobek, O., and Hronec, M., 2021. "Does It Pay to Follow Anomalies Research? Machine Learning Approach with International Evidence." *Journal of Financial Markets* 56, 100588.

UBS, 2021. "Growth Investing via Machine Learning." UBS Global Quantitative Research, 8 February 2021.

Van Binsbergen, J., Han, X., and Lopez-Lira, A., 2022. "Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases." *Review of Financial Studies*, forthcoming.

Wang, J., Zhang, Y., Tang, K., Wu, J., and Xiong, Z., 2019. "AlphaStock: A Buying-Winners-and-Selling-Losers Investment Strategy using Interpretable Deep Reinforcement Attention Networks." arXiv working paper no. 1908.02646.

Zing, F., Cambria, E., and Welsch, R., 2018. "Natural Language Based Financial Forecasting: A Survey." *Artificial Intelligence Review* 50, 49-73.