Routledge Taylor & Francis Group

The European Journal of Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rejf20

Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective

Andreas G. F. Hoepner , David McMillan , Andrew Vivian & Chardin Wese Simen

To cite this article: Andreas G. F. Hoepner , David McMillan , Andrew Vivian & Chardin Wese Simen (2021) Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective, The European Journal of Finance, 27:1-2, 1-7, DOI: 10.1080/1351847X.2020.1847725

To link to this article: https://doi.org/10.1080/1351847X.2020.1847725

THE EUROPEAN JOURNAL OF FINANCE
2021, VOL. 27, NOS. 1–2, 1–7
https://doi.org/10.1080/1351847X.2020.1847725

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective

Andreas G. F. Hoepner [a,b], David McMillan [c], Andrew Vivian [d] and Chardin Wese Simen [e]

[a]Michael Smurfit Graduate Business School & UCD Lochlann Quinn School of Business, University College Dublin. Dublin, Ireland, EU; [b]Technical Expert Group on Sustainable Finance, DG FISMA, Brussels, Belgium, EU; [c]Department of Accounting and Finance, University of Stirling, Stirling, UK; [d]School of Business and Economics, Loughborough University, Loughborough, UK; [e]Management School, University of Liverpool Liverpool, UK

ABSTRACT
Although machine learning is frequently associated with neural networks, it also comprises econometric regression approaches and other statistical techniques whose accuracy enhances with increasing observation. What constitutes high quality machine learning is yet unclear though. Proponents of deep learning (i.e. neural networks) value computational efficiency over human interpretability and tolerate the 'black box' appeal of their algorithms, whereas proponents of explainable artificial intelligence (xai) employ traceable 'white box' methods (e.g. regressions) to enhance explainability to human decision makers. We extend Brooks et al.'s [2019. 'Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics?' *European Journal of Finance* 25 (17): 1627–36.] work on significance and relevance as assessment critieria in econometrics and financial data science to contribute to this debate. Specifically, we identify explainability as the Achilles heel of classic machine learning approaches such as neural networks, which are not fully replicable, lack transparency and traceability and therefore do not permit any attempts to establish causal inference. We conclude by suggesting routes for future research to advance the design and efficiency of 'white box' algorithms.

ARTICLE HISTORY
Received 22 October 2020
Accepted 26 October 2020

KEYWORDS
explainability; explainable artificial intelligence (xai); neural networks; relevance; regressions; significance

1. Introduction

Machine learning occurs whenever a 'computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E' (Mitchell 1997, 14). According to this common definition, machine learning would occur if an econometric regression model operationalised in a computer code (e.g. in Python or R) had identified an authentic economic relationship T that it could forecast with increasing accuracy P as the amount of data / degrees of freedom E increases. In other words, many econometric regression models constitute machine learning. Hence, it is unsurprising that regressions repeatedly feature in reviews of machine learning algorithms although the majority of machine learning research is conducted with neural networks, support vector machines and other classification algorithms (Basha and Rajput 2019; Mahdavinejad et al. 2018; Portugal, Alencar, and Cowan 2018).

Recently, machine learning has experienced a strong trend towards explainable artificial intelligence, so called XAI, which classifies neural networks, support vector machines or tree ensembles as so called 'black box' machine learning algorithms, since they are hard to interpret and practically impossible to trace. In contrast, 'white box' machine learning methods such as regressions or Bayesian models have gained popularity recently, which balance computational power and human interpretability and allow for attempts of causal inference (Barredo Arrieta et al. 2020; Yang et al. 2019). In other words, XAI substantially increases attention on

CONTACT Andreas G. F. Hoepner andreas.hoepner@ucd.ie

regression type models well known to econometricians and financial data scientists (Barredo Arrieta et al. 2020). In fact, '[t]here has been a trend of moving away from blackbox models towards white-box models, particularly for critical industries such as healthcare [and] finance' (Loyola-Gonzalez 2019, 154096). While recent papers on XAI explore and define over a dozen concepts such as trustworthiness, informativeness or decomposability (e.g. Barredo Arrieta et al. 2020), we are to the best of our knowledge not aware of a single paper published on the topic in a finance journal.

Hence, in introducing this special issue, we extend the work of Brooks et al. (2019) and discuss the concepts of significance, relevance and, in particular, explanainability from a statistical and societal (e.g. economic) perspective. While econometricians and financial data scientists routinely assess the statistical significance of their models and increasingly measure the economic significance by examining the sign and magnitude of the parameter estimate, the concept of relevance is gaining some traction via a desire to support evidence-based policy making or research impact assessments such as the UK's Research Excellence Framework.[1]

In statistical terms, relevance and significance are seamlessly separatable, with statistical significance describing the confidence in the meaningfulness of the coefficient estimate and relevance measuring and comparing the extent of variation explained by multiple coefficients.[2] Statistical relevance thereby indicates how commonly the respective coefficient occurs as explanation of the dependent variable's behaviour. In other words, while coefficients compete for statistical relevance, their statistical significance is, in principle, indifferent of each other[3] and largely dependending on the available degrees of freedrom (Brooks et al. 2019).

In economic terms, we observe a similar differentiation, whereby the coefficients' economic significance is, in principle, indifferent of each other, while economic relevance allows for a direct comparison. Economics significance, also called economic substance, usually describes the meaningfulness of the coefficient in terms of impact on the dependent variable, preferably via scaling the coefficient by the dependent variable's standard deviation. Scaling multiple regression coefficients by the dependent variable's standard deviation in a linear regression setting assesses how meaningful a one unit change in the respective coefficient would be for the dependent variable. However, it does not allow to compare the coefficients' relevance among each other, as it makes no statement regarding how often such a one unit change occurs. Hence, economic relevance can be measured by scaling the multiple regression coefficients in the standard deviation of the respective independent variable. Consequently, economic relevance allows to compare coefficients on the extent to which they can commonly impact the dependent variable.[4] We summarise this classification visually in Table 1.

Explainability, however, might appear an obvious virtue to applied econometricians. In the context of machine learning and its prime method 'neural networks', which have been popularised in finance academia by Gu, Kelly, and Xiu (2020) and in finance practice by López de Prado (2019), explainability is far from easy, as the explainable artificial intelligence (xai) research movement demonstrates (e.g. Barredo Arrieta et al. 2020; Rai 2020). While neural networks are expected to replace humans for any mental task which 'a typical person can do ... with less than one second of thought' (Ng 2016, 2), neural networks have substantial weaknesses too, especially when compared with econometric regression approaches. Four weaknesses are effectively the Achilles heel of neural networks.

First, neural networks have a random seed and are hence not entirely replicable. Every repetition of a neural network with a different random seed will lead to slightly diverging results, substantially different results if the gradient descent does not converge (near) optimal. Second, neural networks lack transparency by design, as Gary Marcus explains:

> In their current incarnation, deep [neural network] learning systems have millions or even billions of parameters, identifiable to their developers not in terms of the sort of human interpretable labels ... but only in terms of their geography within a complex network (e.g. the activity value of the ith node in layer j in network module k). ... The transparency issue ... is a potential liability when using deep learning for problem domains like financial trades or medical diagnosis, in which human users might like to understand how a given system made a given decision. (Marcus 2018, 10–11)

Consequently, statistical explanations of the decision processes resulting from neural networks are (very) limited.

Third, given this lack of practical transparency, human users cannot trace individual aspects of the decision process. Such inability for spot checking is unlikely to build trust towards users in credence services such as

Table 1. Assessment of significance, relevance and explainability of machine learning algorithms in statistical and economic/societal terms.

Discipline (horizontal) | Concept (vertical)	Statistical	Economic/Societal
Significance	Conventional statistical significance levels of 1, 5% and 10% may need to strengthen to 0.1%, 0.5% and 1% given the vastly increasing statistical power of big data. (see Brooks et al. 2019, Table 1)	The effect size of estimated coefficients scaled by the standard deviation of the dependent variable assesses the economic or societal significance of the coefficients, since it measures the coefficients' impact on the dependent variable.
Relevance	Since individual coefficients' probabilities of being statistically significant increase with statistical power, their statistical relevance becomes crucial, which can be measured using Shapley Values.	The effect size of the estimated coefficient scaled by the standard deviation of the respective independent variable assesses the relevance that a feasible change (i.e. one standard deviation) in the respective independent variable has on the dependent variable.
Explainability	To ensure explainability from a statistical perspective, a machine learning algorithm needs to be fully transparent and entirely replicable.	To ensure explainability from an economic/societal use case, a machine learning algorithm needs to be entirely traceable to ensure that human decision makers can interpret each (relevant) analytical decision step.

Notes: This table extends Table 2 of Brooks et al. (2019) by the concept of explainability and further refines some of its content to reflect advances in econometrics and financial data science. Since machine learning applications represent economic as well as broader societal use cases, we amended the name of the column presented on the right. If Shapley Values are computationally too intensive for a specific use case (e.g. more than 10–15 independent variables), we recommend Quasi Shapley values, which are computed like Shapley Values with the exception that they drop all regression combinations for computation efficiency except the first and final step (i.e. they are computed based (i) on each variables' stand alone ability to explain the variation of the outcome variable and (ii) the loss in explanatory power when dropping the respective variable from the final model).

finance or health care, where the customer can often only ex-post assess the quality of the service and hence requires substantial ex-ante trust. Fourth, as a consequence of this lack of transparency and traceability, neural network are not fully interpretable for human experts and do not allow for attempts at causal inference (Barredo Arrieta et al. 2020; Marcus 2018).[5]

In this context, we structure our manuscript around the concepts of significance, relevance and explainability before we conclude suggesting routes for future research. Specifically, the papers in this special issue of *The European Journal of Finance* inform academics, practitioners and policymakers regarding developments in Financial Data Science and Econometrics and provide insight on how cutting-edge data and/or cutting-edge statistical techniques can be applied to enhance our understanding of financial phenomena. Three broad themes are covered in these research papers: financial forecasting and the related statistical significance (McMillan 2020; Meligkotsidou et al. 2020), financial risk modelling and related statistical relevance (Kim, Park, and Yoon 2020; Lin, Kolokolova, and Poon 2020; McGee and Olmo 2020), and explaining novel datasets (Asimakopoulos, Asimakopoulos, and Zhang 2020; Dechezleprêtre, Muckley, and Neelakantan 2020; Petukhina, Reule, and Härdle 2020),

2. Significance in financial forecasting

Standard statistical significance levels can be easily obtained as a result of the sheer scale of the observations. This can be partly addressed by using far more stringent significance levels, but moreover, it means much more emphasis needs to be placed on the incremental explanatory power of a variable (statistical relevance) as well as the economic magnitude (i.e. significance) of the effect. Thus, with huge datasets, we need to be examining: (i) is the predictor variable of interest statistically significant at a confidence level of 1% or better (statistical significance)? (ii) how much the predictor variable of interest generates extra explanatory power (statistical relevance)? and (iii) how big is the response of the dependent variable to a change in the predictor variable, both scaled in standard deviations of the predicted variable (economic significance) as well as in one standard deviation of the predictor variable (economic relevance)?

Financial forecasting is an area which continues to attract huge interest. Much of the literature focuses on the real-time prediction of the US equity premium. The debate was reignited by Welch and Goyal (2008) who, in a comprehensive study, argue that most existing predictors of the US equity premium fail to outperform the historical average. This conclusion was soon challenged, notably by (i) Campbell and Thompson (2007) who argued that placing economically motivated restrictions on the model lead to small improvements in forecast power which could lead to substantial economic gains and (ii) by Rapach, Strauss, and Zhou (2009) who illustrate that combining forecasts can lead to larger improvements in forecast power. The papers by McMillan (2020) and Meligkotsidou et al. (2020) revisit this crucial topic of forecasting the US equity premium and contribute new predictors, new methods and new results. The tone of McMillan (2020) is broadly in-line with Welch and Goyal (2008) that finding models that consistently beat the historical average is very difficult and few models, if any, can do this consistently. However, McMillan (2020) highlights that the purchasing managers' index is a new variable that generally performs well across a vast array of tests conducted and that information contained in the term structure is also valuable for investors. These measures are both forward-looking measures of future economic conditions which can theoretically lead to time-variation in expected returns.

Meligkotsidou et al. (2020) build on the line of research stimulated by Rapach, Strauss, and Zhou (2009), which lead to the development of the complete subset linear regression framework (Elliott, Gargano, and Timmermann 2013). They demonstrate that by (i) applying quantile forecasts to the complete subset approach and then (ii) combining these quantile forecasts into a point forecast can lead to statistically significant improvements in forecast accuracy and to economically substantial gains to an investor, thereby demonstrating both statistical and economic significance (Meligkotsidou et al. 2020). This research highlights the importance and usefulness of advances in modelling which can help us to both more accurately represent financial phenomena as well as generate superior outcomes for end-users.

3. Statistical relevance in risk modelling

Research that can explain twice as much variation as previous work is a great example for statistical relevance, especially in the field of financial risk modelling. Lin, Kolokolova, and Poon (2020) explore two novel endogenous systematic risk factors explaining CDS spreads. Their slow-moving factor is based on a CreditGrades (CG) model to price CDS, while their fast-moving factor represents the estimation errors adjusted cross-sectional mean absolute deviation between the observed CDS spread and the respective fitted value from the CG model. Analysing the 5-year CDS spreads of US non-financials, they find the statistical relevance of their factors to exceed their peers substantially. Specifically, they observe in terms of statistical relevance that individual firm-specific factors only explain 11% of individual CDS spreads and individual illiquidity only explains about half of this, whereas their novel factors can account for twice the variation than the firm-specific factor (i.e. 22%). Curiously, adding their new endogenous systematic risk factors to a model of determinants of CDS spreads yields established systematic risk factors such as VIX and term spread to become insignificant (Lin, Kolokolova, and Poon 2020).

McGee and Olmo (2020) provide an interesting analysis of the size anomaly. This anomaly dates back at least to Banz (1981), who finds that stocks with a small market capitalisation outperform large stocks on a risk-adjusted basis (Banz 1981). McGee and Olmo (2020) set out to analyse the extent to which lottery-like features are relevant in explaining this anomaly. They begin with a simple analysis where, at each point in time, they exclude the best-performing stocks from the sample of investable securities. Interestingly, the authors document that the trading strategy that captures the size anomaly, by purchasing small firms while simultaneously taking short positions in large firms, disappears once the best performing assets are excluded. Pursuing their analysis, the authors explore the asset pricing implications of this result. The paper shows that the exclusion of the best performing assets does not materially affect the exposure to the size factor. However, it significantly reduces the magnitude of the size risk premium.

The paper by Kim, Park, and Yoon (2020) examines the role of time-variation and asymmetries in correlations for portfolio allocations. In particular, the paper derives the optimal consumption and investment strategies allowing for regime-dependent correlations. The paper derives the utility function under different conditions and conducts an empirical exercise based on US data for the market (S&P500) and five industry portfolios. The results reveal that failure to account for asymmetries and regime dependency leads to misallocation of portfolio weights and result in a loss of wealth (Kim, Park, and Yoon 2020). Most notably, while performance during 'normal' market times is comparable to other portfolio allocation approaches, the 2007–9 financial crisis increases the relevance of accounting for time-variation and asymmetry becomes more prominent.

4. Explainability of novel data sets

Many Financial Data Scientists have a keen interest in novel data and its exploration. Interest maybe sparked, for example, through there being a new data source that is now being examined in a financial context or through there being a much bigger dataset that is being analysed than has hitherto been examined, which may now be feasible due to advances in computational power. Thus, breaking away from datasets which have been heavily analysed, data-mined even, such as CRSP and Compustat is a natural feature of Financial Data Science. The use of novel datasets can also lead to the opportunity to explain relationships that have the potential for huge societal impact (e.g. climate change, Covid-19, fintech). The creation of new markets also inevitably leads to the availability of new data.

One area that has received great interest both in the media and more recently in academia has been the development of cryptocurrency markets. There are important differences between cryptocurrency markets and traditional financial markets; for example, traders in cryptos are mainly anonymous and usually trade directly with each other in a peer-to-peer digital network with extremely low transaction costs. A timely overview of cryptocurrencies and how to explain them is provided by Härdle, Harvey, and Reule (2020). The novel research topic investigated by Petukhina, Reule, and Härdle (2020) is if trading in these new cryptocurrency markets leads to a domination of machines, especially algorithmic traders. Algorithmic trading has grown in prominence within traditional financial markets over recent years. However, little is known about their use for trading in crypto markets, although we are aware that huge computing power is harnessed in the mining process of cryptos. Focusing on high-frequency trading, Petukhina, Reule, and Härdle (2020) explain the patterns in this data, which are characterised by (very) high levels of kurtosis, fluctuating levels of volatility and modest cross-correlations between the cryptos. The authors examine trading activity during the week and find a clear drop off in activity on Saturdays and Sundays (Petukhina, Reule, and Härdle 2020). This is consistent with human trading in the market. However, there is no reason why algorithms should not be active in these markets at the weekends. Thus, while these crypto markets appear ideal environments for machines to operate, in terms of trading, it seems that humans still play a major role.

Dechezleprêtre, Muckley, and Neelakantan (2020) demonstrate how novel data can be used to address a new research question and explain the results of societal importance in an interpretable manner. To be specific, they use a global patent database (PATSTAT) to examine filings by more than 15,000 firms and distinguish between (i) 'clean' innovation (e.g. renewable energy and electric cars) and (ii) 'dirty' (e.g. fossil fuels and combustion engines) innovation. Their main finding is that the stock market values firms which engage in 'clean' innovation more highly than firms which engage in 'dirty' innovation (Dechezleprêtre, Muckley, and Neelakantan 2020). These results are robust across various alternative specifications and an extensive set of controls. A key implication of the study is that there is no deterrent from the equity market for firms that engage in 'clean' innovation which is aligned with recent findings of no deterrent to 'clean' entrepreneurship (Cojoianu et al. 2020). On the contrary, there appears to be support from the market for such firms given a valuation premium is given to 'clean' innovators. Thus, the market seems to be behaving in a way that is consistent with global de-carbonisation in the long-term and the European Union supporting the sustainable finance sector through its green taxonomy (Slevin et al. 2020) and its Paris Agreement aligned equity index concepts (Hoepner et al. 2019).

Asimakopoulos, Asimakopoulos, and Zhang (2020) examine the impact of credit rating changes on firm dividend behaviour. They explain that firms listed in the S&P500 typically operate a dividend smoothing policy. The paper then proceeds to consider how credit rating down- and up-grades affect such dividend smoothing. The paper reveals that credit downgrades lead to a lower degree of dividend smoothing. Notably, the empirical results in the paper report that while the payout ratio has declined from earlier work (reported at around 25%, approximately half the previously reported value), the degree of smoothing remains broadly constant (with an AR(1) parameter of 0.7). The paper argues this effect is a desire of firm management to hold onto cash to ensure normal operations.

5. Conclusion and directions for future research

The field of econometrics and financial data science as outlined in Brooks et al. (2019) is rapidly developing with researchers exploring new data sources and methods while assessing results in terms of statistical significance (e.g. Meligkotsidou et al., 2020), relevance (e.g. Lin, Kolokolova, and Poon 2020) and societal explainability (Dechezleprêtre, Muckley, and Neelakantan 2020). Simultaneously, a debate between 'black box' and 'white box' approaches has commenced in the related and currently very prominent field of machine learning. Although we, as econometricians and financial data scientists, might not have immediately recognised our involvement, we are considerably engaged as white box machine learners have adopted a substantial part of our algorithmic tool kit. Hence, it is important for our community to contribute. To do so, we extend Brooks et al.'s (2019) work on significance and relevance as assessment critieria in econometrics and financial data science to contribute to this debate. Specifically, we identify explainability as the Achilles heel of neural networks, which are not fully replicable, lack transparency and traceability and therefore do not permit any interpretations as to causal inferences.

We conclude with three suggested routes for future research to advance the design and efficiency of 'white box' algorithms. First, we recommend to define clear boundaries when statistical approaches lose their traceability. For instance, ratios or orthogonalizations are fully traceable to their conceptual origins while this is harder for principle components. Second, it would be suitable to develop a framework for non-linear regressions which introduced the non-linearity in a controlled manner that is less at risk of overspecifying on any given sample. Third, to enhance the computational efficiency of 'white box' approaches where relevant, variable pre-processing may be suitable to reduce the temptation of including an unnecessarily large set of variables and address potential multicollinearity obstacles.

Notes

1. For more details on evidence-based policy making https://www.kcl.ac.uk/aboutkings/facts/WorldStatisticsDay/policymaking.
2. We recommend operationalising statistical relevance as relative explanatory power measurable by (Quasi) Shapley Value.
3. Multicollinearity due to independent variable interdependencies can, of course, create spurious significance in a poorly specified model.
4. Scaling multiple regression coeffcients by the standard deviations of the dependent variable and the respective independent variable allows consequently a simultaneous assessment of economic significance and relevance.
5. For a more extensive criticism of (deep) neural networks, see Marcus' (2018) ten criticisms.

Acknowledgements

We are very grateful to Chris Adcock for providing us with the opportunity to empower and develop relevant research using methods from the field of financial data science and econometrics. Our thinking has benefited from discussions with Alexander Arimond, Damian Borth, Theodor Cojoianu, Sergio Garcia Vega, Georgiana Ifrim, Max Lin, Yanan Lin, Markus Koch, Markus Leippold and Gabija Zdanceviciute. Hoepner acknowledges funding from the European Union's Horizon 2020 research and innovation programme for research on Fintech (grant number H2020-ICT-825215), Science Foundation Ireland (Award 19/FIP/AI/7539) and from Bank of Ireland, Citibank Europe, Deloitte Ireland and Institute of Banking for research on Operational Risk. Authors are alphabetically listed. All remaining errors are our sole responsibility.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Horizon 2020 Framework Programme: [grant number H2020-ICT-825215]; Science Foundation Ireland: [grant number 19/FIP/AI/7539].

ORCID

Andreas G. F. Hoepner http://orcid.org/0000-0002-4311-2403
David McMillan http://orcid.org/0000-0002-5891-4193

Andrew Vivian http://orcid.org/0000-0002-6036-4477
Chardin Wese Simen http://orcid.org/0000-0003-4119-3024

References

Asimakopoulos, Panagiotis, Stylianos Asimakopoulos, and Aichen Zhang. 2020. "Dividend Smoothing and Credit Rating Changes." *The European Journal of Finance* 1–24.

Banz, Rolf. 1981. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9: 3–18.

Barredo Arrieta, A., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, A. Tabik, A. Barbado, S. Garcia, et al. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." *Information Fusion* 58: 82–115.

Basha, S. M., and D. S. Rajput. 2019. "Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap." In *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, edited by A. K. Sangaiah, 153–164. St. Louis: Elsevier.

Brooks, Chris, Andreas G. F. Hoepner, David McMillan, Andrew Vivian, and Chardin Wese Simen. 2019. "Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics?" *European Journal of Finance* 25 (17): 1627–1636.

Campbell, John Y., and Samuel B. Thompson. 2007. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *The Review of Financial Studies* 21 (4): 1509–1531.

Cojoianu, Theodor F., Gordon L. Clark, Andreas G. F. Hoepner, Paolo Veneri, and Dariusz Wójcik. 2020. "Entrepreneurs for a Low Carbon World: How Environmental Knowledge and Policy Shape the Creation and Financing of Green Start-Ups." *Research Policy* 49 (6): 103988.

Dechezleprêtre, Antoine, Cal B. Muckley, and Parvati Neelakantan. 2020. "Is Firm-Level Clean or Dirty Innovation Valued More?" *The European Journal of Finance* 1–31.

Elliott, Graham, Antonio Gargano, and Allan Timmermann. 2013. "Complete Subset Regressions." *Journal of Econometrics* 177 (2): 357–373.

Gu, S., B. Kelly, and D. Xiu. 2020. "Empirical Asset Pricing Via Machine Learning." Forthcoming in Review of Financial Studies.

Härdle, Wolfgang Karl, Campbell R. Harvey, and Raphael C. G. Reule. 2020. "Understanding Cryptocurrencies∗." *Journal of Financial Econometrics* 18 (2): 181–208.

Hoepner, A. G. F., P. Masoni, B. Kramer, D. Slevin, S. Hoerter, C. Ravanel, H. Viñes Fiestas, et al. 2019. *TEG Final Report on Climate Benchmarks and Benchmarks' ESG Disclosure*. Brussels: European Commission.

Kim, Myeong Hyeon, Seyoung Park, and Jong Mun Yoon. 2020. "Industry Portfolio Allocation with Asymmetric Correlations." *European Journal of Finance* 0 (0): 1–21.

Lin, Ming-Tsung, Olga Kolokolova, and Ser-Huang Poon. 2020. "Slow-and Fast-Moving Information Content of CDS Spreads: New Endogenous Systematic Factors." *European Journal of Finance 1–22*.

López de Prado, M. M. 2019. *Advances in Financial Machine Learning*. 1st ed. Hoboken, New Jersey: John Wiley and Sons, Inc.

Loyola-Gonzalez, O. 2019. "Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses from a Practical Point of View." *IEEE Access* 7: 154096–154113.

Mahdavinejad, M. S., M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth. 2018. "Machine Learning for Internet of Things Data Analysis: a Survey." *Digital Communications and Networks* 4 (3): 161–175.

Marcus, G. 2018. "Deep Learning: A Critical Appraisal". Working Paper, New York University.

McGee, Richard J., and Jose Olmo. 2020. "The Size Premium as a Lottery." *The European Journal of Finance* 1–20.

McMillan, David G. 2020. "Forecasting US Stock Returns." *The European Journal of Finance* 1–24.

Meligkotsidou, Loukia, Ekaterini Panopoulou, Ioannis D. Vrontos, and Spyridon D. Vrontos. 2020. "Out-of-Sample Equity Premium Prediction: A Complete Subset Quantile Regression Approach." *European Journal of Finance 1–26*.

Mitchell, T. M. 1997. *Machine Learning*. New York: McGraw–Hill.

Ng, A. 2016. "What Artificial Intelligence Can and Can't Do Right Now." Harvard Business Review, November 2014.

Petukhina, Alla A., Raphael C. G. Reule, and Wolfgang Karl Härdle. 2020. "Rise of the Machines? Intraday High-Frequency Trading Patterns of Cryptocurrencies." *European Journal of Finance*.

Portugal, I., P. Alencar, and D. Cowan. 2018 "The use of Machine Learning Algorithms in Recommender Systems: A Systematic Review." *Expert Systems with Applications* 97: 205–227.

Rai, A. 2020. *Explainable AI: From Black box to Glass box Journal of the Academy of Marketing Science* 48: 137–141.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2009. "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *The Review of Financial Studies* 23 (2): 821–862.

Slevin, D., S. Hoerter, N. Humphreys, H. Viñes Fiestas, S. Lovisolo, J.-Y. Wilmotte, P. Latini, et al. 2020. *Taxonomy: Final Report of the Technical Expert Group on Sustainable Finance*. Brussels: European Commission.

Welch, Ivo, and Amit Goyal. 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies* 21 (4): 1455–1508.

Yang, J. H., S. N. Wright, M. Hamblin, D. McCloskey, M. A. Alcantar, L. Schrübbers, A. J. Lopatkin, et al. 2019. "A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action." *Cell* 177 (6): 1649–1661.e9.