SCHOOL OF
ECONOMICS AND
MANAGEMENT

LUND UNIVERSITY

Is the Federal Reserve causing funds to underperform?

A causal machine learning analysis

by

Oliver Serenhov

Willian Trindade Leite

May 2022

Master's Programme in Data Analytics and Business Economics
DABN01

Supervisor: Krzysztof Podgórski

Abstract

Macroeconomic conditions heavily influence financial markets, and the leading corpus of theory in the field lays out some general principles observed by investors and asset managers alike. However, while the theory is sound, it is hard to measure how much effect these conditions have. To better understand this scenario, this study uses a novel technique that combines traditional causal inference with machine learning techniques to measure macroeconomic policy effects in the financial market. Specific to this study, the double machine learning framework measures the average treatment effect of the US Federal Reserve System's interest rate's growth rate on fixed income and equity funds returns. The period studied is from January 1986 to December 2021, as it includes data from many relevant events that caused interest rates to change around the world. Furthermore, the data is separated into two main clusters, i.e., passively and actively managed funds, since finance theory indicates that the latter should be less affected by a central bank's interest rate changes.

As the double machine learning framework can use virtually any statistical learning procedure as a learner, two different techniques are tried in this study. Linear regression sets a baseline, and gradient boosting is used to assess what technique would produce better results. The results show some evidence of gradient boosting being a worthy technique for predicting returns and, therefore for the double machine learning procedure. The actively managed funds dataset results indicate that a 1% increase in the US Federal Reserve System's interest rate led to a -11.97% decrease in actively managed fund returns. However, this effect must be considered alongside all others that might affect financial markets. The results indicate a rich field for future research that can serve investors and managers through data-driven decision-making.

Keywords: Causal inference, actively managed funds, passively managed funds, gradient boosting, double machine learning

Acknowledgements

Academic research is seldom easy, especially when one has so little time to perform it. Concluding this work without the guidance and advice from Krzysztof Podgórski would have made it much harder. We also feel fortunate that we could get the support of professors who, even if not officially supervising us, still devoted much of their time to solving our doubts, exchanging ideas, offering articles, and sometimes even introducing us to new perspectives on our results. Therefore, we would like to thank Joakim Westerlund and Simon Reese for taking us to the next level regarding time-series analysis and the double machine learning framework. We are also grateful to Alessandro Martinello and Isaiah Hull for the exquisite articles and workshops. These were inspiring, and in many ways shaped the direction the research took.

Finally, we must also mention the moral support from so many classmates and friends who, in one way or another, helped us conclude this work. From the several friends who just motivated us through the hardship, from those who lent us their hardware for parameter tuning (without which the results would not have been concluded in able time), or even from those who were just patient enough to hear us ventilate our frustrations: Thank you.

Table of Contents

List Of Tables

List Of Figures

1 Introduction

Forecasting returns or features that indicate the direction of a return is the object of many studies (Pierdzioch, Döpke & Hartmann, 2005; Leung, Daouk & Chen, 2000; Neely, Rapach, Tu & Zhou, 2014; Paye, 2012; Chen, 2009). From a practical perspective, investors and asset managers use these forecasts to decide on their investments (Hünermund, Kaminski & Schmitt, 2022; Athey, 2017). This type of strategy can bear fruits, but it does not help investors and managers to understand what causes these movements (Heiss, 2020; Athey, 2017). This paper advocates for causal inference to better understand the actors responsible for the erratic trends in financial markets and make data-driven decisions that could lead to better returns.

While traditional econometrics offer some causal methods, Granger causality being one example, these are often not robust enough to support a valid causality claim (Granger, 1969; Ding, Chen & Bressler, 2006; Stokes & Purdon, 2017). Additionally, traditional causal inference methods have limitations regarding the size and properties of the data (Heiss, 2020; Spirtes, 2000; McElreath, 2020; Ahrens, Aitken & Schaffer, 2021). New techniques that combine statistical learning with conventional causal inference methods offer solutions to deal with highly complex data and models while still producing conceptually similar conclusions to traditional causal inference (Athey, 2017; Hünermund, Kaminski & Schmitt, 2022).

Although recent, these new methods have been widely applied to socio-economic questions (Dube, Jacobs, Naidu & Suri, 2020; Knaus, 2021; Grodecka-Messi & Hull, 2019; Davis & Heller, 2017; De Neve, Imbert, Spinnewijn, Tsankova & Luts, 2021; Bertrand, Crépon, Marguerie & Premand, 2017; Athey & Imbens, 2016). However, only a few attempts have been made to use this type of causal inference in finance questions (Wasserbacher & Spindler, 2021; Varaku, 2021). Moreover, few have used real-world data instead of simulated data (Wang, 2020). Thus, it is possible to say that there is a lack of studies in this area, especially studies that can aid investors and managers alike in their decision-making.

This study focuses on actively and passively managed funds created since January 1986, and the period studied closes in December 2021. Approximately seven thousand funds are used in the study, considering both data from actively and passively managed funds. The choice for actively and passively managed funds relates to their popularity among investors and relevance as reflections of the whole market (Xu, 2021; Das, 2009; Huang, Pilbeam & Pouliot, 2021; U.S. Securities and Exchange Commission, n.d.). First, two methods are used to model the complex behavior of financial markets; namely, linear regression (Hastie, Tibshirani & Friedman, 2009) and gradient boosting (Schapire, 1990; Hastie, Tibshirani & Friedman, 2009; James, Witten, Hastie & Tibshirani, 2013). These are then incorporated into double machine learning (DML), a method created by Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) that proposes a novel way to measure the average treatment effect, i.e. the causal effect, of a variable.

A series of macroeconomic variables are used to capture the complexities of the market, with their choice following relevant examples from academia (Asgharian, Hou & Javed, 2013; Malmendier & Nagel, 2011; Asgharian, Christiansen & Hou, 2015; Birz & Lott, 2011; Hamilton & Lin, 1996; Baker & Wurgler, 2007; Glosten, Jagannathan & Runkle, 1993; Whitelaw, 1994; Brandt & Kang, 2004) and common sense choices considering macroeconomic theory. One such variable is the Central Bank's interest rate, which is of particular interest in this study as it serves as a treatment for several symptoms described by macroeconomics. Moreover, it is recognized by academia and specialized media alike as a cause of different movements in financial markets, which makes it a relevant predictor for fund returns (Glosten, Jagannathan & Runkle, 1993; Whitelaw, 1994; Brandt & Kang, 2004; The Economist, 2022; Tillier, 2021; O'Connell & Curry, 2022; Domm, 2022). In light of this, two questions guide this research:

Question 1: *"How feasible is using the double machine learning framework for measuring causal effects in fund returns?"*

Question 2: *"What effect, on average, does the US Central Bank's interest growth rate have on fund returns?"*

As will be discussed, the DML framework seems like a reasonable choice for causal inference in finance questions, although some assumptions and workarounds are necessary. However, the necessity for these workarounds is expected since financial data is often complex and with many challenging aspects. Furthermore, as one of the requirements for the DML framework is the use of reasonably good prediction models, this paper also offers evidence of the use of gradient boosting for predicting fund returns with arguably high precision. For example, the statistically significant gradient boosting model presented here achieved an R^2 of approximately 0.73 and a mean squared error of roughly 0.0007.

The DML procedure also shows results that agree with the macroeconomics theory regarding a Central Bank's interest rate change. Specifically for the active funds set, a 1% increase in the current rate, can produce an average reduction of approximately 11% in the returns. Moreover, this framework is devised to easily change the variable of interest, which one may want to measure the average treatment effect. Thus, another of the aims of this study is to offer a framework that is easy to implement and interpret that can aid investors and asset managers alike in their decision-making.

Section 2 of this study covers the research and core literature necessary to understand the concepts and methods presented here. Followed by section 3 where a description of how the DML is implemented, with particular attention to the tuning procedure for the gradient boosting method. A detailed account of the data collection and preprocessing is then presented in section 4. This thesis results are presented and discussed in section 5 as well as the study's known limitations and ideas for future research. Finally, in section 6 the authors conclusions of the study are offered to summarize the study.

2 Literature Review

2.1 Funds and Macroeconomics

2.1.1 Conceptual differences between actively and passively managed funds

Central to answering this paper's research questions are defining the concept of a fund and how it will be used throughout the thesis. Das (2009) defines a fund as the pooling of capital of several investors, usually a large number of them, to make collective investments in various assets. Investing in funds has two general characteristics; outsourcing investment management to the fund's money managers and increased diversification, as funds are often invested in diversified assets (Das, 2009; U.S. Securities and Exchange Commission, n.d.; Gruber, 2010). There are also costs commonly associated with selling, buying, and owning funds, for example, fees to pay the funds' employees (U.S. Securities and Exchange Commission, 2020). The fact that individual investors do not have to make their own investment decisions and thus save time and effort and the relative ease with which diversification can be achieved are two reasons why investing through a fund has become as popular as it is today (Xu, 2021; Huang, Pilbeam & Pouliot, 2021; U.S. Securities and Exchange Commission, n.d.). Das (2009) further clarifies that various types of funds exist, and individual investors often choose which fund to invest in based on these stated goals and strategies.

The type of a fund is defined by its stated structure, investment objective, and strategy employed (Das, 2009; U.S. Securities and Exchange Commission, 2020). First, the U.S. Securities and Exchange Commission (2020), the SEC, discloses that the structure of a fund clarifies the rules regarding the sale, creation, and discontinuation of shares of the fund. Two examples the SEC lifts are open-end or closed-end funds in which the former shares are sold and created continuously. In contrast, in the latter case, a fixed amount of shares are made in the fund's initial public offering, which can be bought and sold on the market. Secondly, a fund's investment objective affects what the fund can invest in, and there exists vast amounts of objectives funds may pursue (U.S. Securities and Exchange Commission, 2020). However, this paper solely focuses on equity and fixed-income funds, i.e. stocks and debt, respectively.. Lastly, funds have different strategies that inform investors how a fund will manage capital. The most pertinent for this paper is the distinction between an active or passive management strategy (U.S. Securities and Exchange Commission, 2020).

Xu (2021) claims that the core distinction between an active and a passive strategy is their fundamental goal. Active strategies allow fund managers to pick the fund's assets based on their information and expertise to reach as high of a return as possible and outperform the overall market or benchmark (U.S. Securities and Exchange Commission, 2020; Xu, 2021). On the contrary, a passive fund, also called an index fund, strives to emulate an underlying index, thus minimizing its so-called tracking error (Das, 2009; Xu, 2021), meaning its returns correlate with its intended index (U.S. Securities and Exchange Commission, 2020). Beyond tracking error, the "Active Share" of a fund can inform an investor how similar the investment of a fund is to the underlying benchmark by comparing the weighting of the fund's assets with that of the benchmark (Xu, 2021). On average, neither active nor passive funds strictly follow their stated strategy based on their tracking error and active share (Xu, 2021); it has even been argued that some active funds are “closet indexing,” meaning that they claim to be actively managed while in truth they are mostly passively managed (Cremers, Ferreira, Matos & Starks, 2016). Index funds have recently surpassed actively managed funds in assets under management (Xu, 2021); this is likely due to the mounting evidence that actively managed funds underperform passive ones (Malkiel, 1995; Elton, Gruber & Blake, 1996; Fama & French, 2010; Sharpe, 1966; Jensen, 1968).

2.1.2 Differences between active management and passive management

The literature on fund returns is abundant, and part of the debate has focused on how active funds' performance differs relative to passive ones (Gruber, 2010; Malkiel, 1995; Elton, Gruber & Blake, 1996; Fama & French, 2010; Sharpe, 1966; Jensen, 1968; Henriksson, 1984; Chang & Lewellen, 1984; Ippolito, 1989; Hendricks, Patel & Zeckhauser, 1993; Kacperczyk, Sialm & Zheng, 2005; Ippolito, 1993; Barras, Scaillet & Wermers, 2010). Sharpe (1966) and Jensen (1968) were some of the first authors to initiate the debate; they claimed that actively managed funds underperformed relative to indexes. These claims have later been supported by other authors who claim that actively managed funds aggregated together as a group underperform relative to passively managed funds, especially when considering the extra cost, they usually charge investors for their services (Xu, 2021; Gruber, 2010; Malkiel, 1995; Fama & French, 2010; Barras, Scaillet & Wermers, 2010).

Another argument against the performance of actively managed funds is that expenditure on investment research and trading does not improve their performance due to the efficient markets hypothesis (Ippolito, 1993), which says that the asset price fully reflects all information at all times (Fama, 1970). Grossman and Stiglitz (1980) challenged the original efficient market hypothesis by claiming that collecting information is costly. Investors who participate in such can expect to earn higher gross profits; however, they can expect the same returns as their uninformed counterparts after expenses. Kacperczyk, Sialm, and Zheng (2005) claim that funds concentrated in a few industries perform better than diversified funds suggesting that fund managers' experience in a specific area makes their information collection more efficient, building on Grossman and Stiglitz's (1980) argumentation.

Other authors have found that funds' current performance can be predicted based on their prior performance (Elton, Gruber & Blake, 1996; Fama & French, 2010). Most of the worst-performing funds perform poorly because of their high expenses. Thus the aggregate of actively managed funds would improve if these high-expense funds were excluded (Elton, Gruber & Blake, 1996). Active versus passive management has also been argued to perform differently in different market conditions, such as the macroeconomic conditions wherein, for example recessions, active management works best (Söderberg & Partners, 2017). Beyond the above-listed arguments, the investment style such as “growth” and “aggressive growth” of the fund have also been claimed to affect whether a particular fund outperforms passive funds or not (Hendricks, Patel & Zeckhauser, 1993; Grinblatt & Titman, 1989; Daniel, Grinblatt, Titman & Wermers, 2012; Kosowski, Timmermann, Wermers & White, 2006; Verbeek & Huij, 2006). However, studies challenge the importance of growth investment styles

(Becker, Ferson, Myers & Schill, 1998). Lastly, survivorship bias among funds may also have made many of the results presented in the literature in the debate on the fund returns of active versus passive spurious, leading to further questioning about any conclusion drawn from the literature (Elton, Gruber & Blake, 1996).

2.2 Macroeconomics in Finance

Extensive support for macroeconomic variables affecting asset returns, and particularly stock returns, exist in the literature (Asgharian, Hou & Javed, 2013; Malmendier & Nagel, 2011; Asgharian, Christiansen & Hou, 2015; Birz & Lott, 2011; Hamilton & Lin, 1996; Baker & Wurgler, 2007; Glosten, Jagannathan & Runkle, 1993; Whitelaw, 1994; Brandt & Kang, 2004). For example, the Capital Asset Pricing Model, CAPM, is commonly used for pricing assets such as stocks or funds by illustrating the relationship between systematic risk and the expected return of the assets (Sharpe, 1964), and macroeconomic variables can affect these factors (Birz & Lott, 2011). Numerous papers have found support for this, e.g., forecasting of stock variance is significantly improved when macroeconomic variables are included (Asgharian, Hou & Javed, 2013), and news of macroeconomic variables such as GDP growth and unemployment significantly affects stock returns (Birz & Lott, 2011), macroeconomic uncertainty affects investments in stocks, and bonds (Asgharian, Christiansen & Hou, 2015), recessions affect the stock market volatility (Glosten, Jagannathan & Runkle, 1993; Schwert, 1989a; Schwert, 1989b). Of particular note for this paper is the interest rates' impact on fund returns, and such a relationship has been shown to exist in multiple papers (Glosten, Jagannathan & Runkle, 1993; Whitelaw, 1994; Brandt & Kang, 2004).

2.3 Causal Inference

Whereas prediction focuses on learning patterns in a system that indicate when something is likely to happen, causation tries to understand what causes the action. This definition is one of many, and, throughout this paper, the primary definition used is that "A variable X is a cause of Y if Y in any way relies on X for its value …[,] X is a cause of Y if Y listens to X and decides its value in response to what it hears." (Pearl, Glymour & Jewell, 2016, p.5, quoted in Heiss, 2020). This definition gives an exciting perspective to the second research question since most treatment effects in financial markets happen via investors (and asset managers) listening to the potential changes in the market (Fama, 1970). One could suppose then that the market listens to Central Bank's intentions and actions, which affect how they behave toward specific assets.

Another common way to express this causal relationship is through directed acyclic graphs (DAGs) (Spirtes, 2000; Kennaway, 2020; Heiss, 2020; Zhao & Hastie, 2021; Hünermund, Kaminski & Schmitt, 2022; McElreath, 2020). These expose the writer's assumptions regarding what variables are independent (or not) and the causal path. Given the time aspect in the problem studied here, one should interpret the DAGs used in this paper as to how the system looks in the current time. The concept behind DAGs demands that there should be no recursion or feedback in the system, yet one should expect that investor behavior may sometimes influence central banks' actions. While this may be the case, this paper focuses on one specific frame of this continuous movement, i.e., how the US

Central Bank affects the market with its interest rate policies, based on an interpretation provided by Lacerda, Spirtes, Ramsey, and Hoyer (2012).

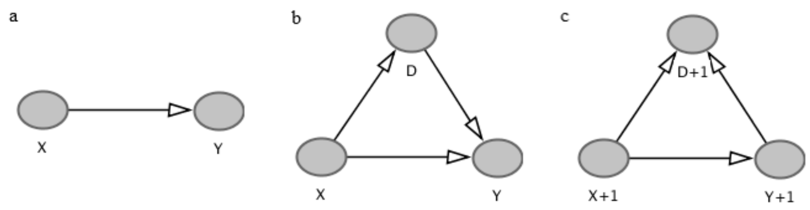

Figure 1: A representation of DAGs. (a) represents a simple model where Y is solely dependent on X. (b) shows a model closer to the one studied here, where Y is dependent on both X and D, but X acts as confounder as it affects both Y and D. (c) simulates the "day after", imagining how the system would react after Y being affected by D and X, and in this hypothetical example D becomes a collider the next day.

Causal inference and prediction can be viewed as two faces of the same coin (Athey, 2017; Mullainathan & Spiess, 2017). The methods used are often similar but follow different assumptions and requirements and have other goals (Athey, 2017). Although with minor adaptations, Mullainathan and Spiess (2017) use a convenient jargon adopted in this paper. They introduce the concept of problems focused on prediction as *y-related* problems (or tasks), i.e., they are interested in predicting some outcome y. They also mention that causality can be analyzed through the same models but instead focus on "discovering" the coefficient related to some variable used in the prediction. These coefficients are often called β, which they use to name the *β* related problems (or tasks). In this work, a similar approach is taken in which the terms *y-task* and *β-task* are used to denote the goal of the analysis at that stage.

As mentioned before, often, the *y-task* and *β-task* models are very similar (not to say the same), but with different assumptions and goals. For example, Mullainathan and Spiess (2017), Young (2019), and Reese (2021) point out that when it comes to β-tasks, there are generally a wide array of assumptions and requirements for the data and the parameter estimates, as they must have acceptable properties. These properties include, for example, classical statistics concepts such as standard errors and bias. On the other hand, y-tasks are generally focused on producing good predictions regardless of the properties of the estimates.

Y-tasks are often the ultimate goal of machine learning methods, especially in financial markets (Dixon, Halperin & Bilokon, 2020; Hoepner, McMillan, Vivian & Wese Simen, 2021). This propensity can be traced back to a series of factors, some of which have to do with the knowledge complexity to extrapolate an analysis beyond correlations (Hünermund, Kaminski & Schmitt, 2022; Mullainathan & Spiess, 2017). Some others because traditional causal inference methods require either randomized controls or broad assumptions prohibiting many complex problems in observational data (Spirtes, 2000; Rosenbaum, 1995; Athey, 2017). Athey (2015) states that the latter is the "fundamental problem of causal inference" since no variable is observed simultaneously in different worlds where one may choose to apply a treatment or not to that variable.

Causal inference is still crucial for business decisions and policymaking despite this fundamental problem (Athey, 2017; Hünermund, Kaminski & Schmitt, 2022). Understanding whether an action causes a given reaction in a system can be essential in many situations. For example, the pharmaceutical industry is heavily dependent on causal inference to prove the efficacy of its drugs, which they do through randomized control trials (RCT). These RCTs, however, often meet ethical barriers when it comes to issues like education and inequality and, in some cases, are simply not possible to be applied, such as in economic problems. There is no way to create an alternate reality where a country simultaneously uses and does not use central bank interest rates.

As experimental scenarios (such as RCTs) are either very hard or plain impossible, one must consider the limitations traditional causal inference methods have regarding observational data. For example, a traditional causal inference method, difference-in-differences, still requires control groups, even if synthetic ones. This method measures the difference between observed treated and control groups (Rosenbaum & Rubin, 1983). Other methods, like instrumental variables (IV) and propensity scores, have been devised to circumvent these clear limitations imposed by the lack of control groups (Rosenbaum & Rubin, 1983; Card, 1999). While these offer some solutions to the fundamental problem of causal inference, they are still limited – either conceptually or applicably – by big data. For example, the presence of many covariates can be hazardous for propensity scores, as they may introduce bias to the analysis (Young, 2019). These methods also share a general cautiousness regarding confounders, often implying hard-to-check assumptions, especially when dealing with large datasets (Spirtes, 2000; McElreath, 2020; Heiss, 2020; Ahrens, Aitken & Schaffer, 2021).

2.3.1 Causal Inference in Finance

Given the challenges presented in the previous section, it becomes evident how complex causal inference is in finance issues. Financial market data can be tricky as it is time-dependent and can be long considering how long some assets have been around. While one can forecast returns using only past values, more variables need to be considered to understand causal effects, and they can have different frequencies. Furthermore, Fama (1970), in his influential article, argues for an efficient market, where one can expect price fluctuations to represent all the information in the market and update according to new information as quickly as they become available. Given how many factors influence financial markets, the efficient market hypothesis imposes a considerable conceptual challenge to traditional causal inference. First, one can never be confident of having good or bad controls (Young, 2019) or even adding all relevant variables to the system. It is also expected that many of the variables added to a model might act as confounders.

Granger (1969) developed a solution that helps with this type of problem that has been widely applied in finance causality analysis (to mention a couple of recent examples: Chang, Ilomäki, Laurila & McAleer, 2020; Mighri, Ragoubi, Sarwar & Wang, 2022). This model studies pairwise covariance to establish whether there is a flow of information from one variable to another (Granger, 1969; Ding, Chen & Bressler, 2006). While helpful to understand possible causal relationships, one must take a cautious stance towards Granger causality since what is measured is whether there is a flow of information and who is the leader and follower in the relationship (Ding, Chen & Bressler, 2006; Stokes & Purdon, 2017). Furthermore, Granger causality has been developed for linear models, and while there are applications for non-linear models (Freiwald, Valdes, Bosch, Biscay, Jimenez, Rodriguez, Rodriguez, Kreiter & Singer, 1999; Chen, Rangarajan, Feng & Ding, 2004; Ancona, Marinazzo & Stramaglia, 2004), the model is still haunted by some limitations one of them being how vulnerable the model is to the choice of variables (Stokes & Purdon, 2017; Granger, 1969). Given the shortcomings, no Granger causality analysis is offered in this work. Still, the conceptual foundations are considered, especially in what concerns how DAGs can be represented with a time component, indicating that a backdoor or feedback is not necessarily problematic if one assumes that the recursion occurs sequentially in respect to the time component.

It is worth considering how causal machine learning has been applied to finance-related questions. Causal inference in machine learning has been a hot topic in academia recently. The applications are, however, often limited to the study of policy impacts on social issues like education, economy, and healthcare, among others (Dube et al. 2020; Knaus, 2021; Grodecka-Messi & Hull, 2019; Davis & Heller, 2017; De Neve et al. 2021; Bertrand et al. 2017; Athey & Imbens, 2016). Furthermore, given the relevance of causal inference to decision-making, one should also expect great academic interest in these methods applied to managerial and financial questions. However, the literature is much thinner when intersecting causal inference, machine learning, and finance. For example, when looking specifically for the use of DML in Finance, only a handful of articles could be found by the authors, and the majority of them used simulated data to showcase a proof of concept rather than study some specific question (Wasserbacher & Spindler, 2021; Varaku, 2021; Wang, 2020).

There are a few possible explanations for this apparent lack of academic interest in the subject. First, academic research on finance is often seen as having little impact beyond academia (Brooks, Fenton, Schopohl & Walker, 2019). Second, the complex concepts and requirements for machine learning and causal inference are a significant obstacle to new researchers that might be accustomed to more straightforward econometrics methods like Granger causality (Hünermund, Kaminski & Schmitt, 2022). Lastly, financial data is complex because it is often continuous, with slight variation, time-dependent, and often with mixed frequencies (when using different exogenous macroeconomic variables). As a result, it may require a considerable amount of preprocessing to guarantee good properties necessary for some methods – like stationarity for autoregressive models (Lopata, Butleris, Gudas, Rudžionis, Rudžionienė, Žioba, Veitaitė, Dilijonas, Grišius & Zwitserloot, 2021; Yadav, Guha & Chakrabarti, 2020; Daniel, 2019).

2.4 Causal Inference and Machine Learning

As hinted earlier, “standard machine learning approaches remain purely correlation and prediction-based, confining them to analytical insights that can only partly address a wide variety of managerial decision problems” (Hünermund, Kaminski & Schmitt, 2022, p.1). Athey (2017) lists practical examples of why extrapolating prediction analysis to decision-making fails. As a simple example, one cannot expect the model selection properties of the LASSO to correctly identify causal relationships (or lack thereof, considering the sparsity generated).

Although there is an explicit limitation in machine learning regarding causal inference, both fields have been researched extensively. This maturity allows researchers to understand underlying assumptions necessary for prediction and causal inference (Zhao & Hastie, 2019). Moreover, as Hünermund, Kaminski, and Schmitt (2022), and Athey (2017) show, some recent methods attempt to reconcile machine learning and traditional causal inference methods. One such method – and the one chosen for the analysis presented here – is double (or debiased) machine learning (DML), developed by Chernozhukov et al. (2018), which can be used for estimating average causal effects (Athey, 2017).

2.4.1 Double Machine Learning

As mentioned in section 2.3, traditional causal inference requires excellent care when it comes to confoundedness, and it has severe limitations regarding the dimensionality of the observed data. Consider, for example, equation 1:

$$Y = D\theta_0 + X\beta + U \tag{1}$$

In this linear example, one may not include all possible covariates ***X*** that may affect ***Y*** and that can affect ***D***. Thus meaning that the treatment variable ***D*** could become correlated with the model error ***U*** that would contain information from covariates not included in the model. This would mean that the treatment effect parameter, θ_0, would get an omitted-variable bias (Heiss, 2020; Reese, 2021). Adding more covariates to the equation may not be as straightforward for traditional causal inference methods either. This scenario is where DML presents clear advantages.

DML allows for using any machine learning method to approximate functions and generate error terms that reduce the variance in the parameters while controlling for bias, especially omitted-variable bias (Chernozhukov et al. 2018). Given that it allows the use of any machine learning method, the dimensionality of the models is only restricted by hardware limitations. However, this hardware limitation is a severely relevant issue to consider, as will become evident in the Methodology section.

This paper presents a simplified version of the explanations given by Chernozhukov et al. (2018) to formally define and introduce the DML method. The reader is advised to read their excellent work for any proof or a more profound understanding of the theorems and requirements of this framework. Chernozhukov et al. (2018) use partially linear regression (PLR) to exemplify their method. In this example, the method is divided into two different steps, one consisting of y-tasks and another of a β-task. The y-tasks can be expressed as:

$$Y = D\theta_0 + g_0(X) + U, \; E[U \mid X, D] = 0, \tag{2}$$

$$D = m_0(X) + V, \; E[V \mid X] = 0. \tag{3}$$

As expected, ***Y*** is the outcome variable, and the PLR formula indicates a model very similar to the standard linear regression presented in equation 1. ***D***, the treatment variable, is assumed to be linearly related to ***Y***. In contrast, the control variables represented by vector ***X*** enter the equation through some unknown, arbitrary, and unobserved function $g_0(.)$. ***U*** is the disturbance or error with an expected value equal to zero. It is also assumed that ***U*** is uncorrelated with ***X*** and ***D*** but correlated with ***Y***. Here, there are no assumptions regarding the size of ***X***, and in fact, ***X*** may be so large that the number of covariates need not be small relative to the sample size, for example (Reese, 2021).

The treatment variable ***D*** is also expressed by an unknown function $m_0(.)$ of the vector ***X*** and a disturbance term ***V***. The error term has zero expected value and is uncorrelated with ***X*** but correlated with $\boldsymbol{D}$. There are no assumptions about whether functions $g_0(.)$ and $m_0(.)$ should use all or only some of the covariates. There are also no requirements for these functions to be linear or not. As these functions are unobserved, they indicate that there are no observable means of understanding how $\boldsymbol{X}$ affects $\boldsymbol{Y}$ or $\boldsymbol{D}$.

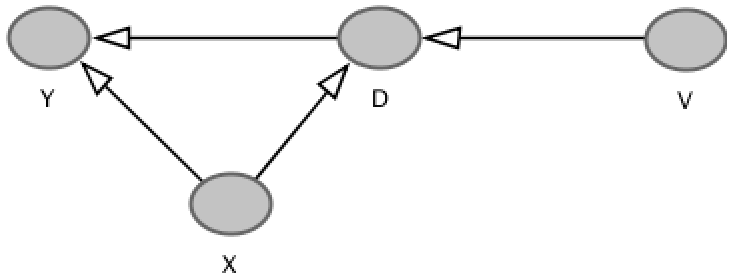

Figure 2: Causal diagram representing the PLR described by equation 2 and 3. Adapted from Bach, Chernozhukov, Kurz & Spindler (2021a).

Equation 2 is the one of interest as the parameter the method is trying to estimate is the θ_0, or the treatment effect estimate. As Chernozhukov et al. (2018, p.C2) explain, equation 3 "keeps track of confounding, namely the dependence of the treatment variable on controls. This equation is not important per se but it is important for characterizing and removing regularization bias". One naive way to estimate θ_0 is to use an estimator $D\hat{\theta}_0 + \hat{g}_0(X)$ to learn the real function $D\theta_0 + g_0(X)$ while splitting the sample to reduce the regularization bias (which may arise from the use of regularization in machine learning). A sample split of $n = \frac{N}{2}$ is considered to keep things simple, where one is used as an auxiliary sample and the other as the main sample. Conveniently, they can be called I_1 and I_2.

This same example explains that regularization bias can be overcome by using Neyman orthogonalization (Neyman & Scott, 1948), obtained by "partialling out the effect of X on D" (Chernozhukov et al. 2018, p.C3). Thus, one may train a machine learning algorithm on the auxiliary sample I_1 for each of the y-tasks and use these trained models in the main sample I_2 to produce out-of-sample prediction errors (adapted from Reese, 2021):

$$\hat{u}_i = y_i - \hat{g}_0(x_i), \tag{4}$$

$$\hat{v}_i = d_i - \hat{m}_0(x_i). \tag{5}$$

These, in turn, are used to solve the β-task in the method and through that, obtain estimates of θ_0:

$$\hat{\theta}_0 = \left(\sum_{i \in I_2} \hat{v}_i^2\right)^{-1} \left(\sum_{i \in I_2} \hat{v}_i \hat{u}_i\right) \tag{6}$$

The main advantages of DML rely on the fact that one may produce critical values and standard errors, which are of considerable importance to causal inference. Proofs for these claims can be found in theorem 4.1 in the original article by Chernozhukov et al. (2018). So, as Hünermund, Kaminski, and Schmitt (2022) and Athey (2017) stated, this method reconciles machine learning concepts and traditional causal inference by exploiting their strong features.

One last caveat must be considered. The DML framework has three main assumptions about the data and the chosen machine learning methods. First, it is assumed that all observations, and therefore their variable combinations, are independent draws "from the same true joint distribution" (Reese, 2021). Secondly, the error terms $\boldsymbol{U}$ and $\boldsymbol{V}$ are assumed to have zero expected value conditioned on $\boldsymbol{X}$, and they should be somewhat random. Lastly, it is assumed that the machine learning methods chosen can estimate the true function $g_0(.)$ and $m_0(.)$ reasonably well. Particular caution is taken in this study regarding the first assumption (as shown in the Methodology section). However, given the nature of the data and the methods chosen, the second assumption is taken as fulfilled. Finally, two methods are chosen to estimate the true functions. Simple linear regression is used first to establish some benchmarks, and gradient boosting is used to check for performance improvement in this framework. A short introduction of autoregressive models is presented next, followed by a brief presentation of gradient boosting, as these were used to improve the models used in this study.

2.4.2 Autoregressive Models

Autoregressive models are widely used for forecasting and are generally perceived to perform well given a series of assumptions (Enders, 2004). These models are called autoregressive because they forecast an outcome variable y with its past information. While helpful, these models face the problem that they do not consider exogenous or multiple endogenous variables. As Enders (2004) explains, one option for this problem is vector autoregressive (VAR) models, a type of dynamic regression model. In these dynamic models, the equation used to forecast the outcome variable y contains past values of y and present and/or past values of the exogenous variables added to the model (Enders, 2004). Specifically for VAR models, these are also considered symmetric; thus, there is an underlying assumption that y can be used for forecasting other endogenous variables.

Furthermore, the viability of VAR models is usually confirmed by checking whether there is Granger causality among the endogenous variables (Enders, 2004). Lastly, the same assumptions made for autoregressive models – stationarity, for example – are necessary for VAR models too. VAR models are relevant for the questions this study tries to answer because 1) they are often considered a benchmark for financial returns forecasting (Leung, Daouk & Chen, 2000), 2) they are widely applied to finance issues given their flexibility and ease of interpretability for time-series problems (Diebold & Yilmaz, 2012; Huang & Mollick, 2020; Hansson, Jansson & Löf, 2005; Kwon, 2022), and 3) they might help in identifying the optimal number of time lags to produce the best forecast. Another advantage of VAR models is that their estimates can be obtained with ordinary least squares (OLS), as with linear regressions (Enders, 2004).

2.4.3 Gradient Boosting

Gradient boosting is a technique that improves decision trees, which themselves are suitable for predictions when there are non-linearities (James et al. 2013). As financial markets are highly complex, one can expect their behavior to be nonlinear (Yadav, Guha & Chakrabarti, 2020), which justifies using some machine learning methods to capture these patterns. Many authors have used boosted tree methods for the prediction of stock returns to various degrees of success (Krauss, Do & Huck, 2017; Chatzis, Siakoulis, Petropoulos, Stavroulakis & Vlachogiannakis, 2018; Basak, Kar, Saha, Khaidem & Dey, 2019; Mittnik, Robinzonov & Spindler, 2015; Zhou, Zhang, Sornette & Jiang, 2019; Gündüz, Çataltepe & Yaslan, 2017). Schapire (1990) introduced boosting as a technique that uses many weak learners to achieve high accuracy. When it comes to decision trees, the core idea is to use many small-sized trees, i.e., trees with only a few terminal nodes, to achieve prediction accuracy only slightly better than a coin flip (James et al. 2013; Schapire, 1990; Hastie, Tibshirani & Friedman, 2009). These weak learners are fitted sequentially and over updated versions of the training data to improve accuracy where it might have failed previously. A better conceptual introduction to the theme can be seen in detail through Hastie, Tibshirani, and Friedman's (2009) work.

3 Implementing the Double Machine Learning Framework

The statistical learning methods chosen for this study were linear regression and gradient boosting. Their implementation in Python was done using the tools provided by the *scikit-learn* library (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot & Duchesnay, 2011). Linear regression is a ubiquitous and easy-to-understand method, and it sets a baseline to compare error scores and the results. However, as shown in the Results section, linear regression fails to capture the variance in the model, which is troublesome since a good prediction tool is a prerequisite for DML. On the other hand, gradient boosting has performed remarkably well for both the passive and active funds set, but it demanded a much more laborious process.

A simple linear regression was fit to the pre-processed data as a way to measure its efficacy. The DoubleML library, created by Bach, Chernozhukov, Kurz, and Spindler (2021b), implements linear regression as a statistical learning tool in Python. As linear regression is very straightforward, one can plug the data into the DML framework on Python and obtain results. The same cannot be said for the gradient boosting framework.

The *xgb* module in the *scikit-learn* library allows the researcher to choose a series of parameters that can, in turn, be tuned to find the ideal parameters to either maximize some score or minimize some error (Pedregosa et al. 2011). As boosted trees were used, the list of parameters to be tuned was quite extensive. Table 1 indicates the parameters used for tuning, which were cross-validated in two folds. At this stage, hardware can be a great limiter, and as Hastie, Tibshirani & Friedman (2009) states, the combination of parameters to tune is likely to be suboptimal. The tuning of the parameters is done then by executing a grid search where the grid is a matrix with all possible combinations of the given parameters. The computational cost of this problem is considerably high. It can achieve long computation times since some of the parameters (like the number of trees or the depth of trees) can significantly increase the number of calculations. The time-complexity of gradient boosting is certainly something one should consider, and given the time constraints of this study, the tuning parameters were chosen arbitrarily and their cross-validation limited to two folds. Initially, there was a plan to use 5-fold cross-validation. However, as each fold adds to the number of times the possible combination of tuning parameters need to be run, one must be mindful. For example, if one full run of the possible parameters takes three hours, a 2-fold cross-validation will take six hours. The fitting of the tuning parameters considered two objective functions, that is, mean squared error (MSE) and R^2. Meaning the tuning looked for the parameters that produced the lowest MSE possible for the highest R^2 possible.

Another factor into the time complexity is the size of the datasets. Whereas the active set was a 906,205x12 matrix at first, after adding the lagged variables and the fixed effect variables the matrix grew to 906,205x88. Thus, all of the tuning time complexity takes even longer considering a larger set. Therefore, the chosen tuning parameters work as an initial guess of the ideal parameters. They serve more as a proof of concept so that future research, with more powerful hardware, can proceed with a more extensive tuning grid search.

Once the tuning is finished, both the ideal combination of parameters and the error and variance scores are obtained. As expected, gradient boosting increased predictive performance enormously, fulfilling the prediction performance prerequisite. This tuning procedure was repeated for each y-task in the passive and active funds set. With the ideal combination of parameters, one can then structure the DML procedure by indicating the statistical learning method used for each of the y-tasks and the tuned parameters. The indicated statistical learning methods at this stage are precisely the functions $\hat{g}_0(.)$ and $\hat{m}_0(.)$ mentioned in the Literature Review. Sample splitting, cross-validation, refitting, and other transformations necessary for the DML framework are done in an "automated" fashion. The authors of the *DoubleML* library envisioned it as an accessible tool to apply the framework explained in Bach et al. (2021). Following their example, all code used in this research, together with the data collected is available in a github repository.

Table 1. Range of parameters used for tuning (Pedregosa et al. 2011). The step size of the learning rate was 0.01. Max depth options are based on the suggestions given in Hastie, Tibshirani, and Friedman (2009).
The tuning was performed in two steps. In the first all parameters except the *learning_rate* were tuned with a *learning_rate* of 0.2. Once those parameters were defined the *learning_rate* was tuned. All set of parameters chosen by the tuning included the highest option for *subsample* and *colsample_bytree*, the maximum tree depth, and some level of L2 regularization.

Parameter	**Range**	**Description**
learning_rate	[0.01 - 0.2]	Step size shrinkage of weights
objective	reg:squarederror	Learning objective. In this case, regression with squared loss
n_estimators	[200, 500, 1000]	Number of gradient boosted trees
reg_lambda	[0, 1, 2]	L2 regularization term on weights
subsample	[0.3, 0.6, 0.9]	Ratio of the training data randomly sampled. Prevents overfitting.
colsample_bytree	[0.3, 0.6, 0.9]	Ratio of columns randomly subsampled in each tree
max_depth	[2, 6, 10]	Maximum tree depth for base learners

4 Data

4.1 Financial Data Collection

4.1.1 Funds data

This study focuses on macroeconomic policy's effects on funds traded in the US market. In the first stage, the data collection focused on filtering what funds were currently active in the market, and as there are 14,816 funds, some filters must be applied to make the data more manageable. Therefore, the Bloomberg terminal at Lund University was used to find and filter the funds.

The Bloomberg terminal has some inherent limitations regarding the amount of data it can show. For example, while there are 14,816 funds available in their database, only 5,000 are displayed at any given time. It is also unclear what criteria Bloomberg uses to select the 5,000 funds shown. Based on this, the funds were filtered according to criteria that would reduce their maximum number to 5,000 or less, which would simultaneously guarantee the data to be more manageable and that all the funds shown are, in fact, all the available ones. The criteria chosen for filtering the funds and their impact can be seen in Table 2.

Table 2. Filtering criteria and actively managed funds available.
These criteria were used to select the fund tickers from a Bloomberg terminal. The choice of filtering criteria is arbitrary to keep the number of funds in the set limited to the total number of funds the Bloomberg terminal can show at once.

Filtering criteria	**Number of funds**
Market Status: Active:	426,366
Fund Primary Share Class = Yes:	139,777
Country/territory of Domicile: United States:	14,816
Inception Date >= 1/1/1985:	14,007
Fund Asset Class Focus: Fixed Income, Equity:	10,996
Fund Actively Managed = Yes:	5,645
Parent Company Name:	5,645
Fund Industry Focus:	5,645
Fund Total Assets (mil) >= 20M:	4,976

The same criteria were applied later to passively managed funds for consistency. Due to an overall lower number of passively managed funds, this dataset resulted in 1,948 funds in total. For both actively and passively managed funds, the data collected consisted of the tickers, formal name, parent company, fund asset class, inception, assets under management, i.e., most of the filters used to select the data. Therefore, the data collected from Bloomberg can be considered metadata.

With this metadata set in hand, we used Python and the *yfinance* library to download the data from the Yahoo Finance database. The data was collected based on the tickers available from the metadata, and the range chosen was from December 1985 to December 2021. The choice for this period is somewhat arbitrary but based mainly on what other articles have used previously (Mighri et al. 2022; Chang et al. 2020; McMillan, 2016). For this study specifically, this period offers some exciting possibilities since it will allow us to study the behavior pre and post three different financial crises and the creation of new types of actively and passively managed funds.

4.1.2 Macroeconomic variables

The choice for independent variables used in this study is primarily based on what common sense and basic economic theory would say influences the market: GDP growth rate, inflation, unemployment levels, money supply, and the central bank interest rate, which is the treatment variable of interest here. Most of this information is available quarterly or monthly in the Federal Reserve Economic Data (FRED, 2022). At the same time, other indexes of interest, such as the anxious index and consumer sentiment, are available through university databases (University of Michigan, 2022; Federal Reserve Bank of Philadelphia, 2022). The choice for these variables has some academic base as they are used in similar studies in other works (or at least similar versions of this data) (McMillan, 2016; Chen, 2009; Neely et al. 2014; Paye, 2012; Altibas & Biskin, 2015). Table 3 lists the variables and their descriptions.

Table 3. Macroeconomic variables and their statistics before pre-processing
These are the macroeconomic variables collected in the start of the research. A short description given by their source is offered together with summary statistics. The treatment variable is the first difference of the *interest_rate* variable.

Variable	Description	Min	Mean	Max	SD
anxious_index	The probability of a decline in real GDP in the quarter following the quarter in which the survey is taken. (Federal Reserve Bank of Philadelphia, 2022)	2.16	18.50	74.77	14.84
consumer_sent	The Index of Consumer Expectations focuses on three areas: how consumers view prospects for their financial situation, how they view prospects for the general economy over the near term, and their view of prospects for the economy over the long term (University of Michigan, 2022).	51.7	85.94	112.0	12.47
inflation	The Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL) is a price index of a basket of goods and services paid by urban consumers. Percent changes in the price index measure the inflation rate between two time	21.48	114.57	284.18	80.82

	periods. It can also represent the buying habits of urban consumers. (FRED, 2022)				
gpd_growth	Real gross domestic product is the inflation-adjusted value of the goods and services produced by labor and property in the United States. (FRED, 2022)	-0.08	0.007	0.0754	0.009
hpi	The FHFA House Price Index (FHFA HPI®) is the nation's only collection of public, freely available house price indexes that measure changes in single-family home values based on data from all 50 states and over 400 American cities that extend back to the mid-1970s (FRED, 2022).	60.01	240.9	557.73	119.40
interest_rate	The federal funds rate is the interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight. (FRED, 2022)	0.05	4.63	19.1	3.62
m2	M2 includes a broader set of financial assets held principally by households. M2 consists of M1 plus: (1) savings deposits (which include money market deposit accounts, or MMDAs); (2) small-denomination time deposits (time deposits in amounts of less than $100,000); and (3) balances in retail money market mutual funds (MMMFs) (FRED, 2022). (In billions)	286.6	4,717	21,811	4,798.49
nrou	The natural rate of unemployment (NAIRU) is the rate of unemployment arising from all sources except fluctuations in aggregate demand. (FRED, 2022)	4.27	5.31	6.23	0.5898
recession	This time series interprets US Business Cycle Expansions and Contractions data provided by The National Bureau of Economic Research (NBER). Our time series is composed of dummy variables representing periods of expansion and recession. (FRED, 2022)	0	0.288	1	0.453
rou	The unemployment rate represents the number of unemployed as a percentage of the labor force. (FRED, 2022)	2.5	5.759	14.7	1.687

These macroeconomic variables will, from now on, be collectively called the variable ***X*** as they will be treated as the independent variables of this study. Furthermore, when this paper refers to the returns of a single fund, it shall be referred to as **y**; however, when the collection of funds is mentioned, it will be referred to as ***Y***. With these conventions defined, the frequency of both ***X*** and ***Y*** variables must be discussed.

4.1.3 On the frequency of the data

The frequency of ***Y*** is monthly; the main reason is that the treatment variable in question does not change very frequently and that the dataset itself has a monthly format. Additionally, many of the FRED datasets are offered in a monthly format. Some of the ***X*** data had a quarterly form. These were transformed accordingly by either repeating one month's information through the subsequent two months or by bridging the data by interpolating the two points. While this method is somewhat usual, it can add unwanted noise to the model (Foroni & Marcellino, 2013). Although there is a risk of adding unwanted noise, only four variables were in quarterly format. Thus the noise added would be minimal. Considering a monthly structure, the ***Y*** set would be a 432x4,976 matrix for the actively managed funds and a 432x1,948 for the passively managed funds. The number of columns is slightly smaller because some of the data was unreachable using *yfinance*, but the data loss is minimal. The ***X*** set would then be a 432x11 matrix.

There is also a considerable amount of missing data regarding the ***Y*** set. This is because not all funds started simultaneously, and not all funds existed at the same time. In a closer inspection, it is possible to see that between 1985 and 1996, there was a minimal number of actively managed funds. However, since 1996 the number of funds increased drastically, and from around 2008, there seems to be another great leap in the number of funds available in the market. In other words, missing data is less of a problem in later years. The ***X*** set has no issues with missing data for the period studied.

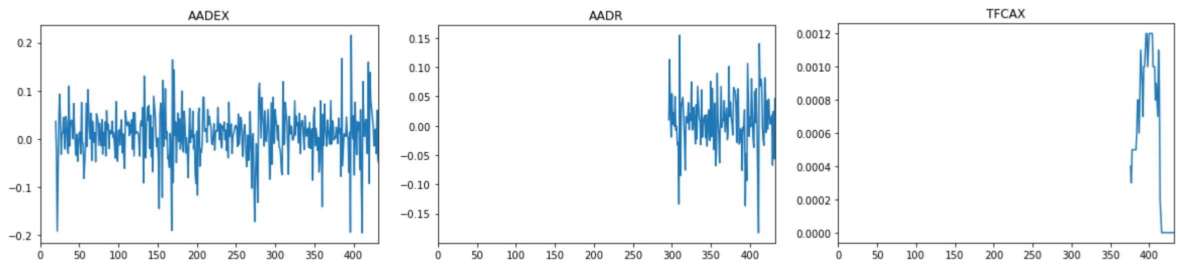

Figure 3: Monthly data for three actively managed funds. The AADEX fund has data for almost the whole period, whereas AADR has data starting later. The TFCAX fund began at a late date and lasted for only a few years (hence the zeros in the later months). This shows how the quantity of information depends on the fund.

The missing data problem is very challenging because it dramatically reduces the robustness of the data when trying to develop a method that can model any **y.** For example, some funds only existed for less than a year, meaning that in a given **y,** there could be as low as ten observations for the whole period. Furthermore, only a few of the funds had data for all 432 timestamps; therefore, developing a method to predict each **y** could generate inconsistencies as a model could work well for those **y**s with much data and probably return dubious results for those **y**s with very few observations.

4.2 Data Preprocessing

4.2.1 Correlation analysis

Before applying any machine learning method, a simple correlation breakdown helps understand how the ***X*** variables might be related to each other and to ***Y***. Given ***Y*** is composed of thousands of different funds the average value for ***Y*** at each point in time is used instead. Some variables are expected to have a high correlation (such as inflation and real GDP), and the analysis confirms this. However, highly correlated variables can indicate backdoors and other problems for causal inference (Heiss, 2020).

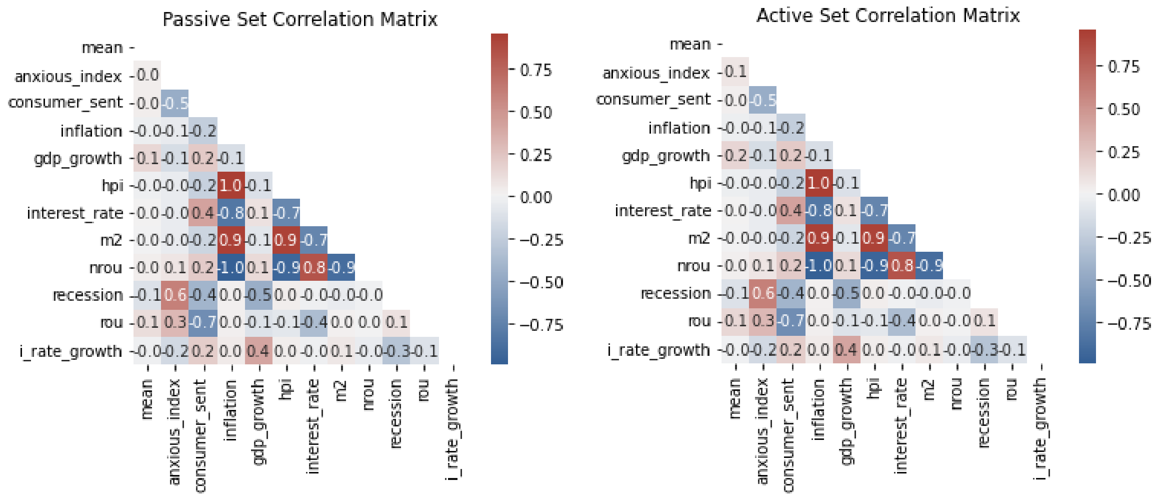

Figure 4: Correlation matrix of the passive and active datasets prior to detrending. Values are rounded to the first decimal. "*mean*" is the average value of ***Y*** for point in time in the set. Some of the variables like "*interest_rate*" and "*nrou*" are non-stationary at this point. The difference in correlation between the two sets is negligible.

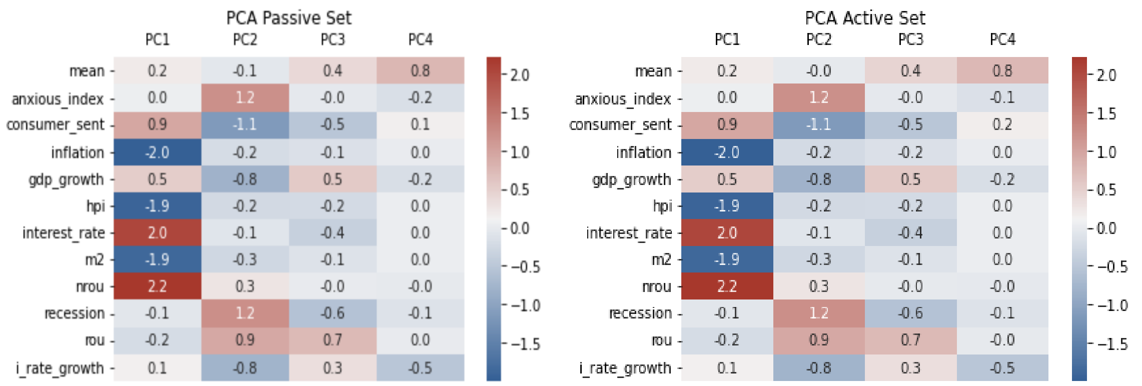

Figure 5: Principal component analysis of the correlation matrix. The first four components represent 95% of the data.

4.2.2 Checking for stationarity

VAR models generally assume the data to be detrended, i.e., stationary around a mean (Enders, 2004). A solution for this is to take the first difference of the non-stationary variables and, in some cases, even higher-order differences. Another problem that detrending can solve is the high correlation among the variables. For example, real GDP and inflation are highly correlated, but their growth rates (the first difference) are not. After checking for stationarity using an Augmented Dickey-Fuller (ADF), it was possible to see that some variables needed detrending while others did not. Therefore, for the sake of consistency, all variables were detrended. The rationale behind this choice follows that used in McMillan's (2016) work. It also significantly reduced the correlation among the variables, and when checking for the variance inflation factor (VIF), almost all variables were within acceptable levels after detrending. The only variable which was not stationary after detrending was the natural rate of unemployment, which was then discarded from the study for simplicity. The recession dummy variable was also not detrended since it is a dummy variable. Given that almost all variables were detrended, it is possible to say then that all of the variables mentioned earlier should be interpreted from now on as the growth rates of the original data concerning the previous month.

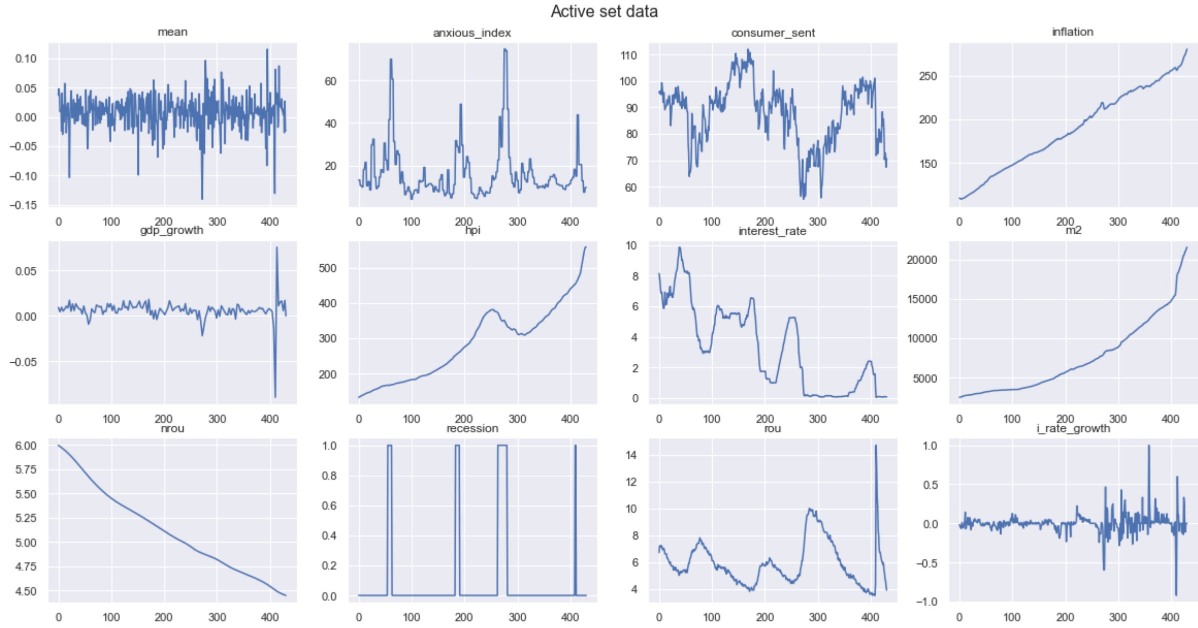

Figure 6: Average ***Y*** for the active set together with the untreated predictors. Average ***Y*** is named "*mean*" in the plot.

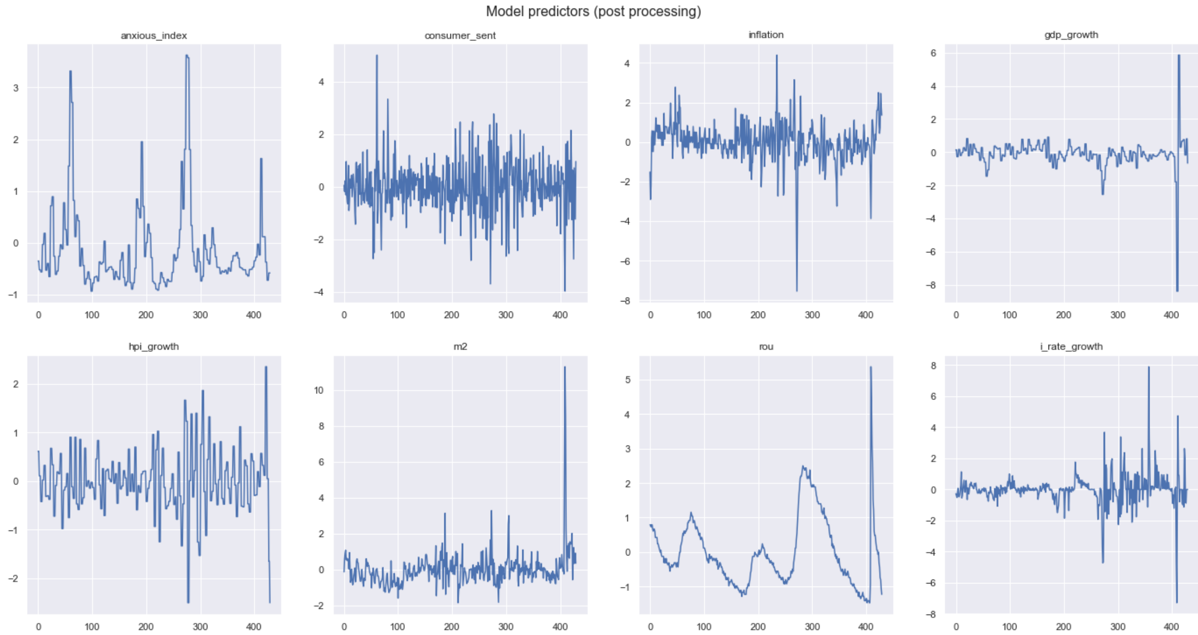

Figure 7: Model predictors after preprocessing. Only stationary variables were left (thus interest_rate and nrou were removed). As *recession* is a dummy variable no preprocessing was necessary and its shape did not change from the one in Figure 6. Data transformations involved detrending and standardization only.

4.2.3 Time as an explanatory variable

While this work does not use VAR models to measure causality directly, VAR models still offer some utility in adding a time component to other methods. As the amount of observations per fund varied greatly, running an individual iteration per fund would likely result in inconsistent performance for each iteration (as was the initial plan for this work). This inconsistency would make aggregating the individual results an untrustworthy technique. A solution to this would entail using a panel format for the data, which, at first, would mean losing the original time component in the data, especially considering that some randomization would be used to divide the set into two samples for the DML framework. The solution was to add lagged versions of all variables as new variables to add the time component back to the panel format. A VAR model was used to aid in choosing the ideal number of lags where the average **Y** for every date was used as the **y** variable. All other variables were included as endogenous variables. Using Python and the *statsmodels* library (Seabold & Perktold, 2010), the VAR model was constructed with a maximum number of twenty-four lags (twenty-four months into the past), and the optimal number of lags was chosen using AIC scores. Based on the lowest AIC score, looking seven months into the past was ideal for forecasting these variables.

However, it is worth mentioning that the average ***Y*** can differ greatly from the actual observations, which can cause issues with the VAR model as it bases its predictions on the average **Y**. Figure 8 shows that even when considering only the first 100 funds in the active set the minimum and maximum values are considerably far from the mean, and when considering all funds it is possible to observe the presence of considerable outliers. While one should approach this technique with caution, the results show that using the number of lags suggested by this technique achieved a sufficiently good result for the selected machine learning frameworks.

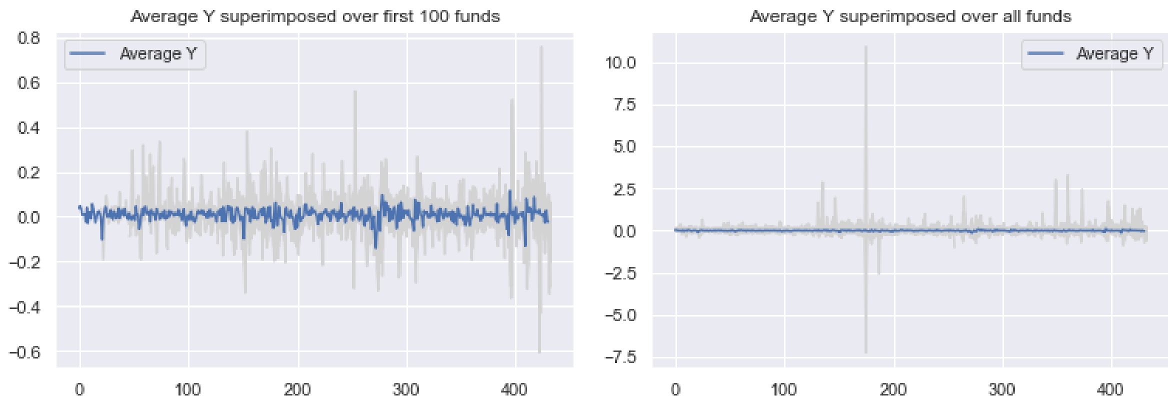

Figure 8: Average Y superimposed over all funds. The gray line are several lines on top of each other. As the funds' returns converge to zero they overlap and form the illusion of just one line. Left: The average ***Y*** in the active set compared to the first 100 funds in the set. Right: The average ***Y*** in the active set compared to the all the funds in that set.

Choosing to lag the data for up to seven months has practical implications. A filter removed all funds with less than eight observations. Additionally, from the funds left after filtering, some unique observations were lost since they did not contain all seven lags, i.e., only those **y**s with enough observations to create seven lags were kept in the set. The loss of the seven observations resulting from the lagging does, however, not entail too much loss of information as the removed observations were simply moved into the lagged versions of the data.

4.2.4 Paneling the data

The data follows a Time Series Cross-Section (TSCS) format, but it has too many inconsistencies. For example, TSCS data has a number of variables larger than the observations, which is precisely the original data format (Urdinez, 2020). By pivoting this data into one column, one can have many observations. Thus, in the active funds set, the data went from a 432x4,987 matrix (4,976 funds plus 11 variables) to 906,205x12.

This implicates partly changing the conceptual perspective through which the data is analyzed. While the inference is conditioned on each fund in the TSCS format, panel data disregard this and focus on the whole population (Urdinez, 2020). Conceptually this does not impose a problem on the research since this study is interested in the average treatment effect, regardless of how heterogenic this effect might be among the individual funds. The problem here is that panel data assumes all observations to be independent of each other, which is not the case, given that there are thousands of different funds (or factors) (Urdinez, 2020).

4.2.5 Adding fixed effects

One of the assumptions in the DML framework is that the observations should be independently drawn from the same true joint distribution (Chernozhukov et al. 2018). This assumption could be a problem with the panel data format since not all variables are random. For example, the fund ticker variable is a fixed effect along with many different observations. One way to solve this would be adding a numeric fund identifier variable to the model. While the DML framework could handle this – after all, it does not have any assumptions over how many variables can be added to the model – doing one-hot encoding or a numerical identifier could be computationally costly since, in the active fund set, there are almost five thousand unique funds.

Johannemann, Hadad, Athey, and Wager (2021) introduce a few methods that can be used to add these fixed effects to panel data. These consist of using techniques that encode most of the unique information to a specific group without generating new individual variables for each group. The simplest method, and the one chosen for this study, is means-encoding which consists of adding the mean of every variable for each specific group. The number of new variables is limited to the number of variables in the model. The combination of means will be unique to each group (or fund in this case). Johannemann et al. (2021) also show that this type of encoding conserves more information than simply doing a one-hot encoding (or other traditional fixed effect techniques). This study applied the means encoding, considering only the detrended unlagged variables and recession; thus, only nine more variables were added to the other variables and lagged variables already in the model.

Urdinez (2020, p.158) states that "[b]y having data that varies over time and/or among the individuals, [a] model will have more information than in cross-section models, thus obtaining more efficient estimators." Thus, applying some type of encoding to add fixed effects and using lagged variables to add the time component effects can help reduce omitted-variable bias. In addition, it allows the model to consider the patterns that only exist when observing either the time component, the fund identifier, or both. This is comparable to the assumption made by Johannemann et al. (2021, p.2) in which no “group membership … has causal effects over Y”, but the group membership may be associated with some latent state that has causal effects over Y.

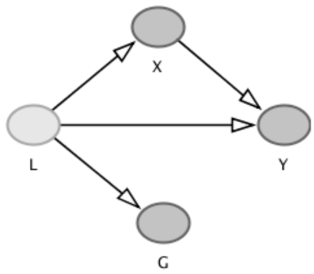

Figure 9: DAG adapted from Johannemann et al. (2021, p.2). Vector ***G*** represents the collection of all unique groups (or funds). These groups have no effect over ***Y*** but they might contain information from some unobserved latent state ***L*** that has influence over ***X***, and consequently, over ***Y***.

5 Results and Discussion

5.1 Analyzing the models' predictive performance

The first result presented in this paper investigates the ability to predict the data of the two types of models used, i.e., linear regression and gradient boosting. As mentioned in the Literature Review, the motivation for including the individual models' ability to make predictions is due to the assumption double machine learning makes about the utilized models' ability to estimate the true functions $g_0(.)$ and $m_0(.)$ well enough. The model's predictive ability is described in table 4, presented as the model's R^2. A new term is introduced here to aid in the analysis. Since there is a pair of y-tasks for each statistical learning method chosen, they will be defined as y-task and d-task. The d-task is the y-task for predicting the treatment variable, i.e., equation three from the Literature Review. In contrast, the y-task refers to equation two in the Literature Review.

Predictive performance is arguably the most complex challenge of the three main requirements presented in the Literature Review concerning DML. Good predictive performance demands good models, data, and frameworks. Some of the problems it may try to estimate can be highly complex and computationally costly. The challenge becomes harder because worse predictive power introduces noise to the estimates of the average treatment effect. If the error terms $\boldsymbol{U}$ and $\boldsymbol{V}$ are too large, the PLR used to estimate the average treatment effect will likely produce large coefficients, which may not be as precise as one would like. In policy-making or asset management, these errors can compound disastrous consequences. In this study, the gradient boosting model produced a considerably large R^2, one considerably larger than other research that attempted to predict stocks and funds' returns. However, whether the predictive performance achieved here is sufficient for the precise estimation of average treatment effects is a matter for future research.

Table 4. R^2 of the models on the entire dataset. In the gradient boosting case, the statistics were obtained through cross-validation. The rounded adjusted R^2 results are the same values as the R^2 and are, therefore, not shown.

	Linear Regression		Gradient Boosting	
	y-task	d-task	y-task	d-task
R^2 (passive):	0.19	0.60	0.81	0.99
R^2 (active):	0.13	0.54	0.73	0.99

These results show that the linear regression performs much worse than the gradient boosting, indicating that the relationship in the data is more complex than what a linear regression can model.

This is an expected result as predicting returns of funds is very complex, as exemplified in the Literature Review where competing opinions and numerous arguments for what and how macroeconomic variables affect funds were shown. This result contributes to this work's first research question by showing that more complex models as learners likely are required for double machine learning. It also contributes specifically to what concerns finance questions or financial data. Given that this type of data is usually time-dependent, there was some doubt at the start regarding how well DML could deal with it, given that it proposes a type of sample splitting that works better with randomization. One possible workaround for this was adding lags of all variables to the model, which DML can handle very well, and it did not seem to affect the well functioning of the framework.

The results presented in table 4 show that the gradient boosting models have high predictive ability, likely due to the combination of multiple schools of thought when it comes to prediction and forecasting. Firstly, the autoregressive school has its time dependency influence from which the lagged variables were inspired. Secondly, the machine learning school with its ability to learn and estimate potentially any high-dimensional complex function given enough observations. The machine learning aspects of the creation of this model were greatly enhanced due to the vast amount of observations made available through the conversion of the data into panel data. The transformation into panel data was only made possible due to the fulfillment of the independence assumption of the observations granted by the addition of the fixed effects. This likely enhanced the ability of the model to learn extensively as the observations increased dramatically compared to, for example, an aggregate version of the fund data where there would at most be 432 observations, or if the model instead were to only look at individual funds where the observations would instead be from 8 to 432 observations per fund.

Lastly, from table 4, one can see that the d-task model has a very high R^2 value for both the active and passive funds, this is a sign that the models might be overfit. This result is also expected due to the high correlation between the $\boldsymbol{X}$ and the treatment variable, as evidenced in figure 4. The possible overfitting comes as a surprise, though, since many of the parameters selected through tuning are used for controlling overfitting. It is important to note that the correlation was reduced considerably once most of the data were detrended. However, some of the variables still had reasonably high correlation even after detrending. The correlation is likely due to the modern monetary theory, which informs central banks on how to adjust the interest rates in response to many of the included variables of the model.

5.2 Analyzing the Double Machine Learning models

The results of the double machine learning models are presented for both the passively and actively managed funds. In tables 5 and 6, one can see that in the linear regression models, the interest rate’s growth rate has a significant negative causal effect on both management styles. On the other hand, the results for the gradient boosting models only show a significant negative causal relationship between the actively managed funds and the interest rate’s growth rate. Due to the treatment variable, interest rate’s growth rate, one should interpret the coefficient in tables 5 and 6 as the effect on the fund return caused by one unit change in the treatment variable. Under this perspective, a one-unit increase can be interpreted as a 100% increase in the interest rate. However, one should also consider that hikes of this magnitude are somewhat rare, meaning the most likely observed effect would be much smaller. In light of this, a more practical interpretation would be to divide the coefficient by 100, as seen in table 7. This shows the effect that a 1% increase in the interest rate has on fund returns.

To aid in comprehending these results, one could imagine a hypothetical situation where funds' returns are only dependent on the interest rate's growth rate. In this scenario, the returns would be precisely the same as $D \cdot \hat{\theta}_0$. The Federal Reserve would be very wary of sharp increases in the interest rate, given the economic importance of funds. Luckily, there are many more effects that influence funds' returns, and the estimates presented here need to be considered together with other market-defining events. Even if other effects might counterbalance the effect of interest rate hikes, these are powerful enough to produce negative returns for investors and managers alike. This is undoubtedly more attenuated because the financial market immediately exercises the effects of interest rate hikes. In contrast, other sectors of the economy are likely to feel them at a slower pace. The speed with which inflation changes is an excellent example of this phenomenon.

Table 5. Result of the Double Machine Learning of the passively managed funds.
Coefficients are the estimated average treatment effect. Here only the linear regression produces significant results, however, the linear regressions have very low predictive power which is a requirement for DML.

	$Coef.$	SE	t	$P > \|t\|$	2.5%	97.5%
Linear Reg.	-0.025	0.001	-23.226	$2.289e-119$	-0.027	-0.023
Grad. Boosting	0.229	0.221	1.034	0.301	-0.205	0.663

Table 6. Result of the Double Machine Learning of the actively managed funds.
Coefficients are the estimated average treatment effect. Here both models produce significant results, but the estimated effect is much larger for the gradient boosting model.

	$Coef.$	SE	t	$P > \|t\|$	2.5%	97.5%
Linear Reg.	-0.019	0.001	-32.671	$3.989e-234$	-0.021	-0.018
Grad. Boosting	-11.970	2.522	-4.747	$2e-6$	-16.913	-7.028

Table 7. Adjusted coefficients of the DML, transformed for the effect of 1% (instead of 100%) change.
As 100% changes are extremely rare, it makes sense to appreciate the results in a more likely scenario, i.e., by calculating the effect under a 1% change.

	$Passive$	$Active$
Linear Reg.	-0.00025	-0.00019
Grad. Boosting	0.00229	-0.1197

5.2.1 Linear regression in Double Machine Learning

The results of the DML models utilizing the linear regression as its learners were to be expected when considering the literature on how the management style affects the fund's performance given changing interest rates. Unsurprisingly, the results agree with the most influential macroeconomic theories. For example, the CAPM suggests that macroeconomic events which negatively affect the overall economy also affect fund returns. The literature also suggests that active management funds perform better in poor macroeconomic conditions. Recessions, for example, are a symptom of such poor macroeconomic conditions, and interest rate hikes are often the remedy applied to solve these. Financial markets are highly complex, and they are affected constantly (and arguably in real-time) by many different events that are also related to overall macroeconomic conditions. As these effects tend to synergize - for good or bad - one could almost compare what happens to compounding in that events feed on each other and thus increase in magnitude. Under this interpretation, the significant adverse effect on the funds can reflect the combination of many adverse events. This type of uncertainty created by poor macroeconomic conditions can foster the ground for opportunities for increased returns or hedging against downturns that can only be reaped by active management, which explains the more prominent negative effect on passive funds relative to the active ones.

However, the validity of the results of these DML models using linear regression as learners can be questioned due to the relatively low R^2 of these models (Table 4). The low R^2 suggests that, perhaps, the assumption DML has about the learners being able to capture the true functions well enough might not have been fulfilled. This would mean that noise would be introduced in the model, which could have led to the small coefficients relative to the gradient boosting model's coefficients. The results can also be questionable due to the rigidity of the linear models as learners as the variables in the data can not interact with each other as is possible with the tree-based method of gradient boosting. With decision trees, the variables interact depending on where the splits in the data are done in each tree. The inflexibility of the linear regression models potentially introduces omitted-variable bias as the data does not interact as is possible with more flexible models.

5.2.2 Gradient boosting in Double Machine Learning

The expectation for the results of these models, as with the linear regression models, was that the interest rate's growth rate would negatively impact the funds' returns, with the active funds having a minor negative effect relative to the passive. The significant adverse effect on the active funds was thus expected; however, the insignificant positive effect of the passive funds was not. The relationship between active funds and the treatment variable is likely the most trustworthy in this research since it is generated by a model that does not suffer from the same shortcomings as the linear regression-based models and since the result is also significant.

These results open up the discussion about whether the previous literature on how funds perform in response to changes in macroeconomic variables is correct or flawed. The negative relationship suggested by the active fund data supports the findings of the previous literature. However, the result of the passive data questions it. To further discuss and make conclusions about whether this paper's findings support or contradict previous research, one needs to look at the models to see which is more trustworthy and thus come to conclusions about how to utilize DML in this field further to determine the effect the interest rate's growth rate has on fund performance. The difference in significance, 2e-6 for the active data result relative to the 0.301 of the passive data result, clearly suggests that the result of the active fund data is more reliable than the passive one.

The relatively poor significance of the passively managed funds result might be due to many potential reasons. Firstly, the difference in the number of observations between the passive and active data might have affected the results as machine learning methods become better at emulating true functions as the number of observations increases. The relatively low number of observations of passively managed fund data, which were about one-fourth as large as the active set, might have caused the model not to perform as well, resulting in a lower significance level. Secondly, the data collection of the passive funds might have been flawed since, as was mentioned earlier in the Literature Review, even funds classified as passive or active exist on a spectrum of activeness versus passiveness. Thus, the collected passive fund data might not have been homogeneous in its passiveness which could have hurt the results of that model. Say that, for example, the interest rate's growth rate significantly affects the returns of passively managed funds. However, since the data classified as passive is heterogeneous, this result would get distorted by including data to which there might be a different relationship. Thirdly, the nature of index funds is to follow a predetermined investment objective that can limit, for example, which industries or types of assets it invests in. The effect of the interest rate's growth rate might be different on the different investment objectives, which again could distort the model's results and lead to the insignificance of the result. A quick exploration of the metadata shows that many of the funds follow very few indices. Thus, it can also be argued that the effective amount of information in the set might be much smaller since a considerable portion of the funds follows a tiny number of indices. However, this and the other hypothesis concerning the high p-values of the passive set would require further research to better understand the driving force behind this phenomenon.

5.2.3 A comment about the quality of the gradient boosting models

An increased effort was put into assessing the quality of the gradient boosting models since the passive set did not produce significant results and the effect estimated for the active set seemed somewhat high. The first thing to check was the variance among the months for both sets. This was done to check whether there was a considerable difference between the sets that could explain the not significant result, but this method's results did not show anything abnormal. Furthermore, the predicted values from the DML method were used to generate residuals which were then plotted against the ***D*** variable, and against the fitted values.

A few cues indicate whether there are issues in the model when it comes to residual plots. For example, a cone-shaped pattern can indicate heteroskedasticity, making coefficient estimates less precise. A U-shape, for instance, reveals some uncaptured nonlinearity in the model. A healthy residual plot usually has no identifiable pattern. Figure 10 shows that for both the passive and the active sets there seems to be no recognizable pattern. For example, the residuals from the passive set show a completely random scenario with no outliers. The active set has a similar situation where most of the variables seem to be randomly distributed around the same region, however, the perspective is warped by the presence of outliers. These outliers can be observed also in figure 8. It is possible to observe that the passive set is more precise – as indicated by the R^2 – since the residuals seem to cluster around a shorter range than in the active set residuals plot.

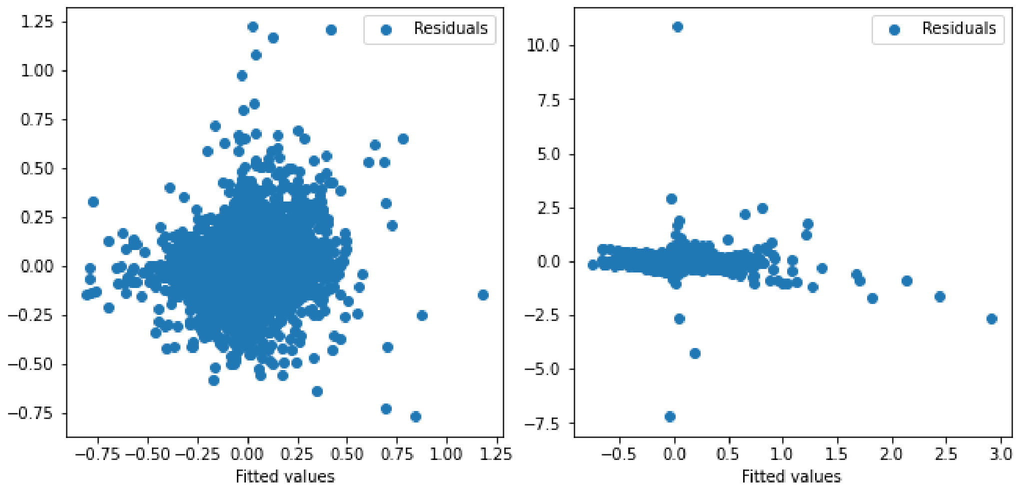

Figure 10: Residuals vs fitted values. The residuals generated by function $\hat{g}_0(.)$ are plotted against the predicted values to investigate possible problems with the model. The left side plot represents the passive set, and the right side plot represents the active set.

Some interesting insights can be obtained when plotting the residuals of the functions $\hat{g}_0(.)$ and $\hat{m}_0(.)$ against the treatment variable. First, the plots for the function $\hat{m}_0(.)$ show a cross-shaped pattern. The active funds' plot also shows an extra vertical line, as shown in figure 11. This does not need to mean that the model is untrustworthy. As explained by K. Podgórski (personal communication, 22 May 2022) the cross shape indicates there could be a combination of two models. The extra vertical line to the left of the cross in the active set residuals plot could be signs of yet another model. Thus, an analysis of the different clusters of residuals could aid the understanding of this mixture-based model and help assess whether there was some issue with the gradient boosting models related to the d-task. However, this goes beyond the scope of this work and should be pursued in future research.

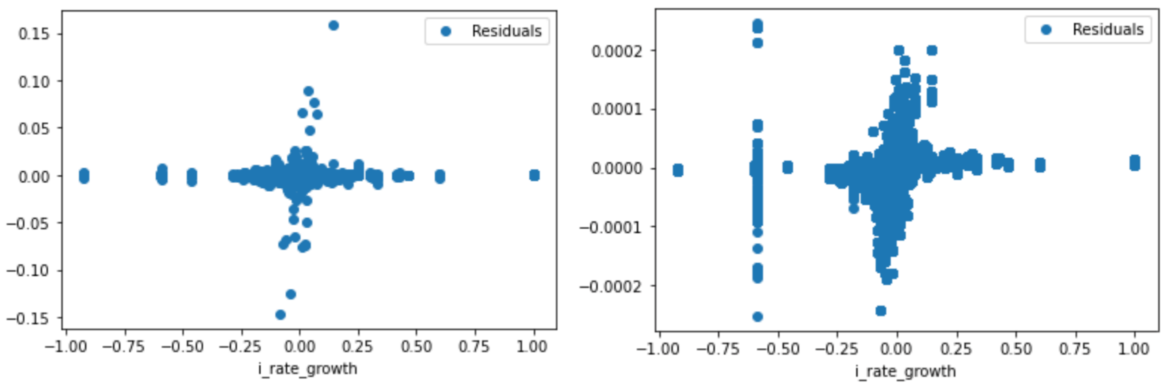

Figure 11: Residuals vs treatment variable. The left plot show the residuals from the passive set for function $\hat{m}_0(.)$ against the treatment variable. The right plot shows the same information for the active set.

Despite the shape, most residuals seem to be centered around zero, which is a good sign given that both the ***Y*** and ***D*** variable had most values around zero. It also shows that for the active funds, there is a specific point in the ***D*** variable where the residuals seem to vary greatly (the vertical line to the left). The residuals also show that even the outliers in the ***D*** variable seem to deviate very little from the predictions. Most of the error is concentrated where the ***D*** variable is precisely or close to zero. This could be due to the interest rate not changing that often. There is no change to the interest rate for several months every year, which probably is a more challenging situation to predict.

Finally, the same plots for function $\hat{g}_0(.)$ show a similar interpretation but with new insights (Figure 12). For example, they also confirm there is less precision in the active funds set, but for both plots most of the residuals converge around zero. The parallel vertical lines present in both plots are likely due to the limited variation in the treatment variable. These plots signalize arguably healthy models, but they also show that controlling for outliers could be a good way to improve some results.

The plot for the passive set in figure 12 also gives some clue as to why it did not produce significant results. For example, the variation in the predictions for the rarer values of the treatment variable is considerably large, even though the actual range of predictions is smaller. It seems like the passive set is generally better at predicting, but this improved performance is limited to the most common values. Under this hypothesis the active set is marginally better at predicting returns for rare events, and this could be the reason why the passive set model did not produce significant results.

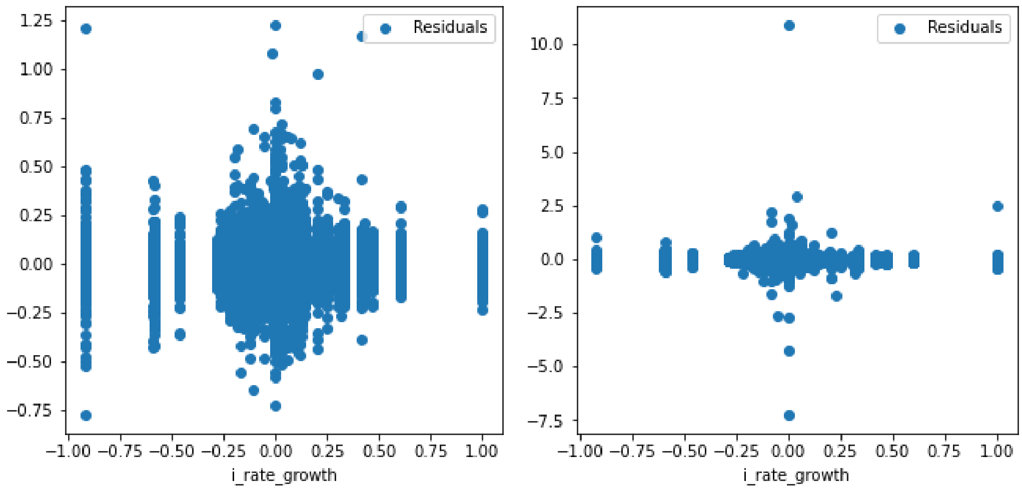

Figure 12: Residuals vs treatment variable. The left plot represents the residuals for function $\hat{g}_0(.)$ against the treatment variable for the passive set, while the right plot shows the same information for the active set. There seems to be more randomness in the active set but the general pattern is the same for both sets, which is expected since they share many similarities. Parallel straight lines are likely due to the low variation in the treatment variable across all months in a year.

5.3 Double Machine Learning Feasibility

Related to the research question about the feasibility of using DML in investigating causal effects on fund returns, the results of this paper are promising, but many shortcomings need to be taken into consideration. Three of the four models created for this paper show significant results, which as a baseline can be considered as good evidence of the feasibility of this framework in further research. Many of the shortcomings encountered are specific to fund returns prediction. While they can be circumvented, they most likely will add a great degree of complexity to the models and demand a cautious, almost skeptical, approach to the results. However, this type of cautiousness is natural for causal inference. For example, the literature indicates that the lines that separate actively from passively managed funds are becoming blurrier. This ambiguousness certainly creates problems with the assumed differences between actively and passively managed funds. However, one way to control it is by adding yet another predictor to the model, i.e., the active share of each fund. Additionally, the computational requirements of using DML can be pretty high depending on the chosen learners.

As discussed, it is difficult to pinpoint exactly why we see the treatment effects presented in tables 5 and 6. There are many different aspects to the DML process that could have affected the results. Nevertheless, we can confidently say that the chosen learners have a significant impact on the result, both in terms of the estimated treatment effect and the significance of the findings. Thus, it would be wise to compare the performance of gradient boosting as a learner against other learners that can model highly complex systems. Measuring the actual effect of the interest rate's growth rate on returns might be impossible. However, if other academically well-accepted machine learning frameworks produce similar estimates, one could assess how well the DML framework performed for this specific question. Moreover, considering the theoretical background offered by the DML procedure and the results obtained here, there is no doubt that gradient boosting offers a better way to estimate average treatment effects in fund returns than linear regressions.

The size of the treatment effects resulting from this analysis also required discussion. For the models utilizing linear regression as learners, the treatment effect is relatively small compared to the ones based on gradient boosting. From an abductive standpoint, the treatment effect of the linear regression models seems more likely as they say that the if the interest rate increases by 1%, then fund returns decrease by -0.025% or -0.019% depending on the category of fund, given all else remain the same, while the decrease based on the significant result of the gradient boosting model would be -11.97% and -0.229% for the insignificant result. Two hypotheses are offered as to why a relatively large effect of the significant gradient boosting model is observed. Firstly, it could be that some of the DML assumptions or that some part of the analysis was overlooked or mishandled by the authors. The other suggestion is that the average treatment effect is accurate and other variables moderate the effect of the interest rate such that the actual effect on fund returns is not as large.

The insignificant results of the gradient boosting model for the passive set also have a possible explanation. It could also be true that there is no causal relationship between the grouping of passive funds and the interest rate's growth rate. However, as discussed above, the granularity of the group allocation could be a potential issue for finding a causal effect if subgroups within the passive fund group are affected differently by the treatment variable relative to the other subgroups.

5.4 Limitations

Some factors limited the results of this work. As with any machine learning method, the conclusions drawn from this work are limited to the quality of data utilized. The amount of data available for the two groups of interest, the active and passive funds, differed. This might have affected both the learners and the DML procedure.

Another limitation related to the data is the potential of subgroups within the active or passive funds classification. Such subgroups could have impacted the results as the treatment effect of interest rate growth rate could have been different for each group. If this is true, then the significance of the finding or the accurate results could have been affected negatively. This study focused on two main clusters related to the management style of funds covering fixed income and equities. Many other subgroups could be studied; for example, there are good reasons to believe that the treatment effects would differ across funds highly specialized in specific industries since the literature suggests that this is the scenario where actively managed funds consistently perform better than passively managed funds. The list of subgroups is virtually infinite, and given their diversity, there is no reason to believe that treatment effects would be homogenous across all these groups.

Additionally, it should be noted that not all potential funds were included in the study due to the initial filtering of funds, which might have caused essential data to be left out. The final limitation of the data is that some of the macroeconomic variables utilized were collected in quarterly data and then transformed into monthly to fit with the other data. The transformation of these variables might have introduced noise into the models and made them less accurate.

There have also been some limitations in this paper concerning the modeling. Perhaps, the most significant limiting factor of the modeling in this study is the high computational cost of the models. The study results could perhaps have been improved if more parameters could have been added to the tuning procedure in the gradient boosting. The model could have been improved, also by including additional predictors. This paper focused on using macroeconomic variables for the predictions. However, the literature is just as rich – if not more – when using other financial scores or more of the financial information contained in the data itself (like with different ratios). Combining both methods could result in a much better model for predicting returns.

Finally, the DML procedure allows for combining any machine learning models as learners in each of the y-tasks. This study opted for the most straightforward approach by using the same learner for each method. However, a regularized version of the linear regression could have been just as good for predicting the d-task – if not better, given there are signs of overfitting for that task when using gradient boosting. Doing so would have significantly reduced the computational cost and the time complexity that comes with it. The computational cost issue could have been solved if there was no time pressure for the conclusion of this research or if the authors had access to better hardware. The feasibility of double machine learning could also have been further explored if more modeling could have been conducted. For example, the causal effects of additional macroeconomic variables could have been investigated so that that question was not so dependent on the causal effect of just one variable. Another advantage of doing so is that one could potentially have a better analysis of what policies should receive more attention from investors and asset managers.

5.5 Future research

There is an immense breadth of potential future research for DML in measuring causal effects in funds returns and other finance-related questions. Future research could potentially utilize the learnings gained from this work and its limitations to improve the results. Such improvements could be the inclusion of all available funds, the subdivision of the included funds according to, e.g., the fund's degree of activeness or their investment objectives, the inclusion of other variables than macroeconomic ones, the testing of and combinations of other learners, or the investigation of other variables causal relationships to fund returns. Some of these paths will likely require increased computing power since some learners have an enormous appetite for memory and processing power.

Potential future research that expands beyond this paper's scope is numerous. Investigating the treatment effects of variables on fund returns in other markets than the US would investigate if the relationships discovered in this study are specific to the US or applicable in other markets. The effect the learners' predictive performance has on the results is another field of future research. Such research could illuminate the importance of and to which degree the DML assumption about learners' abilities to estimate their true functions is required when doing DML with fund data. Investigating the assumption of DML’s two error terms, see equations two and three, would also be illuminating for future research on fund data. Checking that the assumptions hold would further strengthen the use case and trustworthiness of results of DML in the field of fund causal inference. Lastly, more causal research should focus on its applicability in business and management as it could popularize the method and make decision-making a less risky event.

6 Conclusion

Given the specific challenges financial data entail, predicting returns or anything related to finance is arduous. However, as challenging as it may be, the methods and frameworks have achieved a mature status, and they generally disregard factors that are hard to control or check. Nevertheless, these factors are vital to traditional causal inference. The cumbersome work that comes with them is possibly why causality does not play a more prominent role in finance.

One of the goals of this study was to first introduce the reader to a method that could change this scenario. DML has few assumptions one needs to check, and it has working frameworks available in two of the most prominent programming languages related to this type of research. However, some of its procedures do not necessarily synergize well with financial data. For example, sample splitting can be somewhat hard when there is a time component to the data. This study offers a simple solution to the sample splitting for data with time components, although computationally costly. The other goal is related directly to the first research question in this work:

Question 1: *"How feasible is using the double machine learning framework for measuring causal effects in fund returns?"*

There are two perspectives through which this question can be answered. First, feasibility can be interpreted as how easy or whether it is even possible to implement the DML framework to this type of question. From this perspective, the answer is obvious, the method is easy to implement, and very few workarounds were necessary to adapt it to finance-related questions. The other perspective pertains to the interpretation of feasibility as trustworthiness. Unfortunately, the answer to this question is unclear because of the machine learning frameworks used as learners in the DML procedure. Only one fulfilled all of the requirements, and out of these, only the results for the active funds set were significant.

There are no reasons to believe that the significant result is not trustworthy, given that the requirements were fulfilled and the framework behaved as expected. However, it is hard to check how precise the results were since it is impossible to run randomized experiments for the second question in this research. Thus, as proposed in the future research section, more machine learning frameworks should be tested as learners and their results compared to understand how good (or bad) the results are when using gradient boosting. Nevertheless, the results presented here show that the DML framework might be relevant to many other fund-related questions.

The other goal of this work was to measure the average treatment effect of interest rates' growth rate in fund returns, as evidenced by question two:

Question 2: *"What effect, on average, does the US Central Bank's interest growth rate have on fund returns?"*

As mentioned in the Results section, only one of the learners fulfilled all requirements for the framework, and out of the results produced by this learner, only one was significant. Based on the information given by this significant result, the average treatment effect of the US Central Bank's interest growth rate on fund returns is -11.97 for every one-unit change in the interest rate's growth rate. However, it is worth mentioning that one-unit changes are infrequent; thus, it might be wiser to interpret this from an example closer to reality. As both the independent and treatment variables are measured as growth rates, a one-unit increase can be interpreted as a 100% increase in the interest rate. A 1% increase in the interest rate would represent a -11.97% decrease in the returns. While there are no reasons to distrust this result from a technical perspective, the authors advise a skeptical attitude towards it since the only way to "confirm" these results would be further tests with other learners, which was not done.

The powerful effect suggested by the result is the main reason for skepticism, but one must also consider that this is most likely not the only event affecting fund returns. Therefore, it is not unlikely that these results are somewhat precise but counterbalanced by other effects. As mentioned in the Future Research section, more research would be necessary to understand this question more profoundly.

The answer to the two research questions suggests many practical implications. The first practical implication is narrowing the academic and business relationship regarding fund returns. Financial research on a business level could use a more academic approach to evolve beyond just using predictions and forecasts to understand what drives the market changes. Investors can use causal machine learning to understand better market conditions and how they are affected by macroeconomic changes. Asset managers can use this knowledge to strategize better and reap opportunities they would otherwise miss if they only focused on predictions. Government officials can also use this type of approach to understand the impact of their policy-making, especially considering that financial markets exercise changes almost instantaneously. The possibilities are virtually infinite, and the new methods that combine traditional causal inference with machine learning techniques offer solutions to many problems in finance-related questions. They make a traditionally complex field much easier to deal with, and these results can undoubtedly be used for portfolio (or financial) planning.

References

Ahrens, A., Aitken, C., & Schaffer, M.E. (2021). Using Machine Learning Methods to Support Causal Inference in Econometrics, *Behavioral Predictive Modeling in Economics. Studies in Computational Intelligence*, vol. 897, pp.23-52, Available online: https://pure.hw.ac.uk/ws/portalfiles/portal/42102613/Ahrens_et_al_PDS_lasso.pdf [Accessed 17 May 2022]

Altibas, H., & Biskin, O.T. (2015). Selecting Macroeconomic Influencers on Stock Markets by Using Feature Selection Algorithms, *Procedia Economics and Finance*, vol. 30, pp.22-29, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S2212567115012514&site=eds-live&scope=site [Accessed 17 May 2022]

Ancona, N., Marinazzo, D., & Stramaglia, S. (2004). Radial Basis Function Approach to Nonlinear Granger Causality of Time Series, *Physical Review E*, vol. 70, no. 5, pp.56221-1 - 56221-7, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=inh&AN=8250961&site=eds-live&scope=site [Accessed 17 May 2022]

Asgharian, H., Christiansen, C., & Hou, A.J. (2015). Effects of Macroeconomic Uncertainty on the Stock and Bond Markets, *Finance Research Letters*, vol. 13, pp.10-16, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S1544612315000367&site=eds-live&scope=site [Accessed 17 May 2022]

Asgharian, H., Hou, A.J., & Javed, F. (2013). The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS approach, *Journal of Forecasting*, vol. 32, no. 7, pp.600-612, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsswe&AN=edsswe.oai.DiVA.org.su.116075&site=eds-live&scope=site [Accessed 17 May 2022]

Athey, S. (2015). Machine Learning and Causal Inference for Policy Evaluation, Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.5-6, Available online: https://dl.acm.org/doi/10.1145/2783258.2785466 [Accessed 17 May 2022]

Athey, S. (2017). Beyond Prediction: Using big data for policy problems, *Science*, vol. 355, no. 6324, pp.483-485, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edswsc&AN=000393183100034&site=eds-live&scope=site [Accessed 17 May 2022]

Athey, S., & Imbens, G. (2016). Recursive Partitioning for Heterogeneous Causal Effects, Proceedings of the National Academy of Sciences of the United States of America, vol. 113, no. 27, pp.7353-7360, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26470691&site=eds-live&scope=site [Accessed 17 May 2022]

Bach, P., Chernozhukov, V., Kurz, M.S., & Spindler, M. (2021a). 3. Models: 3.1. Partially linear regression model (PLR), DoubleML documentation, Available online: https://docs.doubleml.org/stable/guide/models.html#partially-linear-regression-model-plr [Accessed 17 May 2022]

Bach, P., Chernozhukov, V., Kurz, M.S., & Spindler, M. (2021b). DoubleML -- An Object-Oriented Implementation of Double Machine Learning in Python, *arXiv*, *Working Paper*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.2104.03220&site=eds-live&scope=site [Accessed 17 May 2022]

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market, *National Bureau of Economic Research Working Papers*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat02271a&AN=atoz.ebs14211003e&site=eds-live&scope=site [Accessed 17 May 2022]

Barras, L., Scaillet, O., & Wermers, R. (2010). False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *The Journal of Finance*, vol. 65, no. 1, pp.179-216, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.25656289&site=eds-live&scope=site [Accessed 17 May 2022]

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S.R. (2019). Predicting the Direction of Stock Market Prices Using Tree-based Classifiers, *North American Journal of Economics and Finance*, vol. 47, pp.552-567, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S106294081730400X&site=eds-live&scope=site [Accessed 17 May 2022]

Becker, C., Ferson, W., Myers, D., & Schill, M. (1998). Conditional Market Timing with Benchmark Investors, *National Bureau of Economic Research Working Papers*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat02271a&AN=atoz.ebs14220736e&site=eds-live&scope=site [Accessed 17 May 2022]

Bertrand, M., Crépon, B., Marguerie, A., & Premand, P. (2017). Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire, *World Bank*, Available online: https://openknowledge.worldbank.org/handle/10986/28460 [Accessed 17 May 2022]

Birz, G., & Lott, Jr.J.R. (2011). The Effect of Macroeconomic News on Stock Returns: New evidence from newspaper coverage, *Journal of Banking and Finance*, vol. 35, no. 11, pp.2791-2800, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0378426611001117&site=eds-live&scope=site [Accessed 17 May 2022]

Brandt, M.W., & Kang, Q. (2004). On the Relationship Between the Conditional Mean and Volatility of Stock Returns: A latent VAR approach, *Journal of Financial Economics*, vol. 72, no. 2, pp.217-257, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0304405X03002435&site=eds-live&scope=site [Accessed 17 May 2022]

Brooks, C., Fenton, E., Schopohl, L, & Walker, J. (2019). Why Does Research in Finance Have so Little Impact?, *Critical Perspectives on Accounting*, vol. 58, pp.24-52, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S1045235418301102&site=eds-live&scope=site [Accessed 17 May 2022]

Card, D. (1999). Chapter 30 The Causal Effect of Education on Earnings, *Handbook of Labor Economics*, vol. 3, no. Part A, pp.1801-1863, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S1573446399030114&site=eds-live&scope=site [Accessed 17 May 2022]

Chang, E.C., & Lewellen, W.G. (1984). Market Timing and Mutual Fund Investment Performance, *The Journal of Business*, vol. 57, no. 1, pp.57-72, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2352888&site=eds-live&scope=site [Accessed 17 May 2022]

Chang. C.-L., McAleer, M., Ilomäki, J., & Laurila, H. (2020). Causality Between CO2 Emissions and Stock Markets, *Energies*, vol. 13, no. 11, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselc&AN=edselc.2-52.0-85086380671&site=eds-live&scope=site [Accessed 17 May 2022]

Chatzis, S.P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting Stock Market Crisis Events Using Deep and Statistical Machine Learning Techniques, *Expert Systems with Applications*, vol. 112, pp.353-371, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=inh&AN=18016727&site=eds-live&scope=site [Accessed 17 May 2022]

Chen, S.-S. (2009). Predicting the Bear Stock Market: Macroeconomic variables as leading indicators, *Journal of Banking & Finance*, vol. 33, no. 2, pp.211-223, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0378426608001544&site=eds-live&scope=site [Accessed 17 May 2022]

Chen, Y., Rangarajan, G., Feng, J., & Ding, M. (2004). Analyzing Multiple Nonlinear Time Series with Extended Granger Causality, *Physics Letters A*, vol. 324, pp.26-35, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.nlin%2f0405016&site=eds-live&scope=site [Accessed 17 May 2022]

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters, *Econometrics Journal*, vol. 21, no. 1, pp.C1-C68, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=127932375&site=eds-live&scope=site [Accessed 17 May 2022]

Cremers, M., Ferreira, M.A., Matos, P., & Starks, L. (2016). Indexing and Active Fund Management: International evidence, *Journal of Financial Economics*, vol. 120, no. 3, pp.539-560, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0304405X16300083&site=eds-live&scope=site [Accessed 17 May 2022]

Daniel, F. (2019). Financial Time Series Data Processing for Machine Learning, *arXiv*, *Working Paper*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.1907.03010&site=eds-live&scope=site [Accessed 17 May 2022]

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (2012). Measuring Mutual Fund Performance with Characteristic-Based Benchmarks, *The Journal of Finance*, vol. 52, no. 3, pp.1035-1058, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2329515&site=eds-live&scope=site [Accessed 17 May 2022]

Das, S. (2009). Perspectives on Financial Services, New Delhi: Allied Publishers Pvt. Ltd.

Davis, J.M.V., & Heller, S.B. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An application to summer jobs, *The American Economic Review*, vol. 107, no. 5, pp.546-550, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.44250458&site=eds-live&scope=site [Accessed 17 May 2022]

De Neve, J.-E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2021). How to Improve Tax Compliance? Evidence from population-wide experiments in Belgium, *Journal of Political Economy*, vol. 129, no. 5, pp.1425-1463, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=150007194&site=eds-live&scope=site [Accessed 17 May 2022]

Diebold, F.X., & Yilmaz, K. (2012). Better to Give Than to Receive: Predictive directional measurement of volatility spillovers, *International Journal of Forecasting*, vol. 28, no. 1, pp.57-66, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S016920701100032X&site=eds-live&scope=site [Accessed 17 May 2022]

Ding, M., Chen, Y., & Bressler, S.L. (2006). Granger Causality: Basic theory and application to neuroscience, in B. Shelter, M. Winterhalder, & J. Timmer (eds), *Handbook of Time Series Analysis*, Weinheim: Wiley-VCH Verlag, pp.451-474, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.q-bio%2f0608035&site=eds-live&scope=site [Accessed 17 May 2022]

Dixon, M.F., Halperin, I., & Bilokon, P. (2020). Machine Learning in Finance: From theory to practice, Cham: Springer, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselc&AN=edselc.2-52.0-85089331413&site=eds-live&scope=site [Accessed 17 May 2022]

Domm, P. (2022). The Fed is Expected to Raise Rates by a Half Point. Investors wonder if it will get more aggressive, *CNBC*, Available online: https://www.cnbc.com/2022/05/03/the-fed-is-expected-to-raise-rates-by-a-half-point-investors-wonder-if-it-will-get-more-aggressive.html [Accessed 17 May 2022]

Dube, A., Jacobs, J., Naidu, S., & Suri, S. (2020). Monopsony in Online Labor Markets, *American Economic Review: Insights*, vol. 2, no. 1, pp.33-46, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=ecn&AN=1815636&site=eds-live&scope=site [Accessed 17 May 2022]

Elton, E.J., Gruber, M.J., & Blake, C.R. (1996). Survivorship Bias and Mutual Fund Performance, *The Review of Financial Studies*, vol. 9, no. 4, pp.1097-1120, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2962224&site=eds-live&scope=site [Accessed 17 May 2022]

Elton, E.J., Gruber, M.J., & Blake, C.R. (1996). The Persistence of Risk-Adjusted Mutual Fund Performance, *The Journal of Business*, vol. 69, no. 2, pp.133-157, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2353461&site=eds-live&scope=site [Accessed 17 May 2022]

Enders, W. (2004). Applied Econometrics Time Series, 2nd edn., Hoboken: John Wiley & Sons

Fama, E.F. (1970). Efficient Capital Markets: A review of theory and empirical work, *Journal of Finance*, vol. 25, no. 2, pp.383-417, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=4660197&site=eds-live&scope=site [Accessed 17 May 2022]

Fama, E.F., & French, K.R. (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns, *The Journal of Finance*, vol. 65, no. 5, pp.1915-1947, Available through: LUSEM Library

website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.40864991&site=eds-live&scope=site [Accessed 17 May 2022]

Federal Reserve Bank of Philadelphia. (2022). Surveys & Data, Available online: https://www.philadelphiafed.org/surveys-and-data [Accessed 17 May 2022]

Foroni, C., & Marcellino, M. (2013). A Survey of Econometric Methods for Mixed Frequency Data, *Norges Bank Research*, *Working Paper*, Available online: https://www.norges-bank.no/contentassets/5f6ee92052ad49918658826720a8bed9/norges_bank_working_paper_2013_06.pdf [Accessed 17 May 2022]

FRED. (2022). Federal Reserve Bank of St. Louis, Available online: https://fred.stlouisfed.org/ [Accessed 17 May 2022]

Freiwald, W.A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J.C., Rodriguez, L.M., Rodriguez, V., Kreiter, A.K., & Singer, W. (1999). Testing Non-linearity and Directedness of Interactions Between Neural Groups in the Macaque Inferotemporal Cortex, *Journal of Neuroscience Methods*, vol. 94, no. 1, pp.105-119, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0165027099001296&site=eds-live&scope=site [Accessed 17 May 2022]

Glosten, L.R., Jagannathan, R., & Runkle, D.E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *The Journal of Finance*, vol. 48, no. 5, pp.1779-1801, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2329067&site=eds-live&scope=site [Accessed 17 May 2022]

Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, vol. 37, no. 3, pp.424-438, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.1912791&site=eds-live&scope=site [Accessed 17 May 2022]

Grinblatt, M., & Titman, S. (1989). Mutual Fund Performance: An analysis of quarterly portfolio holdings, *The Journal of Business*, vol. 62, no. 3, pp.393-416, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2353353&site=eds-live&scope=site [Accessed 17 May 2022]

Grodecka-Messi, A., & Hull, I. (2019). The Impact of Local Taxes and Public Services on Property Values, *Sveriges Riksbank Working Paper Series*, vol. 374, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsswe&AN=edsswe.oai.lup.lub.lu.se.f4aaabc7.953d.4bfe.bcbf.7f2b52d2aa50&site=eds-live&scope=site [Accessed 17 May 2022]

Grossman, S.J., & Stiglitz, J.E. (1980). On the Impossibility of Informationally Efficient Markets, *The American Economic Review*, vol. 70, no. 3, pp.393-408, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.1805228&site=eds-live&scope=site [Accessed 17 May 2022]

Gruber, M.J. (1996). Another Puzzle: The growth in actively managed mutual funds, *The Journal of Finance*, vol. 51, no. 3, pp.783-810, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2329222&site=eds-live&scope=site [Accessed 17 May 2022]

Gündüz, H., Çataltepe, Z., & Yaslan, Y. (2017). Stock Daily Return Prediction Using Expanded Features and Feature Selection, *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, no. 6, pp.4829-4840, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=inh&AN=17787939&site=eds-live&scope=site [Accessed 17 May 2022]

Hamilton, J.D., & Lin, G. (1996). Stock Market Volatility and the Business Cycle, *Journal of Applied Econometrics*, vol. 11, no. 5, pp.573-593, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2285217&site=eds-live&scope=site [Accessed 17 May 2022]

Hansson, J., Jansson, P., & Löf, M. (2005). Business Survey Data: Do they help in forecasting GDP growth?, *International Journal of Forecasting*, vol. 21, no. 2, pp.377-389, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0169207004001086&site=eds-live&scope=site [Accessed 17 May 2022]

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data mining, inference, and prediction, 2nd edn., New York: Springer

Heiss, A. (2020). Causal Inference, in F. Urdinez & A. Cruz (eds), *R for Political Data Science*, New York: CRC Press, pp.235-273

Hendricks, D., Patel, J., & Zeckhauser, R. (1993). Hot Hands in Mutual Funds: Short-run persistence of relative performance, 1974-1988, *The Journal of Finance*, vol. 48, no. 1, pp.93-130, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2328883&site=eds-live&scope=site [Accessed 17 May 2022]

Henriksson, R.D. (1984). Market Timing and Mutual Fund Performance: An empirical investigation, *The Journal of Business*, vol. 57, no. 1, pp.73-96, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2352889&site=eds-live&scope=site [Accessed 17 May 2022]

Hoepner, A.G.F., McMillan, D., Vivian, A., & Wese Simen, C. (2021). Significance, Relevance and Explainability in the Machine Learning age: An econometrics and financial data science perspective, *European Journal of Finance*, vol. 27, no. 1/2, pp.1-7, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=147988427&site=eds-live&scope=site [Accessed 17 May 2022]

Huang, R., Pilbeam, K., & Pouliot, W. (2021). Do Actively Managed US Mutual Funds Produce Positive Alpha?, *Journal of Economic Behaviour & Organization*, vol. 182, pp.472-492, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=psyh&AN=2019-24102-001&site=eds-live&scope=site [Accessed 17 May 2022]

Huang, W., & Mollick, A.V. (2020). Tight Oil, Real WTI Prices and U.S. Stock Returns, *Energy Economics*, vol. 85, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S014098831930369X&site=eds-live&scope=site [Accessed 17 May 2022]

Hünermund, P., Kaminski, J.C., & Schmitt, C. (2021). Causal Machine Learning and Business Decision Making, Academy of Management Annual Meeting Proceedings, Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3867326 [Accessed 17 May 2022]

Ippolito, R.A. (1989). Efficiency With Costly Information: A study of mutual fund performance, 1965-1984, *The Quarterly Journal of Economics*, vol. 104, no. 1, pp.1-23, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2937832&site=eds-live&scope=site [Accessed 17 May 2022]

Ippolito, R.A. (1993). On Studies of Mutual Fund Performance, 1962-1991, *Financial Analysts Journal*, vol. 49, no. 1, pp.42-50, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.4479612&site=eds-live&scope=site [Accessed 17 May 2022]

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R, New York: Springer

Jensen, M.C. (1968). The Performance of Mutual Funds in the Period 1945-1964, *The Journal of Finance*, vol. 23, no. 2, pp.389-416, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2325404&site=eds-live&scope=site [Accessed 17 May 2022]

Johannemann, J., Athey, S., Hadad, V., & Wager, S. (2019). Sufficient Representations for Categorical Variables, working paper, Stanford Graduate School of Business, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=140105122&site=eds-live&scope=site [Accessed 17 May 2022]

Kacperczyk, M., Sialm, C., & Zheng, L. (2005). On the Industry Concentration of Actively Managed Equity Mutual Funds, *The Journal of Finance*, vol. 60, no. 4, pp.1983-2011, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.3694859&site=eds-live&scope=site [Accessed 17 May 2022]

Kennaway, R. (2020). When Causation does not Imply Correlation: Robust violations of the faithfulness axiom, in W. Mansell (ed.), *The Interdisciplinary Handbook of Perceptual Control Theory: Living control systems IV*, San Diego: Academic Press, pp.49-72

Knaus, M.C. (2021). A Double Machine Learning Approach to Estimate the Effects of Musical Practice on Student's Skills, *Journal of the Royal Statistical Society: Series A*, vol. 184, no. 1, pp.282-300, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=148280271&site=eds-live&scope=site [Accessed 17 May 2022]

Kosowski, R., TImmermann, A., Wermers, R., & White, H. (2006). Can Mutual Fund "Stars" Really Pick Stocks? New evidence from a bootstrap analysis, *Journal of Finance*, vol. 61, no. 6, pp.2551-2595, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=23646353&site=eds-live&scope=site [Accessed 17 May 2022]

Krauss, C., Do, X.A., Huck, N. (2017). Deep Neural Networks, Gradient-boosted Trees, Random Forests: Statistical arbitrage on the S&P 500, *European Journal of Operational Research*, vol. 259, no. 2, pp.689-702, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0377221716308657&site=eds-live&scope=site [Accessed 17 May 2022]

Kwon, D. (2022). What Drives Emerging Stock Market Returns? A Factor-Augmented VAR Approach., *Emerging Markets Finance & Trade*, vol. 58, no. 5, pp.1215-1232, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=bth&AN=156055948&site=eds-live&scope=site [Accessed 17 May 2022]

Lacerda, G., Spirtes, P.L., Ramsey, J., & Hoyer, P.O. (2012). Discovering Cyclic Causal Models by Independent Components Analysis, Working paper, *arXiv*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.1206.3273&site=eds-live&scope=site [Accessed 17 May 2022]

Leung, M.T., Daouk, H., & Chen, A.-S. (2000). Forecasting Stock Indices: A comparison of classification and level estimation models, *International Journal of Forecasting*, vol. 16, no. 2, pp.173-190, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0169207099000485&site=eds-live&scope=site [Accessed 17 May 2022]

Lopata, A., Butleris, R., Gudas, S., Rudžionis, V., Rudžionienė, K., Žioba, L., Veitaitė, I., Dilijonas, D., Grišius, E., & Zwitserloot, M. (2021). Financial Data Preprocessing Issues, In: Lopata, A., Gudonienė, D., Butkienė, R. (eds) *Information and Software Technologies. ICIST 2021. Communications in Computer and Information Science*, vol 1486. Cham: Springer, Available online: https://link.springer.com/chapter/10.1007/978-3-030-88304-1_5 [Accessed 17 May 2022]

Malkiel, B.G. (1995). Returns from Investing in Equity Mutual Funds 1971 to 1991, *The Journal of Finance*, vol. 50, no. 2, pp.549-572, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2329419&site=eds-live&scope=site [Accessed 17 May 2022]

Malmendier, U., & Nagel, S. (2011). Depression Babies: Do macroeconomic experiences affect risk-taking?, *National Bureau of Economic Research Working Papers*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat02271a&AN=atoz.ebs14190399e&site=eds-live&scope=site [Accessed 17 May 2022]

McMillan, D.G. (2016). Which Variables Predict and Forecast Stock Market Returns?, Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2801670 [Accessed 17 May 2022]

Mighri, Z., Ragoubi, H., Sarwar, S., & Wang, Y. (2022). Quantile Granger Causality Between US Stock Market Indices and Precious Metal Prices, *Resources Policy*, vol. 76, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0301420722000460&site=eds-live&scope=site [Accessed 17 May 2022]

Mittnik, S., Robinzonov, N., & Spindler, M. (2015). Stock Market Volatility: Identifying major drivers and the nature of their impact, *Journal of Banking and Finance*, vol. 58, pp.1-14, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0378426615000795&site=eds-live&scope=site [Accessed 17 May 2022]

Neely, C.J., Rapach, D.E., Tu, J., & Zhou, G. (2014). Forecasting the Equity Risk Premium: The role of technical indicators, *Management Science*, vol. 60, no. 7, pp.1772-1791, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.42919633&site=eds-live&scope=site [Accessed 17 May 2022]

Neyman, J., & Scott, E.L. (1948). Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, vol. 16, no. 1, pp.1-32, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.1914288&site=eds-live&scope=site [Accessed 17 May 2022]

O'Connell, B., & Curry, B. (2022). What Happens When The Fed Raises Interest Rates?, *Forbes*, Available online: https://www.forbes.com/advisor/investing/fed-raises-interest-rates/ [Accessed 17 May 2022]

Paye, B.S. (2012). 'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables, *Journal of Financial Economics*, vol. 106, no. 3, pp.527-546, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0304405X12001316&site=eds-live&scope=site [Accessed 17 May 2022]

Pearl, J., Glymour, M., & Jewell, N.P. (2016). Causal Inference in Statistics: A primer, Wiley

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python, *Journal of Machine Learning*, vol. 12, pp.2825-2830, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.1201.0490&site=eds-live&scope=site [Accessed 17 May 2022]

Pierdzioch, C., Döpke, J., & Hartmann, D. (2008). Forecasting Stock Market Volatility With Macroeconomic Variables in Real Time, *Journal of Economics and Business*, vol. 60, no. 3, pp.256-276, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselp&AN=S0148619507000240&site=eds-live&scope=site [Accessed 17 May 2022]

Reese, S. (2021). Lecture 8: ML for causal inference I, DABN14, powerpoint presentation, LUSEM Lund, 13 December 2021

Rosenbaum, P.R. (1995). Observational Studies, New York: Springer

Rosenbaum, P.R., & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, vol. 70, no. 1, pp.41-55, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2335942&site=eds-live&scope=site [Accessed 17 May 2022]

Schapire, R.E. (1990). The Strength of Weak Learnability, *Machine Learning*, vol. 5, no. 2, pp.197-227, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselc&AN=edselc.2-52.0-0025448521&site=eds-live&scope=site [Accessed 17 May 2022]

Schwert, G.W. (1989a). Why Does Stock Market Volatility Change Over Time?, *The Journal of Finance*, vol. 44, no. 5, pp.1115-1153, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2328636&site=eds-live&scope=site [Accessed 17 May 2022]

Schwert, G.W. (1989b). Business Cycles, Financial Crises, and Stock Volatility, *Carnegie-Rochester Confer. Series on Public Policy*, vol. 89, no. C, pp.83-125, Available through: LUSEM Library

website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselc&AN=edselc.2-52.0-0001817296&site=eds-live&scope=site [Accessed 17 May 2022]

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python, Proceedings of the 9th Python in Science Conference, pp.92-96, Available online: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf [Accessed 17 May 2022]

Sharpe, W.F. (1964). Capital Asset Prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance*, vol. 19, no. 3, pp.425-442, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2977928&site=eds-live&scope=site [Accessed 17 May 2022]

Sharpe, W.F. (1966). Mutual Fund Performance, *The Journal of Business*, vol. 39, no. 1, pp.119-138, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2351741&site=eds-live&scope=site [Accessed 17 May 2022]

Spirtes, P.L. (2020). The Limits of Causal Inference from Observational Data, Available online: https://www.researchgate.net/publication/2634723_The_Limits_of_Causal_Inference_from_Observational_Data [Accessed 17 May 2022]

Stokes, P.A., & Purdon, P.L. (2017). A Study of Problems Encountered in Granger Causality Analysis from a Neuroscience Perspective, Proceedings of the National Academy of Sciences of the United States of America, vol. 114, no. 34, pp.E7063-E7072, Available through: LUSEM Library website
http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.26487324&site=eds-live&scope=site [Accessed 17 May 2022]

Söderberg & Partners. (2017). Valet Mellan Aktiv och Passiv Fondförvaltning, White Paper från Söderberg & Partners, Available online:
https://www.soderbergpartners.se/globalassets/newsroom/aktiv-och-passiv-fondforvaltning/soderberg-partners-white-paper-valet-mellan-aktiv-och-passiv-fondforvaltning.pdf [Accessed 17 May 2022]

The Economist. (2022). The Fed Causes Gyrations in Financial Markets: Despite a sudden rally, pain lies ahead, *Leaders*, *The Economist*, Available online:
https://www.economist.com/leaders/the-federal-reserve-is-causing-pain-in-financial-markets/21809132 [Accessed 17 May 2022]

Tillier, M. (2021). What 'Tighter Monetary Policy' Means and What Investors Should Do About It, *Nasdaq*, Available online:
https://www.nasdaq.com/articles/what-tighter-monetary-policy-means-and-what-investors-should-do-about-it [Accessed 17 May 2022]

U.S. Securities and Exchange Commission. (2020). Mutual Funds and ETFs: A guide for investors, *Office of Investor Education and Advocacy*, Available online:

https://www.investor.gov/sites/investorgov/files/2020-04/mutual-funds-ETFs_2_0.pdf [Accessed 17 May 2022]

U.S. Securities and Exchange Commission. (n.d.). Mutual Funds, *Introduction to Investing*, Available online: https://www.investor.gov/introduction-investing/investing-basics/investment-products/mutual-funds-and-exchange-traded-1 [Accessed 17 May 2022]

Urdinez, F. (2020). Panel Data, in F. Urdinez & A. Cruz (eds), *R for Political Data Science*, New York: CRC Press, pp.147-171

Varaku, K. (2021). Essays on Causal Inference and Treatment Effects in Productivity and Finance: Double robust machine learning with deep neural networks and random forests, PhD thesis, Rice University, Available online: https://scholarship.rice.edu/handle/1911/110418 [Accessed 17 May 2022]

Verbeek, M., & Huij, J. (2006). Cross-Sectional Learning and Short-Run Persistence in Mutual Fund Performance, *ERIM Research Paper Series*, Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=567701 [Accessed 17 May 2022]

Wang, S. (2020). Adding Data-driven Modeling to Causal Inference And Financial Economics, PhD thesis, University of Western Ontario, Available online: https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=9982&context=etd [Accessed 17 May 2022]

Wasserbacher, H., & Spindler, M. (2021). Machine Learning for Financial Forecasting, Planning and Analysis: Recent developments and pitfalls, *arXiv*, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.2107.04851&site=eds-live&scope=site [Accessed 17 May 2022]

Whitelaw, R.F. (1994). Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns, *The Journal of Finance*, vol. 49, no. 2, pp.515-541, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.10.2307.2329161&site=eds-live&scope=site [Accessed 17 May 2022]

Xu, N. (2021). Essays on Index Funds and Actively Managed Funds, PhD thesis, University of California, Irvine, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edssch&AN=edssch.oai%3aescholarship.org%2fark%3a%2f13030%2fqt5097d1rm&site=eds-live&scope=site [Accessed 17 May 2022]

Yadav, G.S., Guha, A., & Chakrabarti, A.S. (2020). Measuring Complexity in Financial Data, *Frontiers in Physics*, vol. 8, pp.339-348, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edswsc&AN=000586009200001&site=eds-live&scope=site [Accessed 17 May 2022]

Young, C. (2019). The Difference Between Causal Analysis and Predictive Models: Response to "Comment on Young and Holsteen (2017)", *Sociological Methods and Research*, vol. 48, no. 2, pp.431-447, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edselc&AN=edselc.2-52.0-85049686108&site=eds-live&scope=site [Accessed 17 May 2022]

Zhao, Q., & Hastie, T. (2021). Causal Interpretations of Black-Box Models, *Journal of Business & Economic Statistics*, vol. 39, no. 1, pp.272-281, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=inh&AN=21585462&site=eds-live&scope=site [Accessed 17 May 2022]

Zhou, F., Zhang, Q., Sornette, D, & Jiang, L. (2019). Cascading Logistic Regression Onto Gradient Boosted Decision Trees for Forecasting and Trading Stock Indices, *Applied Soft Computing*, vol. 84, pp.646-658, Available through: LUSEM Library website http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=inh&AN=19541779&site=eds-live&scope=site [Accessed 17 May 2022]