Stock Return Predictability:

Review of Empirical Asset Pricing via Machine Learning

Raúl Correa Ocañas

Essex Business School, University of Essex

BE352: Asset Pricing

Dr. Lazaros Symeonidis

January 22, 2026

Word count: 2403

1. Introduction

One of the main objectives in the field of asset pricing is to explain the cross-section of expected returns using economically meaningful sources of risk. Traditional empirical frameworks typically approach this task through linear factor models, in which expected excess returns are expressed as a combination of compensations for exposure to a small number of systematic risk factors. While this approach has provided influential benchmarks, such as the Capital Asset Pricing Model and multifactor models, it has struggled to account for the growing body of documented, empirical trends. Over time, the literature has identified many return predictors based on stock characteristics, leading to a rise of concerns about *false discovery* of factors and model misspecification. “Given the plethora of factors, and the inevitable data mining, many of the historically discovered factors would be deemed “significant” by chance” (Harvey, Liu & Zhu, 2016, p. 36).

A key challenge is that expected returns are highly dependent on the abundance of quality information. Firm characteristics, macroeconomic variables, and their interactions together may determine both risk exposures and prices of risk. However, often empirical implementations impose strong restrictions on its functional form, such as the assumption of linearity and low dimensionality. Cochrane (2005) described in “20.2 The Cross Section: CAPM and Multifactor Models” how poor empirical performance may emerge from restrictive modelling choices rather than fundamental failures of asset pricing theory.

Gu, Kelly, and Xiu (2020) address this challenge by proposing a general empirical asset pricing framework in which expected returns are modelled via an unknown conditional pricing function. Instead of exploring new economic factors, the paper reframes the empirical task as one of estimating this function while preserving the no-arbitrage logic of stochastic discount factor models. This way, machine learning methods are employed as tools for spanning a rich space of transformations of observable characteristics.

The central research question of this review is: **How can empirical asset pricing models approximate the conditional pricing function when expected returns depend on high-dimensional information, and what does this imply for factor models and asset pricing anomalies?**

This essay examines the methodology and findings of Gu, Kelly, and Xiu (2020) within the broader asset pricing literature. Section 2 outlines the asset pricing framework and its relation to stochastic discount factor models. The empirical design and main findings are then discussed and critically evaluated. Finally, the implications for factor models and anomalies are explored, followed by finishing remarks and scope for further empirical analysis.

2. Asset Pricing Framework and Methodology

2.1 Expected Returns and the Stochastic Discount Factor

Modern asset pricing theory is built around the existence of a stochastic discount factor that prices all assets in the economy. For any asset i, excess returns satisfy the no-arbitrage condition $E\left[m_{t+1} r_{i,t+1} \middle| F_t\right] = 0$, where m_{t+1} denotes the SDF and F_t represents the information set available at time t. Under standard regularity conditions, this restriction implies that expected excess returns can be expressed as functions of conditional covariances between returns and the SDF. However, the SDF is unobserved in empirical settings, and asset pricing tests are typically conducted by modelling expected returns directly as conditional expectations (Cochrane, 2005, Ch.1 & Ch.6).

Gu, Kelly, and Xiu (2020) adopt this empirical perspective by focusing on the conditional mean of excess returns $r_{i,t+1} = E\left[r_{i,t+1} \middle| F_t\right] + \varepsilon_{i,t+1}$, where $E\left[\varepsilon_{i,t+1}, \middle| F_t\right] = 0$. This decomposition emphasizes that asset pricing models explain expected returns rather than realized outcomes. Approaching empirical asset pricing from this formulation shows that pricing errors arise from misspecification of the conditional expectation, and not from predictable innovations or violations of no-arbitrage restrictions.

2.2 Conditional Pricing Function and Relation to Factor Models

The main methodological contribution is to formalize empirical asset pricing as a problem of approximating an unknown conditional pricing function. Expected excess returns can then be expressed as $E\left[\mathrm{r}_{i,t+1} \middle| F_t\right] = \mathrm{g}^*\left(\mathrm{z}_{\mathrm{i,t}}\right)$, where $z_{i,t}$ denotes the observable information set constructed as $z_{i,t} = x_t \otimes c_{i,t}$. Here, $c_{i,t}$ includes firm characteristics such as size, momentum, and valuation ratios, while x_t captures macroeconomic conditions. This structure reflects the economic intuition that firm characteristics act as a proxy for time-varying risk exposures, while macroeconomic variables capture the time variation in prices of risk.

Although the framework allows for flexible pricing functions, it remains grounded in standard asset pricing logic. Since conditional expectations are linear operators, this implies that any admissible pricing function can be represented as a linear combination of enhanced transformation of the features. When nonlinearity is discussed regarding expected returns, it is regarding nonlinear transformations of observable variables, not through departures from the linear pricing paradigm itself.

The proposed pricing formulation encapsulates traditional linear factor models as a special case restricted to being linear in a small number of factors: $E\left[r_{i,t+1}\middle|F_t\right] = \beta'_{i,t}\lambda_t$. Under further assumptions that risk exposures and prices of risk are linear functions of characteristics and macroeconomic variables respectively, the general pricing function $g^*\left(z_{i,t}\right)$ reduces to a linear function of the combined information set. In this sense, empirical factor models impose strong restrictions on how information enters expected returns (Fama & French, 1993, Section 2). Poor empirical performance of such models may stem from these restrictions rather than from the absence of priced risk.

2.3 Machine Learning as Function Approximation

Within this setting, machine learning methods are introduced as tools for approximating the unknown pricing function $g^*(\cdot)$ in high-dimensional environments. Specifically, tree-based methods and neural networks generate nonlinear transformations and interactions while controlling overfitting through regularization and validation.

Crucially, these methods do not change the underlying economic object of interest. Expected returns remain conditional expectations consistent with an SDF representation, and model evaluation focuses on pricing errors rather than trading profitability. Any improvements in out-of-sample prediction performance are interpreted as evidence of reduced approximation error in the pricing function, not as a market efficiency infringement. This interpretation positions the contribution within empirical asset pricing, with machine learning as a flexible estimation strategy that is sustained via economic theory.

3. Data and Empirical Design

The empirical analysis uses monthly U.S. equity data from CRSP ranging from 1957 to 2016. The sample includes all common stocks listed on NYSE, AMEX, and NASDAQ exchanges. Excess returns are measured with respect to the one-month Treasury bill rate.

The information set consists of two components. First, a large set of firm-level characteristics drawn from the cross-sectional asset pricing literature, covering categories such as size, value, momentum, profitability, investment, liquidity, and return volatility. Second, a small set of aggregate macroeconomic state variables commonly used to capture time variation in expected returns, including valuation ratios, interest rate spreads, and measures of market volatility. All variables are lagged appropriately to reflect realistic information availability and to avoid look-ahead bias (Gu, Kelly & Xiu, 2020, Section 2).

The sample is divided into training, validation, and testing subsamples that respect temporal ordering. Model parameters and regularization hyperparameters are selected using rolling validation windows, and performance is evaluated over a long out-of-sample testing period. Models are re-estimated periodically rather than recursively each month.

Model performance is assessed using out-of-sample R^2 measures that quantify the extent to which estimated conditional expectations explain variation in realized excess returns. As specified before, improvements in these metrics are interpreted as reductions in pricing errors associated with misspecification of the conditional pricing function. Portfolio-level analyses provide complementary evaluations, assessing whether improvements at the stock level translate into economically meaningful reductions in pricing errors for aggregated portfolios.

4. Review and Critical Evaluation of Findings

4.1 Return Predictability and Model Misspecification

A central empirical finding is that flexible estimation of the conditional pricing function leads to economically meaningful improvements in out-of-sample explanatory power compared to traditional linear models. Unrestricted linear regressions that include the full set of firm characteristics perform poorly out of sample, exhibiting strongly negative R^2 values, thus reinforcing earlier concerns about severe overfitting. In contrast, constrained linear models that impose regularization or dimensionality reduction (ENET, PLS, PCR) achieve modest but positive out-of-sample R^2. However, their explanatory power remains limited, suggesting that restricting dimensionality alone is insufficient. Nonlinear methods, specifically tree-based models and neural networks, deliver the strongest performance, with substantial increases in monthly out-of-sample R^2 relative to both unrestricted and regularized linear benchmarks. From an asset pricing perspective, these results support the interpretation that traditional models fail primarily due to overly restrictive functional form assumptions.

The best performing model for all stocks is a neural network with 3 hidden layers, which obtained an OOS coefficient of determination of 0.40% at the month-level. At the annual-level, a neural network of 4 hidden layers has the highest R^2, being 3.60%.

4.2 Feature Importance and Economic Interpretation

The paper provides insight into which types of information are most relevant for explaining expected returns. Using a variable-importance procedure based on setting individual characteristics to zero and measuring the resulting deterioration in performance, the authors identify several groups of characteristics that consistently contribute to return predictability. The most important predictors fall into four broad categories: price trends and momentum, liquidity measures, risk proxies such as volatility and market beta, and valuation signals. These findings are broadly consistent with established empirical regularities (Fama & French, 2015, Section 2). An important distinction arises between linear and nonlinear models in how information is weighted. Linear models emphasize momentum and reversal characteristics, while nonlinear models distribute importance more evenly across variables. This suggests that nonlinear estimators capture interactions and conditional effects suppressed in linear specifications. The results of the feature importance analysis open the discussion of whether anomalies may be the result of complex interactions among characteristics, or isolated sources of mispricing.

4.3 Portfolio-Level Implications

To assess the economic relevance of stock-level improvements, the authors examine portfolio-level performance. Since portfolios aggregate individual assets, they provide a more rigorous test of whether estimated expected returns capture systematic variation over idiosyncratic noise. Across pre-specified portfolios, nonlinear models achieve substantial improvements in out-of-sample explanatory power. These gains translate into higher Sharpe ratios and lower return volatility, indicating that improvements are economically meaningful when aggregated. However, it is important to assess with caution portfolio-level success, because dependence across stock returns means that strong individual predictions do not mechanically guarantee portfolio-level accuracy. The results demonstrate that flexible aggregation of information can improve empirical approximations to expected returns at economically meaningful levels of aggregation.

4.4 Critical Assessment and Limitations

Despite its strengths, the empirical evidence raises several questions. First, while neural networks often outperform linear benchmarks, differences between tree models and neural networks are not always statistically significant. The implication of this is that the key improvement comes from allowing for nonlinear interactions on the factors. Second, the analysis primarily focuses on monthly horizons, and statistical comparisons at annual frequencies are less extensively documented. Given that asset pricing implications often concern long-horizon expected returns, further evaluation at lower frequencies would strengthen the interpretation. Third, while the feature-importance analysis provides valuable insights, it still is model-dependent. The interpretation of importance scores is less straightforward in nonlinear settings, and causal conclusions about the pricing role of individual characteristics remain elusive. Future work could explore if unsupervised learning methods can uncover any hidden structure in the data that further clarifies the economic sources of predictability.

5. Implications for Factor Models and Anomalies

The findings have important implications for how empirical asset pricing models should be interpreted and evaluated. Traditional factor models impose strong restrictions on how information enters the pricing relation. By expressing expected returns as linear functions of a limited number of factors, these models implicitly assume that risk exposures and prices of risk are both low-dimensional and linearly related to observable characteristics. The empirical success of flexible estimation methods suggests that such restrictions may be overly restrictive. Poor performance of linear factor models may therefore reflect misspecification of the pricing function rather than the absence of economically meaningful risk compensation (Cochrane, 2005, Ch. 20).

From this perspective, there may be several asset pricing anomalies that would not need to be interpreted as evidence of mispricing or market inefficiency. Instead, they may arise because linear factor models fail to adequately aggregate information contained in firm characteristics and macroeconomic conditions. The fact that nonlinear models capture return predictability without introducing new economic variables supports the view that anomalies reflect complex interactions among known predictors rather than distinct sources of priced risk. This interpretation aligns with concerns raised in the *factor zoo literature* about distinguishing genuine risk factors from statistical artifacts (Harvey, Liu & Zhu, 2016).

Critically, the framework does not reject factor models outright but nests them as special cases within a more general conditional pricing framework. Factor models remain useful as summaries of systematic variation, particularly when interpretability and economic intuition are prioritized. However, their empirical limitations should be understood considering restrictive assumptions they impose. Standard linear pricing tests may understate the explanatory power of observable information if the true pricing function is nonlinear. By framing empirical asset pricing as a problem of approximating a high-dimensional conditional pricing function, the paper provides a coherent explanation for why return predictability can coexist with no-arbitrage restrictions and stable risk premia.

6. Conclusion and Scope for Extension

This essay reviewed Gu, Kelly, and Xiu (2020) through the lens of empirical asset pricing, focusing on how expected returns can be modelled when they depend on a high-dimensional information set. Rather than proposing new risk factors or alternative asset pricing theories, the paper reframes empirical asset pricing as the problem of approximating an unknown conditional pricing function implied by the stochastic discount factor. Within this framework, machine learning methods are introduced as flexible estimation tools that relax restrictive functional form assumptions while preserving the no-arbitrage logic of standard asset pricing models.

The empirical evidence reviewed suggests that the weak performance of traditional linear factor models may reflect functional form misspecification rather than a lack of economically meaningful information. Flexible estimators are shown to substantially reduce pricing errors, particularly for large and liquid stocks and at the portfolio level, without relying on return innovations or trading strategies. Importantly, the results are broadly consistent with established empirical regularities, reinforcing the interpretation that machine learning complements rather than overturns existing asset pricing insights.

A natural direction for further analysis is to examine how sensitive these conclusions are to alternative information sets, estimation horizons, or restrictions on model flexibility. The empirical extension in the Appendix explores one such dimension by comparing linear and nonlinear approximations to the conditional pricing function under a controlled and transparent design. This extension aims to assess whether the key insights of Gu, Kelly, and Xiu persist under simplified settings and to further clarify the economic interpretation of flexible empirical asset pricing models.

7. References

Cochrane, J.H. (2005). *Asset Pricing: (Revised Edition)*. Princeton: Princeton University Press.

Fama, E.F. and French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), pp.3–56. doi:https://doi.org/10.1016/0304-405X(93)90023-5.

Fama, E.F. and French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), pp.1–22. doi:https://doi.org/10.1016/j.jfineco.2014.10.010.

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, [online] 33(5), pp.2223–2273. doi:https://doi.org/10.1093/rfs/hhaa009.

Harvey, C.R., Liu, Y. and Zhu, H. (2016). … and the Cross-Section of Expected Returns. *Review of Financial Studies*, [online] 29(1), pp.5–68. doi:https://doi.org/10.1093/rfs/hhv059.

Appendix A: Empirical Extension

You can find the documentation and scripts for the empirical extension here.