This chapter is to provide a survey of how machine learning techniques – in excess of standard econometric approaches – can be used for the benefit of empirical asset pricing.

The basic asset pricing equation is one of the linchpins of financial research:

\[ \mathbb{E}_t[R_{t+1}^i m_{t+1}(\delta)] = 1, \]  

where \( R_{t+1}^i \) refers to the gross return of asset \( i \) and \( m_{t+1}(\delta) \) denotes the stochastic discount factor parameterized by \( \delta \). When specified correctly, \( m_{t+1}(\delta) \) should be able to price all assets. Empirical analyses based on Eq. (1) are hampered by the fact that (a) the above expression refers to conditional moments, (b) \( m_{t+1}(\delta) \) might be a highly complex function, and (c) this function might change over time. Furthermore, even after restricting oneself to a particular functional form of \( m_{t+1}(\delta) \), estimation of \( \delta \) – for example using the Generalized Method of Moments (GMM) – and subsequent testing of the model requires the choice of a set of test assets.

The search for specifications of the stochastic discount factor – or at least components thereof – has led to a veritable factor zoo that has been described by Harvey et al. (2015) and Feng et al. (2020). The challenges listed above, combined with this vast set of variables that might or might not contribute in some way to the stochastic discount factor, have brought forth a vibrant literature on empirical asset pricing that promotes a (financial) theory-guided use of machine learning techniques to either support or replace classical econometric analyses. To provide an comprehensive overview of this potential of including machine learning techniques in the empirical
assessment of asset pricing models, the chapter is to cover the following topics:¹

1. *Introduction*

   E.g., introduce the basic asset pricing equation, the factor zoo, concepts like managed portfolios, no-arbitrage condition, and estimation techniques conventionally used (e.g., GMM) to create a unified framework for the following subsections to rely on.

2. *How machine learning techniques can help identify the stochastic discount factor*

   One particular example for this line of literature would be Chen et al. (2021) who use a combination of three different neural networks (Long-Short-Term-Memory, Feed-Forward, and an Adversarial Network). The Adversarial Network allows them to build on Hansen’s (1982) GMM and focus on those moments that are associated with the strongest mispricing. The authors compare their identification strategy favorably to standard econometric approaches and stress the importance of economic guidance of machine learning methods by means of the no-arbitrage condition. Another example of a study belonging to this strand of literature (and also relying on the no-arbitrage condition) is presented by the Autoencoder-based approach by Gu et al. (2021).

3. *How machine learning techniques can be used to test/evaluate asset pricing models*

   For example, focusing on model evaluation, Bryzgalova et al. (2020) argue that the test assets conventionally used in empirical asset pricing studies do not only fail to provide enough of a challenge for the models under consideration but also contribute to the growing factor zoo. To circumvent this issue, they

¹ For illustrative purposes, the following overview contains brief examples of papers that should be mentioned in the respective subsections.
propose to use tree-based methods for the purpose of constructing interpretable and hard-to-price test assets.

4. How machine learning techniques can be applied for prediction problems in the context of empirical asset pricing

The focus of Chapter 12 will be on the preceding subsections. However, one of the studies mentioned here should be that by Gu et al. (2020).

References


URL: http://dx.doi.org/10.2139/ssrn.3493458


URL: http://dx.doi.org/10.2139/ssrn.3350138


URL: https://doi.org/10.1093/rfs/hhaa009


URL: https://ideas.repec.org/a/eee/econom/v222y2021i1p429-450.html


URL: http://www.jstor.org/stable/1912775

**URL:** [https://doi.org/10.1093/rfs/hhv059](https://doi.org/10.1093/rfs/hhv059)