DISSERTATIONES
RERUM
OECONOMICARUM
UNIVERSITATIS
TARTUENSIS
91

LUCA ALFIERI

Forecasting business cycles and financial crises with machine learning

DISSERTATIONES RERUM OECONOMICARUM
UNIVERSITATIS TARTUENSIS
91

DISSERTATIONES RERUM OECONOMICARUM
UNIVERSITATIS TARTUENSIS
91

LUCA ALFIERI

Forecasting business cycles and financial crises with machine learning

UNIVERSITY OF TARTU
Press
1632

School of Economics and Business Administration, University of Tartu, Estonia

This dissertation is accepted for the defence of the degree of Doctor Philosophy (in Economics) on the 4th of April 2025, by the Council of the School of Economics and Business Administration, University of Tartu.

Supervisors: Associate Professor Jaan Masso (PhD),
University of Tartu, Estonia

Associate Professor Lenno Uusküla (PhD),
University of Tartu, Estonia,
and Chief Economist, Luminor Bank, Estonia

Opponents: Dr. Dmitry Kulikov (PhD),
Chief Specialist, Eesti Pank (Bank of Estonia), Estonia

Associate Professor Christian Brownlees (PhD),
Universitat Pompeu Fabra, Spain

The public defence of the dissertation is on the 19th of May 2025 at 14.00 in room 1018, University of Tartu, Narva Rd 18.

The publication of this dissertation is granted by the Doctoral School of Economics and Innovation of University of Tartu created under the auspices of European Social Fund and by University of Tartu School of Economics and Business Administration.

European Union
European Regional
Development Fund

Investing
in your future

ISSN 1406-1309 (print)
ISBN 978-9916-27-864-2 (print)
ISSN 2806-254X (pdf)
ISBN 978-9916-27-865-9 (pdf)

Copyright Luca Alfieri, 2025

Tartu University Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF AUTHOR'S PUBLICATIONS AND CONFERENCE PROCEEDINGS 7
I. Papers in international journals 7
II. Other research publications 7
III. Working papers and pre-prints 7
INTRODUCTION 9
Motivation for the research 9
Aim and research tasks of the thesis 13
Research methodology 14
Contributions 15
Novelties of the study 17
Delineation and limitations of the study 18
Structure of the study 18
Acknowledgements 19
1 AN OVERVIEW OF BIG DATA AND MACHINE LEARNING 21
1.1 Predictions in Macroeconomics 21
1.2 Big Data and Machine Learning in Economics 26
1.3 Machine Learning Methods in Macroeconomics 31
1.4 Literature Overview of Forecasting Business Cycles and Financial Crises with Machine Learning Methods 33
2 OVERVIEW OF THE MAIN MACHINE LEARNING METHODS 45
2.1 Regression methods in Machine Learning 45
2.1.1 Shrinkage methods 45
2.1.2 Regression trees 48
2.1.3 More trees to avoid overfitting: ensemble methods 49
2.1.4 Other regression methods 55
2.2 Classification methods in Machine Learning 55
2.2.1 Classification trees and AdaBoost 55
2.2.3 Other classification methods 57
2.3 Summary 57
3 MACROECONOMIC APPLICATIONS USING MACHINE LEARNING 58
3.1 Forecasting US industrial production using a Big Fat dataset 58
3.1.1 Data and data treatment 59
3.1.2 Methodology 61
3.1.3 Main Results 63
3.1.4 Additional Results 70
3.1.5 Evaluations of the results 73
3.2 AdaBoost vs logit for identifying periods of crisis 73
3.2.1 The data and data treatment 73

3.2.2 Methodology 78
3.2.3 Main Results 80
3.2.4 Evaluations of the models 88
4 DISCUSSION AND CONCLUSIONS 91
4.1 Discussion of the first empirical study: forecasting industrial production 91
4.2 Discussion of the second empirical study: predicting crises 93
4.3 General discussion and future investigations 96
REFERENCES 101
APPENDIX 113
Tuning of the hyperparameters and Robustness Checks – first empirical study 113
FRED-MD Data Description and Results of all the estimations 115
Other results and descriptive statistics 137
SUMMARY IN ENGLISH 147
SUMMARY IN ESTONIAN 161
CURRICULUM VITAE 175
ELULOOKIRJELDUS 176

LIST OF AUTHOR'S PUBLICATIONS AND CONFERENCE PROCEEDINGS

I. Papers in international journals

1. **Alfieri, L.** (2021). "Heterogeneity of Financial Institutions in the Process of the Economic and Monetary Integration in East Asia". *The World Economy*, 44 (4), p. 1053–1076. https://doi.org/10.1111/twec.13029
2. Pavlenkova, I., **Alfieri, L.**, Masso, J. (2023). "Effects of Automation on the Gender Pay Gap: the case of Estonia". *Industrial and Corporate Change*, Special Section, Technology and the Labor Market, 00, p. 1–25. https://doi.org/10.1093/icc/dtad065
3. **Alfieri, L.,** Gabrielyan, D. (2024). "The Communication Reaction Function of the European Central Bank. An Analysis using Topic Modelling". *Baltic Journal of Economics*, 24(1), p. 58–87. https://doi.org/10.1080/1406099X.2024.2303904

II. Other research publications

1. Micek, G., Bednar, P., Rafaj, O., Belvončikova, E., Paas, T., **Alfieri, L.**, Małochleb, K., Matoškova, J. (2022). "Independently-run coworking spaces and effects of the COVID-19 pandemic". In I., Mariotti, M., Di Marino (eds.) *The Covid-19 pandemic and the Future of Working Spaces*. Routledge-RSA Regions and Cities Book Series – Regions, Cities and Covid-19. Milton Park, Abingdon, Oxfordshire, UK: Routledge, p. 39–51. https://doi.org/10.4324/9781003181163-5
2. Piirsalu-Kivihall, K., Sinitsyna, A., **Alfieri, L.**, Paas T. (2023). "Acceleration of the Remote Work and Coworking Practices in Estonia During Covid-19". In Akhavan, M., Hölzel, M., Leducq D., (eds.) *European Narratives on Remote Working and Coworking During the Covid-19 Pandemic.* SpringerBriefs in Applied Sciences and Technology PoliMI SpringerBriefs. Springer, Berlin, May, p. 19–27. https://doi.org/10.1007/978-3-031-26018-6_3

III. Working papers and pre-prints

1. **Alfieri, L.** (2018). "Forecasting US Industrial Production with Boosting in a Data-rich Environment". Unpublished manuscript, University of Tartu. Available here: https://www.researchgate.net/publication/343007898_Forecasting_Industrial_Production_with_Boosting_in_a_Data-rich_Environment

2. **Alfieri, L.** (2019). "Boosting Crises Over a Century – A Multi-Country Perspective". Unpublished manuscript, University of Tartu. Available here: https://www.researchgate.net/publication/342947183_Boosting_Crises_over_a_Century_A_Multi-Country_perspective
3. **Alfieri, L.,** Kokashvili, N. (2019). "The Financial Safety Nets in East Asia and Europe: A Political Economy Assessment". University of Tartu Working Paper Series, No. 120.
4. Ashyrov G., Tverdostup, M., Paas, T., **Alfieri, L.** (2019). "A Cross-Country Input-Output Analysis of Estonian and Finnish Blue Industries". (previous version: University of Tartu Working Paper Series, No. 109). University of Tartu.
5. **Alfieri, L.,** Jõgi, K., Krassavina, V., Lamson, L., Vettik-Leemet, P., Sinitsyna, A., Towner, M., Trabskaia, J., Vadi, M. (2019). "Analysis of The Estonian Debate on R&D Funds Using a Scenario Building with Predetermined Decision-Making Constraints". Unpublished manuscript, University of Tartu. Available here: https://www.researchgate.net/publication/371500465_Analysis_of_The_Estonian_Debate_on_RD_Funds_Using_a_Scenario_Building_with_Predetermined_Decision-Making_Constraints
6. **Alfieri, L.,** Sinitsyna, A. (2022). "Effects of Financial Similarity on Remittances in Nordic-Baltic Region". University of Tartu, unpublished Manuscript.
7. **Alfieri, L.**, Eratalay, M., Lapitskaya, D., Sharma, R. (2022). "Effects of the ECB Communications on Financial Markets before and during Covid". University of Tartu Working Paper Series n° 140.
8. **Alfieri, L.**, Mariotti, I., Rossi, F. (2024). "The Twin Transition and Flexible Work Arrangements: A Systematic Literature Review". Politecnico di Milano, unpublished Manuscript.

INTRODUCTION

Motivation for the research

Making accurate predictions about the future is one of the main tasks for economists and policymakers in international organizations, governments (Wieland and Wolters, 2013), think tanks and private institutions (Fouliard et al., 2021). Businesses rely on economic forecasts to anticipate market conditions, adjust production levels, and allocate resources efficiently. Forecasts are also vital for making forward-looking monetary and fiscal policy decisions and ensuring financial stability.

Monetary policy decisions are typically made every six weeks at the European Central Bank (ECB). The ECB needs forecasts at least monthly to prepare for the discussions and have all up-to-date information. For example, the possibility of an economic slowdown in advanced economies can affect the consumption choices of households and investment decisions of firms, which, in turn, requires an early policy response.

In this sense, quarterly forecasts based on quarterly data are not sufficient to direct monetary policy decisions. Instead, only forecasting more frequently than on a quarterly basis provides sufficient information for central banks. This is why the first study in the thesis employs monthly data.

While a central bank's forward guidance and policy predictability do not necessarily suggest a rigid and unalterable plan, they play a crucial role in shaping expectations (Sutherland, 2023). The quality of present-day information and the accuracy of future predictions collectively provide a general signal to households and businesses regarding the objectives of central banks and the strategies they intend to employ to achieve these goals (Woodford, 2005).

Understanding the evolution of business cycles through indicators that are available more frequently than once a quarter and published swiftly, such as industrial production, can be very useful given the relevance of industrial production as an indicator of the business cycles during periods of elevated uncertainty and rapidly changing monetary policy, as in 2022–2023.[1] The information available can change quickly, and possible revisions or adjustments can be necessary. In particular, downturns in the economy are usually sharp and difficult to predict. Recovery from deep recessions that lack a sufficient and prompt policy response can take years (Bordo and Haubrich, 2017).

Even one-off events, such as specific acquisitions of banks; for example, the case of UBS bank buying Credit Suisse in a period of high volatility, can require an update on policy responses. Rare events, such as financial crises and non-economic shocks (e.g., Covid-19, the war in Ukraine), can require substantial

[1] Indeed, Ahir et al. (2022) show how the world uncertainty index (WUI) has increased on average since 2012 and has reached its peak during the Covid-19 pandemic.

adjustments in forecasts taking into account the latest data and understanding the dynamics at higher frequency than the standard quarterly business cycle. In this sense, policy decisions taken in February 2020 because of Covid-19 differ from those in March. Similarly, the decisions made in early February 2022 before the Russian invasion had to be revised in March.

In contrast to the business cycle, when forward decisions consider financial cycles, the different length of the cycles also plays a role. Large changes in credit cycles and house prices can take time to mature, and sudden changes may be of less relevance.[2] Hence, low frequency data, such as yearly data, can be a better choice when scholars want to study systemic financial crises. The comprehensive examination of intricate issues predicting rare events, such as systemic financial crises, necessitates rigorous analysis and use of the most suitable frequency of data.

The perceived failure (Shiller and Shiller, 2011) to accurately warn against and prepare for the financial crisis of 2007/2008 still resonates with the general public more than 15 years after the event. Indeed, the consequences of the Great Recession had a costly effect on the economies involved. The creation of new institutions is another consequence of the Great Recession.

The Financial Stability Board (FSB) is the successor of the Financial Stability Forum (FSF), and it is in charge of monitoring and giving recommendations related to global financial stability to the main economies of the world and international organizations, which are members of these new institutions. At the European level, the European Systemic Risk Board (ESRB) was created in 2010, and its scope is to implement macro-prudential regulations for financial markets in the European Union (EU) to prevent or mitigate financial systemic risk. In this sense, empirical studies focusing on early warning indicators are relevant for these organizations and policymakers.

More recently, the central banks have started to review how they can improve economic forecasting under uncertain times when pandemics and wars can cause unexpected shocks that make it hard to forecast future inflation and growth. Among other recommendations, Bernanke (2024) suggests a revision of the forecasting infrastructure (data, software, and models) and the importance of hiring data specialists to help in the maintenance of this infrastructure and the collection of Big Data, especially updated financial data. Moreover, he adds that artificial intelligence, that is mostly based on machine learning methods, can play a role in the future of economic forecasting. An example in this sense is the recent model proposed by ECB economists to forecast headline and core inflation (Lenza et al., 2023).

Business cycles and financial cycles correlate, but their properties can differ greatly (Rünstler, 2016). For example, Rünstler and Vlekke (2018) show that the standard business cycles of a time frame of less than eight years and the housing price cycles are uncorrelated. However, there are signs of correlation when GDP medium-term cycles of more than eight years are considered.

---

[2] Even if the large downturn of the Great Recession happened in 2008, initial issues started to be already visible in 2007.

Financial crises occur much more seldom than business cycles downturns. However, in economic terms they are often more costly. Claessens et al. (2010) suggest that output drops and cumulative losses are much larger when recessions are combined with credit crunches and house price busts. In particular, recessions without bust and credit crunches between 1960 and 2007 in 21 OECD countries have shown on average an output decline equal to 1.6 percent and corresponding cumulative losses equal to 2.3 percent. Instead, recessions with crunches and busts have had a decline in output equal to 2.5 percent and cumulative losses equal to 4.8 percent. Therefore, forecasting problems in the financial sector is a task on its own.

Simultaneously, it is imperative to recognise that both business cycles and financial crises often exhibit nonlinear dynamics. Existing literature demonstrates that business cycles can manifest nonlinear characteristics attributable to transient deviations from the norm (Hamilton, 2005) and the pro-cyclicality inherent to the credit market (Bouvatier et al., 2014). Financial crises represent even more extreme instances of nonlinear behaviour, often resulting in severe economic downturns (Akinci and Queralto, 2022). The asymmetrical nature of financial cycles is intricately linked to corresponding asymmetries in the variables associated with them, notably credit-related metrics (Akinci and Queralto, 2022) and stock prices (Sarantis, 2001).

Given the inherent complexity and the ongoing nature of forecasting, economists remain steadfast in their efforts to refine their methodologies, often incorporating novel techniques from the field of data science. In this context, a prominent set of approaches collectively categorised as machine learning (ML) has emerged as a valuable resource for addressing linear and nonlinear relationships among macroeconomic variables (Mosavi et al., 2020; Athey, 2018).

Hence, just the presence of nonlinearities can motivate the usage of machine learning methods. Many machine learning methods express their potential in the presence of nonlinear continuous variables. Some machine learning methods have great potential to predict specific events, such as financial crises (Casabianca et al., 2022; Bluwstein et al., 2020).

A second motivation for why economists are interested in ML methods is the availability of lots of data, which is often described under the term Big Data. Historically, economists primarily relied on smaller, meticulously curated datasets, often originating from national statistical offices and various public and private entities as reported in subsection 1.1. Notably, a disaggregation of these datasets falls within the purview of what is now within the Big Data paradigm. However, the sheer volume of data has surpassed the practicality of traditional econometric techniques for analysis.

Extensive datasets offer valuable and novel insights for forecasting future macroeconomic trajectories, including events of significance, such as business cycles and financial crises, which are, by nature, infrequent and significant occurrences. ML methods have demonstrated their capacity to handle these types of data. Moreover, non-traditional data, also referred to as unstructured data, coming from texts (Hansen et al., 2019; Hansen et al., 2018), satellite images, and

telecommunications (Cerdeiro et al., 2020; Donaldson and Storeygard, 2016) require ML methods in order to be transformed into quantitative data.

This has led to the development of many of the methods used in this monograph, even if all the potential of Big Data cannot be examined here. Currently, economists and econometricians study when and how ML methods can be used in their empirical exercises and what their limitations are. Goulet Coulombe et al. (2022) observe how the presence of nonlinearities is the true game changer that makes ML methods convenient tools for forecasting macroeconomic indicators, especially in a data-rich environment.[3]

The need for nonlinear methodologies in economic analysis is underscored by the prevalence of intricate, nonlinear relationships among independent variables in the context of macroeconomic uncertainty, financial stress, and housing market fluctuations (Goulet Coulombe et al., 2022). This underscores the practicality of employing nonlinear machine learning methods, such as regression tree-based techniques. Notably, Doğan and Midiliç (2019) and Heinisch and Scheufele (2018) provide evidence of the usefulness of incorporating diverse financial and macroeconomic indicators in forecasting exercises.

Doğan and Midiliç (2019) exemplify how including an extensive array of financial series (204) enhances their predictions and nowcasts for Turkish GDP. In contrast, Heinisch and Scheufele (2018) investigate various methodologies, notably mixed-data sampling (MIDAS) regression and factor models, demonstrating their adeptness in extracting valuable insights from financial and macroeconomic indicators to enhance forecasts of German GDP.

Machine learning methods prove particularly valuable in data-rich environments due to their capacity to effectively address nonlinear relationships and a plethora of covariates. They play a pivotal role in refining forecasts, especially in the context of nonlinear economic cycles such as business and financial cycles. Complex machine learning models offer a robust solution by simultaneously mitigating overfitting concerns while delivering improved forecasts. For instance, Lehmann and Wohlrabe (2016) harness the potential of boosting models, which combine simpler machine learning models, to predict German industrial production. They leverage 175 macroeconomic indicators to explore how boosting characteristics can enhance forecasting accuracy. Buchen and Wohlrabe (2011) affirm the competitiveness of boosting models against factor models, further underscoring the potential of machine learning in the realm of economic forecasting.

In conclusion, the motivation behind this thesis is rooted in the imperative need for accurate predictions in economics and policymaking. Economists, policymakers and various organizations are responsible for forecasting to make informed decisions related to monetary and fiscal policies, and financial stability. Monetary policy decisions require up-to-date forecasts at a frequency that matches the rapid changes in the global economic landscape, particularly in the face of potential economic slowdowns or significant events. Nonlinearities in economic

[3] A data environment with a large number of predictors.

cycles, along with the unique properties of financial crises, call for advanced forecasting methods such as machine learning to address the intricacies of contemporary economic challenges.

Moreover, the advent of Big Data has expanded the horizons of economic data sources, necessitating new methodologies to harness the wealth of information. Machine learning has proven itself invaluable in handling these vast datasets, offering the potential to predict rare events like financial crises and business cycle dynamics (Casabianca et al., 2022; Lehmann and Wohlrabe, 2016). Additionally, non-traditional data sources from texts, satellite images and telecommunications (Cerdeiro et al., 2020; Hansen et al., 2019; Donaldson and Storeygard, 2016) have opened new avenues for economic analysis, driven by the capabilities of machine learning. As economists explore the utility of machine learning methods and their limitations, they discover that nonlinear methodologies are essential in addressing intricate relationships among independent variables in data-rich environments.

Furthermore, the relevance of machine learning is evident in its capacity to handle nonlinear relationships, and it has emerged as a powerful tool in forecasting macroeconomic indicators (Goulet Coulombe et al., 2022). Researchers have demonstrated how integrating various financial and macroeconomic indicators in forecasting exercises can lead to improved predictions and nowcasts, and machine learning's ability to mitigate overfitting concerns while delivering more accurate forecasts has been demonstrated. Hence, the integration of machine learning into economic forecasting not only accommodates nonlinearities but also addresses the demands of a data-rich world, positioning it as a critical tool for contemporary economic analysis.

Aim and research tasks of the thesis

The aim of this thesis is to evaluate the effectiveness of machine learning, particularly boosting methods, in economic forecasting across different data environments[4]. By comparing ML-based predictions with standard econometric benchmarks, the study assesses their performance in forecasting both continuous variables, such as industrial production (in the first empirical study of the thesis), and discrete ones, like financial crises (in the second empirical study of the thesis). Through this analysis, the thesis seeks to determine the conditions under which machine learning models offer advantages over traditional approaches in economic prediction.

The tasks of the thesis can be summarised as:

1. Predict economic variables using ML, concentrating on boosting models, and standard econometric methods;

[4] In this thesis, the term "data environment" encompassing considerations related to linearity, non-linearity, and the volume of data (e.g., Big Data or data-rich environments). A similar conceptualization can be observed in the work of Medeiros et al. (2021).

2. Compare the predictions made by ML approaches with different standard benchmark methods;
3. Assess how boosting models can help to predict continuous variables such as industrial production;
4. Explore how different boosting models can be compared to other machine learning methods in a continuous variables context;
5. Assess how boosting models can help to predict discrete variables, such as financial crises, and pre-crisis and crisis periods;
6. Take into account the importance of different data environments in the empirical exercises, and provide conclusions on when machine learning models can be helpful.

Research methodology

The focus of the thesis is on machine learning techniques. The monograph first provides an overview of the literature related to Big Fat Data,[5] machine learning in economics and, more specifically, in macroeconomics. Moreover, the methodological section describes the ML methods used in the thesis. Finally, the thesis applies a large variety of ML methods for two empirical macroeconomic studies using ML regression and classification methods. In the first study, the empirical models are evaluated in terms of root mean squared errors. In the second study, the models are evaluated with different criteria, such as the F1 score.[6]

The first study focuses on boosting algorithms applied in regression analysis. These algorithms hold significant promise for making better predictions. The results of the boosting algorithms are compared with pertinent benchmark models and other machine learning methodologies.

The second study identifies crisis-periods through AdaBoost (Freund and Schapire, 1997; Casabianca et al., 2022) as a classification method. AdaBoost efficacy is compared with two benchmark models in different data environments. Moreover, the most important variables per country for identifying crises, and pre-crisis and crisis-periods are detected using the algorithm.

This second study employs the Macro History dataset from Jordà et al. (2017). The dataset includes 25 yearly time series of 17 developed economies. The data include a systemic financial crisis dummy and different possible economic and financial variables relevant for predicting crises.

[5] Data where the number of the predictors is greater than the number of observations.

[6] F1 score is a measure of accuracy for classification exercises.

Contributions

This thesis provides significant contributions to several aspects of the existing literature. First, it advances the field of boosting techniques in macroeconomics, expanding both classification (Ward, 2017; Papadopoulos et al., 2019; du Plessis, 2022; Casabianca et al., 2022; Torky et al., 2023) and regression exercises (Exterkate et al., 2016; Maehashi and Shintani, 2020; Syed and Lee, 2021; Yousuf and Ng, 2021; Zhang et al., 2023) using machine learning. More specifically, it introduces a diverse array of boosting methods in regression, highlighting how specific characteristics of these techniques can impact predictions.

Previous papers such as Medeiros et al. (2021), Lehmann and Wohlrabe (2016), Buchen and Wohlrabe (2011), and Bai and Ng (2009) have studied only one or two boosting models without sufficiently expanding their discussion. The thesis addresses this research gap in the literature by presenting the different characteristics and performances of boosting models. It examines a range of boosting methods in greater detail than previous studies and recognises the importance of various aspects of boosting models, including loss functions and the methods of iteration, in determining the success of a machine learning method in forecasting, as suggested by Goulet Coulombe et al. (2022).

In classification, the thesis demonstrates the application of AdaBoost to forecast country-specific crisis periods, subject to certain conditions, emphasizing the need for further exploration. This part of the thesis takes into account papers such as Casabianca et al. (2022) but explores a different method of oversampling. Moreover, in contrast to studies such as those by Casabianca et al. (2022) and Papadopoulos et al. (2019), the thesis does not use the data of all the countries in the dataset to train the model to identify systemic financial crises or the pre-crisis or crisis-periods of one specific country (Papadopoulos et al., 2019) or group of countries (Casabianca et al., 2022) but the data of one specific country is used to identify its own systemic financial crises, or pre-crisis or crisis-periods, filling a gap in this specific literature.

In addition, considerable attention is paid to applying machine learning in forecasting exercises both with and without predictor pre-selection (Cepni et al., 2019; Jardet and Meunier, 2022; Borup et al., 2023) and addressing datasets with numerous predictors (Ward, 2017; Panagiotelis et al., 2019; Liu, 2023). Indeed, the first exercise includes a large amount of indicators consistent with a high dimensional space, as in previous literature (Goulet Coulombe et al., 2021a; Yousuf and Ng, 2021; Cepni et al., 2019; Panagiotelis et al., 2019), recurring on an already established dataset in the literature, such as the FRED-MD dataset (Goulet Coulombe et al., 2021a; Yousuf and Ng, 2021). While the second exercise employs the dataset from Jordà et al. (2017) that considers specific variables related to systemic financial crises.

Moreover, the second exercise adds more variables to the estimations through quadratic and cubic transformations. This makes it possible to observe both the effects of a pre-selection of predictors and the presence of a high dimensional space environment. The results underscore the benefits of variable pre-selection

for noise reduction, while emphasizing that the full potential of machine learning emerges with a large number of predictors, especially in the presence of non-linearities and imbalanced data.

Furthermore, the thesis provides valuable insights into applying machine learning to business cycles (Maehashi and Shintani, 2020; Yousuf and Ng, 2021; Richardson et al., 2021) and systemic financial crises (Alessi and Detkon, 2018; Beutel et al., 2019; du Plessis, 2022; Casabianca et al., 2022). In forecasting industrial production, the research points out specific aspects to consider when using boosting techniques, such as the procedure of splitting the trees and the introduction of factors. Furthermore, the research emphasizes the relevance of using specific time periods as training sets, which is line with previous findings in the literature, such as in Guerrón-Quintana and Zhong (2023) and Yousuf and Ng (2021).

Regarding systemic financial crises, it explores AdaBoost's application in predicting crises, and pre-crisis and crisis periods using a specific theoretically based dataset. In this sense, this exercise is a combination of the findings of Casabianca et al. (2022) plus the usage of the dataset from Jordà et al. (2017), as in Tölö (2020) and Torky et al. (2023), but in a country-specific and high dimensional space environment. Furthermore, the thesis contributes to the literature on machine learning and its connection with nonlinear behaviours (Maehashi and Shintani, 2020; Kauppi and Virtanen, 2021; Goulet Coulombe et al., 2021a) but with a specific focus on boosting methods and financial cycles. It highlights the increased effectiveness of AdaBoost in the presence of more nonlinear predictors and its ability to discern relevant variables and their nonlinear transformations.

The thesis stands out by employing an extensive array of boosting methods compared not only with standard benchmarks but also classic machine learning methods showing in which occasions some of the boosting methods are superior. In this sense, the thesis follows previous papers, such as Chu and Qureshi (2022) and Goulet Coulombe et al. (2021a), where boosting methods are compared with other ensemble methods and shrinkage methods. In addition, it introduces the use of AdaBoostRT and AdaBoostR2 in this strand of applied macroeconomic literature showing the potential of AdaBoostRT contributing to a neglected area of research. Moreover, it considers factor models as a benchmark, such as in Syed and Lee (2021) and Soybilgen and Yazgan (2021).

In the classification study, the thesis goes beyond comparing AdaBoost with a conventional logit model, as seen in prior studies, such as Casabianca et al. (2022), Bluwstein et al. (2020), and Beutel et al. (2018). It also contrasts AdaBoost with a stepwise logistic model, recognizing that the latter selection method is often more interpretable and parsimonious than benchmark models or complex machine learning approaches, which consider all indicators or create rankings of importance without eliminating variables, as the AdaBoost model does. Furthermore, the thesis offers a novel comparison of AdaBoost against standard logit benchmarks and stepwise logit regression in systemic financial crisis contexts taking into account different sampling methods for the dependent variable (Casabianca et al., 2022).

Finally, the exploration of machine learning across various timeframes and frequencies is a unique feature. Monthly time series data is analysed within distinct samples, including data from the Great Recession and predating it. The second empirical exercise uses yearly data, innovatively predicting crises, and pre-crisis and crisis periods using only country-specific predictors in a date frame of more than one century. This thesis meticulously considers specific yet differentiated datasets, illustrating how boosting techniques can be applied to analyse the configurations of these two datasets effectively.

Novelties of the study

This thesis offers a detailed exploration of the capabilities and limitations of various boosting methods in both regression and classification contexts. It evaluates their performance by comparing their predictions with established benchmarks found in the literature, including references such as Medeiros et al. (2021) and Casabianca et al. (2022). To provide a more comprehensive assessment, this study introduces more complex benchmark models, such as factor models, which serve as stringent evaluative standards for the forecasting prowess of novel methods.

Additionally, this thesis conducts a comparative analysis of regression boosting with other machine learning methods, such as shrinkage methods, employing hyperparameter tuning in the style of Medeiros et al. (2021), which relies on information criteria instead of standard cross-validation methods. A comprehensive evaluation of boosting models necessitates a comparison with other machine learning techniques, particularly those within the shrinkage model family, known for their simplicity and interpretability (James et al., 2013).

Finally, this study distinguishes itself as the first to employ the dataset provided by Jordà et al. (2017) using machine learning methods, focusing on each country's specific time series. This dataset presents a unique opportunity due to its extensive temporal coverage, inclusion of numerous variables, and incorporation of data from multiple countries. The dataset, featuring yearly data and a more equitable distribution of crises across countries, spans a wide timeframe from 1870 to 2016, encompassing 17 advanced economies and 92 crises. In comparison, previous studies, such as Beutel et al. (2018) covered 15 countries, but only 22 crises over the last 45 years.

Casabianca et al. (2022) included a larger number of countries (100), but the average number of crises per country was less than two, the variables used were limited, and the time frame was shorter, with only eight country-specific crises. These constraints hindered a comprehensive analysis of country-specific dynamics. Barthélémy et al. (2020) approached a dataset of 276 crises over 104 countries from 1973 to 2017, but this dataset included various types of crises, requiring different predictors. Therefore, by employing a dataset with specific indicators related to systemic financial crises, this study facilitates cross-country comparisons when systemic financial crises occur.

Delineation and limitations of the study

The work will not consider excessively complex and computationally expensive deep learning methods. Zheng et al. (2023) show how the problems of interpretability of results and the relationship between the accuracy and efficiency of these models must be overcome. At the same time, connecting economic theory and deep learning models is still difficult, even if some attempts in macroeconomics, such as Goulet Coulombe (2022) deserve to be considered for future research.

Moreover, empirical exercises on unstructured Big Data, such as text analysis, are not considered in this thesis as it focuses only on publicly available structured data. Text analysis is one of the main topics where machine learning methods can help macroeconomists extract new information and solve identification issues through publicly available texts.

Scholars studying monetary policy and central banking have started to use text analysis more and more frequently (Alfieri and Gabrielyan, 2024; Hansen et al. 2019; Hansen et al., 2018). Furthermore, the thesis is not aimed at checking what the best methods are among all machine learning methods, given the vastness of the machine learning methods available, or what the best ways are to find causal relationships (Bluwstein et al., 2020; Chernozhukov et al., 2018). Finally, even if variables connected with the literature on systemic financial crises are included in the second study, it does not aim to test any particular macroeconomic theory.

Structure of the study

The thesis has four chapters besides the introduction. Chapter 1 considers the importance of predictions in macroeconomics and the difficulties of making accurate forecasts in a nonlinear world. The chapter illustrates the concept of Big Data, which can be interpreted in different ways considering the different fields of research. Moreover, it provides an overview of the literature on using machine learning methods in economics. Finally, the chapter explores how machine learning methods are currently used in the macroeconomic field and in forecasting business cycle and financial crises through a systematic literature review.

Chapter 2 illustrates the ML models used in the thesis in both regression and classification analysis. In particular, the chapter describes the shrinkage models – the lasso, ridge and elnet models. Furthermore, it refers to the regression trees with their strengths and limitations. The chapter shows how ensemble methods of trees can overcome problems related to the overfitting of standard regression trees. Moreover, the chapter explains why other ML models are not used in the thesis. Concerning the classification methods, Chapter 2 illustrates the classification tree method and AdaBoost as the most used ensemble method related to classification trees. Finally, the chapter discusses why other classification methods are not used in the thesis.

Chapter 3 presents two empirical macroeconomic exercises. The first is a forecasting study of US industrial production using the FRED-MD dataset and ML regression models against benchmark models, such as autoregressive methods, random walk, and factor models. The variables used are the first four lags of more than 120 macroeconomic indicators from the FRED library. The exercise is a standard "horse race"[7] and forecasts different short, medium and long-term horizons (one month, two months, three months, four months, eight months and twelve months ahead).

The second empirical exercise aims to identify systemic financial crises, and pre-crisis and crisis periods and their main drivers. The dataset used is the Macrohistory dataset.[8] The data span from 1870 to 2016, and the variables of the dataset derive from literature related to systemic financial crises. AdaBoost is the machine learning method employed against logistic regression and step-wise logistic regression. Moreover, the results of the variable importance of AdaBoost are compared with the results of the previous theoretical and empirical literature. Different data environments are tested.

Chapter 4 discusses the results of the two empirical investigations and discusses how ML methods can be useful in different data environments. Subsequently, the thesis concludes by taking into account the general findings of the thesis and possible future areas of research.

Acknowledgements

This thesis has been an exciting and extenuating journey with many people that deserve my gratitude. I want to thank my girlfriend, Ekaterina Bogdanova, who shared the burden and has sacrificed part of her time since my first year in Tartu. The pride of having such a woman at my side cannot be explained in simple words. I wish to thank my family for their support during these years, as well as my recently passed away great-grand aunt Ines Tebaldi. She is greatly missed.

I would like to thank my PhD supervisors, Jaan Masso and Lenno Uusküla. I am very grateful for their support and help. I hope I did not disturb them too much. I can be very annoying sometimes. I would like to thank Karsten Staehr and Dmitri Kulikov for reading and commenting on the thesis. Their feedback has significantly contributed to its improvement.

I aim to thank all my co-authors for their great work in all the papers we worked on together. Many thanks to my friends Alessandra Morrone, Alessandra Dezi and Lisa Zorzato for creating a de facto specialised group of support for Italians doing a PhD at the University of Tartu. I wish also to thank my dear Brazilian friend Luiza Schuch de Azambuja.

[7] The horse race is a standard exercise in forecasting literature when a series of models are compared in their predictive capacities over different horizons.

[8] The data are from multiple sources such as central banks, international organisations, and other scholars' research.

I am extremely grateful to my PhD colleagues as well. In particular, Gaygysyz Ashyrov, Diana Gabrielyan, Anastasia Sinitsyna, Ilona Pavlenkova, Maria Chebotareva, Laura Kivi, and Darya Lapitskaya. Furthermore, I thank all my colleagues from the Chair of Economic Modelling. In particular, I want to thank Tiiu Paas, Hakan Eratalay, Ricardo Alfredo Mendes Pereira Vicente, Helen Poltimäe, Eneli Kindsiko, and all the other colleagues from the Chair for creating a very friendly environment in which to work.

I also want to express my gratitude to my many colleagues from the School of Economics and Business Administration. In particular, I am thankful to Maaja Vadi, Kadri Ukrainski, Raul Eamets, Priit Vahter, Urmas Varblane, and Anne Reino for their valuable research suggestions. My deepest thanks also go to Riina Viilup, Anneli Kütt, Ülle Maidla, Tiia Haldma and Katrin Tamm for their help with administrative matters. I really think that Estonian academia should value their work much more.

I want to thank my dear friends Tania Masi and Umut Türk for always believing in me over the last eleven years. I would like to thank the research unit at Eesti Pank for the wonderful visiting period I spent there, even if it was a difficult time for everybody as it was the pandemic period. I am also grateful to the people I worked with at the Vienna Institute for International Economic Studies for their suggestions about the thesis. In particular, Maryna Tverdostup is a great friend and an amazing economist. Last but not least, I want to thank the Advanced Study Program at the Kiel Institute of World Economy for having trusted me and giving me an opportunity to recover and continue my career after the worst period of my life.

1 AN OVERVIEW OF BIG DATA AND MACHINE LEARNING

1.1 Predictions in Macroeconomics

Considering forecasting in general terms, Hawkins (2005) observes how ancient Egyptians looked at the level of the Nile to understand the future and foretell harvests. Modern economic forecasting had its infancy during the period 1910–1930. The Great Depression put a stop to this first wave of forecasting research, since the forecasters were accused of failing to predict the 1929 crisis (Hawkins, 2005). It is after World War II that the literature on economic forecasting has definitely grown. Only one year after the end of WWII, Burns and Mitchell (1946) published their book "Measuring Business Cycle" where the modern definition of the business cycle can be found.[9] Forecasting economic fluctuations becomes the main issue for macroeconomists aiming to help policymakers.

After the end of the Keynesian revolution, the Lucas critique (Lucas, 1976) and the new microfounded models, such as the Real Business Cycle (RBC) model (Kydland and Prescott, 1982) and its New Keynesian version, the Dynamic stochastic general equilibrium (DSGE) model (Smets and Wouters, 2007; Galí, 2018), have provided a new level of sophistication for forecasting exercises.[10] Nonetheless, the Great Recession prompted a paradigm shift in economic modelling, leading to the development of models such as DSGE models with heterogeneous agents, Heterogeneous Agents New Keynesian (HANK) models (Kaplan et al., 2018), and models incorporating financial frictions (Bernanke et al., 1999).

From a more data-driven perspective, Stock and Watson (1998) observe the relations between business cycle frequency and macroeconomic variables, such as production, prices and employment. Moreover, Stock and Watson (2006) show how multivariate models with a large number of predictors can be helpful in time series forecasts. It can be noticed how macroeconomic forecasters face major challenges in predicting financial crises. This is due to the differences that characterise these events. Crises have similarities (Reinhart and Rogoff, 2009) – periods of boom and bust in the credit cycles, as well as relations between crises and household debt can shape a financial crisis. In this sense, even if forecasters cannot really predict financial crises, they have learnt what important variables to consider. Most of these variables, as suggested by Fouliard et al. (2021), can already be found in classical studies such as Fisher (1933), Kindleberger (1978) and Minsky (1986).

Many institutions across the world are involved in forecasting business cycles and macroeconomic variables. Table 1 shows examples of key public institutions and private companies which provide different forecasts for the general public of

[9] Even if some earlier studies on the business cycle can be traced to De Sismondi (1819).

[10] However, the DSGE model has not been used very much for forecasting outside academia and research institutions.

some of the main economies or of the entire world economy. Various institutions specialise in forecasting variables related to monetary policies, while others concentrate on those associated with fiscal policies. Some institutions, as well as private enterprises, direct their forecasting efforts towards variables like credit or trade. The list is by no means exhaustive, but it illustrates the pervasive role of forecasting in contemporary society and the extensive array of organizations engaged in this endeavour.

Table 1 – Examples of main private and public independent institutions, central banks and international organizations that forecast macroeconomic variables and financial crises

	Fiscal Variables	Monetary Variables	General Variables
U.S.	CBO: Budget and Economic Outlook Moody's Analytics Forecasts JP Morgan Forecasts Morgan Stanley Forecasts Barclays Forecasts Goldman Sachs Forecasts	FRB (staff): Greenbook/Tealbook FOMC: FOMC minutes Moody's Analytics Forecasts JP Morgan Forecasts Morgan Stanley Forecasts Barclays Forecasts Goldman Sachs Forecasts	Philadelphia FED: Survey of professional forecasters Moody's Analytics Forecasts JP Morgan Forecasts Morgan Stanley Forecasts Barclays Forecasts Goldman Sachs Forecasts
Euro Area	European Commission: European Economic Forecasts Société Générale Forecasts Intesa San Paolo Forecasts Oxford Economics Forecasts	Eurosystem (staff): Monthly Bulletin (ECB +National Central Banks) Economist Intelligence Unit Forecasts Société Générale Forecasts Intesa San Paolo Forecasts Oxford Economics Forecasts	ECB: ECB Survey of Professional Forecasters Economist Intelligence Unit Forecasts Société Générale Forecasts Intesa San Paolo Forecasts Oxford Economics Forecasts
UK	Office for Budget Responsibility: Economic and Fiscal Outlook Beacon Economics Forecasts Citigroup Forecasts Economist Intelligence Unit Forecasts Oxford Economics Forecasts	Bank of England: Inflation Report Beacon Economics Forecasts Citigroup Forecasts Economist Intelligence Unit Forecasts Oxford Economics Forecasts	Beacon Economics Forecasts Citigroup Forecasts Economist Intelligence Unit Forecasts Oxford Economics Forecasts

	Fiscal Variables	Monetary Variables	General Variables
Switzerland	Federal Finance Administration (FFA): Federal Financial Statements KOF Swiss Economic Institute Forecasts Zürcher Kantonalbank Forecasts Citigroup Forecasts	SNB: Monetary Policy Assessment + Quarterly Bulletin KOF Swiss Economic Institute Forecasts Zürcher Kantonalbank Forecasts Citigroup Forecasts	KOF Swiss Economic Institute Forecasts Zürcher Kantonalbank Forecasts
Japan	Cabinet Office: Economic Outlook and Basic Stance for Economic and Fiscal Management Oxford Economics Forecasts ITOCHU Institute Forecasts	Bank of Japan: Outlook for Economic Activity and Prices Oxford Economics Forecasts ITOCHU Institute Forecasts	Oxford Economics forecasts ITOCHU Institute Forecasts
International	IMF: Fiscal Monitor, Regional Economic Outlook IMF Staff Country Report OECD: Country Report	Consensus Economics Forecasts: World Activity IMF: World Economic Outlook Regional Economic Outlook IMF Staff Country Report	IMF: World Economic Outlook Regional Economic Outlook IMF Staff Country Report OECD: Economic Outlook, Country Report World Bank: Global Economic Prospects Consensus Economics Forecasts: World Activity

Source: Wieland and Wolters (2013), Consensus Economics and author's adjustments.

Central banks, international organizations, investment banks, and governmental institutions are all involved in forecasting the future and handling negative consequences in case of recessions or financial crises. In this sense, forecasters are continuously looking for new data and methods to improve their predictive capacity, especially in the case of many predictors, and to handle possible nonlinear behaviour of the business and financial cycles.

There are two different types of forecasting situations (Fouliard et al., 2021). The first type concerns continuous variables, such as variables that describe the business cycle (e.g. industrial production). The second type is related to specific events that are binary variables, such as financial crises. The analysis of binary variables, such as crises, need to consider how rare the positive events (crises) are compared to the negative ones (absence of crises). This is the problem of having imbalanced data (Fernández et al., 2018), which needs to be addressed in these kinds of studies.

Economic recession can be related to financial crises, but this is not always the case, given that the structures of the two events are different (Rünstler, 2016). There are business cycle troughs without financial weakness and not every financial crisis results in an economic downturn. An overview of the business cycle measured as GDP growth and shown as a solid line, and systemic financial crises shown by grey vertical bars is given in Figure 1. The figure summarises the two empirical studies of the thesis: forecasting continuous and binary variables.

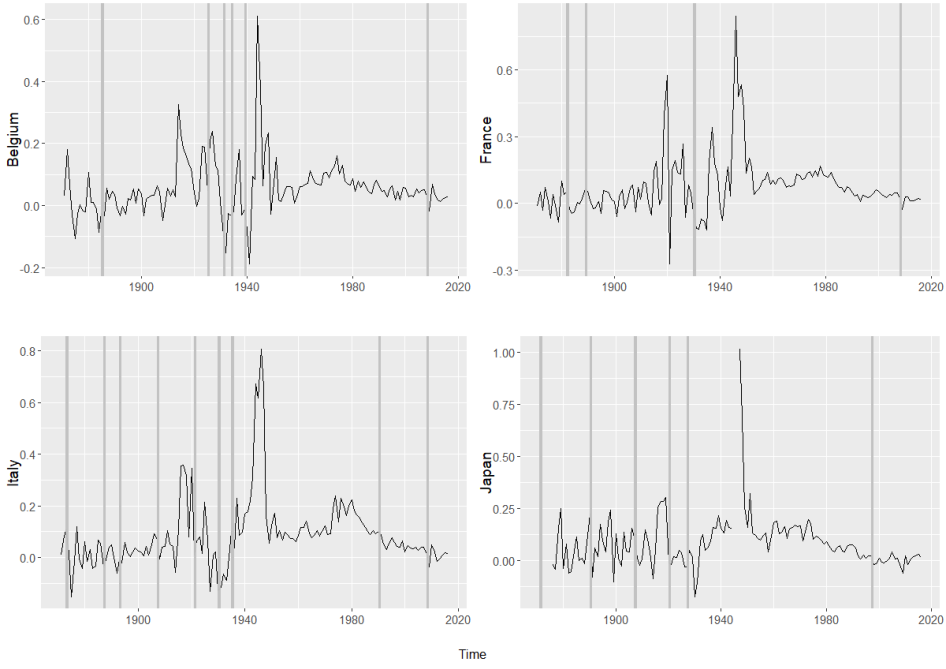

Figure 1. GDP growth and systemic financial crises in selected developed countries from 1870–2016 (Jordà et al. 2017).

Legend: black line – GDP growth; grey vertical lines – systemic financial crises

It should be noted that forecasting the business cycle considering GDP growth is not the only option. Industrial production is the most used variable at monthly frequency for measuring business cycles (Bulligan et al., 2010). It is a measure that is available much faster than GDP to measure the state of the economy. In particular, industrial production plays a pivotal role in nowcasting and forecasting GDP in the most advanced economies (Golinelli and Parigi, 2007). Industrial production is a variable that can be forecasted on its own. At the same time, it can be used to forecast GDP, so it is a perfect variable to detect changes in the business cycle earlier.

Given the importance of industrial production as a macroeconomic indicator to show the state of the economy, many scholars work on forecasting US industrial production. Stock and Watson (2003) show that the results of models can depend on the presence of the latest revisions of macroeconomic indicators. GDP and

other macroeconomic indicators can be subject to revisions from statistical institutions. This can have important consequences on the forecast results (Bulligan et al., 2010; Diebold and Rudebusch, 1991).[11]

The same authors in another work (Stock and Watson, 2006) forecast monthly US industrial production with 130 predictors using dynamic factor models. They observe how predicting one leading indicator, such as industrial production, with many predictors can provide information not previously available. Moreover, they can solve the structural instability of low-dimension forecasts. Finally, they indicate that simple factor models calculated with the first few principal components can improve on Bayesian model averaging (BMA) methods.

Forecasting extends beyond considerations of business cycles and macroeconomic indicators, encompassing the critical domain of financial crises. The literature on financial crises predominantly emphasizes the identification of primary predictors that precipitate systemic financial crises before the actual forecasting process. Given the breadth of literature in this field, this section selectively highlights studies closely aligned with the thesis at hand. These references primarily pertain to studies that employ machine learning methods or to studies employing methods that may serve as competitors to machine learning approaches. In particular, it can be observed how the literature highlights certain specific variables that have turned out to be crucial for forecasting financial crises.

Richter et al. (2021) and Davis et al. (2016) posit that current account deficits, especially when accompanied by significant capital flow fluctuations, can trigger financial crises. Schularick and Taylor (2012) underscore how a surge in credit growth can lay the foundation for a financial crisis. Jordà et al. (2015) emphasize the critical role of scrutinizing credit growth and asset prices in comprehending the channels of financial crises.

Acharya and Richardson (2009) offer an illustrative example by highlighting the connection between the credit boom and soaring house prices, which culminated in the Great Recession. Their analysis underscores the unsustainable real estate bubble created by an unprecedented annual growth rate of 11% in house prices from 2002 to 2007, eventually leading to the 2008 crisis.

In addition to these factors, scholars must also consider the implications of weak public finances when endeavouring to predict financial crises. Casabianca et al. (2022) and Manasse and Roubini (2009) stress the relationship between a nation's public debt level and its vulnerability to crises. Researchers like Reinhart and Rogoff (2009) draw attention to the common elements found in financial crises across different time periods, including credit collapses and unsustainable public finances. Consequently, empirical studies aimed at identifying recurrent patterns within an extended time frame are justified, notwithstanding inherent limitations. As argued by Fouilard et al. (2021), the ramifications of financial crises, with their potential for economic and political disruptions, render the study of financial crises a central theme in macroeconomic literature.

[11] FRED-MD dataset is not composed by real-time data. A study that aims to perform similar exercises with real-time data can employ the ALFRED dataset.

1.2 Big Data and Machine Learning in Economics

Big Data is at the centre of the new technological evolution of the world economy and society. This term is even overused in the last decade and sometimes turns out to be a buzzword (Acharya et al., 2018). Ng (2017) suggests two distinct periods – pre- and post-Google – where the quantity of data on Earth exploded in the post-Google period. Indeed, the combination of social media, web search data, online prices, and transaction data brought the total digital space to 100 zettabytes in the twelve years to 2022 (Figure 2) and it will reach 180 zettabytes in 2025. The growth of data is exponential, and there seems to be no end to it. The classification of these large amounts of data is pretty complex, depending on what discipline the scholar is from.

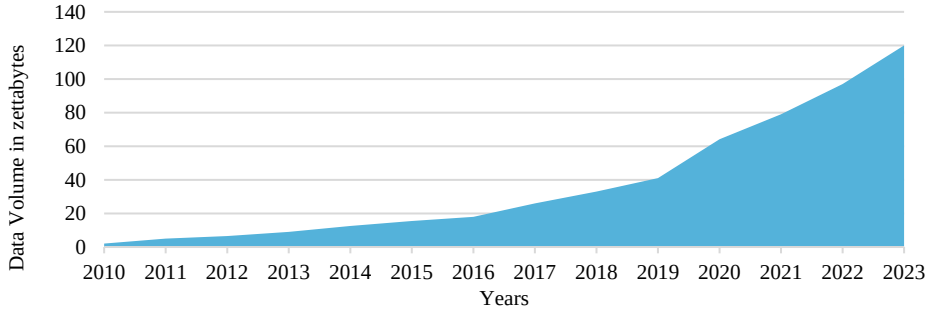

Figure 2. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2023 (in zettabytes)

Notes: The figures from 2021 to 2023 were calculated by Statista based on the 2020 forecast figure and the five-year compound annual growth rate of 23 percent of the data provided by Statista. The figures prior to 2020 are based on an International Data Corporation (IDC) forecast from late 2018.
Source: Statista (2022)

In general, Big Data is usually defined by the 4 V's of Big Data by IBM (2019): volume, velocity, veracity and variety. "Volume" concerns the scale of data created every day. "Velocity" considers the speed of the data streamed. "Veracity" is related to the data's reliability problem and the poor quality of some of the data when they are not appropriately treated. "Variety" regards the different typologies of Big Data. In addition, some scholars add another V, "Value", which refers to the value obtained by the use of Big Data (Furht and Villanustre, 2016). Khan et al. (2018) even arrive at 10 Vs.[12] Moreover, Varian (2014) describes Big Data as "Huge", "Fat" and "Tall". In a Huge Big Data environment, observations and predictors are numerous. When Big Fat Data are involved, the predictors are much more than observations, and when Big Tall Data are studied, the observations are much more than predictors.

[12] Volume, Velocity, Variety, Value, Variability (data flow inconsistency), Veracity, Validity (data properly understandable), Volatility (data durability), Viability (data activeness), and Viscosity (data complexity).

In contrast to Varian (2014), Ng (2017) defines Big Data as the large and complex datasets that are not gathered via surveys, are not backed by government organizations, and require a change in the computer environment before they can be stored or studied. This definition is more related to the number of gigabytes or terabytes employed. As Ng (2017) suggests, standard computers can easily handle some gigabytes of data but the capacity of desktop computers to analyse terabytes of data is limited even considering unlimited financial resources. Moreover, the time to complete a task on a computer depends on both software efficiency and hardware performance. Key factors include the speed of data transfer between storage, RAM, and the CPU, as well as queuing time based on the amount requested by the RAM (Ng, 2017).

Another division is between structured and unstructured Big Data (the variety aspect of Big Data). Structured Big Data refers to numerical data, while unstructured Big Data includes text, video, satellite data, and audio, which are forms of data that must be transformed into numerical data using different machine learning methods or algorithms from text mining analysis. Unstructured Big Data is a large part of Big Data worldwide, and the relative importance of unstructured data over structured data is growing exponentially (Azad et al., 2020). For this reason, it is necessary to use specific algorithms or machine learning methods to treat and compute them.

Machine Learning is considered part of the artificial intelligence field and consists of the creation of algorithms based on the capacity to *learn* by themselves without specific instructions from the creator (Koza et al., 1996; Samuel, 1959). In this sense, this relates to the division between the *training part* and the *test part* of the data analysed by the machine, where the training part is used by the machine to learn and improve its performance. Subsequently, the test part is utilised to evaluate the model's performance on unseen data.

In principle, it is possible to use ML methods with default values for the hyperparameters without testing, but this can lead to suboptimal results and/or overfitting. At the same time, the creator of the algorithms can enhance the learning capacity of the machine through the choice of its hyperparameters. Hyperparameters are parameters that are not internal to the machine's training process and need to be determined separately. Scholars can refer to different statistical methods to find the best values for the hyperparameters, a process commonly referred to as *tuning the hyperparameters*, or derive them from previous literature.

One of the main goals of tuning the hyperparameters of ML models is to find the optimal number of predictors and hyperparameter values to obtain the optimal combination that will reduce the bias and variance of the chosen model (James et al., 2013). Most of the time, especially with cross-sectional data, tuning the hyperparameters is achieved using K-fold cross-validation. In K-fold cross-validation, the training data are split into multiple random k validation sets to find the optimal values of the hyperparameters. Once the optimal values of the hyperparameters are found, a model can be applied over a test set and be evaluated

using different possible evaluation measurements (misclassification error[13], root mean squared error[14], ROC curve[15], F1-score[16], and others).

Machine learning methods can be divided into supervised machine learning and unsupervised machine learning methods (Hastie et al., 2009). The general idea of supervised machine learning methods is to derive the importance of a certain amount of independent variables to predict or detect a specific continuous or discrete dependent variable. Supervised machine learning can be divided into regression and classification methods. The first aims to find relationships between a continuous dependent variable and its predictors, while the second wants to identify/classify a discrete dependent variable given determined predictors. Unsupervised ML can find hidden patterns and connections in the data and are more used when performing classification exercises. Clustering, principal component analysis and their derivations are possible examples of unsupervised machine learning methods.

Machine learning methods can be derived from concepts related to statistical learning theory (James et al., 2013) and their subsequent applications and developments. Statistical learning deals with using data to make predictions in a way that the algorithms are able to learn by themselves how to improve their predictions. Some of the leading models are in common with econometrics, such as principal component analysis or clustering methods. Other methods include derivations of standard statistical methods, such as linear regressions. Examples are shrinkage methods, which are fundamentally regressions using penalization hyperparameters. Lasso (least absolute shrinkage and selection operator) and ridge regressions are part of this type of ML. These models are explained in more detail in the subsection 2.1.1 of the thesis.

Another important family of ML methods are the tree-based ones, shown in the subsection 2.1.2 and the most famous ensemble methods[17] based on regression trees, described in the subsection 2.1.3, such as random forest and boosting. The trees (regression or classification trees) are ML methods able to partition the data by splitting it into different branches. To increase their performance, scholars

[13] The percentage of incorrect predictions over the total number of predictions in a classification exercise (James et al., 2013).

[14] The root mean square error (RMSE) is one of the most used measurements to evaluate accuracy in forecasting exercises (Hyndman and Koehler, 2006).

[15] The acronym means “receiver operating characteristic”. The ROC curve (Pepe, 2000) is a graphical plot that shows the ability of a classification method to detect a binary dependent variable. The threshold can have a different value and it depends on the probability of correct predictions against the probability of false alarms.

[16] F1-score is an accuracy test usually used in case the dependent variable of the classification method is constituted of imbalanced data (Jeni et al., 2013). The test is based of the harmonic mean of precision and recall, where precision is the number of true positives divided by the number of all positive results and recall is the number of true positives divided the sum of the true positive results and the false negative ones.

[17] The ensemble methods are a combination of multiple simpler models to obtain better predictions. See James et al. (2013).

use random forest or boosting trees algorithms. The former creates a high number of trees through bootstrapping and randomly selects some of the predictors to be used to construct the different trees.

The boosting trees group (gradient boosting, AdaBoost, XGBoost, boosting component-wise) differs from the random forest, especially in one aspect – the information constructed by the previously created trees is internalised in the new trees. In this manner, they become more sophisticated and precise in their prediction and classification capacities. More on the different boosting models is explained in subsection 2.1.3.

During the last decades, an exponential growth in interest in machine learning methods from economists in different fields has been witnessed (see Figure 3). Studies try to understand when and how machine learning methods can be used to extract relevant information from Big Data or try to solve identification problems in different ways from non-traditional data sources. This section illustrates how machine learning methods can be useful in related studies by economists and econometricians.

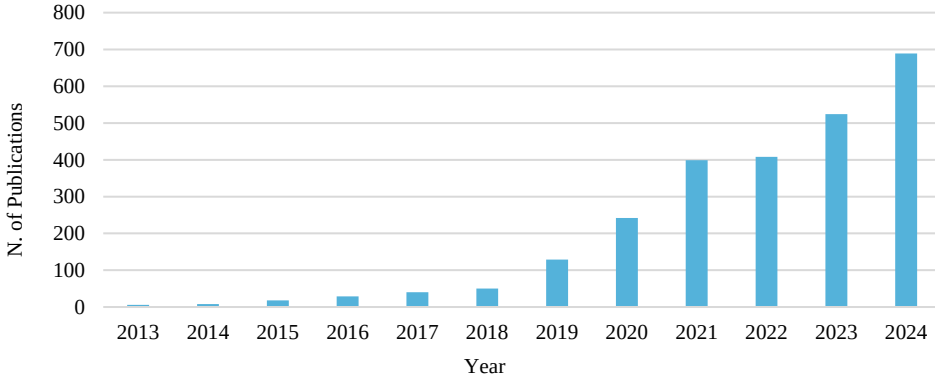

Figure 3. Number of research articles in economics including machine learning issues from 2013 to 2024 in the Web of Science

Notes: Source Web of Science, topics "economics"

For instance, machine learning methods have garnered attention from economists and econometricians due to their ability to address complexities associated with structured and unstructured Big Data. The *curse of dimensionality* is one of the main issues where the use of Big Data becomes challenging. The curse states that the more the model becomes complex due to the number of predictors and parameters involved, the worse the model is due to overfitting (James et al., 2013). This concept is connected with the bias-variance trade-off. If the model's complexity increases, the model's bias decreases and the variance increases. On the other hand, if the model's complexity decreases, the bias increases and the variance decreases.

Lasso and its derivations (i.e. post-lasso[18] or adalasso[19]) receive attention from economic and econometric scholars, given their potential use as selection models (Chernozhukov et al., 2015). Moreover, Chernozhukov et al. (2018) use the studies over debiased-lasso (Belloni et al, 2013; Belloni et al., 2014a; Belloni et al., 2014b) to derive their method of double (or debiased)[20] machine learning methods to obtain causal inference with cross-sectional data using different machine learning methods (lasso, random forest, boosting, deep neural nets). Most of the ensemble methods using trees are used by economists and econometricians either in combination with econometric methods or only for forecasting exercises.

One of the few examples in economic literature of using ML methods for causal inference is the study by Wager and Athey (2018). The authors use a combination of heterogeneous treatment effects and random forest to derive a "causal forest" methodology. The same authors applied the new method to the National Study of Learning Mindsets to derive essential insights into mindset intervention over student achievements and methodological novelties (Athey and Wager, 2019). Most of these studies rely on cross-sectional data for their analysis, and they are still limited in number in comparison to the ones related to mere predictions. Studies on machine learning methods for causal inference in time series are still in their infancy but deserve attention for future research (Joseph, 2020; Bluwstein et al., 2020).

A possible example of a forecasting exercise using ensemble methods can be found in Ng (2014). The author uses AdaBoost to predict economic recessions in the USA, considering 1,500 potentially relevant predictors composed of 132 real and financial time series and their lags. AdaBoost performs well in predicting a recession in a Big Fat dataset. Moreover, the Aaa spread is the most relevant predictor of recessions in the short (3 months) and medium horizon (6 months). The risky bond and 5-year spreads are the best predictors for the long horizon (12 months). More details on AdaBoost are illustrated in the subsection 2.2.1.

More intricate machine learning methods, particularly deep learning techniques, continue to be prominently employed in forecasting exercises (Zheng et al., 2023). These methods, however, necessitate extensive datasets to yield optimal results. Moreover, their inherent complexity often results in the *black box* issue, where the inner workings of the models are challenging to interpret (Varian, 2014). As a consequence, the relevance of these methods within the domain of economic research extends predominantly to scholars engaged in forecasting or nowcasting exercises, rather than those pursuing broader economic inquiries.

In macroeconomics and macroeconometrics, the structured data analysed are either very high-frequency financial data or Big Fat dataset/high dimensional

[18] Post-lasso is a method consisting in running an OLS with only the variables selected by lasso (Belloni and Chernozhukov, 2013).

[19] In adalasso, weights are applied to each parameter related to the lasso's constraint (Zou, 2006).

[20] This indicates that the method applied reduces the bias error that can emerge using ML methods on causal inference-related exercises.

data/data-rich environment,[21] given the lack of many observations of real macroeconomic variables. In some studies, the number of predictors is augmented by scholars by adding lagged variables (Ng, 2014) or using polynomial transformations. Regarding unstructured Big Data, the most frequent currently employed is text data. These data types need to be treated through specific machine-learning methods related to text mining analysis.

This thesis uses mainly structured data in the presence of a large number of predictors. As explained by Bok et al. (2018), handling such large and complex data is not entirely new for economists, especially those engaged in real-time and massive data flows. Reichlin (2015) stresses how large macro series are affected by a serious collinearity issue. This does not allow us to obtain differentiated results nor to decide what useful methods are.

The standard OLS method is not optimal or even feasible (Kapetanios et al., 2016). Indeed, most of the studies using Big Data focus on forecasting and nowcasting exercises rather than on finding causal relationships.[22] Scholars in economics have applied machine learning methods, such as lasso (Tibshirani, 1996) or random forest, advanced econometric tools, like Bayesian vector autoregressions (BVAR), and compared them, as for example, in Plakandaras et al. (2015).

Lim and Zohren (2021) and Torres et al. (2021) explore how exceptionally complex deep learning methods can be used for time series forecasting and their current limits in terms of interpretability and computationally-expensive estimations. Goulet Coulombe et al. (2022) illustrate how machine learning methods can be useful for forecasting and in which environments. The authors stress how the presence of data-rich environments and nonlinearities can be relevant for determining the performances of the models. Forecasting mixed-frequency time series[23] (Babii et al., 2022) studies in high-dimensional spaces investigate the potential that machine learning methods offer.

1.3 Machine Learning Methods in Macroeconomics

In macroeconomics compared to microeconomics, interest in machine learning methods is more recent. This is related to the larger amount of data that microeconomists can usually access relative to macroeconomists. In addition, the massive use of more classical models handling large amounts of data, such as factor models in macroeconomics, have slowed down the growth of the number of studies using

[21] The last term is used by some authors more to stress the diversity of data exploited together such as data derived from text analysis together with standard macroeconomic indicators or adding different data in mixed-frequency exercises (Babii et al., 2022).

[22] Athey and Imbens (2015) is a good example of supervised machine learning technique (i.e. regression trees) with proper modifications that tried to obtain better predictions and evaluate causal effects. More recently, Chernozhukov et al. (2018) applied a new methodology called double machine learning method to learn the coefficient of an endogenous variable in a partially linear instrumental variables model.

[23] Forecasting future low-frequency data adding current and lagged high-frequency predictors.

ML methods. These models have been so successful that the importance of ML methods have been underestimated until the last decade. However, machine learning models allow new interpretations of the data and keep information that can be lost when using factor models. Furthermore, ML methods permit handling nonlinearities quite well. In this section, studies related to the use of machine learning in macroeconomics are discussed.

Analysing nonlinearities can be relevant in a more and more complex macroeconomic environment. Franses and Van Dijk (2005), in their work on 18 OECD countries, stress the importance of nonlinearity issues when forecasting industrial production at long-term horizons. However, using nonlinear models does not show improvements in forecasting short-term out-of-sample in comparison with simple linear models.

Indeed, machine learning methods seem to perform better than factor models, especially for long-term horizons.[24] Franses and Van Dijk (2005) also point out the importance of possible transformations of the variables involved. Nonlinearities and the increase of the number of variables considered in the model due to the possible transformations cause scholars to consider the possibilities offered by boosting and other machine learning methods described in Chapter 2 of the thesis. Döpke et al. (2017) used boosting regression trees, more specifically the AdaBoost method (see subsection 2.2.1), to analyse indicators that can predict recessions using German data.

Econometric methods retain their capacity to forecast leading variables, such as industrial production, even in the presence of many indicators. Still, methods such as factor models can be difficult to interpret, and furthermore, they can be integrated with new methods such as machine learning methods. Scholars illustrate how ML methods perform against or in combination with factor models. Li and Chen (2014) compare and combine dynamic factor models and lasso-based models to forecast different US macroeconomic indicators. Lasso-based models perform better than dynamic factor models in the majority of the forecasted indicators, industrial production among them.

Heij et al. (2011) forecast the predictive ability of the Conference Board's Composite Leading Index (CLI) to forecast US industrial production growth rates from 1989 to 2009. However, using simple principal components provides more accurate forecasts of industrial production than CLI. Kock and Teräsvirta (2014) study macroeconomic indicators during the Great Recession in G7 countries plus four Scandinavian countries. They find that their neural network models beat the linear AR model 12 periods ahead in forecasts of US industrial production.

The regression tree methodologies, described in the subsections 2.1.2 and 2.1.3, are useful for handling nonlinear relationships and outliers (Döpke et al., 2017). Moreover, the authors remark on how boosting can resolve the problem of the high variance of the parameter estimated in simple regression trees. Ng (2014) uses boosting to predict recession with US data using more than 1,500

[24] However, some studies like the one of Robinzonov et al. (2012) find good performance using boosting for German industrial production growth rates also for short and medium term.

possible predictors derived from 132 economic and financial variables. Buchen and Wohlrabe (2011) examine the performance of the component wise boosting method in forecasting US industrial production by comparing it with dynamic factor models. Boosting seems to perform better in the long run. The authors suggest using cross-validation to compare the models instead of normal information criteria like AIC (Akaike information criterion) or BIC (Bayesian information criterion).[25]

The utility of boosting methods for making predictions is also observed by Lehmann and Wohlrabe (2015). The authors use component wise boosting with 253 monthly indicators from German states to forecast regional GDP. They find that boosting is a viable method for forecasting regional GDP, especially one and two quarters ahead. Lehmann and Wohlrabe (2016) exploit boosting and 175 monthly indicators to forecast German industrial production for four different forecasting horizons, showing how boosting can also be useful when forecasting different macroeconomic indicators apart from GDP.

Bai and Ng (2009) use boosting to select predictors in factor-augmented regressions for forecasting four key economic variables. They propose a stopping rule to avoid overfitting problems with the estimated predictors. They consider two approaches for handling the lags: component wise and block wise. Carriero et al. (2011) imply what they consider the three most important alternatives to forecast large datasets: factor model, bayesian VAR and multivariate boosting. They forecast three key economic variables, including industrial production growth, inflation and the federal funds rate. They find that the combination of shrinkage and reduction rank methods substantially improves the accuracy of the predictions in the case of both univariate and multivariate forecasting.

1.4 Literature Overview of Forecasting Business Cycles and Financial Crises with Machine Learning Methods

This section presents a systematic literature review of studies on forecasting business cycles and financial crises, focusing on papers that use machine learning methods. The systematic literature review includes articles published from 2015 to the present and follows the approach used by of Nazareth and Reddy (2023) to conducting a systematic literature review. The repositories used are Scopus and Science Direct. The selection of articles from the two repositories follows these rules:

- For Science Direct: Only review and research articles are selected, with English as the chosen language; only research conducted from 2015 to 2024 is considered. The research areas included “economics, econometrics, and finance,” “business, management, and accounting,” “social sciences,” and “computer science.”For Scopus: Only review and research articles in English are chosen,

[25] In contrast, Medeiros et al. (2016) affirm that BIC can be useful for the choice of the tuning parameters, such as the shrinkage parameters for shrinkage models.

covering research conducted from 2015 to 2024. The research areas are restricted to "economics, econometrics, and finance," "business, management, and accounting," "social sciences," and "computer science."

The following keywords with boolean operators are introduced in the two search systems of the repositories:

- For searching articles concerning business cycle and machine learning studies: "machine learning" OR "deep learning" OR "neural networks" OR "support vector machine" OR "LSTM"[26] OR "regression tree" OR "random forest" OR "boosting method" OR "boosting algorithm" AND "business cycle" OR "GDP growth" OR "industrial production" OR "recession. "
- For searching articles concerning financial crises and machine learning studies: "machine learning" OR "deep learning" OR "neural networks" OR "support vector machine" OR "LSTM" OR "decision tree" OR "random Forest" OR "boosting method" OR "boosting algorithm" AND "financial crisis. "

Due to the limit to the maximum number of boolean operators available in each search system, multiple search executions are necessary. The search systems provided access to several thousands of articles. A further selection is performed to exclude articles that are clearly unrelated to the aim of the thesis. For example, studies concerning only stock crashes or inflation are immediately discarded. As a result, in total 461 articles from Science Direct and 106 articles from Scopus were downloaded. Subsequently, the abstracts of these articles are analysed, and a further selection resulted in 29 articles about machine learning and business cycles from Scopus, 24 articles about machine learning and business cycles from Science Direct, 25 articles about machine learning and financial crises from Scopus, and 12 articles about machine learning and financial crises from Science Direct.

Further analysis of the articles excludes those using classification methods to forecast the business cycle, papers only related to neural networks or deep learning methods, and studies analysing different crises than systemic financial crises, like currency crises, stock crises, or sovereign debt crises. Microeconomic applications related to predicting bank defaults or credit scoring are also excluded. The final papers selected are shown in Table 2 and 3. Table 2 covers the papers on machine learning regression methods and the analysis of the business cycle.

Table 3 lists papers on machine learning classification methods and financial crises. It can be observed how the studies on business cycles are more concentrated in specific journals than those on financial crises. Indeed, 44% of the papers on business cycles are from the International Journal of Forecasting, while the Journal of Financial Stability, the journal where studies on machine learning and systemic financial crises are mostly published, accounts for 27% of the total number of articles published on the topic.

[26] Long Short-Term Memory (LSTM) network. This is a recurrent neural network that has acquired a certain relevance in computer science in the last decade and for this reason has been included by Nazareth and Reddy (2023) in their search keywords.

Table 2. Studies with ML applications to forecast business cycle

Authors	Title	Journal	Data	Models	Time period
Exterkate et al. (2016)	Nonlinear Forecasting with Many Predictors using Kernel Ridge Regression	International Journal of Forecasting	132 US macroeconomic and financial variables from Stock and Watson (2005)	Kernel ridge regression, principal component regressions, random walk, AR(1), self-exciting threshold autoregressive model (SETAR)	January 1959 – January 2010
Cepni et al. (2019)	Nowcasting and Forecasting GDP in Emerging Markets using Global Financial and Macroeconomic Diffusion Indexes	International Journal of Forecasting	Sets of economic indicators, consisting of 103, 103, 117, 110 and 88 economic series for Turkey, Brazil, Mexico, South Africa, and Indonesia respectively. Various survey variables, such as the Markit PMI survey, are also included in the dataset. Data from Bloomberg	Sparse principal component analysis (SPCA), the elastic net, the least absolute shrinkage operator, and least angle regression with latent global macroeconomic and financial factors (diffusion indexes) in a dynamic factor model (DFM), linear regression	January 2003 – June 2018
Panagiotelis et. al. (2019)	Macroeconomic Forecasting for Australia using a Large Number of Predictors	International Journal of Forecasting	Australian dataset containing observations on 151 aggregate and disaggregate economic series as well as 185 international variables from the Australian Bureau of Statistics (ABS) and the Reserve Bank of Australia (RBA)	DFM, ridge, least angular regression (LAR), bagging least angular regression (BLAR), bayesian VAR (BVAR), AR	Q4 1984 – Q2 2015
Maehashi and Shintani (2020)	Macroeconomic Forecasting using Factor Models and Machine Learning: an Application to Japan	Journal of the Japanese and International Economies	219 Japanese macroeconomic time series	AR, lasso, ridge, elastic net, bagging, random forests, boosting, feedforward neural networks, convolutional neural networks, recurrent neural networks	January 1973 – June 2018
Soybilgen and Yazgan (2021)	Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors	Computational Economics	FRED-MD and ALFRED dataset	Bagging, random forests, and stochastic gradient tree boosting, dynamic factor models, random walk.	Q2 2000 – Q4 2018

Authors	Title	Journal	Data	Models	Time period
Syed and Lee (2021)	Macroeconomic Fore-casting for Pakistan in a Data-rich Environment	Applied Economics	161 Pakistani macroeconomic time series	Dynamic factor model (DFM), ridge regression, the lasso, the elastic net and different models of bagging, naïve mean model, and the AR model	July 2007 – July 2017
Yousuf and Ng (2021)	Boosting High Dimensional Predictive Regressions with Time Varying Parameters	Journal of Econometrics	FRED-MD	Two L2 boosting algorithms. The first method uses component wise local constant estimators as base learner, while the second relies on component wise local linear estimators.	January 1960 – August 2018
Kauppi and Virtanen (2021)	Boosting Nonlinear Predictability of Macroeconomic Time Series	International Journal of Forecasting	FRED-MD	Two boosting models vs linear models	January 1959 – December 2016
Goulet Coulombe et al. (2021a)	Macroeconomic Data Transformations Matter	International Journal of Forecasting	FRED-MD	Adalasso, elnet, boosting linear, boosting trees, random forest	January 1960 – December 2017
Richardson et al. (2021)	Nowcasting GDP using Machine-Learning Algorithms: A Real-time Assessment	International Journal of Forecasting	41 real-time vintages from the Reserve Bank of New Zealand. Total amount series between 532 and 634 per vintage	Gradient boosting, factor model, shrinkage models, neural network, AR	Q1 2009 – Q1 2019
Mogliani and Simoni (2021)	Bayesian MIDAS Penalised Regressions: Estimation, Selection, and Prediction	Journal of Econometrics	FRED-MD and a set of weekly and daily financial data coming from different sources.	Group lasso, bayesian estimations, oracle B-MIDAS, autoregression, random walk	Q1 1980 – Q4 2017
Jardet and Meunier (2022)	Nowcasting World GDP Growth with High-frequency Data	Journal of Forecasting	151 monthly variables and 39 weekly series out of 500 monthly and 250 weekly potential regressors, different national and international datasets used.	Factor-augmented mixed data sampling (FA-MIDAS), with and without pre-selection of the variables versus lasso-MIDAS	Q1 2001 – Q1 2021

Authors	Title	Journal	Data	Models	Time period
Chu and Qureshi (2022)	Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast U.S. GDP Growth	Computational Economics	225 variables from FRED-QD and other sources	Lasso, ridge regression, XGBoost, Gradient Boosting, Random Forest, deep neural networks and the long short-term memory model (LSTM)	Q1 1959 – Q4 2020
Babii et al. (2022)	Machine Learning Time Series Regressions With an Application to Nowcasting	Journal of Business & Economic Statistics	34 series used in the Federal Reserve Bank of New York nowcast model, discarding two series; 8 financial time series; and 76 news attention series.	Sparse-group lasso, lasso, AR(1)	Q1 2002 – Q2 2017
Zhang et al. (2023)	Nowcasting Chinese GDP in a Data-rich Environment: Lessons from Machine Learning Algorithms	Economic Modelling	89 macroeconomic variables from CEIC data manager	Ridge regression, lasso, elnet, feed-forward, neural network, k-nearest neighbour, classification tree, macroeconomic random forest (Goulet Coulombe, 2020), two-stage ridge regression, mixed data sampling regression (MIDAS) – Ghysels et al. (2016), mixed data sampling regression with AR –Ghysels et al. (2016), autoregression, different factor models	Q1 1995 – Q4 2020
Borup et al. (2023)	Targeting Predictors in Random Forest Regression	International Journal of Forecasting	FRED-MD	Random forest with or without target predictors	January 1970 – December 2018
Guerróon-Quintana and Zhong (2023)	Macroeconomic Forecasting in Times of Crises	Journal of Applied Econometrics	13 monthly US macroeconomic from FRED-MD and financial data series	Nonparametric nearest neighbour (NN) methods; ARMA, Markov-switching autoregression model, self-exciting threshold autoregression (SETAR) model	January 1959 – June 2015
Lima et al. (2020)	Quantile Forecasting with Mixed-frequency Data	International Journal of Forecasting	132 macroeconomic series Jurado et al. (2015), for ADS index (Aruoba et al., 2009)	Quantile combination approach (QCA) of Lima and Meng (2017) in situations with mixed-frequency data, AR(4) with factors and or lasso/elnet.	January 1963 – September 2017

Table 3. Studies with ML applications and financial crises

Authors	Title	Journal	Data	Models	Time period
Ward (2017)	Spotting the Danger Zone: Forecasting Financial Crises with Classification Tree Ensembles and Many Predictors	Journal of Applied Econometrics	Jordà et., (2017) – old versions till 2011, used as panel data. Furthermore, two post-1970 samples (annual and quarterly) with 162 countries from Laeven and Valencia (2013) a previous version of Laeven and Valencia (2020)	Random forest, bagging, simple classification trees, logit	1870 – 2011
Papadopoulos et al. (2019)	Systemic Early Warning Systems for EU14 Based on the 2008 Crisis: Proposed Estimation And Model Assessment for Classification Forecasting	Journal of Banking Regulation	14 EU countries; dataset composed by the authors using different sources.	Linear regression, logistic regression, probit, discriminant analysis, classification tree, random forest and the C5.0 algorithm	Q1 2001 – Q1 2014
Alessi and Detken (2018)	Identifying Excessive Credit Growth and Leverage	Journal of Financial Stability	28 EU countries, from a project of the European Systemic Risk Board described and used in Detken et al. (2014)	Random forest and logit	Q1 1970 – Q4 2013
Beutel et al. (2019)	Does Machine Learning Help us Predict Banking Crises?	Journal of Financial Stability	Dataset composed by the authors from various sources (i.e. dataset from European System of Central Banks and the European Systemic Risk Board).	Trees, knn, svm, neural network, random forest, logit	1970 – 2016 (quarterly data)
Liu et al. (2022)	Predicting Financial Crises with Machine Learning Methods	Journal of Forecasting	Laeven and Valencia (2018) and older version of Laeven and Valencia (2020)	Lasso, logit, knn, svm, artificial neural network, classification tree, gradient boosting, AdaBoost, random forest. Shaply value employed.	1970 – 2017

Authors	Title	Journal	Data	Models	Time period
Tölö (2020)	Predicting Systemic Financial Crises with Recurrent Neural Networks	Journal of Financial Stability	Jordà et al., (2017)	Long-short term memory (LSTM) and the gated recurrent unit (RNN-GRU) neural nets, logit	1870 – 2016
du Plessis (2022)	Multinomial Modelling Methods: Predicting Four Decades of International Banking Crises	Economic Systems	Sample of 49 countries. Data from different sources	artificial neural networks, gradient boost, k-nearest neighbours, random forest, multinominal logit, probit	Q1 1971 – Q4 2017
Casabianca et al. (2022)	A Machine Learning Approach to Rank the Determinants of Banking Crises over Time and Across Countries	Journal of International Money and Finance	100 countries (emerging and advanced ones). Dataset composed by the authors	AdaBoost, logit	1970 – 2017 (yearly data)
Lin and Jin (2023)	Does ESG Predict Systemic Banking Crises? A Computational Economics Model of Early Warning Systems with Interpretable Multi-Variable LSTM based on Mixture Attention	Mathematics	Laeven and Valencia (2020), focus on G20 members excluding the European Union	Logistic, gaussian naïve bayes, classification tree, SVM, random forest, multi-layer perceptron, a specific LSTM model	1993 – 2017
Liu (2023)	Towards Better Banking Crisis Prediction: Could an Automatic Variable Selection Process Improve the Performance?	Economic Record	23 OECD countries, dataset mostly based on Laeven and Valencia (2018) and older version of Laeven and Valencia (2020)	Lasso and logit	Q1 1970 – Q3 2018
Torky et al. (2023)	Explainable AI Model for Recognizing Financial Crisis Roots Based on Pigeon Optimization and Gradient Boosting Model	International Journal of Computational Intelligence Systems	Jordà et al. (2017)	Gradient boosting classifier in the pigeon inspired optimiser (PIO) algorithm, random forest	1870 – 2016

Concerning the analysis of business cycles with machine learning techniques, certain recurring themes are prevalent in the literature. These themes revolve around how machine learning can be effectively utilised to forecast business cycles across different countries and the conditions under which machine learning methods excel.

Most of the studies focus on the following key themes:

- Comparison among machine learning models and/or univariate benchmarks (Exterkate et al., 2016; Maehashi and Shintani, 2020; Syed and Lee, 2021; Yousuf and Ng, 2021; Zhang et al., 2023), for example, autoregression or random walk.
- Comparison with factor models (Cepni et al., 2019; Soybilgen and Yazgan, 2021; Syed and Lee, 2021).
- Importance of the combination of factor models and machine learning methods (Panagiotelis et. al., 2019; Maehashi and Shintani, 2020; Goulet Coulombe et al., 2021a).
- Relevance of using many predictors and high-dimensional spaces, with or without mixed-frequency data (Panagiotelis et. al., 2019; Mogliani and Simoni, 2021; Babii et al., 2022).
- Pre-selection of predictors vs. automatic selection through algorithms (Cepni et al., 2019; Jardet and Meunier, 2022; Borup et al., 2023).
- Exploring the presence of nonlinearities (Maehashi and Shintani, 2020; Kauppi and Virtanen, 2021; Goulet Coulombe et al., 2021a).
- Understanding variations in results when considering sub-samples (Panagiotelis et al., 2019; Soybilgen and Yazgan, 2021; Zhang et al., 2023; Guerrón-Quintana and Zhong, 2023).

Moreover, even though studies utilise different datasets, it is evident that the FRED-MD, along with its quarterly version, FRED-QD, are the predominantly used datasets (Kauppi and Virtanen, 2021; Goulet Coulombe et al., 2021a; Mogliani and Simoni, 2021; Borup et al., 2023).

To dig into the analysis of U.S. business cycles, notable research by Exterkate et al. (2016) showcases the efficacy of kernel ridge regression in predicting U.S. monthly macroeconomic time series spanning from 1970 to 2009. Their comparison of kernel ridge with standard methods, including principal components and benchmarks such as random walk and autoregression, highlights the former's superior handling of nonlinearities. However, statistical significance in the results is often lacking in the predicted series.

In a broader examination of machine learning performance across diverse data environments, Guerrón-Quintana and Zhong (2023) demonstrate the superiority of a nonparametric nearest neighbour (NN) method over parametric counterparts and combination forecasts, particularly during the Great Recession. Their approach involves NN segmentation of data into clusters, notably excelling in forecasting macroeconomic time series, including US industrial production. It can be observed

how analyses of US macroeconomic indicators is continuing to be investigated with different methods and data environments in the last decade.

The literature explores how ML methods based on regression trees can be employed in forecasting US macroeconomic data. Soybilgen and Yazgan (2021) investigate the effectiveness of tree-based machine learning methods in forecasting U.S. GDP using over 100 macroeconomic and financial variables, revealing superior performance over factor models, albeit with some limitations during the Great Recession. Meanwhile, Borup et al. (2023) elucidate how targeting predictors enhances the forecast accuracy of random forest models, emphasizing the impact of weak predictors on model efficacy. These studies highlight a significant interest in tree-based methods for forecasting the US business cycle within the existing literature.

Taking into account the literature on ensemble methods and, in particular, boosting methods, Yousuf and Ng (2021) study the nuances of boosting methods by employing modified models to forecast US macroeconomic variables from 1971 to 2018, revealing varying performance across different periods. Kauppi and Virtanen (2021) similarly explore boosting models, highlighting their effectiveness in capturing nonlinear predictability in the FRED-MD dataset. These studies are examples of how boosting methods, as specific ensemble methods based on regression-trees, are currently under scrutiny by scholars. However, a comprehensive analysis of various boosting methods is still missing in the literature. Moreover, some boosting models, such as AdaBoostRT, are still not considered in macroeconomic forecasting studies.

Moreover, a comparison of different ML methods and benchmarks as well as experiments with different types of data can be observed in another strand of the literature. For example, Goulet Coulombe et al. (2021a) adopt a comprehensive approach integrating shrinkage, random forest and boosting models to forecast US macroeconomic time series, emphasizing the importance of complex transformations in enhancing forecast accuracy. Mogliani and Simoni (2021) expand the FRED-MD dataset with daily financial data, revealing limited additional information for short-term predictions but notable performance improvements using group-lasso models.

Chu and Qureshi (2022) explore the effectiveness of various machine learning models in forecasting US GDP growth, highlighting the superiority of ensemble methods over sparsity-based approaches for short-horizon forecasts. Babii et al. (2022) contrast sparse-group lasso[27] with traditional methods, illustrating the former's outperformance in estimating US GDP growth. Lima et al. (2020) showcase the efficacy of a quantile combination approach in forecasting industrial production, emphasizing the reduction of bias error without significant variance increase. The studies underscore ongoing efforts to address the bias-variance tradeoff in machine learning methods and how different kinds of machine learning

[27] Sparse-group is a combination of a standard lasso and group lasso. Group lasso is an extension of lasso capable of making a selection on groups of variables. Sparse-group lasso in this sense is able to provide a selection of variables between and within groups.

methods have to be considered for a proper comparison and evaluation of their performances.

Studies examining business cycles in various countries provide valuable insights into machine learning's performance beyond the specific US environment. Cepni et al. (2019) demonstrate superior results using machine learning methods combined with factor models in forecasting GDP, while Panagiotelis et al. (2019) find benchmark models outperform machine learning methods, particularly as model complexity increases after the Great Recession. Maehashi and Shintani (2020) reveal machine learning's superior forecasting accuracy over factor models and autoregression for Japanese macroeconomic variables.

Syed and Lee (2021) show the superior performance of machine learning methods in most cases, although traditional models perform better for short-term GDP growth[28] forecasts. Richardson et al. (2021) and Jardet and Meunier (2022) demonstrate the better performance of machine learning methods in nowcasting New Zealand GDP growth and world GDP, further illustrating the benefits of integrating domestic and international predictors. Zhang et al. (2023) highlight the effectiveness of machine learning methods in nowcasting real Chinese GDP growth, underscoring the crucial role of data characteristics in determining model suitability. They emphasize the need for tailored approaches based on data characteristics, with shrinkage models emerging as particularly successful.

The studies in Table 2 show that data from countries other than the US have also been tested. While the findings of these studies are valuable for methodological comparisons, it is notable that US data are the most frequently used, enabling evaluations based on a consistent data framework.

A number of recurrent themes emerge in the literature on using machine learning for business cycle predictions, with a primary emphasis on model comparison, data complexity, and the influence of the predictors used. Numerous studies compare various machine learning techniques to factor-based methods and standard benchmark models, such as autoregression models, demonstrating the benefits of ML in managing high-dimensional data fields and nonlinearity. Although approaches such as ensemble techniques, tree-based models, and shrinkage models have shown promising results, their performance frequently differs depending on the data environment. It can also be noticed how the most frequently analysed datasets are FRED-MD and FRED-QD because of the relevance of the US economy and the amount of macroeconomic data provided by this publicly available dataset.

As machine learning techniques prove useful for understanding and forecasting business cycles, researchers have also begun to investigate their potential for predicting financial crises. Nonetheless, investigating financial crises, in contrast to normal business cycles, includes special difficulties, like capturing unexpected market turbulence and spillover effects, which call for careful

[28] Using an industrial production index or index of large-scale manufacturing industries as proxies.

consideration of the machine learning techniques used, the sampling methods, and the data sources.

Studies focusing on financial crises reveal the mixed performance of machine learning methods compared to traditional models. This is due to the limited amount of these studies compared to those on business cycles. Alessi and Detken (2018) find random forest slightly outperforms logit in predicting systemic financial crises in EU countries, while Beutel et al. (2019) emphasize logit's superior stability over machine learning methods.

Liu et al. (2022) highlight the effectiveness of machine learning methods in early crisis prediction, particularly when using the Shapley value for indicator selection. However, evaluation criteria and data characteristics significantly influence results, as seen in Tölö (2020), Du Plessis (2022) and Casabianca et al. (2022).

Lin and Jin (2023) showcase the predictive capabilities of ESG ratings in forecasting financial crises, underscoring the importance of specific country aspects. Similarly, Liu (2023) emphasizes the significance of predictor selection and model simplicity in crisis prediction, highlighting lasso's effectiveness over logit.

Ward (2017) demonstrates that ensemble methods based on classification trees can significantly outperform both logit models and simple classification tree models. In particular, the author emphasizes the strong performance of random forests. However, in his study, the specific characteristics of individual countries are not accounted for, as the datasets are used in a panel data format to forecast the overall number of crises across the group of countries analysed.

Papadopoulos et al. (2019) report relatively good results for tree-based models in forecasting pre-crisis periods in 14 EU countries, utilising both macroeconomic and financial data – a common combination in the literature reviewed in this section. More recently, du Plessis (2022) examines a broader set of countries (49) using machine learning within a multinomial framework, allowing for the analysis of crisis, pre-crisis, and post-crisis periods across four decades of banking crises. The author highlights that boosting models have superior predictive power.

Finally, Torky et al. (2023) compare random forest and gradient boosting methods, revealing the latter's superior performance in predicting systemic crises. This study as well as Casabianca et al. (2022) shows the evolving landscape of ensemble methods and underscores the ongoing quest to optimise crisis prediction methodologies and the interest in the literature on boosting methods for the classification of systemic financial crises.

The literature on predicting financial crises using machine learning methods reflects both promises and challenges. Studies indicate that while machine learning models like random forest, gradient boosting, and general ensemble approaches can sometimes outperform traditional models, factors such as evaluation criteria, predictor selection, and country-specific characteristics play crucial roles in determining model effectiveness. The findings highlight a need for further refinement of ML models tailored to financial crises, especially given the

complexity and rarity of these events, and to focus more on studies on specific countries. Continued advancements in model accuracy, particularly through enhanced indicator selection, and ensemble methods, hold potential for more reliable crisis prediction frameworks. Intuitions related to the treatment of the data, such as those by du Plessis (2022), deserve careful consideration.

In conclusion, the application of machine learning in forecasting both business cycles and financial crises demonstrates considerable potential, although with varying degrees of success and unique challenges inherent to each domain. For business cycle prediction, machine learning techniques capitalise on their ability to handle high-dimensional datasets and capture nonlinear relationships, often outperforming traditional models under specific conditions. However, financial crises have been hard to predict, as the infrequency and structural complexity of crises complicate the design of models demonstrating high predictive accuracy.

2 OVERVIEW OF THE MAIN MACHINE LEARNING METHODS

In this chapter, the ML methods concerning regression and classification tasks are discussed. This chapter aims to clarify the specific characteristics and procedural steps of the various algorithms employed in the thesis. The choice of hyper-parameters and loss functions plays a significant role in achieving accurate forecasts (see subsection 1.4). Additionally, ascertaining similarities and differences between machine learning methods and traditional econometric techniques may help explaining the results obtained (see subsections 1.2 and 1.3).

Furthermore, shared features across machine learning models illustrate their potential usefulness in a broader context. Shrinkage methods such as ridge and lasso derivatives are based on penalization of the estimated coefficients. Furthermore, different models based on regression and classification trees are examined. The limitations of models based on simple trees can be overcome using different bootstrapping approaches (i.e. bagging, random forest, boosting). In particular, boosting methods are reviewed in detail. Boosting methods can vary in different ways, for example, by their loss functions or by how the branches of the regression trees are split. The main focus of this chapter is the varieties of boosting models and their possible application with Big Fat Data (see subsection 1.2).

2.1 Regression methods in Machine Learning

2.1.1 Shrinkage methods

This section summarises the regression methods from the ML used in this monograph. The first group of methods described is related to the shrinkage methods. The name of this group stems from their capacity to shrink the coefficient of the independent variables in a regression. The shrinkage allows us to rank different variables (ridge model) or even select them by dropping the less important ones (lasso model). These models can be very helpful in forecasting exercises with many predictors, especially when the number of predictors exceeds the number of observations. Shrinkage models allow better predictions due to more parsimonious regressions. The most common methods are ridge, lasso and elnet.

Ridge (Hoerl and Kennard, 1970) shrinks the coefficients of the predictors toward zero but does not force them to zero. It applies a penalty to the regressors, but it does not eliminate them from the models. The formula of the model can be described as:

$$\sum_{i=0}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = RSS + \lambda\sum_{j=1}^{p}\beta_j^2 \qquad (1),$$

where λ is the penalization parameter, y_i is the dependent variable of interest, x_{ij} refers to the possible covariates, β is the vector of the coefficients of independent

variables, RSS is the residual sum of squares, *i* is the *i*th observation, *j* is the j^{th} predictor, n is the total amount of observations, and *p* is the total amount of the predictors.

When λ is sufficiently large, it can force the coefficients of the regressors towards but not equal to 0. As can be observed, the formula is exactly the same as the standard OLS equation plus the penalization term. This method allows for the visualization of a ranking of independent variables based on the magnitudes of their coefficients. The researcher can take these results into account and select the variables that should be used to find correlation or causation in the subsequent steps of the analysis.

Lasso (Least Absolute Shrinkage and Selection Operator) *and its derivatives* are getting more attention compared to other the shrinkage models from applied economists and econometricians for shrinking the coefficients towards zero. Lasso was introduced in the seminal work of Tibshirani (1996). Lasso is a variable selection method where the choice of the penalization hyperparameter λ is essential for obtaining a meaningful result.

The $\widehat{\beta_\gamma^L}$ is a vector of lasso coefficients estimated after the shrinkage; it can be found minimizing the following estimation:

$$\sum_{i=0}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j| \qquad (2),$$

When the coefficients of the predictors are shrunk, the variance of the model is reduced with a small increase in bias. This means that the shrinkage improves the accuracy of the forecasting. It can be noticed how the main difference between ridge and lasso concerns the penalization method of the β coefficients.

Lasso is such a successful model that various derivatives have been created, such as adalasso (Adaptive Least Absolute Shrinkage and Selection Operator), post-lasso, double-lasso, group lasso and so on.[29] The *adalasso* is an extension of the lasso method (Zou, 2006) and the principal difference with the standard lasso is the inclusion of additional information concerning the relevance of each variable that gives consistency to the model. From an analytical perspective, the adalasso includes adding a weighting parameter derived by a first-step model, which can be a lasso or a simple linear regression, to the standard lasso penalty. The adalasso penalty is given by:

$$\lambda\sum_{i=1}^{n} p(\beta_i;\lambda,\omega_i) := \lambda\sum_{j=1}^{n}\omega_i|\beta_i| \qquad (3),$$

The weighting parameter ω_i is derived by:

$$\omega_i = |\beta_i^*|^{-\gamma} \qquad (4),$$

[29] Hereby, only adalasso is described. Other methods are more related to cross-section data studies (Chernozhukov et al., 2018) or mixed-frequency data (Babii et al, 2022).

Where β_i^* is the set of coefficients of the first-step model and γ is a positive costant for the adjustment of the weight's vector.

This parameter adjusts the penalization for each coefficient. In this way, adalasso penalises lower estimates of the initial step more.

The last standard shrinkage technique applied in the study is the *elastic net* (Zou and Hastie, 2005) or elnet. The elastic net is a generalization that contains the lasso and ridge as special cases. The elastic net regularises the coefficients linearly, combining its ridge penalties and its lasso penalty. The hyper parametrization of elnet concerns the values of the λ penalty and α penalty mixing parameter. The value of α is important to determine the balance between lasso and ridge methods.[30] Lasso and ridge are contained as special cases of elnet because if α is equal to 1, only lasso is applied, and if it is zero, only ridge is applied. Elnet usually selects more variables than the lasso if the same value of λ for both of the variable selection methods is used. The elnet estimates $\hat{\beta}$ can be derived by:

$$\hat{\beta} = argmin\,(y - X\beta)^2 + \alpha\,\lambda \sum_{i=1}^{n} \beta_i^2 + (1-\alpha)\lambda \sum_{i=1}^{p} |\beta_i| \qquad (5),$$

where $\alpha \in [0, 1]$.

In the case of potentially correlated predictors, standard lasso regression typically retains only one of them while discarding the others. The inclusion of the ridge regularization component in the elnet framework is aimed at mitigating this concern. However, the use of a double shrinkage method like elnet can lead to an increase in coefficient bias and potentially result in less accurate predictions.

As can be seen, the shrinkage models are de facto ordinary least squared regression with specific penalization parameters. The choice of the methods depends on the final goal of the research. For example, a researcher who wants to focus on finding the best variable selection procedure can consider lasso as an adequate one[31]. However, lasso can have negative sides, such as discarding relevant variables. In this sense, other shrinkage models can offer a valid alternative. The interest in lasso on the part of economists should not imply that other shrinkage methods are not noteworthy.

[30] In this work the default value of α is kept equal to 0.5 given the presence of ridge and lasso to have a proper differentiation. However, α can variate between whatever value between 0 and 1. The decision depends on whether the lasso or ridge part of the penalization is considered more important or this can be driven by the data.

[31] The shrinkage models, particularly lasso, can aid in selecting the optimal set of variables by shrinking the estimated coefficients of predictors or even reducing them to zero, achieving this in a completely data-driven manner.

2.1.2 Regression trees

A framework that is distinct from shrinkage models involves regression trees, which is one of the most well-known methods in machine learning due to its ease of interpretability. Regression trees can handle more complex data without discarding variables or shrinking the estimated coefficients. The main idea of a regression tree is to divide the predicting space into different regions (left side of Figure 4). Graphically, the regression tree can be represented as a diagram resembling a tree (right side of Figure 4). The tree can be split into different *leaves* through splitting rules. The *branches* are the conjunctions between the different intermediate splits (internal nodes) in the trees. The final output of the trees are called *terminal nodes* or *leaves*. The splitting rule for regression trees can be defined by the researcher, but usually, the minimization of the residual sum of squares is the default rule.

The example by Hastie et al. (2009) suggested by Medeiros et al. (2021) can be helpful to better understand how a regression tree works. In this example presented in Figure 4 there are two predictors $X_1$and X_2 where each takes values within a specified interval. The area to the left (right) of $X_1 = z_1$ is split at $X_2 = z_2$ ($X_1 = z_3$). At $X_2 = z_4$, the area to the right of $X_1 = z_3$ is further divided. The end result is a partitioning of the set of predictors into five sections, as shown in right the plot in Figure 4: R_j, where $j = 1,\ldots,5$. In each region R_j, the model predicts the dependent variable with a constant that can be estimated by the sample average of the realisations of the dependent variable inside region R_j.

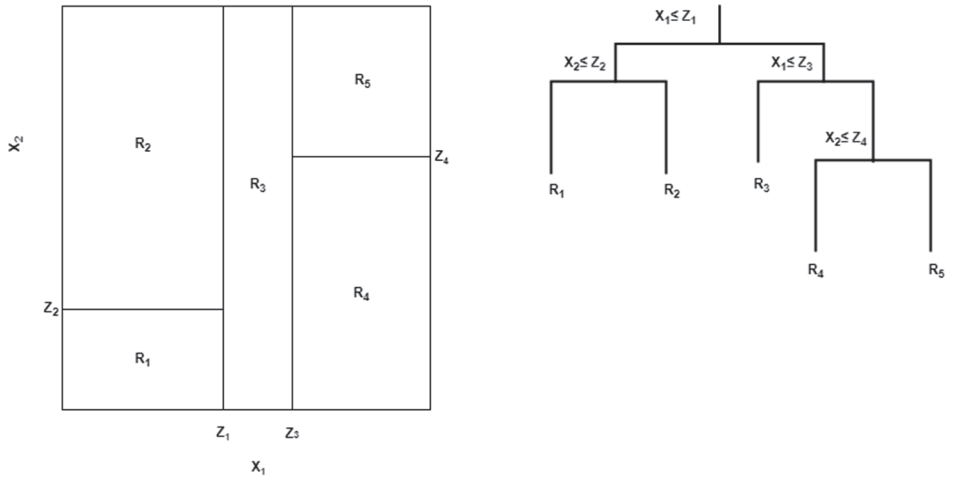

Figure 4: Example of regression trees

Source: reproduction of Medeiros et al. (2021). In this example, there are two explanatory variables, X_1 and X_2, and 5 regions.

The aim is to find regions that minimise the residual sum of squares:

$$\text{RSS} = \sum_{J=1}^{J} \sum_{i \in R_j} \left(y_i - \hat{y}_{R_j}\right)^2 \quad (6),$$

where $\hat{y}_{R_j}$ is the mean estimated response for the training observations within the j^{th} box. The first split is derived with the first predictor X_J selected, and it divides the predictor spaces into two regions in the following way:

$$\{X|X_J < s\} \text{ and } \{X|X_J \geq s\} \quad (7),$$

The variable s is the cut point and splits the predictor region to cause the minimization of the RSS.[32] The process continues to look for a new best predictor that minimises the RSS and splits the newly created regions into the other new regions until the process reaches a stopping limit. This happens when a further split does not provide any extra information.

Regression trees can better predict nonlinear relationships (James et al., 2013). Unfortunately, simple regression trees are not always accurate in their prediction capacities, especially in the presence of too many predictors because a more complex tree leads to a high variance of the results and overfitting issues. Contrary, a very complex tree loses its interpretability. For these reasons, the literature exploits bootstrap processes to create ensembles of trees (James et al., 2013; Breiman, 2001).

2.1.3 More trees to avoid overfitting: ensemble methods

Ensemble methods are algorithms that combine simpler models, sometimes called weak learners, to obtain better predictions, and the methods are mostly used together with regression trees (Breiman et al., 1984). As mentioned in the previous subsection, the bootstrapping processes in the following algorithms help to increase the accuracy of trees-based models. All these models are connected in the continuous struggle to reduce overfitting and, at the same time, trying to keep a certain level of interpretability.

The model called *bagging*, also called bootstrap aggregation, is a procedure for reducing the high variance related to the standard regression tree, and its basic algorithm can be found in Breiman (1996). This method takes different samples from a training dataset (James et al., 2013). The result is the average of the predictions provided by each tree from the different bootstrapped training sets. The model for bagging can be presented as follows:

[32] The cut point is usually decided through a cross-validation process where each predictor has a sufficient amount of observations.

$$\hat{f}_{bag}(\mathrm{x}) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \quad (8),$$

where $\hat{f}^{*b}(x)$ is the prediction of the b^{th} bootstrap training set generated by a regression tree, and *B* is the total amount of bootstraps training set. The goal is to get $\hat{f}_{bag}(\mathrm{x})$, which is the average of the predictions of the trees for each bootstrap. The *random forest* (Breiman, 2001) is the best known ensemble of regression trees. The method creates bootstrap training samples but randomly selects m predictors among the total amount of predictors p. The split can use only one of those m predictors. The number of splits is usually equal to one third of the total number of predictors in the regression random forest such that $m = p/3$.

The *Boosting* model is an ensemble method similar to bagging and random forest, where the trees involved as weak learners (base models), exploit information from past regression trees. In this text, the following boosting models are described: gradient boosting, component wise boosting, AdaBoostR2, AdaBoostRT and XGBoost. Even if they are part of the same group of ensemble methods, they differ in various respects such as: parameter regularization, loss function, number of hyperparameters to tune, as well as the mechanisms for splitting the trees.

In the *gradient boosting* ensemble method (James et al., 2013; Friedman, 2001), the trees use information from previous regression trees. First, it has to set a weak learner, such as $\hat{f}(x) = 0$ and the residuals $r_i = y_i$ for all i observations in the training set. Subsequently, over all trees b, where $b = 1,2,\dots,B$, and considering *X* as the full set of predictors and *r* as the vector of the residuals, the following process has to be repeated:

- fitting a tree f^b in the training data (X, r) considering *d* splits,
- modifying $\hat{f}$ by adding to it the shrunken version of the created new tree: $\hat{f}(x) \leftarrow \hat{f}(x) + \eta\hat{f}^b(x)$,
- conversely, recalculate the residuals $r_i: r_i \leftarrow r_i - \eta\hat{f}^b(x)$, where η is a shrinkage parameter.

All these steps give rise to a stronger learner:[33] the boosted model. The final model can be described as follows:

$$\hat{f}(x) = \sum_{b=1}^{B}\eta^b\hat{f}^b(x) \quad (9),$$

where *B* is the total amount of bootstrap training sets, $\hat{f}^b(x)$ is the prediction provided by the b^{th} weak learner. In this way, the final prediction of the model $\hat{f}(x)$ improves, even if slowly, due to the increasing complexity of the ensemble. The shrinkage parameter η makes the process much slower, which facilitates

[33] Here the underlying general idea of the ensemble methods can be observed: using a certain amount of simpler ML methods (*weak learners*) in order to create a stronger machine able to perform better predictions (*strong learner*).

making the model more flexible. The parameter can have different values, depending on the shrinkage level the researcher aims to reach.

Another boosting model is the *component wise boosting* model suggested by Bai and Ng (2009). Following Bühlmann and Hothorn (2007),[34] the first step is taking a generalised OLS, whereby the component-wise boosting selects the best variables, including the lagged variables, in the following way:

$$\hat{S}^* = argmin_{1<j<p} \sum_{i=1}^{n} \left(r_i - \hat{\beta}^{(j)} X_i^j\right)^2 \quad (10),$$

where X_i^j is the matrix of predictors with *i* observations and *j* predictors, $\hat{\beta}^{(j)}$ is the vector of the estimated predictors' coefficients, r_i are the residuals, *n* is the total amount of observations and p is the total amount of the predictors. Finally, $\hat{S}^*$ represents the selected variable within the context of a linear model fitted through ordinary least squares (OLS).

For every iteration m, where $m = 1, \ldots, K$, the function $\hat{f}^{[m]}$ of generic variable x is updated linearly considering the new variable selected by boosting:

$$\hat{f}^{[m]}(x) = \hat{f}^{[m-1]}(x) + \nu \hat{\beta}^{\hat{S}_m} x^{S_m} \quad (11),$$

where ν is the step-length factor[35] comprised in the interval (0, 1] and $\hat{S}_m$ is the index of the predictor selected during every iteration. Hence, the final fitted model of the stronger learner is updated every time, considering its lag $\hat{f}^{[m-1]}(x)$ and a new selected variable x^{S_m}.

The component wise boosting applied in this study uses factors extracted using principal component analysis following Bai and Ng (2009) and Medeiros et al. (2021). The method by Medeiros et al. (2021) allows the selection of factors by boosting using information criteria. This specific method is also called "factor boosting" by Medeiros et al. (2021).

AdaBoost (Freund and Schapire, 1997; Freund et al., 1999; Friedman et al., 2000) is well-known as an ML classification method based on classification trees, but there also exist alternatives that use regression analysis. Drucker (1997) proposes a variation of the basic method – A*daBoostR2* – that can be used for more general regression problems.

Another AdaBoost model for regressions – *AdaBoostRT* – was suggested by Solomatine and Shrestha (2004). In this study, both AdaBoostR2 and AdaBoostRT models are employed. The following explanations, drawn from Drucker (1997) and Solomatine and Shrestha (2004), are used for each model.

---

[34] More information on the component wise boosting algorithm can be found in Bühlmann and Hothorn (2007).

[35] The step-length factor serves as a shrinkage parameter, and it can play a role in enhancing prediction accuracy. Bühlmann and Hothorn (2007) recommend that this factor should be relatively small to be effective, even though it is generally regarded as a parameter of lesser significance.

The estimations of AdaBoostR2 can be derived through a number of steps that can be summarised as follows:

- given a matrix of predictors *x* and a vector *y* of the dependent variable to be predicted, where *m* represents the number of iterations and *n* is the total number of observations within the sample, it is necessary to specify the number of iterations m,
- the initial uniform distribution D for each observation i is $D_m(\mathrm{i}) = \frac{1}{n}$,
- when the procedure is initialised, the average loss function, denoted as $\bar{L}_m$ is set equal to 0,
- subsequently, the weak learner with distribution D_m is iterated till the average loss function is $\bar{L}_m < 0.5$,
- after every iteration, the algorithm generates a hypothesis such as: $f_m(x) \rightarrow y$,
- the loss for each training sample l_m such as $l_m = |f_i(x_i) - y_i|$, can be calculated in three different functional forms: linear, exponential, and quadratic,
- the average loss is computed $\bar{L}_m = \sum_{i=1}^{n} L_m(\mathrm{i}) D_m(\mathrm{i})$,
- the betas are equal to $\beta_m = \frac{\bar{L}_m}{1-\bar{L}_m}$,
- the new distribution D_m is updated using the following formula:

$$D_{m+1}(i) = \frac{D_m(i)\beta_m^{(1-L_m(i))}}{Z_m} \quad (12),$$

  where Z_m is the normalization factor that ensures D_{m+1} remains a valid function,
- set $n = n + 1$,
- finally, AdaBoostR2 provides the following final hypothesis function $f_{fin}(x)$:

$$f_{fin}(x) = inf\left[\sum_{m:\, f_m(x) \le y}(\log\frac{1}{\beta_m}) \ge \frac{1}{2}\sum_m(\log\frac{1}{\beta_m})\right] \quad (13),$$

Instead, AdaBoostRT has the following steps that have important differences compared to AdaBoostR2:

- the first steps till the beginning of the iterations are the same as with AdaBoostRT,
- the weak learner with distribution D_m can be iterated till the average loss function reaches the maximum number of iterations M,
- after every iteration, the algorithm again generates a hypothesis such as: $f_m(x) \rightarrow y$,
- calculate the error rate, ε_m, of $f_m(x)$:

$$f_m(x): \varepsilon_m = \sum D_m(\mathrm{i})\ where\ \mathrm{i}: \left|\frac{f_m(x_i) - y_i}{y_i}\right| > \phi \quad (14),$$

where ϕ is a threshold used to differentiate between correct and incorrect forecasts,

- setting $\beta_m = \varepsilon_m^2$,
- subsequently, the distribution is updated as follows:

$$D_{m+1}(i) = \frac{D_m(i)}{Z_m} \times \begin{cases} \beta_m & if \ \left|\frac{f_{tm}(x_i) - y_i}{y_i}\right| \leq \phi \\ 1 & otherwise \end{cases} \quad (15),$$

- set $n = n + 1$,
- finally, the AdaBoostRT produces the following final hypothesis function $f_{fin}(x)$:

$$f_{fin}(x) = \sum_m log\left(\frac{1}{\beta_m}\right) f_m(x) / \sum_m log\left(\frac{1}{\beta_m}\right) \quad (16).$$

As can be seen, the variations of AdaBoost for regression differ in two principal respects between each other: how the error term is considered and how the weights of predictors in the new iterations are updated.

Extreme gradient boosting (XGBoost) is considered the most advanced boosting method at the time of writing the monograph (2024) because of its complexity and prediction capacities that can rival some deep learning algorithms. Its implementation using trees as weak learners can be found in Chen and Guesterin (2016). The most important innovations of this algorithm are as follows:

- high-scalability,
- better handling of sparse data,
- the use of a theoretically founded weighted quantile sketch procedure to weight the observations in an approximate tree learning algorithm that allows a better split of the trees,
- a better regularization process,
- a faster implementation through more efficient use of a parallelization system, see Chen and Guesterin (2016).

Here, the regularization process and the relation between the tree learning algorithm and the decision on the best-splits of the trees in XGBoost are briefly illustrated. The regularization process reduces the risk of overfitting, given the complexity of the trees. It can be summarised as:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (17),$$

where $L(\phi)$ is the loss function and l is a convex loss function which differentiates the prediction $\hat{y}_i$ and the actual observation y_i. $\Omega(f_k)$ is the penalization function for the complexity of the model for every independent tree structure f_k.

Every penalization function for each independent tree structure $\Omega\ (f_k)$ can be described as:

$$\Omega\ (f_k) = \gamma T + \frac{1}{2}\lambda\left\|\omega_j\right\|^2 \quad (18),$$

where T is the number of tree nodes, ω_j is the vector of nodes, and γ and λ are the parameters of the penalization function.

The weights of the leaf are calculated as:

$$\omega_j = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \quad (19),$$

$I_j = \{i|q(x_i) = j\}$ is the set of the indices of the data points for every j^{th} leaf where every index i is conditioned by $q(x_i)$, the directing function that assigns every observation to a particular leaf g and h are the first and second order of the loss function and, λ is the regularization parameter. This process allows us to find the best weights for the ω_j leaf. When this process, which can be considered an advanced pruning process,[36] is over, the iteration is stopped and the information gained is calculated.

The gain from the split due to the XGBoost algorithm can be summarised as follows:

$$gain = \frac{1}{2}\left[\frac{\sum_{i\in I_L} g_i}{\sum_{i\in I_L} h_i + \lambda} + \frac{\sum_{i\in I_R} g_i}{\sum_{i\in I_R} h_i + \lambda} - \frac{\sum_{i\in I} g_i}{\sum_{i\in I} h_i + \lambda}\right] - \gamma \quad (20),$$

where I_L and I_R are the set of indices of the left and right nodes after the split, and $I = I_L \cup I_R$ and γ is the parameter of T. The algorithm creates the trees after finding the best split through a recursive process till the maximum depth allowed. Afterwards, XGBoost prunes the trees with a negative gain in bottom-up order. Considering the requirement to tune six hyperparameters for XGBoost within a data-rich environment framework and aiming to maintain comparability with other boosting models in terms of training time,[37] the number of iterations for XGBoost is relatively low in comparison to the other boosting models utilised in this study.

[36] Pruning is a standard process of regression trees. It implies to reduce the number of nodes at the minimum in order to avoid overfitting.

[37] The extensive number of hyperparameters that need to be estimated in XGBoost significantly extends its computational time. In theory, the computational time could further increase depending on the range of values introduced during the cross-validation step. The primary objective here is to ensure that the computational time of XGBoost does not significantly exceed that of other boosting models.

2.1.4 Other regression methods

Deep learning methods are used in different applications concerning forecasts, given their capacity to handle Huge Big Data, both structured and unstructured. As explained in Chapter 1, deep learning models, such as convolutional networks or recurrent neural networks (Lim and Zohren, 2021; Torres et al., 2021) are excluded from the thesis for two reasons. First, they can generally provide noticeably better results in comparison to 'not' deep learning methods in the presence of huge amounts of data,[38] which is not the case in this thesis.

Second, there are difficulties in interpreting *black box* models, such as the typical deep learning methods. The black box is the inexplicable process of the intermediate steps of the deep learning algorithms. This is not only related to mathematical complexities but also to real difficulties in understanding why the intermediate layers of the machine arrived at a certain decision, which, in part, influences decisions in other layers (see subsection 1.2). Of course, it can be argued that even simpler ML methods face black box issues. However, they are still more interpretable than deep learning methods. Finally, these models need a complex tuning hyper parametrization that presents computationally expensive processes.

2.2 Classification methods in Machine Learning

2.2.1 Classification trees and AdaBoost

Classification trees are the counterpart of regression trees. In this sense, the main difference is that these trees try to understand what variables help identify the specific patterns that help classify the dependent variables of interest. Of course, the classification trees have the same issues as the regression trees, such as the risk of overfitting (see subsection 1.3) and decreasing interpretability when the number of predictors is too large. Hence, ensemble methods can be useful also when classification trees are involved.

AdaBoost (Friedman et al., 2000)[39] is the most well-known classification method among the boosting family and, possibly, among ensemble methods as a whole. As discussed in Chapter 1, the main idea of ensemble methods is to amplify the effects of some single weak learners (algorithms) through a bootstrapping process. In particular, AdaBoost is used as a means to amplify classification tree algorithms. The classification tree model allows the segmentation of a predictor space in different regions taking into account the importance of different explanatory variables in the classification of the response variables. The name derives

[38] When both the number of predictors and observations is massive. A huge amount of data is usually measured in terms of terabytes of data.

[39] To see a more detailed description of the model please refer to the Appendix.

from the fact that the splitting rules can be shown in a graph resembling a tree. For a general overview of the classification trees, see James et al. (2013).

AdaBoost facilitates the creation of a multitude of trees sequentially. As Casabianca et al. (2022) explain, the observations that are not correctly classified by the machine in the initial step of each replication receive a higher weight to be correctly estimated in the new iteration. In other terms, the machine assigns more weight to the most difficult observations to classify. This process reduces the model bias and can estimate the different values of the dependent variable (Casabianca et al., 2022).

In more detail, AdaBoost can be described in the following way. The algorithm preserves the weights in the training set (Freund and Schapire, 1997; Freund et al., 1999). The weight is initialised using the distribution $D_m\ (i) = \frac{1}{n}$ on which the weak learner is trained and each observation is equally weighted, where *n* is the total number of observations within the sample. Subsequently, for $i\ =\ 1,\ldots,n$:

- set a candidate model h_i: $X \rightarrow Y$ where Y is {1,–1} that minimises the weighted errors w.r.t. to the distribution D where the error of the predictions, considering the difference between $h_i(x_i)$ and the real value of the dependent variable y_i, is $\varepsilon_i = Pr_{i\sim D_i}[h_i(x_i)\ \neq\ y_i]$,
- Set $\alpha_i\ =\ \frac{1}{2}\log\frac{1-\varepsilon_i}{\varepsilon_i}$. α_i assesses the temporary importance given to h_i,
- Update: $D_{i+1}\ (i) = \frac{D_{i,}(i)\ \exp(-\alpha_i y_i\ h_i(x_i))}{Z_{im}}$,

  where Z_i is the factor that normalises D_i .

In this sense, the algorithm updates the distribution of the weights of the learner considering the set of candidate models. The values of α_i are high when the error is low. In this way, AdaBoost assigns more weight to the hardest observations to identify. The AdaBoost returns the selected model $H(x)$:

$$\mathrm{H(x)} =\ \mathrm{sign}\ (\textstyle\sum_{i=1}^{n} \alpha_i\ h_i(x)) \qquad (21),$$

where the sign depends on the binary target variable's values $(-1, +1)$.

As can be observed, AdaBoost is a valid method for classifying binary variables. The basis of AdaBoost's structure is the classification tree. This helps to handle nonlinearities in the data. At the same time, the different steps of AdaBoost and the process of bootstrapping can be useful to avoid the problem of overfitting that a simple classification tree has. Finally, the algorithm procedure can extract the most relevant independent variables to predict the positive or negative values of the dependent variable.

2.2.3 Other classification methods

Other methods of classification quite well known in machine learning literature are the support vector machine (SVM) and clustering in its different fashions (James et al., 2013). SVM uses hyperplanes (decision boundaries) to classify the observations. The number of hyperplanes is equal to the number of predictors minus one. The main goal of the algorithm is to find a hyperplane that maximises the distances between data points. In this way, the observations are better separated. SVM is a good classifier, and some of its extensions allow it to operate in the context of nonlinearities. However, its efficiency is reduced when the data dimension is very large and clear separation among variables is not possible.

Clustering is another important classification method that can be used even in very complex data environments. However, its hyper parametrization is mostly heuristic and the interpretation of the results is not very clear when many variables are involved, even when using complex hierarchical clustering (James et al., 2013).

Some of the ensemble methods, such as bagging and random forest, described in the regression section, can be employed in classification exercises but, currently, their use in macroeconomics is very limited. Moreover, AdaBoost is considered a better option as an ensemble method for classification exercises (Ng, 2014).

2.3 Summary

When closely examining shrinkage methods, it becomes apparent that their primary distinctions from standard econometric techniques, such as ordinary least squares (OLS), stem from the application of penalization parameters to coefficient estimates varying in complexity. Notably, lasso and ridge regression serve as clear examples in this context, highlighting their connection with OLS, with the hyperparameter lambda being the principal point of departure from OLS regressions. This is why scholars often direct their attention to these methods when exploring the utilization of machine learning in causal inference studies.

However, when the analysis shifts towards tree-based models, it can be argued that the fundamentals of regression estimations undergo transformation. Nonetheless, even within this more intricate environment, some standard econometric principles persist. This is due to the fact that splitting rules can still rely on well-established residual sum of squares rules.

Additionally, certain limitations inherent to trees, such as high variance, can be effectively addressed through the employment of a familiar statistical technique – bootstrapping. The ensemble methods are essentially grounded in this process. What does change, however, is their capacity to handle nonlinearities, largely contingent on the construction of regression/classification trees. Furthermore, the manner in which weights are updated on different variable sets during the iterative learning process plays a crucial role. The use of bootstrapping allows for an artificial and exponential amplification of training sets, contributing to the enhanced learning process.

3 MACROECONOMIC APPLICATIONS USING MACHINE LEARNING

3.1 Forecasting US industrial production using a Big Fat dataset

Forecasting industrial production is essential for economic analysis, as it remains a key variable in predicting short-term GDP evolution in most countries (Golinelli and Parigi, 2007). Despite the growing importance of the service sector, industrial production still plays a significant role in aggregate economic fluctuations, especially since many business services are closely related to industrial activities (Bruno and Lupi, 2004).

The first empirical study forecasts monthly US industrial production using the FRED-MD dataset from 1963 to 2019. De Nicolò and Lucchetta (2017) exploit the data from the FRED-MD dataset for their study on forecasting tail risks of macroeconomic indicators. Caggiano et al. (2017) control for the first-moment macroeconomic shocks[40] with one of the factors extracted from a factor analysis executed on the McCracken and Ng (2015) dataset, using their nonlinear approach to economic policy uncertainty and unemployment in the USA. In their study on external instruments, Stock and Watson (2018) use the four most important factors extracted from the dataset. The dataset was the most important part of the data that Popp and Zhang (2016) used in their paper on the role of financial channels in the macroeconomic effects of uncertainty shocks.

Smeekes and Wijler (2018) exploit the dataset for their work on a macroeconomic forecasting exercise that compares the lasso method and factor models. They found that lasso performs better in the case of cointegrated variables. Ellington et al. (2017) use the FRED-QD quarterly data, the FRED-MD version in quarterly data, for their study on liquidity shocks with bayesian VAR methods.

Hännikäinen (2017) exploits the dataset to investigate the predictive content of the shadow rates for US real economic activity and inflation in a data-rich environment. Peña et al. (2019) deploy the FRED-MD dataset for their study on forecasting multiple time series with one-sided dynamic principal components[41]. Kim and Park (2020) combined the dataset with Korean data to analyse the connection between the US dollar and the Korean bilateral exchange rates with underlying macroeconomic variables using a factor model approach.

This empirical study aims to compare ML models, particularly boosting models, with benchmark models, such as random walk, autoregressive models and a simple factor model in a data-rich macroeconomic environment. This study implements

[40] First-moment shocks refer to declines in confidence, while second-moment shocks are uncertainty shocks. Of course, these different types of shocks can be interrelated (Claeys and Vašíček, 2019).

[41] A one-sided dynamic component for time series is a linear combination of present and past values of the series, that can minimize the reconstruction mean squared error (Peña et al., 2019).

ridge, lasso, elastic net (elnet), and adalasso following the suggestions of Medeiros et al. (2016), Medeiros and Mendes (2015), and Wang et al. (2007). Furthermore, often used ensemble methods like bagging and random forest are estimated. Finally, component wise boosting, gradient boosting, AdaBoostR2, AdaBoostRT, XGBoost are implemented.[42]

This thesis demonstrates how boosting methods can be applied to forecast business cycles and evaluates their performance in a regression exercise. Three benchmarks (autoregression, random walk, and a factor model) are compared with different boosted regression methods and other machine learning methods. The analysis is performed with both fixed window estimations and expanding window estimations. Different monthly horizons are estimated: short-run (h=1), medium-run (h=4), and long-run (h=12). The main goal is to study the performances of these models in a data-rich environment and to understand which boosted regression is better than others in this context.

This chapter presents the dataset used, the initial treatment of the data, the methodology implemented and the results. The subsection 3.1.1 concerns the data used and the data treatment. The subsection 3.1.2 describes the different ML methods used in this first empirical study. The subsection 3.1.3 shows the main results of the forecasting exercise. Finally, the subsection 3.1.4 illustrates the additional results related to comparisons between the different ML methods between each other. Subsection 3.1.5 presents a general evaluation of the results obtained.

3.1.1 Data and data treatment

As discussed in Chapter 1.4, the FRED-MD dataset from McCracken and Ng (2015) have already been used in numerous empirical macroeconomic works and the dataset is publicly available. In the last decade, scholars have used this dataset in their empirical studies, giving them the possibility to access more than 130 macroeconomic variables in a single dataset. Goulet Coulombe et al. (2022) use the FRED-MD dataset to test in which environment machine learning methods forecast better and how different features, such as nonlinearities and alternative loss functions, influence the results of the forecasting exercise. Medeiros et al. (2021) employ the dataset to forecast US inflation using different machine learning methods in a data-rich environment. Smeekes and Wijler (2018) use the dataset for their study on macroeconomic forecasting, comparing the lasso method and factor models.

Furthermore, as suggested by Goulet Coulombe et al. (2022), the time series available in FRED-MD allows a larger test set than other data in the FRED universe of time series. Moreover, as McCracken and Ng (2015) remark, it has three important characteristics: it is designed to be updated every two months,

[42] R packages used to implement machine learning methods are: glmnet, HDeconometrics, rpart, caret, gbm, xgboost, randomForest. The functions for AdaBoostRT and AdaBoostR2 are extracted from the package ReBoost.

facilitating new exercises which could replicate previous studies but with a longer time frame and other possible structural breaks; it is entirely publicly accessible facilitating replication of the research; it remains consistent over time, sparing researchers from dealing with frequent changes and revisions. All the time series are of the same frequency (monthly), avoiding problems of transformation in cases of higher or lower frequencies.

In this study, the variables are first transformed following the instructions of McCracken and Ng (2015) and Bai and Ng (2008), as in Medeiros et al. (2021) and Goulet Coulombe et al. (2022), involving for example log-differentiation, double log-differentiation and others. A complete list of all the transformations per time series is available in the Appendix. Next, the industrial production index (INDPRO) is calculated by taking the average growth rate over the next *t* months where (*t* varies between one and twelve). This transformation considers both the transformation suggested by McCracken and Ng (2015) and the possible problems due to the accumulation of forecast errors (Estrella and Hardouvelis, 1991).

Hence, ΔY_t is the average growth rate of industrial production over the next month *t* computed as $\Delta Y_t = \left(\frac{1200}{k}\right)[log(y_t) - log(y_{t-k})]$, where y_t is a given industrial production index in period *t* and *k* is the step ahead.[43] The variables of the dataset where there are massive amounts of missing data (decades of data missing) are dropped, as suggested by the creators of the dataset. This allows us to create a balanced dataset and avoid problems with missing data to which some machine learning methods are sensitive.[44]

The total number of selected predictors is 123. Hence, as covariates, there are the first four lags of all these 123 predictors plus the four lags of the average growth rate of industrial production.[45] Hence, the total amount of predictor variables is 496.[46] Considering the necessary transformations to make the time series stationary, the data span from February 1963 to April 2019 (674 observations per time series).[47] The most recent revisions of the data are employed.

---

[43] The fomula is derived by Estrella and Hardouvelis (1991). In the case of Estrella and Hardouvelis (1991) 400/k is used, given that they employ quarterly data. Instead, given that the presented exercise is on monthly data, a change to 1200/k is needed because of the different frequency of the data.

[44] The dropped variables suggested by McCracken and Ng (2015) are: Trade Weighted U.S. Dollar Index: Major Currencies(TWEXAFEGSMTHx.), Consumer Sentiment Index (UMCSENTx), New Orders for Nondefense Capital Goods (ANDENOx), and New Orders for Consumer Goods (ACOGNO).

[45] Following Medeiros et al. (2021).

[46] For the complete list of the variables and their transformations refer to the Appendix, Tables A1–A8. This paper employs the May 2021 vintage of the data. More information is available here: https://files.stlouisfed.org/files/htdocs/fred-databases/fredmdchanges.pdf

[47] In certain instances, time series data originate from the year 1959, while others do not share this starting point. Due to the presence of these missing data points and in accordance with the prescribed transformations detailed in the Appendix, it is necessary to adjust these time series. This adjustment entails the removal of observations to ensure that all time series commence at the same point in time and adhere to the same time frame.

3.1.2 Methodology

In order to test and compare different forecasting models, pseudo-out-of-sample estimations using both fixed and expanding window methods are performed. Fixed windows make it possible to account for possible changes in the structure of the data over time. The expanding window method is better for preserving more observations in the out-of-sample dataset. The fixed window method makes it possible to take into account the volatility of the time series in different periods. The windows contain 50% of the observations (337) to have enough data to properly train the machine learning models.

Moreover, as Rossi and Inoue (2012) report, large windows are preferable when the data are stationary, as is also the case in this work. Shorter windows can be more robust to structural breaks but cannot provide as precise estimations as larger windows. Considering the number of steps ahead normally performed by scholars of forecasting (Breitung and Knüppel, 2021; Drechsel and Scheufele, 2012), the forecasting horizons are 1, 2, 3, 4, 8, 12 months ahead.[48] In the case of fixed window estimations, the window is rolling over time.

The initial window equals exactly half of the length of the transformed data, 337 months. Instead, in the expanding window estimations, the window starts from the 338th month and continues to increase till the last month of the test set. In the estimation with only the data before the Great Recession, the estimations stop in January 2008.[49] Given the smaller amount of data, the window size decreases to 270 months. In the general form, the models can be represented as:

$$\Delta Y_{t+h} = z_h(x_t) + u_{t+h}, h = 1, \dots H, t = 1, \dots, T \quad (22),$$

where ΔY_{t+h} is the average growth of industrial production over $t + h$ where h represents the horizon and t is time. $x_t = (x_{1t}, \dots, x_{nt})'$ is an n-vector of regressors containing lags of y_t. While z_h is the mapping between covariates and industrial production lags; and u_t is a zero mean random error. The target function $z_h(x_t)$ can be a single model or an ensemble of different specifications (in this work, bagging, random forest and boosting models). There is a different mapping for each forecasting horizon *h*. The terms *H* and *T* are the final horizon and month available in the set.

The direct forecasting equation is given by:

$$\Delta \hat{Y}_{t+h|t} = z_{h,t-R_h+1:t}(x_t) \quad (23),$$

where $z_{h,t-R_h+1:t}$ is the estimated target function based on data from time $t - R_h + 1$ up to t, and R_h is the window size, which varies according to the forecasting horizon and the number of lagged variables in the model. The first

[48] Only 1, 4 and12 horizons are shown in thesis. The other results do not add relevant information but they are available under request.

[49] As in Drechsel and Scheufele (2012).

observation is dropped in the case of the fixed window method to maintain the fixed window size.

The first benchmark is the random walk (RW) model[50], where for $h = 1,\ldots,12$, the model is computed as follows:

$$\Delta\hat{Y}_{t+h|t} = \Delta y_t, \tag{24}$$

where $\Delta\hat{Y}_{t+h|t}$ is the forecasted average growth rate of industrial production, Δy_t is the actual growth rate of industrial production, and t is time. The second and third benchmarks are two autoregressive models (AR)[51]: an AR(p), selected by a Bayesian information criterion (BIC), and an AR(12). AR(12) can be inefficient but unbiased in the case of monthly data because twelve lags can better capture within year dynamics at monthly frequency but, at the same time, can trigger the problem of overfitting. When the AR model is of order p, the choice of p can differ in each horizon. The two AR models can be described as follows:

$$\Delta\hat{Y}_{\mathrm{t+h|t}} = \sum_{j=0}^{p} \alpha_j \Delta y_{t-j} + e_{t+h} \; |\alpha| < 1, \tag{25}$$

where $\widehat{\Delta Y}_{\mathrm{t+h|t}}$ is the forecasted average growth rate of industrial production, Δy_{t-j} is the growth rate of industrial production at time $t - j$, and the order p is either the order selected by BIC or 12. α_j is the parameter of the j^{th} lag, and e_t is white noise.

The last benchmark model is a simple factor model (FM). The factors are derived using principal component analysis applied on the first lag of the predictors together with the first lag of industrial production growth. Subsequently, the two first factors are selected, taking into account the findings of McCracken and Ng (2015) that show how the first two factors are the most relevant in the FRED-MD dataset. The model is an OLS regression that can be described as:

$$\Delta Y_{t+h|t} = \alpha + \beta_1 f_{1,t-1} + \beta_2 f_{2,t-1} + e_{t+h} \tag{26},$$

where $f_{1,t-1}$ and $f_{2,t-1}$ are the two extracted factors at time $t - 1$. β_1 and β_2 are the parameters of the two factors. The model can be compared with that in Ghysels and Marcellino (2018, p. 447). The models are compared according to the root mean squared error (RMSE).

$$RMSE_{m,h} = \sqrt{\frac{1}{T-T_0+1}} \sum_{T=T_0}^{T} \hat{e}_{t,m,h}^2 \tag{27},$$

[50] Examples of studies using RW as a benchmark model against machine learning methods can be seen in Medeiros et al. (2021) and Goulet Coulombe et al. (2021b).

[51] AR model is included among the benchmarks because it can also show the response dynamics of the forecasted variable that are not visible using a constant growth rate.

where $\hat{e}^2_{t,m,h} = y_t - \hat{y}_{t,m,h}$ and $\hat{y}_{t,m,h}$ is the industrial production forecast for month t made by model m with information up to $t - h$. Moreover, the Diebold-Mariano (DM) test (Diebold and Mariano, 1995) for expanding window estimations and the Giacomini-White (GW) test (Giacomini and White, 2006) for fixed window estimations are performed. Both tests determine whether the compared models exhibit significant differences. The application of the GW test is justified by the existence of nested models, where some models may include a subset of variables found in other regressions. This is particularly relevant for shrinkage models, where selected variables can be subsets of others.

3.1.3 Main Results

This section describes the results of the boosted regression against the benchmarks. Table 4 shows the results of all the ML models against the benchmark models with the full sample using both fixed and expanding window approaches. The horizons shown are the first *(h=1*), the fourth (*h=4*) and the twelfth (*h=12*). The results for the second, the third, and the eighth horizons are not adding particular supplementary information, but they are available on request. RW is generally outperformed by boosting models in Table 4, and the differences between the predictions of ML methods and RW are statistically significant. The best results are obtained by the XGBoost model in the four-steps ahead forecasts, but in the long horizon forecasts (*h=12*), where XGBoost is beaten by RW and other boosting models that perform better than XGBoost. In particular, component wise boosting has the best results in the twelve-step ahead forecasts.

For what concerns the comparisons between the boosting models and the AR (p) model[52] in Table 4, it can be observed that in fixed and expanding window estimations, AdaBoostRT beats the AR (p) model, but its results are not significant. XGBoost beats the AR model in the four-step ahead forecasts. Moreover, the XGBoost's forecast significantly differs from the AR (p) in both expanding and fixed window estimations.

The FM model in Table 4 is beaten by the component-wise boosting model in the fixed and expanding window estimations for all the horizons, but only for the first and the fourth horizons in the fixed window estimations, the forecasts are significantly different from the FM model forecasts. Instead, in the expanding window estimations, all the component wise boosting forecasts are significantly better than the FM. AdaBoostRT beats the FM model in the first horizon in both fixed and expanding window estimations, and the forecast is significant. XGBoost beats the FM model in the four-steps ahead forecasts, and all the results are significant in both fixed and expanding window estimations.

[52] The results against the AR (12) model are omitted in the main text because they do not show particularly relevant differences with the results against AR (p) but they are available in the Appendix.

Table 4. RMSE of the forecasting models for the full sample, relative to benchmark models

	Fixed windows			**Expanding windows**		
Model	**AR (p)**					
Horizon	h=1	h=4	h=12	h=1	h=4	h=12
Lasso	1.04	1.02	1.01	1.01	1.03	1.03
Ridge	1.08***	1.08***	1.07***	1.07***	1.07***	1.09***
Elnet	1.06***	1.07***	1.06***	1.02	1.03	1.03
Adalasso	1.06*	1.03	1.00	1.04	1.03	1.04
RF	0.99	1.03	1.05*	0.98	1.01	1.02
Bag	1.04	1.81	1.42	0.99	1.31	1.39
CBoost	1.01	1.01	1.01	0.99	1.00	1.00
AdaR2	1.12***	1.09***	1.07***	1.08***	1.06***	1.07***
AdaRT	0.99	1.05	1.04	0.99	1.01	1.02
XGB	1.13***	0.61***	1.80***	1.10**	0.61***	1.83***
GradB	1.07***	1.02	1.08**	1.05***	1.03	1.02
	FM					
Lasso	0.99	0.97	0.97	0.98	0.99	1.00
Ridge	1.03	1.03	1.03	1.04	1.04	1.05
Elnet	1.01	1.02	1.02	0.99	1.00	1.00
Adalasso	1.01	0.98	0.97	1.01	1.00	1.01
RF	0.95***	0.98	1.01	.95***	0.98	0.99
Bag	0.99	1.73	1.37	0.96	1.27	1.35
CBoost	0.96**	0.97*	0.97	0.96**	0.96**	0.97**
AdaR2	1.07	1.04	1.04	1.05	1.03	1.04
AdaRT	0.95***	1.00	1.01	0.96**	0.98	0.99
XGB	1.08*	0.58***	1.74***	1.07	0.60***	1.78***
GradB	1.02	0.98	1.05	1.02	1.00	0.99
	RW					
Lasso	0.76***	0.78***	0.65	0.76***	0.79***	0.67***
Ridge	0.79**	0.82***	0.69***	0.80**	0.83***	0.71***
Elnet	0.77**	0.82***	0.69***	0.76***	0.80***	0.67***
Adalasso	0.78***	0.79***	0.65***	0.78***	0.80***	0.68***
RF	0.73***	0.79***	0.68***	0.73***	0.78***	0.66***
Bag	0.76***	1.38	0.92	0.74***	1.02	0.90
CBoost	0.74***	0.77***	0.65***	0.74***	0.77***	0.65***
AdaR2	0.82*	0.83***	0.69***	0.81**	0.82***	0.70***
AdaRT	0.73***	0.80***	0.67***	0.74***	0.78***	0.66***
XGB	0.83*	0.46***	1.16*	0.82**	0.48***	1.19**
GradB	0.79**	0.78***	0.70***	0.78**	0.80***	0.67***

Notes: Numbers are rounded to the second decimal. Significance after DM/GW test: *p<0.1, **p<0.05, ***p<0.01. Ratios of the machine learning methods RMSE over the benchmark models RMSE.

In Table 5, the results of the estimations related to the fixed and expanding window approach are reported with a subsample from before the Great Recession. XGBoost[53] provides better results in the first and fourth horizons against RW. AdaBoostR2 performs better in the fourth and twelfth horizons in fixed window estimations. In the expanding window estimations, the best result of AdaBoostR2 is obtained in the fourth horizon.

Component wise boosting is significantly better than RW in both expanding and fixed window estimations. The best result from component wise boosting against RW is the first horizon in the fixed window estimations. However, the differences by horizons, considering both estimation methods, are negligible. AdaBoostRT has a similar behaviour, and its best result is in the first horizon in the fixed window estimations. As expected, the standard gradient boosting has the worst performance in comparison with other boosting models. Still, gradient boosting performs better than RW.

In Table 5, the only boosting model able to beat AR model forecasts in the fixed window estimations is XGBoost in the first and fourth horizons. The component wise boosting, AdaBoostRT and XGBoost are able to beat the FM model with significant results. Component wise boosting and AdaBoostRT beat the FM in all the horizons in both fixed and expanding window estimations. However, AdaBoostRT performs slightly better than component wise boosting in the expanding window estimations. XGBoost beats the AR model in the first and fourth horizons in the fixed estimations and in the fourth horizons in the expanding estimations.

The forecasting performance of the models does depend on the forecasting horizon. This is well documented also in the literature. Taking into account the findings of Goulet Coulombe et al. (2022) and Buchen and Wohlrabe (2011) on forecasting industrial production in data-rich environments, it becomes evident that machine learning methods – particularly boosting techniques – tend to perform better at medium-term forecast horizons (4–8 months ahead). For instance, Buchen and Wohlrabe (2011) show that most of their boosting models achieve their best performance compared to various benchmarks at horizons between the third and sixth periods.

[53] In the estimations using the subsample of data from before the Great Recession, XGBoost is unable to complete the hyperparameter tuning process. However, the results are still reported for completeness. To determine whether this issue stems from an excessive number of hyperparameters' values evaluated during cross-validation, the process is repeated with a reduced set of hyperparameters' values. Under this adjustment, XGBoost successfully completes the hyperparameter tuning process. Nevertheless, comparing this simplified model with the XGBoost model applied to the full sample is no longer feasible. Therefore, the results from the original XGBoost model, which do not complete cross-validation, have been reported due to their closer alignment with the model used for the full sample. Further details can be found in the Appendix.

Table 5. RMSE of the forecasting models with a subsample period from before the Great Recession, relative to benchmark models

Model	AR (p)					
	fixed windows			expanding windows		
Horizon	h=1	h=4	h=12	h=1	h=4	h=12
Lasso	1.05**	1.05**	1.06***	1.04*	1.04*	1.05**
Ridge	1.08***	1.08***	1.08***	1.08***	1.08***	1.08***
Elnet	1.06***	1.06***	1.05***	1.04*	1.04*	1.05**
Adalasso	1.10***	1.11***	1.11***	1.10***	1.11***	1.12***
RF	1.04	1.04	1.04*	1.04	1.03	1.04*
Bag	1.21***	1.36***	2.18**	1.22**	1.46	1.39
CBoost	1.04	1.04	1.04	1.05*	1.05*	1.05*
AdaR2	1.14***	1.08***	1.09***	1.13***	1.08***	1.10***
AdaRT	1.03	1.04	1.03	1.04	1.04	1.04
XGB*	0.65***	0.77***	7.73***	1.10**	0.79***	2.35***
GradB	1.24***	1.25***	1.26***	1.18***	1.27***	1.21***
Model	**FM**					
Lasso	0.95***	0.95**	0.97	0.94**	0.94**	0.95**
Ridge	0.98	0.98	0.98	0.98	0.98	0.98
Elnet	0.96	0.96**	0.96*	0.94**	0.94**	0.95**
Adalasso	1.00	1.01	1.01	1.00	1.00	1.01
RF	0.94***	0.94***	0.95**	0.94***	0.94***	0.94***
Bag	1.10**	1.23**	1.98**	1.10	1.32	1.26
CBoost	0.94***	0.94**	0.94***	0.95**	0.95***	0.95**
AdaR2	1.03	0.98	0.99	1.02	0.97	1.00
AdaRT	0.94***	0.94***	0.94***	0.94**	0.94***	0.94**
XGB*	0.59***	0.70***	7.02***	1.00	0.71***	2.13***
GradB	1.13***	1.13***	1.14***	1.06***	1.15***	1.10***
Model	**RW**					
Lasso	0.70***	0.73***	0.73***	0.69***	0.72***	0.72***
Ridge	0.72***	0.75***	0.74***	0.72***	0.75***	0.74***
Elnet	0.71***	0.73***	0.72***	0.69***	0.72***	0.72***
Adalasso	0.73***	0.77***	0.76***	0.73***	0.77***	0.77***
RF	0.69***	0.72***	0.71***	0.69***	0.72***	0.72***
Bag	0.81***	0.94	1.49	0.81***	1.01	0.96
CBoost	0.69***	0.72***	0.71***	0.70***	0.73***	0.72***
AdaR2	0.76***	0.75***	0.75***	0.76***	0.75***	0.76***
AdaRT	0.69***	0.72***	0.71***	0.70***	0.72***	0.71***
XGB*	0.43***	0.54***	5.29***	0.73***	0.55***	1.61***
GradB	0.83***	0.86***	0.86***	0.79***	0.88**	0.83***

Notes: rounded to the second decimal. Significance after DM/GW test: *p<0.1, ** p<0.05, ***p<0.01.
Ratios of the machine learning method RMSE over the benchmark model RMSE. For XGB, see the Appendix.

Similarly, although Goulet Coulombe et al. (2022) do not specifically use boosting models, they observe a comparable trend in some of their forecasts using machine learning methods, where relevant improvements of the forecasts are seen at the third and sixth horizons, but these gains are not always confirmed in longer horizons, such as the twelfth or twenty-fourth month.

This pattern of improvement at intermediate horizons is further corroborated and generalised by Lehmann and Wohlrabe (2016), who focus on German data and demonstrate that boosting methods yield the most accurate forecasts between the third and sixth horizons. When closely examining the results of Goulet Coulombe et al. (2022) and Lehmann and Wohlrabe (2016), it can be hypothesised that combining different types of nonlinearities, different choices in terms of loss functions, and employing weak learners could enhance the performance of XGBoost, even at extended forecast horizons. This suggests a potential avenue for further research.

The results of the other ML models against the benchmarks are summarised in this section. In Table 4, the shrinkage models beat RW and are statistically different from it in the fixed window estimations and in the expanding window estimations. However, lasso is not significantly different from RW in the last horizon with fixed estimations. In Table 5, with the data related to the subsample from before the Great Recession, all the shrinkage models show better results than RW and are significantly different from it in both fixed and expanding window estimations.

It can be noticed that the performance of the shrinkage models are not particularly improved if the estimation methods are changed. If different data samples are employed, the performance improved on average in the forecasts before the Great Recession, but the variation is not particularly large. In terms of horizons, the better results are in the one-step ahead forecasts with the data before the Great Recession (Table 5) and in the long-horizon (12-steps ahead) for the full sample (Table 4).

Concerning the AR (p) model, the shrinkage models perform worse than the autoregression model, but not all estimations are statistically different from the AR model. In Table 4, the lasso and adalasso models, and AR (p) results are not statistically different with the exception of the first horizon forecast of adalasso in fixed window estimations. Table 4 shows that ridge is the model with the worst performance compared to the AR model, and the results are statistically significant in the case of both estimation methods. Elnet is worse than AR (p) in both expanding and fixed window estimations in Table 4 and significantly different in the fixed window estimation results. In Table 5, with the data subsample on the period from before the Great Recession, all the shrinkage models are beaten by the AR (p) model and are statistically significant in all the horizons. On average, the performances are worse than the ones in Table 4.

Regarding the comparison between the factor model and the shrinkage models, it can be observed that in Table 4, lasso and some adalasso results are slightly better than the FM. However, the results are insignificant in both estimations methods. On the other hand, in Table 5, lasso and adalasso beat the FM model, and the outcomes are significant in almost all the results. This can be motivated

by the presence of unnecessary information – noise – captured by the factor model. Lasso and adalasso can exclude not relevant variables that can reduce this noise.

The ensemble methods beat RW in Table 4. However, bagging is better than RW and statistically significant only in the short-run forecasts in both windows estimations and both in Tables 4 and 5. Instead, random forest performs better than RW, and its results significantly differ from RW in all the estimations in both Tables 4 and 5. The results in expanding window estimations are, on average, slightly better than those in fixed window estimations. Random forest and bagging are able, especially RF, to compete with the AR model in the short horizons in Table 4, but the ensemble models are not significantly different from the AR (p). In Table 5, the results are completely the opposite: almost all are significant, but neither can the ensemble models beat the AR model, and bagging performs much worse than in the estimations for the full sample. This can be related to the different nonlinearities present in the various decades.

Finally, random forest can beat the FM model in Table 4 in both expanding and fixed window estimations, with significant results in one-step ahead forecasts. Bagging is able to perform slightly better than FM only in the one-step ahead forecasts with fixed and expanding methods with the full data sample (Table 4), but the results are not significant. The performance of the two models become worse in the long horizon forecasts. In Table 5, it can be observed that random forest beats the FM model in both estimations, with all results statistically significant and stable among the horizons. Instead, bagging performance in Table 5 is worse than in Table 4.

Figure 5 shows a simpler data visualization of the performance of the model against the benchmarks in the different horizons. Forecast errors at 4 steps ahead can be seen in Figure 6 (AdaBoostRT vs RW), Figure 7 (AdaBoostRT vs AR(p)), and Figure 8 (AdaBoostRT vs FM).

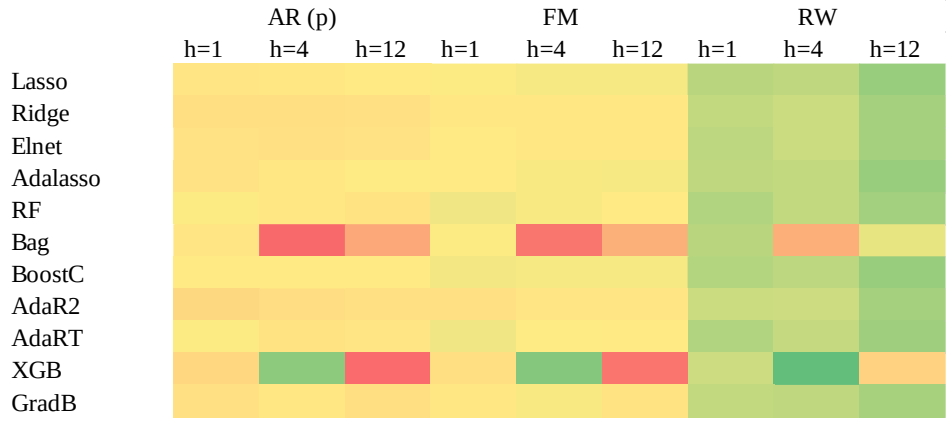

Figure 5. Heatmap performances of the model compared to the benchmarks (fixed window estimations) – Autoregression model (AR(p)), factor model (FM) and random walk (RW).

Note: Colours towards green indicate better results, while colours towards red indicate worse results.

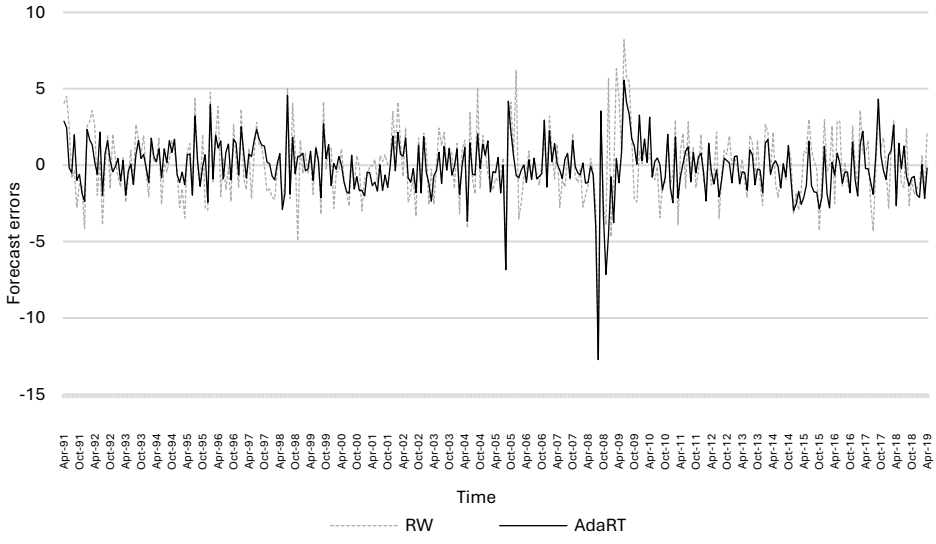

Figure 6. Forecast errors 4 steps ahead – expanding windows – AdaBoostRT vs RW

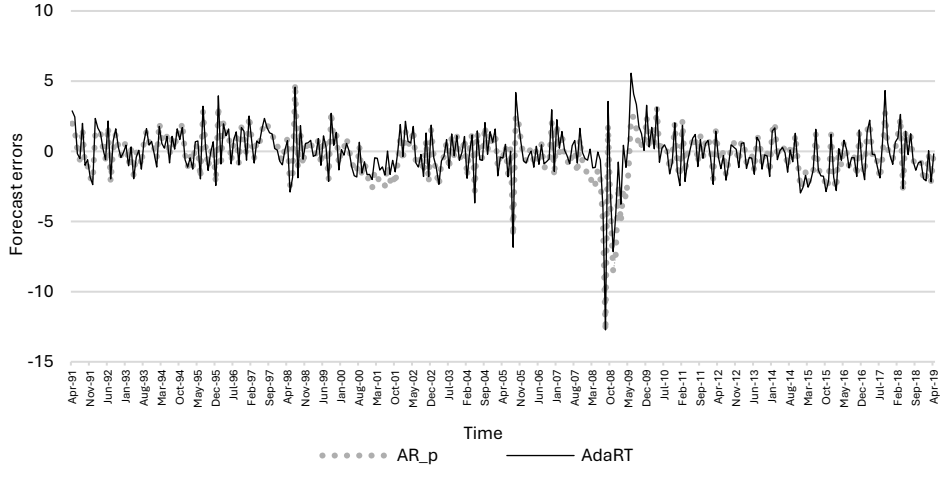

Figure 7. Forecast errors 4 steps ahead – expanding windows – AdaBoostRT vs AR_p

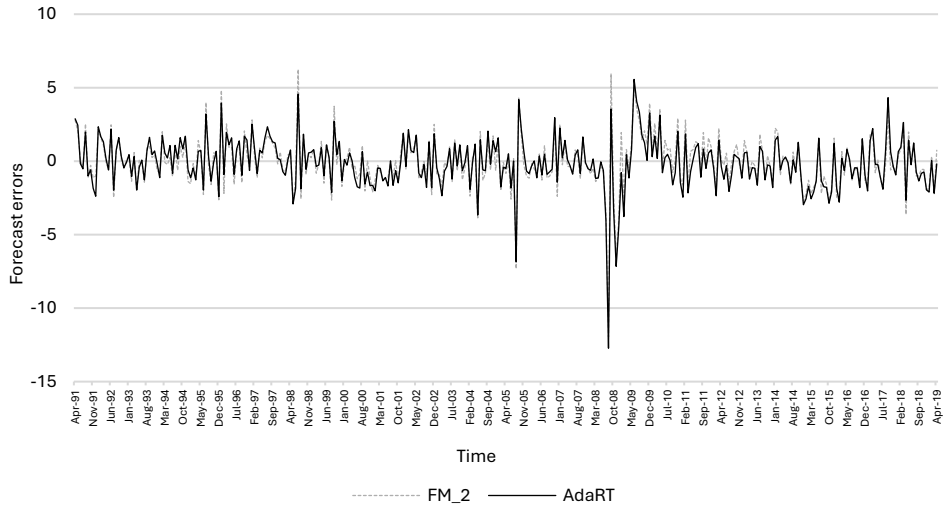

Figure 8. Forecast errors 4 steps ahead – expanding windows – AdaBoostRT vs FM

3.1.4 Additional Results

Comparisons of all models can be found in the Appendix, enabling a comprehensive assessment of various machine learning models. Thus far, only the comparisons between boosting models and benchmark models have been analysed, as well as between other machine learning models and the benchmark. However, a comparative analysis of the performance differences between boosting models and other machine learning methods has not been presented.

This section is dedicated to examining these disparities in performance. Table A9 shows the results of the one-step ahead forecasts with a fixed window approach using the full sample of data. In this table, it can be noticed that component wise boosting and AdaBoostRT are better than shrinkage models and bagging, but they are at the same level of forecast accuracy as the random forest. Looking at AdaBoostRT and component wise boosting, the difference in forecasting in this setting is negligible.

The four-steps ahead forecasts with fixed windows of the full sample is illustrated in Table A10. In this table, AdaBoostRT beats ridge and elnet even if the results are not significant. Instead, component wise boosting beats all the shrinkage models, and the results against ridge and elnet are significant. Gradient boosting is better than all the shrinkage models except lasso, but its results are not significant. The results for XGBoost are largely better than the shrinkage models and are significant in all cases. All the boosting models are better than bagging. Concerning random forest, XGBoost is considerably better than random forest. Component wise boosting is better than random forest, but the difference is not significant. Finally, gradient boosting and random forest have exactly the same level of forecast accuracy in Table A10.

The twelve-step ahead forecasts with fixed windows of the full sample can be observed in Table A11. Component wise boosting is better than ridge and elnet, but only the forecasts between ridge and component wise boosting are significantly different. AdaboostRT is better than ridge and elnet, but its forecasts and those of the shrinkage models are not significantly different. Bagging is worse than the boosting models, except XGBoost. Component wise boosting beats random forest, but the results are not significant. AdaboostRT and random forest have the same accuracy forecast.

The one-step ahead forecasts with the expanding window approach using the full sample is reported in Table A12. As in Table A9, AdaboostRT and component wise boosting are better than the shrinkage models. However, the results for component wise boosting are only significantly different compared to ridge and adalasso, and AdaboostRT is only significantly different from ridge. Gradient boosting is significantly better than ridge and has the same forecasting capacities as adalasso. Random forest, bagging and the boosting models are almost equal in their forecasting capacities and not significantly different in Table A12.

The four-steps ahead forecasts with expanding windows using the full sample of data is shown at Table A13. XGBoost beats all the other ML methods and is significantly different from them. Component wise boosting beat all the shrinkage models, random forest, and bagging, but the differences are significant only in relation to ridge. AdaboostRT is better than the shrinkage models and bagging. Moreover, it has the same forecasting accuracy as the random forest. However, as for component-wise boosting, only the difference between AdaBoostRT and ridge is significant. Finally, gradient boosting and AdaBoostR2 are better than ridge, but the differences in terms of results are not significant. Gradient boosting has equal forecast capacities compared to elnet and adalasso.

The twelve-steps ahead forecasts with expanding windows using the full sample is illustrated in Table A14. Component wise boosting is better than the other ML methods, but the results are only occasionally significant such as against ridge and adalasso. AdaBoostRT is also better than all the other ML methods, but the only significant result is that against ridge. AdaBoostR2 beats ridge and bagging, but the results are not significant. Gradient boosting beats ridge, adalasso, and bagging, but the difference is not significant. Moreover, it has forecast accuracy equal to that of lasso, elnet, and random Forest.

All the rest of the tables from A15 to A20 concern the subsample related to the period before the Great Recession. The one-step ahead forecasts with a fixed window approach can be observed in Table A15. Component wise boosting performs better than shrinkage models and bagging. Moreover, its results and those of random forest are the same. The differences between component wise boosting and ridge, adalasso, and bagging are significant. AdaBoostRT beats all the other ML methods, but are only significantly different to ridge, adalasso, AdaBoostR2, gradient boosting, and bagging. AdaBoostR2 beats bagging, but the results are not significantly different. XGBoost beats all the other models.

The four-steps ahead forecasts using the fixed window method is described in Table A16. Component wise boosting performs better than the shrinkage models

and bagging. The results of component wise boosting and of random forest are the same. This is exactly the same situation as in Table A15. AdaBoostRT beats all the shrinkage models and bagging. The results of random forest and AdaBoostRT are the same. However, the only results that differ significantly from AdaBoostRT are those of ridge, adalasso and bagging. XGBoost significantly beats all the other ML models. Gradient boosting beats bagging, but the difference is not significant. AdaBoostR2 beats bagging, and the difference between the two forecasts is significant.

The results of the fixed window method concerning the twelve-steps ahead forecasts is shown in Table A17. Component wise boosting beats all the other ML methods. The significant results are the ones against ridge, adalasso and bagging. AdaBoostR2 performs better than adalasso and bagging, but only the result against bagging is significant. AdaBoostRT beats all the other ML methods, but its results compared to elnet and random forest are not statistically different. Gradient boosting beats bagging, and the difference between the two forecasts is significant.

The one-step ahead forecasts with the expanding window approach can be observed in Table A18. Component wise boosting performs better than ridge, adalasso and bagging, but only the difference between component wise boosting and bagging is significant. AdaBoostR2 beats bagging but the difference between their forecasts is not significant. AdaBoostRT beats ridge, adalasso and bagging. However, the significant differences are only with adalasso and bagging. Moreover, the forecasts of AdaBoostRT are equal to those of lasso, elnet and random forest. XGBoost and gradient boosting are both better than bagging, but the results are not significantly different.

The results of the four-steps ahead forecasts using expanding windows is illustrated in Table A19. Component wise boosting performs better than ridge, adalasso and bagging, although the only significant difference is with adalasso. AdaBoostR2 beats adalasso and bagging. Moreover, its forecast is equal to ridge. However, none of the results are significant. AdaBoostRT performs better than ridge, adalasso, gradient boosting, and bagging. Only the results between AdaBoostRT and adalasso are significant. Gradient boosting beats bagging, but the forecasts are not significantly different.

The twelve-steps ahead forecasts with expanding windows is described in Table A20. Component wise boosting performs better than ridge, adalasso and bagging. However, only the forecasts of adalasso and component wise boosting are significantly different. AdaBoostRT beats all the other models but only the forecasts against adalasso, gradient boosting, and AdaBoostR2 are significantly different. AdaBoostR2 performs better than adalasso and bagging but the differences in their forecasts are not significant. The estimation results for the other forecast horizons are available on request.[54]

[54] See Table A21 for an example.

3.1.5 Evaluations of the results

Section 3.1 provides a comprehensive analysis of the performance of various machine learning models in forecasting US industrial production across estimations with different horizons and windows. The primary focus is on comparing the performance of boosting models (e.g., XGBoost, component-wise boosting, and AdaBoostRT) with shrinkage models (e.g., ridge and elastic net), ensemble methods (e.g., random forest and bagging), and benchmark models. Results are detailed for forecasts one-step, four-steps, and twelve-steps ahead using both fixed and expanding windows over the full sample period. The findings indicate that boosting models generally outperform random walk, shrinkage methods, factor models and bagging. XGBoost, component-wise boosting and AdaBoostRT perform better compared to other boosting models.

The subsample analysis (pre-Great Recession period) reinforces the results of boosting models under different economic conditions, with component-wise boosting and AdaBoostRT remaining effective models. Overall, the results suggest that boosting models, especially XGBoost, component-wise boosting, and AdaBoostRT, are reliable for complex, data-rich forecasting environments. The findings emphasize the potential of boosting techniques for accurate predictions, particularly when using an expanding window approach for capturing medium to long-term trends.

3.2 AdaBoost vs logit for identifying periods of crisis

The second empirical study in this thesis is a classification exercise to forecast financial crises, and pre-crisis and crisis-periods using AdaBoost[55] and comparing the results with stepwise logistic regressions and logistic regressions. Moreover, the variables selected by AdaBoost are related to the previous literature dealing with financial crises (Jordà et al., 2016, Jordà et al., 2015).

The section can be divided as follows: the subsection 3.2.1 includes data presentation and treatment; the subsection 3.2.2 subchapter describes the methodology; the subsection 3.2.3 presents the results of the classification exercise and the variable importance extracted from AdaBoost; and the subsection 3.2.4 is a general analysis of the results.

3.2.1 The data and data treatment

This work considers the 16 countries of the historical macroeconomic dataset[56] the Jordà-Schularick-Taylor Macrohistory Database created by Jordà et al. (2017) and, for the house prices, the variable by Knoll et al. (2017). The dataset

[55] The ada and rpart packages have been used to implement AdaBoost.

[56] The data used in this paper refers to the previous version of the dataset before the last updates in June 2019; see Figure 9.

comprises 25 nominal yearly time series from 1870 to 2016. Among the three variables related to GDP, only Nominal GDP (in the local currency) is chosen.[57] Table 6 shows the variables used in the analysis.[58] Some of the time series have missing data in the initial observations or in the first part of the 20th century. To handle this issue, the variables are treated and selected with the help of the following principles.

Table 6. Time series of the Jordà-Schularick-Taylor Macrohistory Database used in the exercise

Category	**Variables included**
Real Economy	Nominal GDP (local currency), Real Consumption per capita (index, 2006=100), Investment-to-GDP Ratio, Population
International	Current Account (nominal, local currency), Imports (nominal, local currency), Exports (nominal, local currency), USD Exchange Rate (local currency/USD)
Government	Government Revenue (nominal, local currency), Government Expenditure (nominal, local currency), Public Debt-to-GDP Ratio
Money, Prices, Interest Rates	Narrow Money (nominal, local currency), Broad Money (nominal, local currency), Short-term Interest Rate (nominal, percent per year), Long-term Interest Rates (nominal, percent per year), Consumer Prices (index, 1990=100)
Credit Data	Total Loans to Non-financial Private Sector (nominal, local currency), Mortgage Loans to Non-financial Private Sector (nominal, local currency), Total Loans to Households (nominal, local currency), Total Loans to Business (nominal, local currency)
Asset Prices	House Prices (index, 1990=100), Stock Prices (index, 1990=100)
Crisis Dates	Systemic Financial Crisis (0–1 dummy)

Source: Jordà-Schularick-Taylor Macrohistory Database, 2017 updated version. Categorisation based on Jordà et al. (2017).

[57] The other alternatives are Real GDP per capita (PPP), Real GDP per capita (index, 2005=100). The preference for the first option is due to the fact that most of the other variables used are nominal, aggregate and in local currencies.

[58] The transformations of the time series are available in the Appendix in Tables A22–A24. Table A22 is related to real and international economy variables. Table A23 is about government, money, prices and interest rates. Table A24 is on credit and asset price data

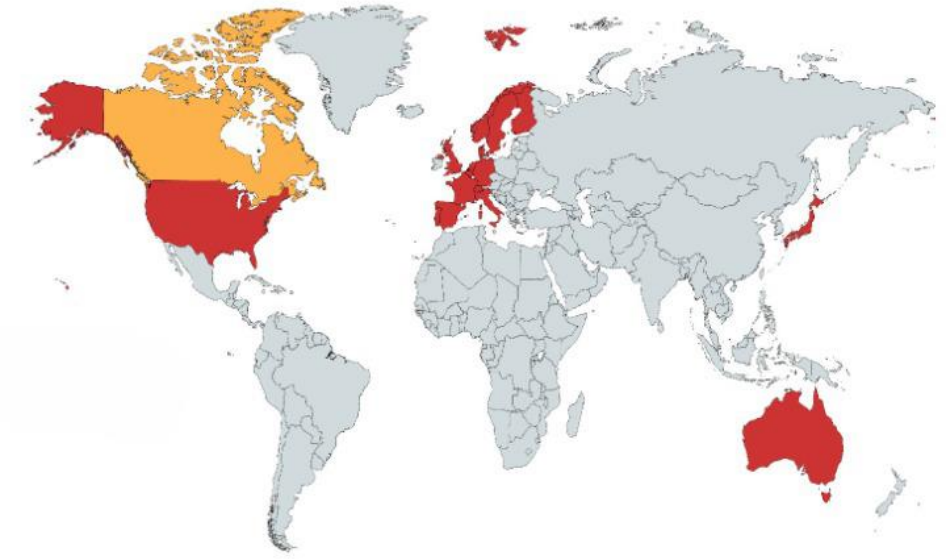

Figure 9. Countries in the Jordà-Schularick-Taylor Macrohistory Database dataset (JSTMD) (2017)

Sources: Jordà et al. (2017); Knoll et al. (2017). Countries in the dataset used (red) and not used (yellow).

First, if the missing values at the beginning of the time frame are sufficiently numerous to lead to the loss of more than one crisis, the whole time series is excluded from the estimation. Second, suppose the missing values in the middle of the time series are equal to more than 14 consecutive years (more than 10% of the total number of years considered). In that case, the time series is also excluded from the estimation. The missing data in the accepted time series are replaced by linear interpolations.[59]

The following part of the treatment concerns the analysis of the stationarity of the time series. The Augmented Dicky-Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test are performed on the time series of all the countries in the dataset. Subsequently, the time series is transformed, taking into account the results of the stationarity testing using log differentiation, differentiation, and double log differentiation. If only one of the two tests does not confirm the stationarity of the time series, the transformation is still performed.

Next, the response variable is defined as suggested by Ng (2014). However, in my work, the response variable is represented by the dummy variable "Systemic

[59] A complete list of the included and excluded time series per country is available in the Appendix. Instead of one table, the description is divided into three parts in tables A.25, A.26 and A.27. Considering that some of the missing data may be in non-standard periods or near structural breaks, robustness checks using different interpolation methods are foreseen in future research.

Financial Crises" and not recessions as in Ng's work.[60] Figure 10 illustrates the distribution of the crises per country in three different periods. The variable signals the presence of a crisis in a given year in a specific country; one is the value that indicates the presence of a crisis; otherwise, it has the variable zero.

Ng (2014) uses US monthly data to forecast up to twelve steps ahead and she employs till the 12th lag of the predictors. Given the lower time frequency of the data used here, a lag of one year is introduced. In addition, to investigate potential variations in model outcomes resulting from the data-rich environment and to examine any alterations in model performance after introducing nonlinearities, new time series are generated by applying standard power transformations, including squared and cubic transformations. A similar approach is also adopted by Bluwstein et al. (2020).

All these operations are repeated in two other sampling methods of the dependent variable. The first refers to the work of Casabianca et al. (2022) and the early warning system. The three pre-crisis years are transformed into positive cases (crises), so they take the value one, and the year of the crisis and three post-crisis years are treated as non-crises, so they are equal to zero. This is done to avoid the post-crisis bias,[61] which refers to the problem of treating years in which the effects of the crisis are still present in the economy as normal times.

The last sampling method applied in the study is the one that considers crisis periods. The crisis periods include the three pre-crisis years, the years of the crisis, and the three post-crisis years. This method of sampling tries to take into account the limits of the algorithm, which is very sensitive to the number of crises in the training set, and the economic meaning of post-crisis periods[62] and the warning in data science literature about classification methods in time series forecasting exercises (Dixon et al., 2017). The latter stresses the risk that standard over-sampling methods could alter the sample correlation structure in a significant way.

[60] Ng (2014) uses recessions as an example to demonstrate how Adaboost can help in a macroeconomic environment. However, nothing prevents the use of Adaboost for predicting financial crises, as Casabianca et al. (2022) show.

[61] For a more insightful analysis of the issue please refer to Bluwstein et al. (2020) and Bussière and Fratzscher (2006).

[62] In a post-crisis period the economy of a country is still recovering and the dynamics of the crisis are still present (Bluwstein et al., 2020).

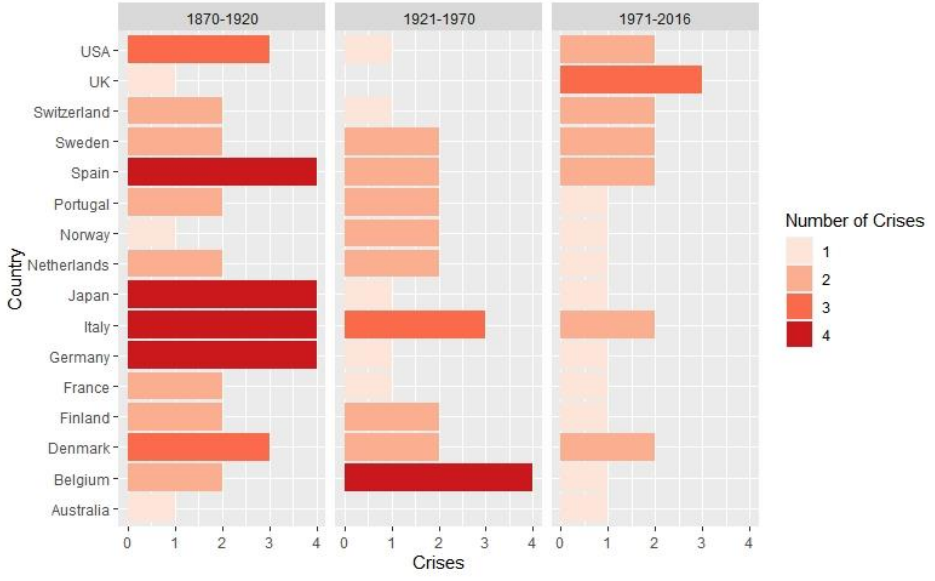

Figure 10. Crises per country and period in the Jordà-Schularick-Taylor Macrohistory Database dataset (JSTMD) (2017)

Sources: Jordà et al. (2017)
Notes: The crises per country are distributed as follows: Australia (2); Belgium (7); Denmark (7); Finland (5); France (4); Germany (6); Italy (9); Japan (6); Netherlands (5); Norway (4); Portugal (5); Spain (8); Sweden (6); Switzerland (5): UK(4); USA(6).

Adding the cubic and squared transformations of the variables in other specifications does not only make it possible to measure nonlinear relationships, but also makes specifications consistent with the definition of a data-rich environment where $N >> T$ N is the number of predictors and T is time.[63] The training set represents 50% of the data, and the test set represents the other 50% in fixed window estimations. In the expanding window estimations, the windows change the dimensions after every iteration. However, the initial window always departs from the middle of the time series.

This more conservative approach in the division of training and test sets compared to the work of Bluwstein et al. (2020) is possible thanks to the different decisions concerning the missing data and the selection of the variables, which makes it possible to retain more observations for each time series. The division of the time frame and the selection of variables is the result of the balance among these different issues.

Lastly, the pseudo-out of sample estimations are performed through the expanding window method for all of the countries, in both the baseline scenarios

[63] In the training set of all the countries $N >> T$.

and the data-rich environment.[64] The pseudo-out of sample estimations with rolling fixed windows are performed only for the countries for which the data allow the algorithm to identify the two values – the presence or absence of a systemic financial crisis in the dependent variable[65]. Expanding windows permit the preservation of more data than the rolling fixed window method. Instead, the rolling fixed window estimations allow the consideration of possible changes in the volatility of the variables.

Moreover, the variable importance of the AdaBoost algorithm is the expression of only the last decades due to the use of the previous ones to train the data. Hence, it can show if the variables detected by the expanding window estimations are still the same or if the change in the economic structure was so deep to bring new variable rankings in the algorithm.

However, the methods suggested by Dixon et al. (2017) are not only related to univariate time series but also perform a sampling that consists of an excessive number of positive cases, in this context an excessive amount of crises, in the training set. This could reduce the economic meaning of the positive events and the distortion of the information provided by the observations.

The idea of crisis-periods allows for balancing these different issues. It permits maintaining a certain distribution of events around economically meaningful periods and, at the same time, gives the algorithm more chances to predict positive events.[66] This sampling method can be considered a logit variant of the multinomial logit method used by Caggiano et al. (2014).[67] This method of oversampling can be called the *area method*. The idea is to consider the *area* around the crisis to have more rare events.

3.2.2 Methodology

The logistic regression used in the study can be illustrated with the following formula:

$$\text{logit}(p_{i,t}) = \alpha_{0,i} + \beta_{i,t}(\text{L})X_{i,t} + e_{i,t} \quad (28),$$

where $\text{logit}(p_{i,t}) = \log(\frac{p_{i,t}}{1-p_{i,t}})$ is the logarithm of the odds ratio, and $p_{i,t}$ is the probability of the dependent variable to be a success/case (crisis) rather than a

[64] The data-rich environment includes non-linear transformations of the variables. This entails to have the number of predictors to be much larger than the observations of the training dataset.

[65] The countries are Denmark, Finland, Italy, Sweden, Spain, USA.

[66] This sampling method is applied to the countries with more crises: Belgium, Denmark, Italy, Spain, and Sweden. Given the interesting results in some of the specifications related to the Casabianca et al. (2022) method with the Japanese data, Japan is included in the subgroup too. This permits Japanese data to show results also for the fixed rolling window estimations.

[67] Caggiano et al. (2014) use a multinominal logit model where the value of zero denotes the absence of crises, one the crises, two the post-crises periods.

failure/non-case (non-crisis) at time t in country i. The matrix $X_{i,t}$ represents the different time series, i denotes the country, L is the lag operator, and $e_{i,t}$ is the error term.

The stepwise logistic model (Efroymson, 1960) selects the variables applying an information criterion, in this study, the Akaike Information Criteria (AIC), following both a forward and backward elimination procedure (Venables and Ripley, 2002). The model is iterated and drops or adds variables for the full model till it reaches the more parsimonious modelization.

The AdaBoost model applied is the one discussed in subsection 2.2.1. The systemic financial crisis dummy is the dependent variable in the model, and the predictors are those selected by the dataset from Jordà et al. (2017). In terms of the parameters of the weak learner – the classification tree – the recommendations put forth by Culp et al. (2006) have been considered. Notably, the maximum depth of the classification trees was a point of focus. However, a slight modification has been introduced. Given the presence of a data-rich environment in some of the model specifications, the maximum depth of the trees is set to 10. This adjustment was made to enhance the model's flexibility in accommodating the complexity of the dataset.[68]

To evaluate the performance of the models, the standard misclassification error rate (ME), the precision, the recall, and the F1 score measures are used. The ME is a standard technique to understand the degree of error of different classifiers.[69] The ME rate measures the average of the incorrect classifications of the model after the prediction in the test set.

$$\mathrm{ME} = \frac{fp+fn}{tp+fp+tn+fn} \tag{30},$$

where fp are the false positives, fn are the false negatives, tp are the true positives, and tn are the true negatives. The importance of the variable in the AdaBoost models is evaluated on the basis of values obtained by variables considering how frequently they are selected for boosting. If a variable is selected more frequently, it means that it contains useful information for classification (Culp et al., 2006). However, given the problem of the imbalanced data of the target variable, evaluating the performance of the different models could result in the wrong conclusions, especially concerning ML methods such as AdaBoost. The algorithm could simply reproduce a large series of non-crisis periods. Given the low number of negative events (crises), this could bring naive conclusions about the performances of the models.

In order to avoid this issue, the F1 score measure, also called the Sorensen-Dice coefficient (Sorensen, 1948; Dice, 1945), is performed. The F1 score is the harmonic mean of precision and recall, and it is one of the most used metrics for classification exercises. It is a simple ratio with a clear meaning, and it is easy to

[68] Attempts with different maximum depth values did not significantly change the results.

[69] See James et al. (2013).

interpret. Even if many other metrics have been made available in recent decades (Wardhani et al., 2019), the F1 score is a commonly used evaluation metric in the literature (i.e. Papadopoulos et al., 2019; Lin and Jin, 2023) that enables, in principle, a comparison with previous studies and it fits the scope of the present work where rare events are present. In fact, the F1 score is a better metric compared with ME when imbalanced data are present. In fact, the latter is not able to properly evaluate data where positive occurrences are much more rare than negative ones. Formally, precision can be defined as:

$$\text{Precision} = \frac{tp}{tp+fp} \tag{31},$$

Recall can be defined as:

$$\text{Recall} = \frac{tp}{tp+fn} \tag{32},$$

Finally, the F1 score is:

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{33},$$

If precision, recall, and the F1 score are equal to one, this implies the best results obtainable by a model. Instead, a result equal to zero, corresponds to the worst result obtainable. The misclassification error operates inversely; the closer the result of the evaluation criteria is to zero, the better the model performs. Conversely, a result nearing one indicates poorer performance.

3.2.3 Main Results

The results for Italy and Japan[70] are reported in this subsection as examples. Every table has four specifications: two specifications with fixed window estimations, with or without US exchange rate variables; and two specifications with expanding windows, with or without US exchange rate variables. The decision to present only these two countries[71] is due not only to the need to reduce the length of the thesis and the number of tables to report, but also because these two countries are where the performances of AdaBoost are more consistent and meaningful both in terms of the results of the classification exercise and variable importance selection and deserve special attention.

Moreover, the literature has a special interest in these two countries. Romer and Romer (2018) focus on these two countries in their study on the effects of macroeconomic policies in the aftermath of a crisis. The failure of other countries in the dataset can be due to different causes: the lack of enough systemic financial

[70] In the main text the results of the new method of sampling are illustrated, while the baseline and Casabianca's methods are reported in the Appendix.

[71] The results of the other countries are available under request.

crises in general for training the model (for instance in the case of Australia), the distribution of the crises that diminishes the power of the oversampling methods, and that the predictors used to train the model are not relevant to predict the crises in the test set.

Italy

This section reports the results with Italian data using the new method of oversampling – the *area method*, on both linear (Table 7) and nonlinear variables (Table 8). Italy is one of the most interesting cases in the study, given the high number of crises in the baseline scenario (9). The high number of crisis-periods in Italian data (56) appearing with the new method of sampling permits us to observe the F1 scores in all the specifications of Table 7 and Table 8, even if the results are unstable. This instability can be related to the issues observed by Ng (2014) connected with the relations among the predictors in time series and how the relevant predictors can evolve over time. The latter is even more important in the analysis, considering that the time frame in this empirical study is indeed long.

For what concerns the MEs, AdaBoost is still the best performer, and logistic regression outperforms stepwise logistic regressions only in the fixed window estimations with linear data. The F1 scores for AdaBoost are the worst among different models. However, it can be noticed that in the fixed window estimations with nonlinear variables (Table 7), the F1 scores for AdaBoost increase considerably in comparison to other specifications.

It can be noticed how the ME and the F1 scores are not stable[72] in most of the specifications in the tables related to the method from Casabianca et al. (2022) and the new method suggested in this work, where crisis-periods (in this thesis it corresponds to crises, and pre-crisis and post-crisis periods as a whole) are considered. Instability issues also arise in the work of Ng (2014) even though her exercise concerns recessions and not crises. For this reason, to extrapolate variable importance, the algorithm is run six times, divided between two different sessions.[73] The variables in the variable importance plots that are constantly among the top three variables are considered as robust results.[74] The rest of the Italian results using the baseline method of sampling and that of Casabianca et al. (2022) are reported in the Appendix. Table A.28 shows the baseline method with only linear variables. The F1 score in the last specifications for logistic regression can be observed thanks to the presence of such a large amount of crises in the Italian data.

---

[72] For this reason they are reported in italics.

[73] This means that R software is first turned off and restarted afterwards. This is a further control implemented to check that there are no other factors causing the repetition of the same results of the estimations only by chance.

[74] The linear and nonlinear transformations of variables are considered as different variables. Instead, the current values and their lags are analysed as the same variable.

Table A.29 presents the baseline method with nonlinearities included. It shows the same results in the first three specifications. The ME of AdaBoost is the lowest among the models in all the specifications in tables A.28 and A.29. Logistic regressions are the worst in terms of ME, and their results become worse when the nonlinear variables are inserted.

Table A.30 illustrates the method from Casabianca et al. (2022) for Italian data. Table A.31 shows the same results with nonlinear variables included. AdaBoost ME is the lowest among all the specifications. In Table A.30 the ME are unstable in three out of four specifications. Instead, inserting the nonlinear variables in Table A.31 stabilises the ME apart in the first specification. Stepwise logistic regressions still perform better than logistic regressions in the specifications with expanding windows in Table A.30 and in all the specifications in Table A.31. A discussion of the results is presented in the next subsection and in the fourth chapter of the thesis.

Table 7. Estimations using the area method of sampling and linear variables, with or without US exchange rate – Italian data

		Fixed est.		**Expanding est.**	
	Models	**no US XR**	**with US XR**	**no US XR**	**with US XR**
Number of independent variables		34	36	34	36
Number of observations		144	144	144	144
Number crisis-periods exploited		56	56	56	56
ME	**logit**	0.36	0.39	0.43	0.43
Precision	**logit**	0.23	0.25	0.19	0.19
Recall	**logit**	0.36	0.5	0.36	0.36
F1score	**logit**	0.28	0.33	0.24	0.24
ME	**stepwise logit**	0.39	0.42	0.36	0.38
Precision	**stepwise logit**	0.21	0.21	0.17	0.16
Recall	**stepwise logit**	0.36	0.42	0.24	0.21
F1score	**stepwise logit**	0.26	0.29	0.19	0.18
ME	**AdaBoost**	*0.3*	*0.3*	*0.3*	*0.3*
Precision	**AdaBoost**	*0.08*	*0.08*	*0.1*	*0.13*
Recall	**AdaBoost**	*0.07*	*0.07*	*0.07*	*0.07*
F1score	**AdaBoost**	*0.08*	*0.08*	*0.08*	*0.09*

Notes: Unstable values in italics. The instability signals that the results are not the same when the models are estimated multiple times. However, some frequent values emerge and the most frequent results are the ones reported.

Table 8. Estimations using the area method of sampling and nonlinear variables included, with or without US exchange rate – Italian data

	Models	Fixed est.		Expanding est.	
		no US XR	with US XR	no US XR	with US XR
Number of independent variables		102	108	102	108
Number of observations		144	144	144	144
Number crisis-periods exploited		56	56	56	56
ME	**logit**	0.58	0.43	0.49	0.46
Precision	**logit**	0.17	0.24	0.23	0.23
Recall	**logit**	0.5	0.57	0.64	0.57
F1score	**logit**	0.25	0.34	0.34	0.33
ME	**stepwise logit**	0.35	0.46	0.29	0.44
Precision	**stepwise logit**	0.28	0.17	0.32	0.09
Recall	**stepwise logit**	0.5	0.26	0.43	0.14
F1score	**stepwise logit**	0.36	0.23	0.36	0.11
ME	**AdaBoost**	*0.2*	*0.2*	*0.2*	*0.2*
Precision	**AdaBoost**	*0.18*	*0.18*	*0.13*	*0.11*
Recall	**AdaBoost**	*0.14*	*0.14*	*0.07*	*0.07*
F1score	**AdaBoost**	*0.16*	*0.16*	*0.09*	*0.09*

Notes: Unstable values in italics. The instability signals that the results are not the same when the models are estimated multiple times. However, some frequent values emerge and the most frequent results are the ones reported.

Japan

Table 9 shows the results for Japanese data with only linear variables and with the new method of sampling explained in the subsection 3.2.1, and Table 10 illustrates the same data with the new method of sampling and nonlinear variables included. It can be observed how AdaBoost outperforms the two benchmark models both in terms of ME and F1 scores in almost all the specifications using the new method of sampling without (Table 9) and with nonlinear variables (Table 10). Furthermore, it is possible, thanks to the new method, to observe the fixed window estimations. Moreover, the F1 score is stable in the second specification of Table 10, around the value of 0.56.

It can be observed that the number of crisis periods exploited is 35, a number inferior to the 56 crisis periods of the Italian data. In addition, the number of crisis periods is bigger in the Italian case, even in relative terms. The number of crisis periods for Japan is equal to 26.92% of the dependent variable's total number of observations. By comparison, the number of crisis periods for Italy is equal to 38.89%.

The number of variables employed is almost the same for both countries, and the proportion of the number of crises in the training sets is still in favour of the Italian data. However, it can be noticed that the number of crisis periods in the Italian test sets is greater than in the Japanese ones, where most crises are con-

centrated in the training sets. It is also possible that the variables used for the estimations are just more relevant for Japan than for Italy.

As for the Italian ones, the other Japanese results concerning the baseline method and the method from Casabianca et al. (2022) are reported in the Appendix. Table A.32 describes the Japanese results in the baseline scenario with linear variables only. Table A.33 shows the Japanese results in the baseline scenario with nonlinear variables included. The number of crises does not allow us to conduct fixed window estimations. The ME of AdaBoost is again the best among the models, and the logistic regression ME is the worst. The logistic regression ME is considerably worse in Table A.33 in all the specifications compared to other models.

In Table A.34 using the method from Casabianca et al. (2022) of sampling and linear variables only, logistic regression performs slightly better in terms of ME when it is in a linear environment and in fixed window estimations. The results with nonlinear variables included (Table A.35) are the same. AdaBoost is still the best performer in terms of ME, but all its results are unstable. What is more intriguing are the two last specifications in Table A.34. Here, even if highly unstable, the F1 scores largely outperform both logistic and stepwise logistic regression F1 scores. These results are the principal motivation to apply the new method also to Japanese data.

Table 9. Estimations using the area method of sampling and linear variables, with or without US exchange rate – Japanese data

		Fixed est.		**Expanding est.**	
	Models	**no US XR**	**with US XR**	**no US XR**	**with US XR**
Number of independent variables		32	34	32	34
Number of observations		130	130	130	130
Number crisis-period exploited		35	35	35	35
ME	**logit**	0.51	0.52	0.62	0.52
Precision	**logit**	0.09	0.15	0.13	0.03
Recall	**logit**	0.43	0.86	0.86	0.14
F1score	**logit**	0.15	0.26	0.23	0.06
ME	**stepwise logit**	0.38	0.22	0.26	0.46
Precision	**stepwise logit**	0.15	0.32	0.27	0.13
Recall	**stepwise logit**	0.57	0.86	0.86	0.57
F1score	**stepwise logit**	0.24	0.46	0.41	0.21
ME	**AdaBoost**	*0.2*	*0.2*	*0.2*	*0.2*
Precision	**AdaBoost**	*0.31*	*0.33*	*0.27*	*0.33*
Recall	**AdaBoost**	*0.57*	*0.86*	*0.43*	*0.43*
F1score	**AdaBoost**	*0.4*	*0.4*	*0.3*	*0.4*

Notes: Unstable values in italics. The instability signals that the results are not the same when the models are estimated multiple times. However, some frequent values emerge and the most frequent results are the ones reported.

Table 10. Estimations using the area method of sampling and nonlinear variables included, with or without US exchange rate – Japanese data

	Models	Fixed est.		Expanding est.	
		No US XR	with US XR	no US XR	with US XR
Number of independent variables		97	103	97	103
Number of observations		130	130	130	130
Number crisis-period exploited		35	35	35	35
ME	**logit**	0.48	0.46	0.46	0.46
Precision	**logit**	0.1	0.07	0.1	0.13
Recall	**logit**	0.43	0.29	0.43	0.57
F1score	**logit**	0.16	0.12	0.17	0.21
ME	**stepwise logit**	0.26	0.25	0.31	0.28
Precision	**stepwise logit**	0.22	0.26	0.22	0.26
Recall	**stepwise logit**	0.57	0.71	0.71	0.86
F1score	**stepwise logit**	0.32	0.38	0.33	0.4
ME	**AdaBoost**	*0.1*	0.12	*0.1*	*0.1*
Precision	**AdaBoost**	*0.45*	0.42	*0.4*	*0.3*
Recall	**AdaBoost**	*0.71*	0.71	*0.43*	*0.43*
F1score	**AdaBoost**	*0.5*	0.56	*0.4*	*0.5*

Notes: Unstable values in italics. The instability signals that the results are not the same when the models are estimated multiple times. However, some frequent values emerge and the most frequent results are the ones reported.

Variable importance in AdaBoost estimations

Tables 11 and 12 show the variable importance of AdaBoost using the three different sampling methods. Table 11 shows them when the estimations are conducted using fixed window estimations and Table 12 when expanding window estimations are conducted. Both results with and without US dollar exchange rate against the currencies of the countries considered are included. It can be observed that the presence of the US dollar exchange rate does not affect the results substantially. Only in some specifications concerning Italy we can observe some changes. The results for the other countries examined are also commented upon and are available on request.

Comparison by method of sampling

When considering the linear and nonlinear specifications for the baseline scenario, two main findings can be observed. First, even if the majority of the specifications of the baseline scenario are not able to produce clear variable importance, nevertheless, some results are robust. Second, the specifications with expanding window estimations and nonlinear variables can more frequently

produce stable results (Table 12) in the sense that the results tend more often to be the same when the estimation is replicated.

This stability is attributed not only to the expanding window estimation method, which provides a substantial number of training observations in the latest windows, but also to the presence of a data-rich environment. Upon examining the variable importance derived from the specifications using the method from Casabianca et al. (2022), it becomes apparent that this sampling method does not increase considerably the total number of specifications in which AdaBoost can identify important variables related to systemic financial crises. Interestingly, in the case of expanding window specifications (see Table 12) and in the presence of a data-rich environment, the number of countries where AdaBoost can distinctly identify crucial variables for crisis detection increases[75].

Table 11. Variable importance by country – fixed window estimations

Data	Only linear terms	Data-rich environ-ment	Only linear terms	Data-rich environ-ment	Only linear terms	Data-rich environment
Oversampling method	No over-sampling	No over-sampling	Casabianca et al. (2022)	Casabianca et al. (2022)	Area method	Area method
Italy without US XR	Unclear	Unclear	*Import*	Unclear	*Import*	*DebtGDP_-cub*
Italy with US XR	Unclear	Unclear	*Import*	Unclear	Unclear	*CA_sqr*
Japan without US XR	–	–	–	–	*Loansto-Nonfin, GDP*	*Loansto-Nonfin, DebtGDP_-sqr*
Japan with US XR	–	–	–	–	*Loansto-Nonfin, GDP*	*Loansto-Nonfin, DebtGDP_-sqr*

Notes: "–" not able to estimate; "Unclear" means that the results are not robust. The endings "sqr" and "cub" denote the squared and cubed variables. Current Account (CA), Debt over GDP (DebtGDP), Total Loans to Non-financial Private Sector (LoanstoNonfin). Data-rich environment includes both linear nonlinear terms.

[75] All these considerations are even more evident after performing the variable selection for all the countries of the dataset.

Table 12. Variable importance by country – expanding window estimations

Data	Only linear terms	Data-rich environ-ment	Only linear terms	Data-rich environ-ment	Only linear terms	Data-rich environment
Oversampling method	No over-sampling	No over-sampling	Casabianca et al. (2022)	Casabianca et al. (2022)	Area method	Area method
Italy without US XR	*Import*	*Export_cub*	*Import*	*Export-cub*	Unclear	*Import_cub*
Italy with US XR	*Import*	*Export_cub Loansto-Nonfin_cub*	*GovExpend*	*Export-_cub*	Unclear	*CPI_cub*
Japan without US XR	Unclear	*Loansto-Nonfin*	*Loansto-Nonfin*	*StockPr-_cub*	*Loansto-Nonfin*	*StockPr-_cub*
Japan with US XR	Unclear	*Loansto-Nonfin_cub*	*Loansto-Nonfin*	Unclear	*Loansto-Nonfin*	*StockPr-_cub*

Notes: "Unclear" means that the results are not robust. The endings "sqr" and "cub" denote the squared and cubed variables Total loans to the non-financial private sector (LoanstoNonfin), Stock prices (StockPr), Government expenditure (GovExpend), Inflation (CPI). Data-rich environment includes both linear and nonlinear terms.

Comparison of the results country by country

After the commonalities among countries have been examined table by table, the study evaluates whether countries with the same variables selected by AdaBoost also present similar results table by table. Tables 11 and 12 show the baseline scenario where AdaBoost identifies the importance of credit variables to detect crises in the case of Japan. Conversely, AdaBoost appears sensitive to international economic variables for Italy. Indeed, Italy is an economy where exports are relevant.[76]

Where the method of oversampling from Casabianca et al. (2022) is used, AdaBoost selects credit variables to detect pre-crisis periods for Japan. Asset prices show a certain importance in some specifications for Japan. This can possibly be due to its highly developed capital market. The results on Italy maintain the importance of international economic variables in the majority of the specifications but it can also be observed how the variable related to government expenditures is selected in expanding window estimations when US exchange rates are introduced in the set of predictors (Table 12).

[76] From World Bank data, Italy has an above OECD average ratio of export over GDP. Moreover, if the ratio of the exports of manufacturing goods over total merchandise export is considered, Italy has the second higher score in Europe among the main economies of the continent after Germany.

The results using the area method of oversampling indicate how international economy variables are still relevant in Italy. It is intriguing to notice that government variables are relevant in the last specification with fixed window estimations for Italy in Table 11, even if they are less relevant than international economy variables (Table 11 and 12). Moreover, when the US exchange rate is introduced, inflation is selected by AdaBoost.

For what concerns the other countries not shown in this thesis, it can be observed that in the baseline estimations, credit variables are relevant for Australia, Finland, Germany, and Switzerland. International economic variables are selected by AdaBoost for Norway and Sweden, which are small open economies. The results for USA, Sweden, Switzerland and the United Kingdom also show asset prices being selected by AdaBoost when the oversampling from Casablanca et al. (2022) is applied. Finally, variables related to money, prices and interest rates are relevant for countries like Denmark, Germany, Norway, Portugal, Spain and Switzerland.

When the area method of sampling the real economy variables are relevant for Spain and Sweden and international economy variables are important for Belgium and Sweden.

3.2.4 Evaluations of the models

It can be observed how the ME for AdaBoost is lower than both stepwise logistic regressions and logistic regressions in all the specifications. However, it is not sufficient to check the mere ME to evaluate the models. Indeed, the results of the F1 score describe the challenges faced by the AdaBoost algorithm, both in the baseline method and in the majority of cases with the method used by Casabianca et al. (2022) to manage imbalanced data in the dependent variable.

This is because tranquil times, when there are no crises, significantly outnumber rare events when crises occur and a standard oversampling of the data cannot be sufficient to improve the predictive performances of the algorithm. Moreover, even in scenarios where AdaBoost performs well in terms of the F1 score due to a sufficient number of positive events, it rarely outperforms logistic regressions and stepwise logistic regressions. AdaBoost typically performs better in high-dimensional spaces where nonlinearities are included (e.g., Table 8 and Table 10). However, it is the combination of the method of estimations, the presence of data-rich environments, and the method of sampling that creates situations where AdaBoost outperforms other methods (e.g., Table 9, the last specification). Table 13 reports a summary of the results.

Table 13. Summary of the findings

Topic	Results
Expanding vs fixed window estimations	Expanding estimations work better for AdaBoost in the present framework.
Data-rich environment	Adding nonlinear terms increases the probability that AdaBoost can identify important variables to detect crises, and pre-crisis and crisis-periods.
Method of sampling	Obtaining significant results with AdaBoost is extremely important, but it is not the only aspect to be taken into account.
AdaBoost vs benchmark models	The combination of the method of sampling, the presence of data rich-environment and the method of estimations is relevant to allow AdaBoost to beat, or at least to compete with logistic regressions and stepwise logistic regressions.
Variable importance by country	Credit variables are important for predicting systemic financial crises but country-specific aspects also have to be taken into account.

Moreover, to further validate the method used to enhance AdaBoost's results, a placebo exercise was conducted. Placebo exercises are mostly employed in medicine but applied social sciences research (Eggers et al., 2024) have started to use them to check the validity of their models. For example, in order to estimate the elasticity of initial unemployment insurance (UI) claims, Aaronson et al. (2022) conduct a state-level event study that focuses on the most relevant hurricanes that have hit the US mainland since 2004. Moreover, they employ a placebo test by designating certain US states as having experienced the rare event (a hurricane), even though they have not.

In this exercise, a randomly generated distribution, reflecting the same percentages of crisis-period occurrences (39% for Italy and 27% for Japan), was used to replace the original dependent variable. In this placebo dependent variable, the rare event (a crisis) can occur in years where it does not show in the actual data of the original dependent variable. The exercise is repeated for 10 randomly generated placebo dependent variables using the nonlinear specifications for both Italian and Japanese data, with fixed and expanding window estimations. To address the issue of instability, each randomly generated dependent variable is forecasted three times, resulting in a total of 30 attempts.

For the Japanese data, in fixed window estimations, the F1-score ranges from a minimum of 0.09 to a maximum of 0.25. In expanding window estimations, the scores range from 0.03 to 0.22. It is also observed that, for the Japanese estimations with placebo dependent variables, AdaBoost produces indeterminate results in 46.7% of the fixed-window attempts and 60% of the expanding-window attempts. For the Italian data, the F1-scores range between 0.05 and 0.33 in fixed-window estimations and between 0.09 and 0.33 in expanding-window estimations.

The differences in the results between the Japanese and Italian data, as also shown in Tables 8 and 10, indicate that the number of crises per country is not the sole factor influencing performance using the area method. As explained in Subsection 3.2.3, the Japanese data include more crises in the training set, which

may play a more significant role. Additionally, it cannot be excluded that more relevant variables should be added to obtain better forecasts with the Italian data. Furthermore, additional hyperparameter tuning was conducted, focusing on the shrinkage parameter of AdaBoost, the maximum depth of any node in the final tree,[77] and the number of iterations by the algorithm. Although this robustness check does not entirely resolve the issue of instability, it does provide some valuable insights.

Overall, it can be observed that, for both the Italian and Japanese data, increasing the maximum depth does not significantly improve results, even when set to 30, which is the maximum value recommended for obtaining meaningful outcomes.[78] Reducing the shrinkage parameter can lead to a slight increase in the F1-score in expanding window estimations, but it also introduces greater instability. For instance, with Italian data, the F1-score can peak at 0.3 but may also drop as low as 0.08. In contrast, when the baseline shrinkage value of 0.1 is maintained, the results remain more stable around 0.1 when the number of iterations is pushed towards 500. Regarding the number of iterations, increasing them beyond 200 does not appear to significantly reduce instability or further improve the F1-scores.

[77] This is the distance along the tree's longest path from the root node to the leaf node that is the furthest away.

[78] See the rpart package in R.

4 DISCUSSION AND CONCLUSIONS

4.1 Discussion of the first empirical study: forecasting industrial production

The number of studies on forecasting industrial production and other macroeconomic indicators using ML methods is growing. Accurate forecasting of industrial production can help quickly understand what kinds of actions should be taken in response to a possible downward trend in the business cycle. However, a clear study using different boosting models that investigates how the distinct parts of the models affect the forecasts of industrial production is still lacking in the literature. A closer analysis of whether boosting can outperform standard econometric methods in forecasting industrial production and why would be relevant for many institutions.

This study forecasts the average growth rate of monthly US industrial production from 1963 to 2019 using different boosting models. The results are compared with other machine learning methods and different econometric benchmarks – an autoregressive model, a random walk, and a factor model. Autoregressive and random walk models are standard univariate benchmarks used in macroeconomic forecasting (Goulet Coulombe et al., 2022; Medeiros et al., 2021). However, Girardi et al. (2017) indicate that when a study analyses what is the best model for selecting or providing better information using a large amount of predictors, a simple comparison with a univariate benchmark is not sufficient. For this reason, a factor model is introduced as a multivariate benchmark. Examples of a similar use of factor models as benchmarks can also be found in Cepni et al. (2019), Soybilgen and Yazgan (2021), and Syed and Lee (2021).

It can be noticed how AR(p) and FM perform fairly similarly, and instead, RW performs worse among the benchmark models (Tables 2 and 3). This is in line with previous literature, where RW is usually the worst among benchmark models (Exterkate et al., 2016; Soybilgen and Yazgan, 2021; Mogliani and Simoni, 2021).

This study shows that boosting models and the other machine learning models generally beat RW in all estimation methods and samples. Boosting models are able to perform better against RW than shrinkage models, or at least, they are good competitors.

Boosting models are also more successful than shrinkage models when competing with the factor model. In particular, AdaBoostRT, component wise boosting, and XGBoost outperform the factor model forecasts. AdaBoostRT beats the factor model only in short-horizon forecasts (such as h=1), while XGBoost is more efficient in medium-horizon forecasts (h=4). Component wise boosting is, instead, better in all the forecast estimations. These results confirm certain previous findings in the literature concerning machine learning methods and business cycles.

The results of AdaboostRT could be related to the study by Yousuf and Ng (2021) that emphasizes the potentialities of the stagewise procedure of boosting

models and the importance of the iterations in boosting models. Indeed, AdaBoostRT can be interpreted as a type of two-stage boosting model. Moreover, even if Yousuf and Ng (2021) argue the relevance of the number of iterations used with a boosting model, considering the results of AdaBoostRT, what it can also count is the iteration method of the parameters of the weights.

The comparison with AR(p) shows that only component wise boosting and AdaBoostRT (in the short-run with expanding window estimations) can beat AR(p). However, the differences in the results are not significant. By contrast, XGBoost beats AR(p) in the medium-run (h=4), and the differences in the results are statistically significant. It can be noticed how the literature about machine learning and business cycles finds that AR models, and in some cases even very simple AR models, are the most challenging standard benchmark models to beat (Maehashi and Shintani, 2020; Yousuf and Ng, 2021).

Considering the results of the boosting models in more detail, the results from the component wise boosting model could be related to the strong factor structure of the data, as suggested by Bai and Ng (2009). Moreover, the results of component wise boosting partially diverge from Buchen and Wohlrabe (2011), given its good performances in the short-run forecasts and not only in the long-run. However, the authors employ a different factor model that could explain this difference. Finally, the literature overall agrees that the combination of machine learning methods and factor models, such as the component wise boosting version, can strengthen machine learning models in forecasting the US business cycle (Goulet Coulombe et al., 2021a), even if this may not hold for all countries (Panagiotelis et al., 2019).

The results of component wise boosting and AdaBoostRT show that the differences in the algorithms of the boosting models can improve or decrease their forecasting performance. In this sense, boosting models deserve more attention from scholars. To the author's best knowledge, this study is the first to use more than two boosting models and the first to employ such a large battery of them to forecast US industrial production. For example, the performances of AdaBoostRT differ substantially from AdaBoostR2 along all of the different estimations, even if both models are derived from AdaBoost. The two models differ in how the error term is considered and how the weights of the predictors in the new iterations are updated. The presence of a threshold in AdaBoostRT could play a role because it is in a way a sort of further penalization in the estimations.

XGBoost is considered the most advanced and complex boosting model. Chen and Guestrin (2016) show how one of the most important characteristics of XGBoost is the splitting method it uses in regression trees. The inferior results of the other boosting methods could be related to overfitting caused by an inefficient method of splitting. This idea is supported by Borup et al. (2023), who explain how placing the splits along stronger predictors can be relevant.

Furthermore, it can be useful to take into account the findings of Chu and Qureshi (2022) that complex models with large degrees of freedom are more sensitive to changes in the parameters. In this sense, there is an increasing risk of overfitting when many predictors are forecasted. XGBoost is the most sensitive

to this issue because it has a large number of hyperparameters to fine-tune (Chu and Qureshi, 2022). The results of XGBoost can be interpreted as follows: XGBoost performs much worse in the long horizon, not only due to the complexity of the model but also due to the reduction of the data used to train the model.

In order to observe potential differences due to the presence of the biggest downturn in the US economy before the Covid-19 pandemic, a subsample of the data before the Great Recession is also analysed. It can be noted how the results for all ML models against AR(p) are worse than when the Great Recession is included. This result could be connected with the findings of Goulet Coulombe et al. (2022). The authors argue that in the presence of recessionary periods, ML methods improve their results against AR(p).

Given that the period under observation starts from the 1960s and finishes before the Covid-19 pandemic, the Great Recession is the largest nonlinear movement of the business cycle in the study. It could be hypothesised that the improvement of ML methods in the full sample is due to the presence of large nonlinearities in turbulent times. These results are in line with those of Soybilgen and Yazgan (2021) and Guerróon-Quintana and Zhong (2023).

Moreover, observing the estimations before the Great Recession, highlights how AdaBoostRT beats the factor model in all the horizons, and these results are significant (Table 5). Considering that to date AdaBoostRT has not been explored using this dataset within the existing literature, there are no comparable results for direct comparison. However, it could be argued that the two-stage procedure framework of the algorithm may contribute to the effectiveness of the machine learning method compared to the factor model.

To test and compare the different models, pseudo-out-of-sample estimations using both fixed and expanding window methods are performed. The expanding window method is better for preserving more observations in the out-of-sample dataset. The fixed window method allows one to take into account the volatility of the time series in different periods and the possible changes in the structure of the data over time. It can be noticed that XGBoost and AdaBoostRT perform worse in expanding window estimations (Table 4). This means that the two models need to consider the changes in the structure of the economy over time to perform more efficiently. However, in general terms, estimating with expanding or fixed windows does not change the results of the ML models significantly.

4.2 Discussion of the second empirical study: predicting crises

The second empirical study aims to understand if AdaBoost can be useful in detecting financial crises, especially in the presence of nonlinearities. The data are taken from the dataset of Jordà et al. (2017). The data includes relevant variables related to financial crises (Jordà et al., 2017). Financial crises have a different structure to the business cycle, and the relationship is highly nonlinear.

Considering previous literature (Casabianca et al., 2022), AdaBoost should be useful for efficiently detecting systemic financial crises when variables derived from the macroeconomic literature are introduced to the model. Indeed, the similarities with logistic regression plus its structure based on trees should make it a valid competitor for standard classification methods in identifying crises, especially when nonlinearities are involved.

In particular, the study attempts to detect systemic financial crises using the AdaBoost machine learning method in sixteen countries from 1870 to 2016. It compares the results with two benchmark models: the logistic and stepwise logistic regressions. AdaBoost can compete with the two benchmark models but only in particular data environments.[79] Indeed, the combination of the method of estimation, the presence of a data-rich environment, and the method of sampling of the dependent variable determine the performance of AdaBoost against the two benchmark models. Some of the results are consistent with previous literature. The improvement in the forecasting capabilities of machine learning methods for predicting systemic financial crises appears to depend on the number of predictors, as Ward (2017) and Liu (2023) have demonstrated.

In terms of the sampling method, it is evident that the issue of dealing with imbalanced data significantly affects the forecasting abilities of any classifier in the existing literature. This is why most authors choose to either not consider country-specific data and estimate models for groups of countries (Casabianca et al., 2022) or employ panel data where the model is trained on the entire dataset and then tested on a single country (Papadopoulos et al., 2019). However, even with these strategies, the results are not consistently satisfactory (Beutel et al., 2019), and they can vary depending on the choice of evaluation criteria (Papadopoulos et al., 2019).

In this empirical exercise, an initial attempt was made to use only each country's predictors to identify systemic financial crises, and pre-crisis and crisis periods. Although the results may not appear highly encouraging at first glance, they do show promise. The literature often lacks suitable evaluation criteria to account for the imbalanced data issue. Furthermore, it is worth noting that this thesis has been used a relatively simple AdaBoost method. More complex AdaBoost variations, as shown in Liu et al. (2022), could potentially yield better results.

Another approach to forecasting systemic financial crises based on country-specific time series is presented by du Plessis (2022). In his multilogit models, crises, and pre-crisis and post-crisis periods, and tranquil times are categorised with values ranging from 1 to 4, making the dependent variable categorical rather than binary. In this framework, the standard AdaBoost may not be suitable, and alternative machine learning methods or adaptations of AdaBoost may need to be explored.

For what concerns the variable importance extracted by AdaBoost, the results related to credit variables in both Table 11 and 12 are in line with previous studies

[79] The particular data environment in this context refers to the presence of non-linearities and/or many predictors, which can be described as a data-rich environment.

on the same dataset, such as Schularick and Taylor (2012), Gourinchas and Obstfeld (2012), and they have recently been confirmed by the research of the Bank of England (Bluwstein et al., 2020). Schularick and Taylor (2012) also suggest that the boom and bust of asset prices in highly developed economies with deep financial systems can trigger systemic financial crises. Rancière and Tornell (2016) stress how excessive financial liberalization is one of the causes of the US housing crisis. The importance of stock price variables related to Japan (Table 12) could be considered as a result in line with those observations.

The iteration between loose monetary conditions, credit, and the housing prices boom could bring on systemic financial crises, while the lack of fiscal space is important in the aftermath of the crises (Jordà et al, 2016; Jordà et al., 2015). The Italian result with a data-rich environment and area method of sampling in Table 11, could be connected with the lack of fiscal space. Certainly, it can be observed that the specification where the debt over GDP ratio is relevant is in fixed window estimations. This means that the algorithm only considers the last decades for computing variable importance. Indeed, the public balance sheet of the Italian Republic has deteriorated in recent decades, and in the case of crises, it is difficult for the government to intervene massively, as in the USA in 2008 (Buera and Nicolini, 2020). Indeed, the results are obtained with the new method that included the post-crisis periods.

Regarding the findings of international economy variables, even if many economic scholars claim that the current account is no longer as relevant as it was in the past to evaluate the macroeconomic and financial stability of a country or predict banking crises, some theoretical and empirical studies suggest the opposite.[80] Indeed, there can be a relationship between systemic financial crises and trade, as observed by Bussière et al. (2013). The results with the new method of sampling in Table 11 and Table 12 for Italy confirm this possibility. The importance of inflation in the specification with the US exchange rates included could be connected with the period of inflation and competitive devaluations which affected Italy during the 70s and 80s in Table 12.

Considering the unreported results of other countries, in terms of variable selection, as expected, a great heterogeneity of the countries can be observed during the last 150 years concerning the importance of the variables in detecting crises. While each country exhibits its unique characteristics, some commonalities emerge. It is important to note that the methodology of the study does not allow for distinctions across different periods due to data constraints. Nevertheless, the analysis effectively compares outcomes across diverse country types, including small, medium and large nations.

It can be observed that the market size of a state can influence its degree of openness towards international trade and financial liberalization (Alesina, 2003). Some possible examples are big economies such as Japan and the USA, where financial structures are more developed but where the openness to international

[80] See Davis et al. (2016), Obstfeld (2012), and Barrell et al. (2010).

trade is less pronounced than small open economies. Moreover, countries with greater production can experience greater productivity (Grossman and Helpman, 1991) due to the presence of a large demand and big enterprises. Such an environment is not only prone to crises provoked from abroad but also domestic ones (see Japan's results using the new method of sampling, Table 11).

4.3 General discussion and future investigations

Contrary to what happened after 1929 crisis (Hawkins, 2005), interest in forecasting has not diminished after the Great Recession,[81] but instead, has grown in sophistication. Scholars and professionals are continuously looking for new models or variables that could help them predict business cycle downturns or an upcoming systemic financial crisis.

Forecasting industrial production or identifying periods of rare events, such as systemic financial crises, are not exercises which can be considered identical, given the different structures of the two phenomena. However, two schools of thought among economic scholars are present in both types of studies. The first believes that the pre-selection of the variables (Girardi et al., 2017; Golinelli and Parigi, 2007) based on macroeconomic theory and previous empirical exercises can guarantee good forecasts. The second is more open to a data-driven perspective (Medeiros et al, 2021; Stock and Watson, 1998) and is convinced that the nonlinearities inside both business cycles and financial crises justify the presence of a large number of variables (Stock and Watson, 2006) in forecasting exercises.

In this thesis, both these sides of the literature are considered. From one side, the variables used in this thesis are selected from both empirical and theoretical macroeconomic literature. From the other side, the use of a large number of predictors is used. Moreover, in the second exercise the thesis introduces some simple nonlinearities. Certainly, the pre-selection of variables is more visible in the second study than in the first. However, it has to be highlighted that the FRED-MD dataset is also based on previous studies, such as that by Stock and Watson (McCracken and Ng, 2015).

This thesis advances the field of macroeconomic forecasting using machine learning methods with a comprehensive approach, integrating a large amount of predictors as well as the theoretically founded pre-selection of variables. It bridges traditional econometric methods and modern machine learning, offering a robust framework for future research, ultimately enhancing forecasting accuracy and economic understanding.

Some implications for future research can be derived from the empirical studies. On the one hand, it can be observed how ML methods, particularly boosting models, have potential usage in the context of Big Fat Data. On the other

[81] In the USA, a forecasting industry was developed in 1910–1930 but much of it was almost erased due to the perceived failure of economists in predicting the crisis of 1929 in the public opinion (Hawkins, 2005).

hand, standard econometric techniques maintain their usefulness with standard data. Types of data environments, such as Big Fat Data or imbalanced data, and the presence of nonlinearities are central to the decisions of investigators to use or not use a simple or more complex machine learning method or employ traditional econometric techniques.

First, the success of component wise boosting models underscores the value of incorporating factors to enhance the performance of machine learning methods in predicting business cycles. These factors provide valuable information that complements the predictive power of machine learning models.

Second, the use of multiple stages in boosting algorithms can significantly improve forecasting capabilities. This indicates that iterative approaches, where the model adjusts iteratively to the data, are effective in capturing complex patterns in business cycles.

Furthermore, complex algorithms like XGBoost, which involve a large number of hyperparameters, may not consistently yield good results when the training data is of limited size, and strong and weak predictors are mixed together. This highlights the importance of careful hyperparameter tuning and the judicious use of complex models.

Lastly, it is essential to consider feature selection and engineering. Overly complex models that include too many weak predictors during splitting operations can lead to poorer performance. Therefore, selecting relevant predictors and ensuring they add value to the model, rather than introducing noise, is crucial.

These insights provide valuable guidance for researchers and practitioners looking to improve the accuracy and effectiveness of machine learning models in predicting business cycles. Taking these factors and model complexities into account can lead to enhanced forecasting performance and a better understanding of economic fluctuations.

The classification exercise dealing with imbalanced data in this thesis underscores the need to handle oversampling or under sampling of the dependent variables carefully. When it comes to identifying systemic financial crises, relying solely on time series estimations can be challenging. However, several key conclusions can be drawn from this thesis.

First, the introduction of nonlinearities and the inclusion of numerous predictors can enhance predictive power, which is also highlighted by studies such as Ward (2017) and Liu (2023). Second, the proper identification and application of suitable evaluation criteria are critical aspects that have often been underestimated in the literature. Only a few authors, such as Papadopoulos et al. (2019) and Casabianca et al. (2022), have recognised the significance of this issue. These findings emphasize the importance of addressing imbalanced data, embracing nonlinear modelling, and selecting appropriate evaluation criteria in the context of forecasting systemic financial crises. Such considerations can significantly improve the effectiveness of predictive models in this domain.

Moreover, further research with different approaches to hyperparameter tuning could bring improvements to the results. Goulet Coulombe (2022) observes that cross-validation is usually the best method for tuning hyperparameters. The main

scope of the present thesis is different from that by Goulet Coulombe (2022), which is more focused on how to optimise generic ML methods from different perspectives, such as hyperparameter tuning. Nevertheless, future research could focus on trying to perform the hyper parametrization tuning of boosting models in other ways, such as using different methods of k-fold cross-validations or implementing the selection of the hyperparameters via various information criteria. Importantly, the authors find that the presence of nonlinearities is the real game changer in favour of ML, and the second exercise shows this.

Huge Big Data exercises, with a large number of observations and predictors, can possibly show even better results for ML methods, especially with XGBoost. However, in macroeconomics, the only possibility to obtain such data, especially as conceived by data scientists, requires the use of either large amounts of unstructured data, such as texts, or the introduction of high-frequency financial data, such as intraday stock market data. This is indeed feasible, as evidenced by studies involving mixed-frequency data (e.g., Babii et al., 2022; Babii et al., 2020). This is particularly pertinent given the potential for financial data to provide valuable insights for forecasting.

Nevertheless, it must be noticed how this will increase the difficulty of tuning the hyperparameters with the risk of being unable to proceed without very powerful tools. Another possibility could be to completely ignore the hyper parametrization and base the values of the parameters on previous literature or reduce the range of possible values to be checked. Of course, there are risks in following this path, such as having higher bias errors or variance, and the researchers have to consider this.

Considering more specifically the two empirical studies, it can be observed that the results of the first study suggest further research is needed on how to introduce nonlinearities in the time series, as in the second empirical study and as suggested by Goulet Coulombe et al. (2022). Moreover, more studies are needed to compare boosting models with more sophisticated factor models with pre-selected variables (Girardi et al., 2017) or bridge models[82] with different specifications as the one suggested in Bulligan et al. (2010). The results of AdaBoostRT should be studied in further research where the aim will be to understand if AdaBoostRT also has significant prediction capacity in tranquil periods.

In addition, financial data could include relevant information for forecasting the business cycle. The use of this kind of data relies on methods connected to mixed-frequency analysis. Studies in this sense focus more on the lasso family of models, such as the group-lasso model (Babii et al. 2022; Babii et al. 2020). However, the boosting family of models could be another useful method if it can be understood which kind of specific boosting model could be better used for this kind of analysis.

[82] In general, bridge models consist in forecasting a low frequency dependent variable (i.e. GDP growth) with higher frequency components of the dependent variable and the lags of the dependent variable on the right-hand side of the model.

Concerning the second empirical study, a possible criticism could be that the systemic financial crises of the 19th century are different compared to the crisis of 2007/2008, and this has to be investigated with a deeper analysis. This could be true even for previous studies like the one by Beutel et al. (2018), who consider the crises before and after the Bretton Wood system together, which were fundamentally different. Nevertheless, it could be possible to check different subsamples of the data.

The first option is reducing the time frame and having higher frequency data. A second way to solve the issue is to use panel data as in Casabianca et al. (2022). In this sense, it could have been possible to use the values of the AdaBoost hyperparameters suggested by Casabianca et al. (2022) in the second empirical study. The authors also suggest how to solve the problem of instability, but the data they use is quite different compared to the one in this thesis. Moreover, passing from time series to panel data would make it difficult to pursue a country-by-country analysis and imply that all the countries have the same mechanism triggering crises.

Another possibility could be to divide the timeframe of the analysis into subperiods. This is a possible solution in principle but entails other limitations. From the thesis, it is clear how the distribution of crises is relevant in order to have enough of them in both the training and test sets. A division in subperiods risks replicating or even worsening the problem of imbalanced data.

Finally, results from new empirical studies with ML methods may lead to new theoretical implications. Certainly, this depends on how possible it is to find causal relationships using ML in time series or panel data. Indeed, these kinds of studies are still in their infancy, but works like the one by Joseph (2020) are encouraging.

From a policy perspective, the thesis' contribution is twofold. First, it shows how the ML methods should be used in particular contexts, helping to forecast complex macroeconomic phenomena. A good economic forecast can better provide information to policymakers and improve the effectiveness of central bank communications. Second, it suggests that some of the current variables used to predict systemic financial crises, such as credit variables, are indeed relevant but that some others, like variables connected with international economy, should be considered more by scholars. It also describes how most of the studies on early warning signals for financial crises should be carefully evaluated, given that they mostly do not take into account country specificities.

In the case of very unexpected events, such as wars and pandemics, economic forecasting can be extremely difficult. Therefore, Bernanke (2024) suggests the use of different scenarios in order to provide multiple possibilities for policymakers. The author also emphasizes the importance of Big Data and the use of new models such as algorithms related to artificial intelligence. Given the complexity of XGBoost, it could be tempting to suggest that this could be the best choice to make economic forecasts in the case of extremely unexpected events, but the issue could be a lack of interpretability of the results obtained that could make it difficult to construct different scenarios as suggested by Bernanke (2024).

A better solution could be to use a combination of standard econometric techniques and ML methods. The model suggested by Lenza et al. (2023) is going in a good direction in this sense.

In evaluating economic forecasts produced using ML methods, it is essential to emphasize that while having current, real-world data, particularly financial data, is critical from a policy perspective, evaluating the performance of ML methods requires the presence of nonlinearities (Goulet Coulombe et al., 2022; Mosavi et al., 2020; Athey, 2018). If nonlinearities are absent, they may need to be simulated.

When outliers are present, the frequency of extreme events must be taken into account. If such events occur frequently, a classification approach, similar to the second empirical exercise in this thesis, could be a practical solution. Conversely, when the frequency of extreme events is low, testing different subsamples and employing regression methods can be an effective strategy.

Regarding the measurement of forecasting performance, researchers must first determine whether the exercise involves classification or regression, then consider the type of data used, and finally, reflect on the intended application of the forecasts. For economic policy purposes, measures that minimise large errors, such as RMSE, might be prioritised to prevent significant missteps, whereas business decision-making may prefer simpler averages of accuracy, like mean absolute error.

The thesis underscores the importance of scholars' assessments concerning the capabilities and limitations of advanced ML methods. Even with the aid of contemporary artificial intelligence tools, the pre-selection of variables, such as those needed for predicting systemic financial crises, remains reliant on expert judgment and theoretical knowledge. Ultimately, given the wide array of ML models available, the decision on which algorithm to use also rests with the researchers.

As demonstrated throughout this thesis, the potential of machine learning methods should neither be underestimated nor overestimated. While ML algorithms can offer valuable new insights for macroeconomic forecasting, their performance often depends on the specific characteristics of the data used and the particular objectives of the researchers. Boosting models, which have been a focal point of this thesis, exhibit similar limitations to other ML methods. However, they also illustrate how variations in elements such as loss functions or the presence of nonlinearities can yield different results.

The prevalence of Big Data and the resulting demand for ML methods to process it should not lead to an uncritical reliance on these techniques or to the complete abandonment of traditional methods. In fact, ML methods can achieve better results when combined with classic econometric approaches, reflecting the shared statistical foundations of machine learning and econometrics. Consequently, new methods are sometimes not as distinct from established ones as they may initially appear.

REFERENCES

Aaronson, D., Brave, S. A., Butters, R. A., Fogarty, M., Sacks, D. W., & Seo, B. (2022). Forecasting Unemployment Insurance Claims in Realtime with Google Trends. *International Journal of Forecasting*, *38*(2), 567–581. doi: 10.1016/j.ijforecast.2021.04.001

Acharya, V. V., and Richardson, M. (2009). Causes of the Financial Crisis. *Critical Review*, *21*(2–3), 195–210. doi: 10.1080/08913810902952903

Acharya, A., Singh, S. K., Pereira, V., and Singh, P. (2018). Big Data, Knowledge Co-creation and Decision Making in Fashion Industry. *International Journal of Information Management*, *42*, 90–101. doi: 10.1016/j.ijinfomgt.2018.06.008

Ahir, H., Bloom, N., and Furceri, D. (2022). *The World Uncertainty Index*. National Bureau of Economic Research, No. w29763, February.

Akinci, O., and Queralto, A. (2022). Credit Spreads, Financial Crises, and Macroprudential Policy. *American Economic Journal: Macroeconomics*, 14(2), 469–507. doi: 10.1257/mac.20180059

Alesina, A. (2003). The Size of Countries: Does it Matter? *Journal of the European Economic Association*, 1(2–3), 301–316. doi: 10.1162/154247603322390946

Alessi, L., and Detken, C. (2018). Identifying Excessive Credit Growth and Leverage. *Journal of Financial Stability*, 35, 215–225. doi: 10.1016/j.jfs.2017.06.005

Alfieri, L., Gabrielyan, D., (2024). The Communication Reaction Function of the European Central Bank. An Analysis using Topic Modelling. *Baltic Journal of Economics*, *24*(1), 58–87. doi: 10.1080/1406099X.2024.2303904.

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time Measurement of Business Conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427. doi: 10.1198/jbes.2009.07205

Athey S., (2018). The Impact of Machine Learning on Economics. In: *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 507–547.

Athey, S., and Imbens, G. W. (2015). *Machine Learning Methods for Estimating Heterogeneous Causal Effects.* Research Papers 3350. Stanford University, Graduate School of Business.

Azad, P., Navimipour, N. J., Rahmani, A. M., and Sharifi, A. (2020). The Role of Structured and Unstructured Data Managing Mechanisms in the Internet of Things. *Cluster computing*, *23*, 1185–1198. doi: 10.1007/s10586-019-02986-2

Athey, S., and Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies*, *5*(2), 37–51. doi: 10.1353/obs.2019.0001

Babii, A., Ball, R. T., Ghysels, E., and Striaukas, J. (2020). *Machine Learning Panel Data Regressions with Heavy-tailed Dependent Data: Theory and Application*. arXiv e-prints, arXiv-2008.

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine Learning Time Series Regressions with an Application to Nowcasting. *Journal of Business & Economic Statistics*, *40*(3), 1094–1106. doi: 10.1080/07350015.2021.1899933

Bai, J., and Ng, S. (2008). Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics*, *146*(2), 304–317. doi: 10.1016/ j.jeconom.2008.08.010

Bai, J., and Ng, S. (2009). Boosting Diffusion Indices. *Journal of Applied Econometrics*, *24*(4), 607–629. doi: 10.1002/jae.1063

Barrell, R., Davis, E. P., Karim, D., and Liadze, I. (2010). Does the Current Account Balance Help to Predict Banking Crises in OECD Countries? *NIESR Discussion Papers* (351).

Barthélémy, S., Binet, M.-E., and Pentecôte, J.-S. (2020). Worldwide Economic Recoveries from Financial Crises through the Decades. *Journal of International Money and Finance*, 105, 102204. doi: 10.1016/j.jimonfin.2020.102204

Belloni, A., and Chernozhukov, V. (2013). Least Squares After Model Selection in High-Dimensional Sparse Models. *Bernoulli 19*(2), 2013, 521–547.

Belloni, A., V. Chernozhukov and C. Hansen (2013). *Inference for High-Dimensional Sparse Econometric Models*. In D. Acemoglu, M. Arellano and E. Dekel (Eds.), Advances in Economics and Econometrics: Tenth World Congress of Econometric Society, Volume III, 245–95. Cambridge: Cambridge University Press.

Belloni, A., V. Chernozhukov and C. Hansen (2014a). *Inference on Treatment Effects After Selection Amongst High-Dimensional Controls*. Review of Economic Studies 81, 608–50. doi: 10.1093/restud/rdt044

Belloni, A., V. Chernozhukov and L. Wang (2014b). Pivotal Estimation via Square-Root Lasso in Nonparametric Regression. *Annals of Statistics*, 42, 757–88. doi: 10.1214 / 14-AOS1204

Bergmeir, C., and Benítez, J. M. (2012). On the Use of Cross-validation for Time Series Predictor Evaluation. *Information Sciences*, *191*, 192–213. doi: 10.1016/j.ins.2011.12.028

Bernanke, B. (2024). *Forecasting for Monetary Policy Making and Communication at the Bank of England: a Review*. Bank of England, Bank of England Independent Evaluation Office, April 12th.

Available here: [https://www.bankofengland.co.uk/independent-evaluation-office/forecasting-for-monetary-policy-making-and-communication-at-the-bank-of-england-a-review/forecasting-for-monetary-policy-making-and-communication-at-the-bank-of-england-a-review]. Last access 20-11-2024.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The Financial Accelerator in a Quantitative Business Cycle Framework. In *Handbook of Macroeconomics*, Vol. 1, ed. by Taylor J. B., and Woodford, M., 1341–93. Elsevier, Amsterdam.

Beutel, J., List, S., and Von Schweinitz, G. (2018). *An Evaluation of Early Warning Models for Systemic Banking Crises: Does Machine Learning Improve Predictions?* Deutsche Bundesbank Discussion Paper No 48/2018, Frankfurt.

Beutel, J., List, S., and von Schweinitz, G. (2019). Does Machine Learning Help us Predict Banking Crises? *Journal of Financial Stability*, 45, 100693. doi: 10.1016/j.jfs.2019.100693

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., and Simsek, Ö. (2020). *Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach.* Bank of England – Staff Working paper No 848, London.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review of Economics*, *10*, 615–643. doi: 10.1146/annurev-economics-080217-053214.

Bordo, M. D., and Haubrich, J. G. (2017). Deep Recessions, Fast Recoveries, and Financial Crises: Evidence from the American Record. *Economic Inquiry*, *55*(1), 527–541. doi: 10.1111/ecin.12374

Borup, D., Christensen, B. J., Mühlbach, N. S., and Nielsen, M. S. (2023). Targeting Predictors in Random Forest Regression. *International Journal of Forecasting*, *39*(2), 841–868. doi: 10.1016/j.ijforecast.2022.02.010

Bouvatier, V., López-Villavicencio, A., and Mignon, V (2014). Short-run Dynamics in Bank Credit: Assessing Nonlinearities in Cyclicality, Economic Modelling, Volume 37, 127–136. https://doi.org/10.1016/j.econmod.2013.10.027.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*(2), 123–140. doi:10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole Statistics-Probability Series, Brooks/ Cole Publishing, California, Monterey.

Breitung, J., and Knüppel, M. (2021). How Far Can We Forecast? Statistical Tests of the Predictive Content. *Journal of Applied Econometrics*, *36*(4), 369–392. doi: 10.1002/jae.2817

Bruno, G., & Lupi, C. (2004). Forecasting Industrial Production and the Early Detection of Turning Points. *Empirical economics*, 29, 647–671. doi: 10.1007/s00181-004-0203-y

Buchen, T., and Wohlrabe, K. (2011). Forecasting with Many Predictors: Is Boosting a Viable Alternative? *Economics Letters*, 113(1), 16–18. doi: 10.1016/ j.econlet.2011.05.040

Buera, F. J., and Nicolini, J. P. (2020). Liquidity Traps and Monetary Policy: Managing a Credit Crunch. *American Economic Journal: Macroeconomics*, *12*(3), 110–38. doi: 10.1257/mac.20170040

Bulligan, G., Golinelli, R., and Parigi, G. (2010). Forecasting Monthly Industrial Production in Real-time: from Single Equations to Factor-based Models. *Empirical Economics*, *39*(2), 303–336. doi: 10.1007/s00181-009-0305-7.

Burns, A. F., and Mitchell W. C. (1946). *Measuring Business Cycles*. New York: National Bureau of Economic Research.

Bussière, M., and Fratzscher, M. (2006). Towards a New Early Warning System of Financial Crises. J*ournal of International Money and Finance*, *25*(6), 953–973. doi: 10.1016/j.jimonfin.2006.07.007

Bussière, M., Imbs, J., Kollmann, R., and Rancière, R. (2013). The Financial Crisis: Lessons for International Macroeconomics. *American Economic Journal: Macroeconomics*, *5*(3), 75–84. doi: 10.1257/mac.5.3.75

Bühlmann, P., and Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, *22*(4), 477–505. doi: 10.1214/07-STS242

Caggiano, G., Calice, P., and Leonida, L. (2014). Early Warning Systems and Systemic Banking Crises in Low Income Countries: A Multinomial Logit Approach. *Journal of Banking & Finance*, *47*, 258–269. doi: 10.1016/j.jbankfin.2014.07.002

Caggiano, G., Castelnuovo, E., and Figueres, J. M. (2017). Economic Policy Uncertainty and Unemployment in the United States: A Nonlinear Approach. *Economics Letters*, *151*, 31–34. doi: 10.1016/j.econlet.2016.12.002

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models. *Journal of Applied Econometrics*, *26*(5), 735–761. doi: 10.1002/jae.1150

Casabianca, E. J., Catalano, M., Forni, L., Giarda, E., and Passeri, S. (2022). A Machine Learning Approach to Rank the Determinants of Banking Crises over Time and across Countries. *Journal of International Money and Finance*, *129*, 102739. doi: 10.1016/j.jimonfin.2022.102739

Cepni, O., Güney, I. E., and Swanson, N. R. (2019). Nowcasting and Forecasting GDP in Emerging Markets using Global Financial and Macroeconomic Diffusion Indexes. *International Journal of Forecasting*, *35*(2), 555–572. doi: 10.1016/j.ijforecast.2018.10.008

Cerdeiro, D. A., Komaromi, A., Liu, Y., Saeed, M. (2020). *World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-Based Nowcasts from Scratch*, IMF Working Paper No 20/57, May.

Chen, T., and Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, San Francisco, August 13–17. doi: 10.1145/2939672.2939785

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/Debiased Machine Learning for Treatmentand Structural Parameters. *The Econometrics Journal*, *21*(1), C1–C68. doi: doi.org/10.1111/ectj.12097

Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach. *Annu. Rev. Econ.*, *7*(1), 649–688. doi: 10.1146/annurev-economics-012315-015826

Chu, B., and Qureshi, S. (2022). Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast US GDP Growth. *Computational Economics*, 62, 1–43. doi: 10.1007/s10614-022-10312-z

Claeys, P., and Vašíček, B. (2019). Transmission of Uncertainty Shocks: Learning from Heterogeneous Responses on a Panel of EU countries. *International Review of Economics & Finance*, 64, 62–83. doi: 10.1016/j.iref.2019.05.012

Claessens, S., Kose, M. A., and Terrones, M. E. (2010). The Global Financial Crisis: How Similar? How Different? How Costly? *Journal of Asian Economics*, *21*(3), 247–264. doi: 10.1016/j.asieco.2010.02.002

Culp, M., Johnson, K., and Michailidis, G. (2006). Ada: An R Package for Stochastic Boosting. *Journal of Statistical Software*, *17*(2), 9. doi: 10.18637/jss.v017.i02

Davis, J. S., Mack, A., Phoa, W., and Vandenabeele, A. (2016). Credit Booms, Banking Crises, and the Current Account. *Journal of International Money and Finance*, *60*, 360–377.

De Sismondi (1819) [2016]. Nouveaux Principes d'Economie Politique. ECONOMICA Publisher.

De Nicolò, G., and Lucchetta, M. (2017). Forecasting Tail Risks. *Journal of Applied Econometrics*, *32*(1), 159–170. doi: 10.1016/j.jimonfin.2015.09.008

Detken, C., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M.M., Castro, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N., Kakes, J., Klaus, B., Lang, J.H., Puzanova, N., and Welz, P. (2014) *Operationalizing the Countercyclical Capital Buffer*. ESRB Occasional Paper, no. 5, June.

Dice, L. R. (1945). Measures of the Amount of Ecologic Association between Species. *Ecology*, *26*(3), 297–302. doi: 10.2307/1932409

Diebold, F.X. and Mariano. R.S., (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253–263. doi: 10.1080/07350015.1995.10524599

Diebold, F. X., and Rudebusch, G. D. (1991). Forecasting Output with the Composite Leading Index: A Real-time Analysis. *Journal of the American Statistical Association*, *86*(415), 603–610. doi: 10.2307/2290388

Dixon, M. F., Klabjan, D., and Wei, L. (2017). *Ostsc: Over Sampling for Time Series Classification in R.* (Available at SSRN 3077767).

Doğan, B. S¸., and Midiliç, M. (2019). Forecasting Turkish Real GDP Growth in a Data-rich Environment. Empirical Economics, 56(1), 367–395. doi: 10.1007/ s00181-017-1357-8

Donaldson, D., and Storeygard, A. (2016). The View From Above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4), 171–98. doi: 10.1257/jep.30.4.171

Drechsel, K., and Scheufele, R. (2012). The Performance of Short-Term Forecasts of the German Economy Before and During the 2008/2009 Recession. *International Journal of Forecasting*, *28*(2), 428–445. doi: 10.1016/ j.ijforecast.2011.04.003

Drucker, H. (1997). Improving Regressors Using Boosting Techniques. In *Machine Learning: Proceedings of the Fourteenth International Conference*, 107–115. doi: 10.5555/645526.657132

Döpke, J., Fritsche, U., and Pierdzioch, C. (2017). Predicting Recessions with Boosted Regression Trees. *International Journal of Forecasting*, *33*(4), 745–759. doi: 10.1016/j.ijforecast.2017.02.003

du Plessis, E. (2022). Multinomial Modeling Methods: Predicting Four Decades of International Banking Crises. *Economic Systems*, *46*(2), 100979. doi: 10.1016/j.ecosys.2022.100979

Efroymson, M. A. (1960). *Multiple Regression Analysis*. In A.Ralston and H. S., Wilf (eds.) Mathematical Methods for Digital Computers, Wiley, New York, p. 191–203.

Eggers, A. C., Tuñón, G., & Dafoe, A. (2024). Placebo Tests for Causal Inference. *American Journal of Political Science*, *68*(3), 1106–1121. doi: 10.1111/ajps.12818

Ellington, M., Florackis, C., and Milas, C. (2017). Liquidity Shocks and Real GDP growth: Evidence from a Bayesian Time-Varying Parameter VAR. *Journal of International Money and Finance*, *72*, 93–117. doi: 10.1016/j.jimonfin.2016.12.002

Estrella, A., and Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, *46*(2), 555–576. doi: 10.2307/2328836

Exterkate, P., Groenen, P. J., Heij, C., and van Dijk, D. (2016). Nonlinear Forecasting with Many Predictors using Kernel Ridge Regression. *International Journal of Forecasting*, *32*(3), 736–753. doi: 10.1016/j.ijforecast.2015.11.017

Fernández, A., Garcia, S., Herrera, F., and Chawla, N. V. (2018). Smote for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. doi: 10.1613/jair.1.11192

Fisher, I. (1933). The Debt-Deflation Theory of Great Depressions, *Econometrica*, *1*(4), 337–357.

Fouliard, J., Howell, M., and Rey, H. (2021). *Answering the Queen: Machine Learning and Financial Crises.* BIS Working Papers No 926, Basel.

Franses, P. H., and Van Dijk, D. (2005). The Forecasting Performance of Various Models for Seasonality and Nonlinearity for Quarterly Industrial Production. *International Journal of Forecasting*, *21*(1), 87–102. doi: 10.1016/j.ijforecast.2004.05.005

Freund, Y., and Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. doi: 10.1006/jcss.1997.1504

Freund, Y., Schapire, R., and Abe, N. (1999). A Short Introduction to Boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(5), 771–780.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Ma-chine. *Annals of Statistics*, *29*(5), 1189–1232. doi: 10.1214/aos/1013203451

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive Logistic Regression:a Statistical View of Boosting (with Discussion and a Rejoinder by the Authors). *The Annals of Statistics*, *28*(2), 337–407.

Furht, B., and Villanustre, F. (2016). *Introduction to Big Data*. In Furht, B., and Villanustre, F. (eds.) Big Data Technologies and Applications. Springer International Publishing, p. 3–11.

Galí, J. (2018). The state of New Keynesian economics: a partial assessment. *Journal of Economic Perspectives*, 32, 3, 87–112. doi: 10.1257/jep.32.3.87

Garcia, M. G., Medeiros, M. C., and Vasconcelos, G. F. (2017). Real-time Inflation Forecasting with High-dimensional Models: The Case of Brazil. *International Journal of Forecasting*, *33*(3), 679–693. doi: 10.1016/j.ijforecast.2017.02.002

Ghysels, E., Kvedaras, V., and Zemlys, V. (2016). Mixed Frequency Data Sampling Regression Models: the R package midasr. *Journal of Statistical Software*, 72, 1–35. doi: 10.18637/jss.v072.i04

Ghysels, E., and Marcellino, M. (2018). *Applied Economic Forecasting Using TimeSeries Methods*. Oxford: Oxford University Press.

Giacomini, R., and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, *74*(6), 1545–1578. doi: 10.1111/j.1468-0262.2006.00718.x

Girardi, A., Golinelli, R., and Pappalardo, C. (2017). The Role of Indicator Selectionin Nowcasting Euro-Area GDP in Pseudo-Real Time. *Empirical Economics*, *53*(1), 79–99. doi: 10.1007/s00181-016-1151-z

Golinelli, R., and Parigi, G. (2007). The Use of Monthly Indicators to Forecast Quarterly GDP in the Short Run: An Application to the G7 Countries. *Journal of Forecasting*, *26*(2), 77–94. doi: 10.1002/for.1007

Goulet Coulombe, P. (2020). *The Macroeconomy as a Random Forest*. arXiv e-prints, arXiv-2006.

Goulet Coulombe, P. (2022). A Neural Phillips Curve and a Deep Output Gap. arXiv preprint arXiv:2202.04146.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2021a). Macroeconomic data transformations matter. *International Journal of Forecasting*, *37*(4), 1338–1354. doi: 10.1016/j.ijforecast.2021.05.005

Goulet Coulombe, P., Marcellino, M., & Stevanović, D. (2021b). Can Machine Learning Catch the Covid-19 Recession? *National Institute Economic Review*, *256*, 71–109. doi: 10.1017/nie.2021.10

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). How is Machine Learning Useful for Macroeconomic Forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964. doi: doi.org/10.1002/jae.2910

Gourinchas, P.-O., and Obstfeld, M. (2012). Stories of the Twentieth Century for the Twenty-First. *American Economic Journal: Macroeconomics*, *4*(1), 226–65. doi: 10.1257/mac.4.1.226

Grossman, G. and Helpman E. (1991). *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.

Guerróon-Quintana, P., and Zhong, M. (2023). Macroeconomic Forecasting in Times of Crises. *Journal of Applied Econometrics*, *38*(3), 295–320. doi: 10.1002/jae.2951

Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, *133*(2), 801–870. doi: 10.1093/qje/qjx045

Hansen, S., McMahon, M., and Tong, M. (2019). The Long-Run Information Effect of Central Bank Communication. *Journal of Monetary Economics*, 108, 185–202. doi: 10.1016/j.jmoneco.2019.09.002

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Prediction, Inference and Data Mining* (2nd ed.). New York: Springer Verlag.

Hamilton, J. D. (2005). What's Real About The Business Cycle? *FRB St. Louis – Review*, v87(4,Jul/Aug), 435–452.

Hawkins, J. (2005). Economic Forecasting: History and Procedures. *Economic Round-up: Journal of the Department of the Treasury*, 1–10.

Heij, C., van Dijk, D., and Groenen, P. J. (2011). Real-time Macroeconomic Fore casting with Leading Indicators: An Empirical Comparison. *InternationalJournal of Forecasting*, *27*(2), 466–481. doi: 10.1016/j.ijforecast.2010.04.008

Heinisch, K., and Scheufele, R. (2018). Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment. *Empirical Economics*, *54*(2), 705–745. doi:10.1007/s00181-016-1218-x

Hoerl, A. E., and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. doi: 10.1080/ 00401706.1970.10488634

Hyndman, R. J., and Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, *22*(4), 679–688. doi: 10.1016/j.ijforecast.2006.03.001

Hännikäinen, J. (2017). The Shadow Rate as a Predictor of Real Activity and Inflation: Evidence from a Data-Rich Environment. *Applied Economics Letters*, *24*(8), 527–535. doi: 10.1080/13504851.2016.1208347

IBM. (2019). *Extracting business value from the 4 V's of Big Data* retrieve from [http://www.ibmbigdatahub.com/infographic/four-vs-big-data]

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

Jardet, C., and Meunier, B. (2022). Nowcasting World GDP Growth with High-frequency Data. *Journal of Forecasting*, *41*(6), 1181–1200. doi: 10.1002/for.2858

Jeni, L. A., Cohn, J. F., and De La Torre, F. (2013). *Facing Imbalanced Data-Recommendations for the Use of Performance Metrics*. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (IEEE), September, 245–251.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2015). Betting the House. *Journal of International Economics*, *96*, S2–S18. doi: 10.1016/j.jinteco.2014.12.011

Jordà, Ò., Schularick, M., and Taylor, A. M. (2017). Macrofinancial History and the New Business Cycle Facts. *NBER macroeconomics annual*, *31*(1), 213–263.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2016). Sovereigns versus Banks: Credit, Crises, and Consequences. *Journal of the European Economic Association*, *14*(1), 45–79. doi: 10.1111/jeea.12144

Joseph, A. (2020). *Parametric Inference with Universal Function Approximators*, arXiv preprint arXiv:1903.04209.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring Uncertainty. *The American Economic Review*, 105(3), 1177–1216. doi: 10.1257/aer.20131193

Kapetanios, G., Marcellino, M., and Papailias, F. (2016). *Big Data and Macroeconomic Nowcasting.* Big Data and Nowcasting, European Commission, CROSCollaboration in Research and Methodology for Official Statistics. Brussels,April 7th.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary Policy According to HANK. *American Economic Revie*w, 108, 3, 697–743.

Kauppi, H., and Virtanen, T. (2021). Boosting Nonlinear Predictability of Macroeconomic Time Series. *International Journal of Forecasting*, *37*(1), 151–170. doi: 10.1016/j.ijforecast.2016.01.006

Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A.A., and Salehian, S. (2018). *The 10 Vs, Issues and Challenges of Big Data*. In: Proceedings of the 2018 International Conference on Big Data and Education, p. 52–56.

Kim, Y., and Park, C. (2020). Are Exchange Rates Disconnected from Macroeconomic Variables? Evidence from the Factor Approach. *Empirical Economics*, 58, 1713–1747. doi: 10.1007/s00181-018-1596-3

Kindleberger, C. P. (1978). *Manias, Panics, and Crashes: A History of Financial Crisis*, 1st ed. New York: Basic Books.

Knoll, K., Schularick, M., and Steger, T. (2017). No Price Like Home: Global House Prices, 1870–2012. *American Economic Review*, *107*(2), 331–53. doi: 10.1257/aer.20150501

Kock, A. B., and Teräsvirta, T. (2014). Forecasting Performances of Three Automated Modelling Techniques During the Economic Crisis 2007–2009. *International Journal of Forecasting*, *30*(3), 616–631. doi: 10.1016/j.ijforecast.2013.01.003

Koza, John R.; Bennett, Forrest H.; Andre, David; Keane and Martin A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design '96*. Springer, Dordrecht, 151–170.

Kydland, F. E., and Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370. doi: 10.2307/1913386

Laeven, L., Valencia, F. (2020). Systemic Banking Crises Database II. *IMF Economic Review*, 68, 307–361. doi: 10.1057/s41308-020-00107-3

Laeven, L., and Valencia, F. (2018). *Systemic Banking Crises Revisited*. IMF Working Paper No. 2018/206.

Laeven L, and Valencia F. (2013). Systemic Banking Crises Database. *IMF Economic Review*, 61, 225–270. doi: 10.1057/imfer.2013.12

Lehmann, R., and Wohlrabe, K. (2015). Forecasting GDP at the Regional Level with Many Predictors. *German Economic Review*, *16*(2), 226–254. doi: 10.1111/ geer.12042

Lehmann, R., and Wohlrabe, K. (2016). Looking into the Black Box of Boosting: the Case of Germany. *Applied Economics Letters*, *23*(17), 1229–1233. doi: 10.1080/13504851.2016.1148246

Lenza, M., Moutachaker, I., & Paredes, J. (2023). *Forecasting Euro Area Inflation with Machine-Learning Models*. ECB Research Bulletin, no 112, October 17th

Li, J., and Chen, W. (2014). Forecasting Macroeconomic Time Series: LASSO-based Approaches and Their Forecast Combinations with Dynamic Factor Models. *International Journal of Forecasting*, *30*(4), 996–1015. doi: 10.1016/j.ijforecast.2014.03.016

Lim, B., and Zohren, S. (2021). Time-series Forecasting with Deep Learning: A Survey. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200209. doi: 10.1098/rsta.2020.0209

Lima, L. R., and Meng, F. (2017). Out-of-sample Return Predictability: A Quantile Combination Approach. *Journal of Applied Econometrics*, *32*(4), 877–895. doi: 10.1002/JAE.2549

Lima, L. R., Meng, F., and Godeiro, L. (2020). Quantile Forecasting with Mixed-frequency Data. *International Journal of Forecasting*, *36*(3), 1149–1162. doi: 10.1016/j.ijforecast.2018.09.011

Lin, S. L., and Jin, X. (2023). Does ESG Predict Systemic Banking Crises? A Computational Economics Model of Early Warning Systems with Interpretable Multi-Variable LSTM based on Mixture Attention. *Mathematics*, *11*(2), 410. doi: 10.3390/math11020410

Liu, L., Chen, C., and Wang, B. (2022). Predicting Financial Crises with Machine Learning Methods. *Journal of Forecasting*, *41*(5), 871–910. doi: 10.1002/for.2840

Liu, X. (2023). Towards Better Banking Crisis Prediction: Could an Automatic Variable Selection Process Improve the Performance?. *Economic Record*, *99*, 325, p. 288–312. doi: 10.1111/1475-4932.12721

Lucas, R. E. Jr. (1976). *Econometric Policy Evaluation: A Critique*. Carnegie-Rochester Conference Series on Public Policy. Vol. 1. North-Hollan.

Maehashi, K., and Shintani, M. (2020). Macroeconomic Forecasting using Factor Models and Machine Learning: An Application To Japan. *Journal of the Japanese and International Economies*, 58, 101104. doi: 10.1016/j.jjie.2020.101104

Manasse, P., and Roubini, N. (2009). "Rules of Thumb" for Sovereign Debt Crises. *Journal of International Economics*, *78*(2), 192–205. doi: 10.1016/j.jinteco.2008.12.002

McCracken, M. W., and Ng, S. (2015). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business and Economic Statistics*, *34*(4), 574–589. Retrieved from https://research.stlouisfed.org/econ/mccracken/sel/ doi: doi.org/10.1080/07350015.2015.1086655

Medeiros, M. C., and Mendes, E. F. (2015). *L1-Regularization of High-Dimensional Time-Series Models with Flexible Innovations.* Textos para discussão, No 636, Department of Economics PUC-Rio (Brazil).

Medeiros, M. C., Vasconcelos, G., and Freitas, E. (2016). Forecasting Brazilian Inflation with High-Dimensional Models. *Brazilian Review of Econometrics*,*36*(2), 223–254. doi: 10.12660/bre.v99n992016.52273.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2021). Forecasting Inflation in a Data-rich Environment: the Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119. doi: 10.1080/07350015.2019.1637745

Minsky, H. (1986). *Global Consequences of Financial Deregulation*. Technical Report 96, Washington University Working Paper.

Mogliani, M., and Simoni, A. (2021). Bayesian MIDAS Penalized Regressions: Estimation, Selection, and Prediction. *Journal of Econometrics*, 222(1), 833–860. doi: 10.1016/j.jeconom.2020.07.022

Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., and Band, S. S. (2020). Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics. *Mathematics*, 8(10), 1640. doi: 10.3390/math8101640

Nazareth, N., and Reddy, Y. Y. R. (2023). Financial Applications of Machine Learning: A Literature Review. *Expert Systems with Applications*, 219, 119640. doi: 10.1016/j.eswa.2023.119640

Ng, S. (2017). *Opportunities and Challenges: Lessons from Analyzing Terabytes of Scanner Data*. National Bureau of Economic Research, Working paper, No. w23673.

Ng, S. (2014). Viewpoint: Boosting Recessions. *Canadian Journal of Economics*, *47*(1), 1–34. doi: doi.org/10.1111/caje.12070

Obstfeld, M. (2012). Does the Current Account still Matter? *American Economic Review*, *102*(3), 1–23. doi: 10.1257/aer.102.3.1

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., and Vahid, F. (2019). Macroeconomic Forecasting for Australia using a Large Number of Predictors. *International Journal of Forecasting*, *35*(2), 616–633. doi: 10.1016/j.ijforecast.2018.12.002

Papadopoulos, S., Stavroulias, P., and Sager, T. (2019). Systemic Early Warning Systems for EU14 Based on the 2008 Crisis: Proposed Estimation and Model Assessment for Classification Forecasting. *Journal of Banking Regulation*, 20, 226–244. doi: 10.1057/s41261-018-0085-0

Pepe, M. S. (2000). Receiver Operating Characteristic Methodology. *Journal of the American Statistical Association, 95*(449), 308–311. doi: 10.1080/01621459.2000.10473930

Peña, D., Smucler, E., and Yohai, V. J. (2019). Forecasting Multiple Time Series with One-Sided Dynamic Principal Components. *Journal of the American Statistical Association, 114*(528), 1683–1694. doi: 10.1080/01621459.2018.1520117

Plakandaras, V., Gupta, R., Gogas, P., and Papadimitriou, T. (2015). Forecasting the US Real House Price Index. *Economic Modelling*, 45, 259–267. doi: 10.1016/j.econmod.2014.10.050

Popp, A., and Zhang, F. (2016). The macroeconomic effects of uncertainty shocks: The role of the financial channel. *Journal of Economic Dynamics and Control, 69*, 319–349. doi: 10.1016/j.jedc.2016.05.021

Rancière, R., and Tornell, A. (2016). Financial Liberalization, Debt Mismatch, Allocative Efficiency, and Growth. *American Economic Journal: Macroeconomics, 8*(2), 1–44. doi: 10.1257/mac.20130190

Reichlin, L. (2015). *Big Data in Macroeconomics*. COEURE Workshop, European Commission, Brussels 3–4 July.

Reinhart, C. M., and Rogoff, K. S. (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton, New Jersey: Princeton University Press.

Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2021). Nowcasting GDP using Machine-Learning Algorithms: A real-time assessment. *International Journal of Forecasting, 37*(2), 941–948. doi: 10.1016/j.ijforecast.2020.10.005

Richter, B., Schularick, M., and Wachtel, P. (2021). When to Lean Against the Wind. *Journal of Money, Credit and Banking, 53*(1), 5–39. doi: 10.1111/jmcb.12701

Robinzonov, N., Tutz, G., and Hothorn, T. (2012). Boosting Techniques for Nonlinear Time Series Models. *AStA Advances in Statistical Analysis, 96*(1), 99–122. doi: 10.1007/s10182-011-0163-4

Romer, C. D., and Romer, D. H. (2018). Phillips Lecture – Why Some Times are Different: Macroeconomic Policy and the Aftermath of Financial Crises. *Economica, 85*(337), 1–40. doi: 10.1111/ecca.12258

Rossi, B., and Inoue, A. (2012). Out-of-sample Forecast Tests Robust to the Choice of Window Size. *Journal of Business & Economic Statistics, 30*(3), 432–453. doi: 10.1080/07350015.2012.693850

Rünstler, G. (2016). *How Distinct are Financial Cycles from Business Cycles?* ECB Research Bulletin, 26(1).

Rünstler, G., and Vlekke, M. (2018). Business, Housing, and Credit cycles. *Journal of Applied Econometrics, 33*(2), 212–226. doi: 10.1002/jae.2604

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development.* 3, 3, 210–229. CiteSeerX doi: 10.1.1.368.2254

Sarantis, N. (2001). Nonlinearities, Cyclical Behaviour and Predictability in Stock Markets: International Evidence. *International Journal of Forecasting*, 17(3), 459–482. doi: 10.1016/S0169-2070(01)00093-0

Schularick, M., and Taylor, A. M. (2012). Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008. *American Economic Review, 102*(2), 1029–61. doi: 10.1257/aer.102.2.1029

Shiller, R. J., and Shiller, V. M. (2011). Economists as Worldly Philosophers. *American Economic Review*, 101(3), 171–175. doi: 10.1257/aer.101.3.171

Smeeks, S., and Wijler, E. (2018). Macroeconomic Forecasting Using Penalized Regression Methods. *International Journal of Forecasting*, 34(3), 408–430. doi: 10.1016/j.ijforecast.2018.01.001

Smets, F., and Wouters R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97, 3, 586–606. doi: 10.1257/aer.97.3.586

Solomatine, D. P., and Shrestha, D. L. (2004). *AdaBoost.RT: A Boosting Algorithm for Regression Problems.* In 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), 2, 1163–1168, Budapest, Hungary, 25–29 July. doi: 10.1109/IJCNN.2004.1380102

Sorensen, T. (1948). A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and its Application to Analyses of the Vegetation on Danish Commons. *Kongelige Danske Videnskabernes Selskab*, *5*(4), 1–34.

Soybilgen, B., and Yazgan, E. (2021). Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors. *Computational Economics*, *57*, 387–417. doi: 10.1007/s10614-020-10083-5

Statista (2022). *Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2020, with Forecasts from 2021 to 2025 (in Zettabytes)*. Available here: https://www.statista.com/statistics/871513/worldwide-data-created/ [retrieve on 01-06- 2022]

Stock, J. H., and Watson, M. W. (1998). *Business Cycle Fluctuations in US Macroeconomic Time Series*. NBER Working Paper, (w6528).

Stock, J. H., and Watson, M. W. (2003). How Did Leading Indicator Forecasts Perform During the 2001 Recession? *FRB Richmond Economic Quarterly*, *89*(3), 71–90.

Stock, J.H., and Watson, M.W. (2005). Implications of Dynamic Factor Models for VAR Analysis. NBER Working Paper No. 11467.

Stock, J. H., and Watson, M. W. (2006). Forecasting with Many Predictors. *Handbookof Economic Forecasting*, *1,* 515–554. doi: 10.1016/S1574-0706(05)01010-4

Stock, J. H., and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics using External Instruments. *The Economic Journal*, *128*(610), 917–948. doi: 10.1111/ecoj.12593

Sutherland, C. S. (2023). Forward Guidance and Expectation Formation: A Narrative Approach. *Journal of Applied Econometrics*, *38*(2), 222–241. doi: 10.1002/jae.2948

Syed, A. A. S., and Lee, K. H. (2021). Macroeconomic Forecasting for Pakistan in a Data-rich Environment. *Applied Economics*, *53*(9), 1077–1091. doi: 10.1080/00036846.2020.1826399

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journalof the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Torky, M., Gad, I., and Hassanien, A. E. (2023). Explainable AI Model for Recognizing Financial Crisis Roots Based on Pigeon Optimization and Gradient Boosting Model. *International Journal of Computational Intelligence Systems*, *16* (50). doi: 10.1007/s44196-023-00222-9

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., and Troncoso, A. (2021).Deep Learning for Time Series Forecasting: A Survey. *Big Data*, *9*(1), 3–21. doi: 10.1089/big.2020.0159

Tölö, E. (2020). Predicting Systemic Financial Crises with Recurrent Neural Networks. *Journal of Financial Stability*, 49, 100746. doi: 10.1016/j.jfs.2020.100746

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. doi: 10.1257/jep.28.2.3

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer.

Wager, S., and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. doi: 10.1080/01621459.2017.1319839

Wang, H., Li, G., and Tsai, C.-L. (2007). Regression Coefficient and AutoregressiveOrder Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(1), 63–78. doi: 10.1111/ j.1467-9868.2007.00577.x

Ward, F. (2017). Spotting the Danger Zone: Forecasting Financial Crises with Classification Tree Ensembles and Many Predictors. *Journal of Applied Econometrics*, *32*(2), 359–378. doi: 10.1002/jae.2525

Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., and Lestantyo, P. (2019). Cross-Validation Metrics for Evaluating Classification Performance on Imbalanced Data. In *2019 International Conference on Computer, Control,Informatics and its Applications (ic3ina)* pp. 14–18.

Wieland, V., and Wolters, M. (2013). Forecasting and Policy Making. In *Handbook of economic forecasting.* Vol. 2, pp. 239–325. Elsevier. doi: 10.1016/B978-0-444-53683-9.00005-0

Woodford M. (2005). *Central Bank Communication and Policy Effectiveness*. Proc. Fed. Reserve Bank Kansas City, August, 399–474

Ye, R. (2016). *Ensemble Time Series Forecasting with Applications in Renewable Energy* (Unpublished doctoral dissertation). doi: 10.32657/10356/66448

Yousuf, K., and Ng, S. (2021). Boosting High Dimensional Predictive Regressions with Time Varying Parameters. *Journal of Econometrics*, *224*(1), 60–87. doi: 10.1016/j.jeconom.2020.08.003

Zhang, Q., Ni, H., and Xu, H. (2023). Nowcasting Chinese GDP in a data-rich environment: Lessons from machine learning algorithms. *Economic Modelling*, *122*, 106204. doi: 10.1016/j.econmod.2023.106204

Zheng, Y., Xu, Z., and Xiao, A. (2023). Deep Learning in Economics: A Systematic and Critical Review. *Artificial Intelligence Review*, 56, 9497–9539. doi: 10.1007/s10462-022-10272-8

Zou, H. (2006). The Adaptive Lasso and its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. doi: 10.1080/ 01621459.2017.1319839

Zou, H., and Hastie, T. (2005). Regularization and Variable Selection via the ElasticNet. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Zou, H., Hastie, T., and Tibshirani, R. (2007). On the "Degrees of Freedom"of the Lasso. *The Annals of Statistics*, *35*(5), 2173–2192. doi: 10.1214/ 009053607000000127

APPENDIX

Tuning of the hyperparameters and Robustness Checks – first empirical study

This section describes the method used to tune the hyperparameters of the different ML methods in the present work.[83] The literature suggests using a cross-validation method to select the proper tuning parameter (Hastie et al., 2009). Cross-validation implied dividing the sample into a training set and a validation set. Although K-fold cross-validation is a sufficient method for cross-sectional data, it is not the most efficient method for tuning hyperparameters in time series, even less if in multivariate exercises. This is due to avoiding problems related to arbitrary choices of the test sets and to preserve the time dependency of the time series.[84] The study considers these observations and uses different approaches from recent literature on the different models applied in time series macroeconomic forecasting.

For what concerns the shrinkage models, Medeiros et al. (2016), Medeiros and Mendes (2015); Wang et al. (2007) and Zou et al. (2007) show that choosing λ with Bayesian Information Criteria (BIC) for both the lasso method and the other shrinkage models is a better and viable alternative for cross-validation, especially when time series is the object of the analysis. For this reason, the shrinkage models are estimated using the method used by Medeiros et al. (2016).[85]

In bagging, the cross-validation error can be calculated considering the out-of-bag (OOB) observations. OOB observations are part of the observations that are not used to fit a given bagged tree (James et al., 2013). According to James et al. (2013), the OOB estimation is a substitute for the classical cross-validation approach. In this study, in order to make bagging more computationally feasible the variables are estimated by groups of ten random variables. The variables change in every iteration. The forecast is the average of the forecasts for all subsets. This method can be found in Garcia et al. (2017).

Component wise boosting extracted a maximum of 10 factors from the principal component analysis. The presented component wise boosting is based on the study by Bai and Ng (2009) and Medeiros et al. (2021) and it used the BIC as information criteria to find the best model. The maximum number of iterations reachable was fixed at 100.[86] The hyperparameters for the random forest,[87]

---

[83] The large part of the codes have been run through cloud computing using Amazon Web Services (AWS).

[84] For issues related to cross-validation with time series, see Bergmeir and Benítez (2012)

[85] The R codes used for these estimations as well as for bagging and component wise boosting were derived principally from the HDeconometrics package

[86] Augmenting the maximum number of iterations to 500 did not show significant changes

[87] The number of trees is fixed at 2,000. Previous attempts with 500 and 1,000 trees has not provided important differences.

gradient boosting, XGBoost,[88] AdaBoostR2 and AdaBoostRT[89] are found through the estimations of rolling fixed window and expanding window estimations applying different sequences of hyperparameters.

The best combination was chosen by taking into account the lowest results of RMSE obtained. XGBoost in the subsample cannot complete the validation process because the number of observations is no longer sufficient to check the same amount of combinations of hyperparameter values. Robustness checks were performed to see if XGBoost is able to run after reducing the amount of possible values of the parameters. The results obtained by this incomplete procedure are reported in Table 3 for completeness, but they should be considered prudently.

[88] The R CARET package and Time slices method are used for random forest, gradient boosting and XGBoost that largely simplified the computations. For XGBoost, the iterations were only from 10 to 25 due to the large amount of hyperparameters to find that made larger iterations computationally prohibitive.

[89] AdaBoostR2 was performed in all the estimations considering three possible loss functions: linear, square and exponential. The possible iterations were either 100 or 500. For AdaBoostR2 the R function in the work of Ye (2016) is used. AdaBoostRT was performed with linear or squared loss function as well as different thresholds: 0.01, 0.1, 0.2, 0.3 and 0.5. The possible iterations were from 100 to 500. The REboost R package is used for AdaBoostRT.

FRED-MD Data Description and Results of all the estimations

Table A1. Data Description: Output and Income – FRED-MD

Group 1: Output and income						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	1	5	RPI	Real Personal Income	M 14386177	PI
2	2	5	W875RX1	Real personal income ex transfer receipts	M 145256755	PI less transfers
3	6	5	INDPRO	IP Index	M 116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M 116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M 116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M 116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M 116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M 116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M 116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M 116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M 116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M 116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M 116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M 116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M 116461275	IP: fuels
16	20	2	CUMFNS	Capacity Utilization: Manufacturing	M 116461602	Cap uti

Note: The column "tcode" denotes the following data transformation for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\ (x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A2. Data Description: Labour Market – FRED-MD

Group 2: Labour Market						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	21*	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22*	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M 110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M 110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M 110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M 110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M 110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed – Less Than 5 Weeks	M 110156527	U ¡ 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	M 110156523	U 5–14 wks
9	29	5	UEMP15OV	Civilians Unemployed – 15 Weeks & Over	M 110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15–26 Weeks	M 110156525	U 15–26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M 110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M 15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M 123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M 123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M 123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M 123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M 123109573	Emp:dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M 123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M 123109193	Emp: services

Group 2: Labour Market						
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M 123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M 123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M 123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M 123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M 123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	M 140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours: Manufacturing	M 123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours: Manufacturing	M 14386098	Avg hrs: mfg
29	127	6	CES0600000008	Avg Hourly Earnings: Goods-Producing	M 123109182	AHE: goods
30	128	6	CES2000000008	Avg Hourly Earnings: Construction	M 123109341	AHE: const
31	129	6	CES3000000008	Avg Hourly Earnings: Manufacturing	M 123109552	AHE: mfg

Note: The column "tcode" denotes the following data transformation for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta\,(x_t/x_{t-1}-1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A3. Data Description: Housing – FRED-MD

Group 3: Housing						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M 110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M 110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M 110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M 110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M 110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M 110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M 110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M 110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M 110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M 110155534	BP: West

Note: The column "tcode" denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta\,(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A4. Data Description: Consumption, orders, and inventories – FRED-MD

Group 4: Consumption, orders, and inventories						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	3	5	DPCERA3M086SBEA	Real personal consumption expenditures	M 123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	M 110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M 130439509	Retail sales
4	65*	5	AMDMNOx	New Orders for Durable Goods	M 14386110	Orders: dble gds
5	67*	5	AMDMUOx	Unfilled Orders for Durable Goods	M 14385946	Unf orders: dble
6	68*	5	BUSINVx	Total Business Inventories	M 15192014	M&T invent
7	69*	2	ISRATIOx	Total Business: Inventories to Sales Ratio	M 15191529	M&T invent/sales

Note: The column "tcode" denotes the following data transformation for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$; (7) $\Delta\ (x_t/x_{t-1} - 1\ )$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A5. Data Description: Money and Credit – FRED-MD

Group 5: Money and Credit						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	70	6	M1SL	M1 Money Stock	M 110154984	M1
2	71	6	M2SL	M2 Money Stock	M 110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M 110154985	M2 (real)
4	73	6	BOGMBASE	Monetary Base; Total	N.A.	N.A.
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M 110155011	Reserves tot
6	75	7	NONBORRES	Reserves Of Depository Institutions	M 110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	78	6	NONREVSL	Total Nonrevolving Credit	M 110154564	Cons credit
10	79*	2	CONSPI	Nonrevolving consumer credit to Personal Income	M 110154569	Inst cred/PI
11	131	6	MZMSL	MZM Money Stock	N.A.	N.A.
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N.A.	N.A.
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N.A.	N.A.
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A.	N.A.

Note: The column "tcode" denotes the following data transformation for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\,(x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI**.**

Table A6. Data Description: Interest and exchange rates – FRED-MD

Group 6: Interest and exchange rates						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M 110155157	Fed Funds
2	85*	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	CPF3M	Comm paper
3	86	2	TB3MS	3-Month Treasury Bill:	M 110155165	3 mo T-bill
4	87	2	TB6MS	6-Month Treasury Bill:	M 110155166	6 mo T-bill
5	88	2	GS1	1-Year Treasury Rate	M 110155168	1 yr T-bond
6	89	2	GS5	5-Year Treasury Rate	M 110155174	5 yr T-bond
7	90	2	GS10	10-Year Treasury Rate	M 110155169	10 yr T-bond
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
10	93*	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS		CP-FF spread
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
18	102*	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M 110154768	Ex rate: Switz
19	103*	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M 110154755	Ex rate: Japan
20	104*	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	M 110154772	Ex rate: UK
21	105*	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M 110154744	Ex rate: Canada

Note: The column "tcode" denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\ (x_t/x_{t-1} - 1\ )$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A7. Data Description: Prices – FRED-MD

Group 7: Prices						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	106	6	WPSFD49207	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	WPSFD49502	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	WPSID61	PPI: Intermediate Materials	M 110157527	PPI: int matls
4	109	6	WPSID62	PPI: Crude Materials	M 110157500	PPI: crude matls
5	110*	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	M 110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and metal products	M 110157335	PPI: nonferrous
7	113	6	CPIAUCSL	CPI: All Items	M 110157323	CPI-U: all
8	114	6	CPIAPPSL	CPI: Apparel	M 110157299	CPI-U: apparel
9	115	6	CPITRNSL	CPI: Transportation	M 110157302	CPI-U: transp
10	116	6	CPIMEDSL	CPI: Medical Care	M 110157304	CPI-U: medical
11	117	6	CUSR0000SAC	CPI: Commodities	M 110157314	CPI-U: comm.
12	118	6	CUUR0000SAD	CPI: Durables	M 110157315	CPI-U: dbles
13	119	6	CUSR0000SAS	CPI : Services	M 110157325	CPI-U: services
14	120	6	CPIULFSL	CPI: All Items Less Food	M 110157328	CPI-U: ex food
15	121	6	CUUR0000SA0L2	CPI: All items less shelter	M 110157329	CPI-U: ex shelter
16	122	6	CUSR0000SA0L5	CPI: All items less medical care	M 110157330	CPI-U: ex med
17	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
18	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	gmdcd	PCE defl: dlbes
19	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	gmdcn	PCE defl: nondble
20	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Note: The column "tcode" denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\ (x_t/x_{t-1} - 1)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A8. Data Description: Stock Market – FRED-MD

Group 8: Stock Market						
	ID	T-code	Mnemonic	Description	Global Insight (GSI)	GSI: description
1	80*	5	S&P 500	S&P's Common Stock Price Index: Composite	M 110155044	S&P 500
2	81*	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M 110155047	S&P: indust
3	82*	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield		S&P div yield
4	83*	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio		S&P PE ratio
5	135*	1	VXOCLSx	VXO		

Note: The column "tcode" denotes the following data transformation for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\ (x_t/x_{t-1} - 1\ )$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Table A9. RMSE ratio among all ML models and AR(12) – one step ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.96**	0.98	0.98	1.04	1.00	1.03	0.93***	1.04	0.92**	0.97*	1.05**
Ridge	1.05**	1.00	1.03*	1.02	1.09**	1.05	1.08**	0.97***	1.09*	0.96	1.01	1.10***
Elnet	1.02	0.98*	1.00	1.00	1.06*	1.02	1.05	0.95***	1.06	0.94*	0.99	1.07***
Adalasso	1.02	0.98	1.00	1.00	1.07***	1.02	1.05**	0.95	1.06*	0.94*	0.99	1.07**
RF	0.96	0.92**	0.94*	0.94***	1.00	0.96	0.99	0.89***	1.00	0.88***	0.93**	1.00
Bag	1.00	0.96	0.98	0.98	1.04	1.00	1.03	0.93	1.04	0.92	0.97	1.05
CBoost	0.97	0.93**	0.95	0.95**	1.01	0.97	1.00	0.90***	1.01	0.89***	0.94**	1.02
AdaR2	1.08***	1.03***	1.06***	1.05	1.12***	1.08	1.11***	1.00	1.12**	0.99	1.04**	1.13***
AdaRT	0.96	0.92*	0.94	0.94**	1.00	0.96	0.99	0.89**	1.00	0.88***	0.93*	1.01
XGB	1.09	1.04	1.07**	1.06*	1.14***	1.09	1.12***	1.01	1.13***	1.00	1.05**	1.14***
GradB	1.03*	0.99	1.01	1.01	1.08	1.03	1.06**	0.96**	1.08*	0.95**	1.00	1.08***
AR (12)	0.95**	0.91***	0.94***	0.93**	1.00	0.95	0.98	0.89***	0.99	0.88***	0.92***	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A10. RMSE ratio among all ML models and AR(12) – four steps ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.94**	0.95*	0.99	0.99	0.56	1.01	0.93***	0.97	1.68***	0.99	1.03
Ridge	1.06**	1.00	1.01	1.05	1.05	0.60	1.07**	0.99	1.03	1.78***	1.05	1.09***
Elnet	1.05*	0.99	1.00	1.04	1.04	0.59	1.06*	0.98	1.02	1.76***	1.04	1.08***
Adalasso	1.01	0.95	0.96	1.00	1.00	0.57	1.02	0.94	0.98	1.69***	1.01	1.04
RF	1.01	0.95	0.96	1.00	1.00	0.57	1.02	0.94	0.98*	1.69***	1.00	1.04
Bag	1.77	1.67	1.69	1.76	1.76	1.00	1.79	1.66	1.73	2.97	1.77	1.83
CBoost	0.99	0.94**	0.95*	0.98	0.98	0.56	1.00	0.93**	0.97	1.66***	0.99	1.02
AdaR2	1.07***	1.01	1.02	1.06	1.06	0.60	1.08**	1.00	1.04	1.79***	1.06	1.10***
AdaRT	1.03	0.97	0.98	1.02	1.02*	0.58	1.04	0.96	1.00	1.72***	1.02	1.06
XGB	0.60***	0.56***	0.57***	0.59***	0.59***	0.34	0.60***	0.56***	0.58***	1.00	0.59***	0.62***
GradB	1.01	0.95	0.96	0.99	1.00	0.57	1.01	0.94	0.98	1.68***	1.00	1.04
AR (12)	0.97	0.91***	0.93***	0.96	0.96	0.55	0.98	0.91***	0.94	1.63***	0.96	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A11. RMSE ratio among all ML models and AR (12) – twelve steps ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.94***	0.95**	1.01	0.96*	0.71	1.00	0.94***	0.97	0.56***	0.93**	1.02
Ridge	1.06***	1.00	1.01	1.07*	1.02	0.75	1.06*	1.00	1.03	0.60***	0.99	1.08***
Elnet	1.05**	0.99	1.00	1.06	1.01	0.75	1.05	0.99	1.02	0.59***	0.98	1.07***
Adalasso	0.99	0.94*	0.94	1.00	0.96	0.70	0.99	0.93*	0.96	0.56***	0.92***	1.01
RF	1.04*	0.98***	0.99	1.04	1.00	0.74	1.04	0.97	1.00	0.58***	0.97	1.06**
Bag	1.41	1.33	1.34	1.42	1.36	1.00	1.41	1.32	1.37	0.79	1.31	1.44
CBoost	1.00	0.94*	0.95	1.01	0.97	0.71	1.00	0.94**	0.97	0.56***	0.93**	1.02
AdaR2	1.07***	1.00	1.01	1.07*	1.03	0.75	1.06**	1.00	1.03	0.60***	0.99	1.09***
AdaRT	1.03	0.97	0.98	1.04	1.00	0.73	1.03	0.97	1.00	0.58***	0.96	1.05*
XGB	1.78***	1.68***	1.70***	1.80***	1.72***	1.26	1.78***	1.67	1.73***	1.00	1.66***	1.82***
GradB	1.08**	1.01	1.02	1.08***	1.04	0.76	1.07**	1.01	1.04	0.60***	1.00	1.10**
AR (12)	0.98	0.92***	0.93***	0.99	0.94**	0.69	0.98	0.92***	0.95*	0.55***	0.91**	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A12. RMSE ratio among all ML models and AR(12) – one step ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.94**	1.00	0.97	1.04*	1.02	1.02	0.93***	1.02	0.92*	0.97**	1.03
Ridge	1.06**	1.00	1.06**	1.03	1.10**	1.08	1.08**	0.99	1.08**	0.97	1.03*	1.08**
Elnet	1.00	0.95**	1.00	0.98	1.04*	1.02	1.02	0.94***	1.03	0.92*	0.97**	1.03
Adalasso	1.03	0.97	1.03	1.00	1.07**	1.05	1.05*	0.96	1.05	0.94	1.00	1.05
RF	0.96*	0.91**	0.96*	0.94**	1.00	0.98	0.99	0.90***	0.99	0.89**	0.93**	0.99
Bag	0.98	0.93	0.98	0.95	1.02	1.00	1.00	0.92	1.00	0.90*	0.95	1.00
CBoost	0.98	0.92**	0.98	0.95*	1.02	1.00	1.00	0.92***	1.00	0.90**	0.95**	1.00
AdaR2	1.07***	1.01	1.07***	1.04	1.11***	1.09	1.09***	1.00	1.10**	0.98	1.04**	1.10***
AdaRT	0.98	0.92**	0.97	0.95	1.01	1.00	1.00	0.91**	1.00	0.90**	0.95*	1.00
XGB	1.09*	1.03	1.09*	1.06	1.13**	1.11*	1.11**	1.02	1.12**	1.00	1.05	1.12**
GradB	1.03**	0.98*	1.03**	1.00	1.07**	1.05	1.05**	0.97**	1.06*	0.95	1.00	1.06***
AR (12)	0.98	0.92***	0.97	0.95	1.01	1.00	1.00	0.91***	1.00	0.90**	0.94***	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A13. RMSE ratio among all ML models and AR(12) – four steps ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.96**	1.00	0.99	1.02	0.78	1.03	0.97*	1.02	1.67***	0.99	1.04**
Ridge	1.04**	1.00	1.04**	1.04	1.06**	0.82	1.08**	1.01	1.06*	1.74***	1.04	1.08***
Elnet	1.00	0.96**	1.00	1.00	1.02	0.78	1.03	0.97*	1.02	1.67***	1.00	1.04**
Adalasso	1.01	0.96	1.00	1.00	1.03	0.79	1.04	0.98	1.02	1.68***	1.00	1.04
RF	0.98	0.94**	0.98	0.98	1.00	0.77	1.01	0.95*	1.00	1.64***	0.98	1.02
Bag	1.28	1.23	1.28	1.27	1.30	1.00	1.32	1.24	1.30	2.14**	1.27	1.33
CBoost	0.97	0.93**	0.97	0.96	0.99	0.76	1.00	0.94**	0.99	1.62***	0.96	1.01
AdaR2	1.03*	0.99	1.03*	1.03	1.05*	0.81	1.06**	1.00	1.05*	1.72***	1.03	1.07***
AdaRT	0.98	0.94*	0.98	0.98	1.00	0.77	1.01	0.95*	1.00	1.64***	0.98	1.02
XGB	0.60***	0.57***	0.60***	0.59***	0.61***	0.47**	0.62***	0.58***	0.61***	1.00	0.59***	0.62***
GradB	1.01	0.96	1.00	1.00	1.02	0.79	1.04	0.98	1.02	1.68***	1.00	1.04
AR (12)	0.96**	0.92***	0.96**	0.96	0.98	0.75	0.99	0.93***	0.98	1.61***	0.96	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A14. RMSE ratio among all ML models and AR(12) – twelve steps ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.95**	1.00	0.99	1.01	0.74	1.03	0.96**	1.01	0.56***	1.00	1.04**
Ridge	1.06**	1.00	1.06**	1.04	1.06*	0.78	1.09**	1.01	1.07*	0.59***	1.06	1.10***
Elnet	1.00	0.95**	1.00	0.99	1.00	0.74	1.03	0.96**	1.01	0.56***	1.00	1.04**
Adalasso	1.01	0.96	1.01	1.00	1.02	0.75	1.04**	0.97	1.03	0.57***	1.02	1.05*
RF	0.99	0.94*	1.00	0.98	1.00	0.74	1.02	0.95*	1.01	0.56***	1.00	1.03
Bag	1.35	1.28	1.35	1.33	1.36	1.00	1.39	1.29	1.36	0.76*	1.35	1.40
CBoost	0.97	0.92**	0.97	0.96**	0.98	0.72	1.00	0.93**	0.98	0.55***	0.98	1.01
AdaR2	1.04**	0.99	1.04**	1.03	1.05*	0.77	1.07**	1.00	1.06**	0.59***	1.05	1.08***
AdaRT	0.99	0.93*	0.99	0.97	0.99	0.73	1.02	0.95**	1.00	0.55***	0.99	1.03
XGB	1.78***	1.69***	1.78***	1.76***	1.79***	1.32*	1.83***	1.71***	1.80***	1.00	1.79***	1.85***
GradB	1.00	0.94	1.00	0.98	1.00	0.74	1.02	0.95	1.01	0.56***	1.00	1.04
AR (12)	0.96**	0.91***	0.96**	0.95*	0.97	0.71	0.99	0.92***	0.97	0.54***	0.97	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A15. RMSE ratio among all ML models and AR (12) before the Great Recession – one-step ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.97	0.99	0.96**	1.01	0.87***	1.01	0.92***	1.02	1.63***	0.84***	1.07***
Ridge	1.03	1.00	1.02	0.98	1.04**	0.89**	1.04**	0.95***	1.05**	1.68***	0.87***	1.10***
Elnet	1.01	0.98	1.00	0.97	1.02	0.87***	1.02	0.93***	1.03	1.64***	0.85***	1.08***
Adalasso	1.05**	1.02	1.04	1.00	1.06*	0.91**	1.06**	0.97	1.06*	1.70***	0.88***	1.12***
RF	0.99	0.96**	0.98	0.95*	1.00	0.86***	1.00	0.91***	1.01	1.61***	0.84***	1.06**
Bag	1.16***	1.12**	1.14***	1.10**	1.17***	1.00	1.17***	1.07	1.17***	1.88***	0.98	1.24***
CBoost	0.99	0.96**	0.98	0.94**	1.00	0.85***	1.00	0.91***	1.00	1.60***	0.83***	1.06**
AdaR2	1.08***	1.05***	1.07***	1.03	1.09***	0.94	1.10***	1.00	1.10***	1.76***	0.91**	1.16***
AdaRT	0.98	0.95**	0.97	0.94*	0.99	0.85***	1.00	0.91***	1.00	1.60***	0.83***	1.05**
XGB	0.62***	0.60***	0.61***	0.59***	0.62***	0.53***	0.62***	0.57***	0.63***	1.00	0.52***	0.66***
GradB	1.18***	1.15***	1.17***	1.13***	1.20***	1.02	1.20***	1.09**	1.20***	1.92***	1.00	1.27***
AR (12)	0.93***	0.91***	0.92***	0.89***	0.94**	0.81***	0.95**	0.86***	0.95**	1.52***	0.79***	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A16. RMSE ratio among all ML models and AR(12) before the Great Recession – four-steps ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.97	1.00	0.95**	1.01	0.78***	1.01	0.97	1.01	1.36***	0.84***	1.07***
Ridge	1.03	1.00	1.03*	0.98	1.05**	0.80***	1.05**	1.00	1.05**	1.40***	0.87***	1.10***
Elnet	1.00	0.97*	1.00	0.95*	1.02	0.78***	1.02	0.97	1.02	1.36***	0.84***	1.08***
Adalasso	1.06**	1.02	1.05*	1.00	1.07**	0.82**	1.07**	1.02	1.07**	1.43***	0.89**	1.13***
RF	0.99	0.96**	0.98	0.93**	1.00	0.76***	1.00	0.96*	1.00	1.34***	0.83***	1.06**
Bag	1.29***	1.25***	1.28***	1.22**	1.31***	1.00	1.31***	1.25***	1.31***	1.75***	1.08	1.38***
CBoost	0.99	0.96**	0.98	0.93**	1.00	0.76***	1.00	0.96**	1.00	1.34***	0.83***	1.06**
AdaR2	1.03	1.00	1.03	0.98	1.05**	0.80***	1.04**	1.00	1.04**	1.40***	0.87***	1.10***
AdaRT	0.99	0.96**	0.98	0.93**	1.00	0.76***	1.00	0.96**	1.00	1.34***	0.83***	1.06**
XGB	0.74***	0.71***	0.73***	0.70***	0.75***	0.57***	0.75***	0.72***	0.75	1.00	0.62***	0.79***
GradB	1.19***	1.15***	1.18***	1.13**	1.21***	0.92	1.21***	1.15***	1.21***	1.61***	1.00	1.27***
AR (12)	0.93***	0.91***	0.93***	0.88**	0.95***	0.72***	0.95**	0.91***	0.95**	1.27***	0.78***	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A17. RMSE ratio among all ML models and AR(12) before the Great Recession – twelve-steps ahead – fixed windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.98	1.01	0.96*	1.02	0.49***	1.03	0.97	1.03*	0.14***	0.85***	1.08***
Ridge	1.02	1.00	1.03*	0.98	1.04*	0.50***	1.05**	0.99	1.05**	0.14***	0.86***	1.11***
Elnet	0.99	0.97*	1.00	0.95*	1.01	0.48***	1.02	0.96*	1.02	0.14***	0.84***	1.07***
Adalasso	1.04*	1.02	1.05*	1.00	1.06*	0.51***	1.07**	1.01	1.07**	0.14***	0.88***	1.13***
RF	0.98	0.96*	0.99	0.94*	1.00	0.48***	1.01	0.95**	1.01	0.13***	0.83***	1.06**
Bag	2.05**	2.01*	2.07**	1.96**	2.09**	1.00	2.10**	1.99**	2.11**	0.28***	1.73*	2.22***
CBoost	0.97	0.95**	0.98	0.93**	0.99	0.48***	1.00	0.95**	1.00	0.13***	0.82***	1.06***
AdaR2	1.03	1.01	1.04*	0.99	1.05	0.50***	1.06**	1.00	1.06**	0.14***	0.87***	1.12***
AdaRT	0.97*	0.95**	0.98	0.93**	0.99	0.47**	1.00	0.94**	1.00	0.13***	0.82***	1.05**
XGB	7.27***	7.13***	7.34***	6.97***	7.42***	3.55***	7.47***	7.07***	7.49***	1.00	6.15***	7.88***
GradB	1.18***	1.16***	1.19***	1.13***	1.21***	0.58*	1.21***	1.15***	1.22***	0.16***	1.00	1.28***
AR (12)	0.92***	0.90***	0.93***	0.88***	0.94**	0.45**	0.95**	0.90***	0.95**	0.13***	0.78***	1.00

Significance after GW test: *p<0.1, **p<0.05, ***p<0.01.

Table A18. RMSE ratio among all ML models and AR(12) before the Great Recession – one-step ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.96**	1.00	0.94**	1.00	0.85**	0.99	0.92*	1.00	0.94*	0.88***	1.06**
Ridge	1.04**	1.00	1.04**	0.98	1.04*	0.89*	1.03	0.95**	1.04	0.98	0.92**	1.10***
Elnet	1.00	0.96**	1.00	0.95**	1.00	0.86**	0.99	0.92*	1.00	0.95*	0.89***	1.06***
Adalasso	1.06**	1.02	1.06**	1.00	1.06**	0.91	1.05	0.97	1.06**	1.00	0.94*	1.12***
RF	1.00	0.96*	1.00	0.94**	1.00	0.85**	0.99	0.92***	1.00	0.94**	0.88***	1.05**
Bag	1.17**	1.13*	1.17**	1.10	1.17**	1.00	1.16*	1.07	1.17**	1.10	1.03	1.24***
CBoost	1.01	0.97	1.01	0.96	1.01	0.86*	1.00	0.93**	1.01	0.96	0.89***	1.07**
AdaR2	1.09***	1.05**	1.09**	1.03	1.09***	0.93	1.08**	1.00	1.09***	1.03	0.96	1.15***
AdaRT	1.00	0.97	1.00	0.95**	1.00	0.86**	0.99	0.92***	1.00	0.95*	0.89***	1.06*
XGB	1.06*	1.02	1.06*	1.00	1.06**	0.91	1.05	0.97	1.06*	1.00	0.94***	1.12***
GradB	1.13***	1.09**	1.13***	1.07*	1.13***	0.97	1.12***	1.04	1.13***	1.07***	1.00	1.20***
AR (12)	0.95**	0.91***	0.94***	0.89***	0.95**	0.81***	0.94**	0.87***	0.94*	0.89***	0.84***	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A19. RMSE ratio among all ML models and AR(12) before the Great Recession – four-steps ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.96**	1.00	0.94**	1.01	0.71	0.99	0.97**	1.00	1.33***	0.82***	1.06**
Ridge	1.04**	1.00	1.04**	0.97	1.04**	0.74	1.03	1.00	1.04	1.37***	0.85***	1.10***
Elnet	1.00	0.96**	1.00	0.94**	1.01	0.71	0.99	0.97**	1.00	1.33***	0.82***	1.06***
Adalasso	1.07**	1.03	1.07**	1.00	1.07**	0.76	1.06*	1.03	1.07**	1.41***	0.87***	1.13***
RF	0.99	0.96**	0.99	0.93**	1.00	0.71	0.98	0.96**	0.99	1.32***	0.81***	1.05**
Bag	1.40	1.35	1.40	1.31	1.41	1.00	1.39	1.36	1.40	1.86*	1.14	1.48
CBoost	1.01	0.97	1.01	0.95*	1.02	0.72	1.00	0.98	1.01	1.34***	0.82***	1.07**
AdaR2	1.03**	1.00	1.03**	0.97	1.04**	0.74	1.02	1.00	1.03	1.37***	0.84***	1.09***
AdaRT	1.00	0.96	1.00	0.94**	1.01	0.71	0.99	0.97	1.00	1.32***	0.82***	1.06**
XGB	0.75***	0.73***	0.75***	0.71***	0.76***	0.54*	0.75***	0.73***	0.75***	1.00	0.62***	0.80***
GradB	1.22***	1.18***	1.22***	1.15***	1.23***	0.87	1.21***	1.19***	1.22***	1.62***	1.00	1.29***
AR (12)	0.94**	0.91***	0.95***	0.89***	0.95**	0.67	0.94**	0.92***	0.95**	1.25***	0.77***	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A20. RMSE ratio among all ML models and AR(12) before the Great Recession – twelve-steps ahead – expanding windows

Model	**Lasso**	**Ridge**	**Elnet**	**Adalasso**	**RF**	**Bag**	**CBoost**	**AdaR2**	**AdaRT**	**XGB**	**GradB**	**AR (12)**
Lasso	1.00	0.97*	1.00	0.94**	1.01	0.75	1.00	0.95***	1.01	0.45***	0.87***	1.07***
Ridge	1.03*	1.00	1.03**	0.97	1.03*	0.77	1.02	0.98	1.04	0.46***	0.89***	1.10***
Elnet	1.00	0.97**	1.00	0.94***	1.00	0.75	0.99	0.95***	1.01	0.45***	0.87***	1.06***
Adalasso	1.06**	1.03	1.07***	1.00	1.07***	0.80	1.06*	1.01	1.08***	0.47***	0.92**	1.13***
RF	0.99	0.97*	1.00	0.93***	1.00	0.75	0.99	0.95**	1.01	0.44***	0.86***	1.06**
Bag	1.33	1.29	1.33	1.25	1.34	1.00	1.32	1.26	1.34	0.59***	1.15	1.42
CBoost	1.00	0.98	1.01	0.94*	1.01	0.76	1.00	0.96*	1.02	0.45***	0.87***	1.07**
AdaR2	1.05***	1.02	1.05***	0.99	1.06**	0.79	1.05*	1.00	1.06**	0.47***	0.91***	1.12***
AdaRT	0.99	0.96	0.99	0.93***	0.99	0.74	0.98	0.94**	1.00	0.44***	0.86***	1.05*
XGB	2.24***	2.18***	2.25***	2.11***	2.26***	1.69***	2.23***	2.13***	2.27***	1.00	1.94***	2.39***
GradB	1.15***	1.12***	1.16***	1.08**	1.16***	0.87	1.15***	1.10***	1.17***	0.51***	1.00	1.23***
AR (12)	0.94***	0.91***	0.94***	0.88***	0.94**	0.71	0.93**	0.89***	0.95*	0.42***	0.81***	1.00

Significance after DM test: *p<0.1, **p<0.05, ***p<0.01.

Table A21. RMSEs forecasting models full sample, relative to benchmark models – fixed estimations – all horizons

Model	**AR (p)**					
Horizon	h=1	h=2	h=3	h=4	h=8	h=12
Lasso	1.04	1.03	1.03	1.02	1.01	1.01
Ridge	1.08***	1.08***	1.08***	1.08***	1.07***	1.07***
Elnet	1.06***	1.05***	1.06***	1.07***	1.06***	1.06***
Adalasso	1.06*	1.04	1.04	1.03	1.02	1.00
RF	0.99	1.00	1.01	1.03	1.06*	1.05*
Bag	1.04	1.14	1.39	1.81	1.27*	1.42
CBoost	1.01	1.01	1.01	1.01	1.02	1.01
AdaR2	1.12***	1.10***	1.09***	1.09***	1.09***	1.07***
AdaRT	0.99	1.01	1.02	1.05	1.06*	1.04
XGB	1.13***	0.58***	0.70***	0.61***	1.20**	1.80***
GradB	1.07***	1.16***	1.06	1.02	1.05	1.08**
	FM					
Lasso	0.99	0.99	0.98	0.97	0.97	0.97
Ridge	1.03	1.03	1.03	1.03	1.03	1.03
Elnet	1.01	1.01	1.01	1.02	1.02	1.02
Adalasso	1.01	0.99	0.99	0.98	0.98	0.97
RF	0.95***	0.96*	0.96	0.98	1.01	1.01
Bag	0.99	1.09	1.32	1.73	1.21	1.37
CBoost	0.96**	0.96*	0.96**	0.97*	0.97	0.97
AdaR2	1.07	1.05	1.04	1.04	1.04	1.04
AdaRT	0.95***	0.97	0.98	1.00	1.02	1.01
XGB	1.08*	0.55***	0.67***	0.58***	1.15**	1.74***
GradB	1.02	1.11***	1.01	0.98	1.00	1.05
	RW					
Lasso	0.76***	0.78****	0.79***	0.78***	0.69	0.65
Ridge	0.79**	0.82***	0.83***	0.82***	0.74***	0.69***
Elnet	0.77**	0.80***	0.82***	0.82***	0.73***	0.69***
Adalasso	0.78***	0.78***	0.80***	0.79***	0.70***	0.65***
RF	0.73***	0.76***	0.78***	0.79***	0.73***	0.68***
Bag	0.76***	0.86	1.07	1.38	0.87	0.92
CBoost	0.74***	0.76***	0.78***	0.77***	0.70***	0.65***
AdaR2	0.82*	0.83***	0.84***	0.83***	0.75***	0.69***
AdaRT	0.73***	0.76***	0.79***	0.80***	0.73***	0.67***
XGB	0.83*	0.44***	0.54***	0.46***	0.83**	1.16*
GradB	0.79**	0.88**	0.82***	0.78***	0.72***	0.70***

Notes: Numbers rounded to the second decimal. Significance after DM/GW test: *p<0.1, **p<0.05, ***p<0.01.

Other results and descriptive statistics

Table A22. Data description: Transformations of real and international economy variables – Macrohistory Database

	Population	Real Con-sumption pc	GDP	Investto-GDP	Current Account	Imports	Exports	EX US dollar
Australia	5	–	5	–	2	5	5	5
Belgium	6	5	6	5	2	5	5	5
Denmark	6	5	6	5	3	5	5	5
Finland	6	5	6	5	2	6	6	5
France	5	5	5	5	2	5	5	5
Germany	5	5	5	5	2	5	5	5
Italy	6	6	6	5	2	5	5	5
Japan	6	6	5	5	2	5	6	5
Netherlands	6	5	6	5	2	5	5	5
Norway	5	5	5	5	2	5	5	5
Portugal	6	6	6	–	2	5	5	5
Spain	6	5	6	5	–	6	6	5
Sweden	5	5	5	5	2	5	5	5
Switzerland	5	5	6	5	–	5	5	5
UK	6	5	6	5	3	5	5	5
USA	6	5	5	5	2	5	5	–

Note: The numbers are related to the transformations for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4$\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta (x_t/x_{t-1} - 1)$.

Table A23. Data description: Transformations of government, money, prices and interest rates' variables – Macrohistory Database

	Debt-to-GDP	**Gov. Revenue**	**Gov. Expenditure**	**CPI**	**Narrow Money**	**Broad Money**	**Short term Int. rates**	**Longterm Int. rates**
Australia	2	–	–	6	5	5	2	6
Belgium	2	6	5	5	6	–	2	6
Denmark	2	–	6	5	5	5	2	6
Finland	–	6	6	6	6	6	2	6
France	2	6	6	6	6	6	2	6
Germany	2	5	5	5	5	5	2	6
Italy	2	6	6	5	6	6	2	6
Japan	2	6	6	6	6	6	–	2
Netherlands	2	6	6	5	6	5	2	6
Norway	2	5	5	5	5	5	2	5
Portugal	3	5	6	5	5	6	2	5
Spain	2	6	6	6	6	6	2	5
Sweden	2	6	5	5	6	5	2	6
Switzerland	2	5	5	5	5	6	2	2
UK	2	5	5	5	5	5	2	6
USA	2	5	5	5	5	5	5	5

Note: The numbers are related to the transformations for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta\,(x_t/x_{t-1} - 1)$.

Table A24. Data description: Transformations of credit and asset price variables – Macrohistory Database

	Total Loans Non fin. pr. sector	Mortgage Loans Non fin. pr. sector	Total Loans to Household	Total Loans to Business	Stock Prices	House Prices
Australia	6	5	5	5	5	5
Belgium	–	–	–	–	5	5
Denmark	6	6	–	–	–	5
Finland	6	–	–	–	–	–
France	–	6	–	–	–	6
Germany	5	5	–	–	5	–
Italy	6	6	–	–	–	–
Japan	5	–	–	–	5	–
Netherlands	6	6	–	–	5	5
Norway	5	5	–	–	5	5
Portugal	–	–	–	–	6	–
Spain	–	–	–	–	–	–
Sweden	5	5	–	–	–	5
Switzerland	5	6	6	5	–	5
UK	6	6	6	5	5	–
USA	5	5	–	–	5	5

Note: The numbers are related to the transformations for a series *x*: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; (7) $\Delta (x_t/x_{t-1} - 1)$.

Table A25. Variables included and excluded in the estimations – part 1

Country\ Variables	GDP	POP	RCONPC	InvesttoGDP Ratio	Current account	Import	Export	DebttoGDP	GovRev	GovExp	CPI
Australia											
Belgium											
Denmark											
Finland											
France											
Germany											
Italy											
Japan											
Netherlands											
Norway											
Portugal											
Spain											
Sweden											
Switzerland											
United Kingdom											
USA											

Source: Jordà-Schularick-Taylor Macrohistory Database. Notes: The time series in green are the ones included in the estimations. The yellow ones are the time series excluded because of the presence of too many consecutive missing years. The red ones are excluded because they show a large amount of missing data at the beginning of the time frame and their inclusion in the estimation would lead to a loss of more than one crisis.

Table A26. Variables included and excluded in the estimations – part 2

Country\Variables	M0	M2	Short term Int Rate	Long term Int Rate	Loans to Non fin Priv	Mort Loans to Non fin Priv	Loans to Households
Australia							
Belgium							
Denmark							
Finland							
France							
Germany							
Italy							
Japan							
Netherlands							
Norway							
Portugal							
Spain							
Sweden							
Switzerland							
United Kingdom							
USA							

Source: Jordà-Schularick-Taylor Macrohistory Database. Notes: The time series in green are the ones included in the estimations. The yellow ones are the time series excluded because of the presence of too consecutive missing years. The red ones are excluded because they show a large amount of missing data at the beginning of the time frame and their inclusion in the estimation would lead to a loss of more than one crisis.

Table A27. Variables included and excluded in the estimations – part 3

Country\Variables	Loans to Business	Stock Prices	House Prices	Ex rate US
Australia				
Belgium				
Denmark				
Finland				
France				
Germany				
Italy				
Japan				
Netherlands				
Norway				
Portugal				
Spain				
Sweden				
Switzerland				
United Kingdom				
USA				

Source: Jordà-Schularick-Taylor Macrohistory Database. Notes: The time series in green are the ones included in the estimations. The yellow ones are the time series excluded because of the presence of too consecutive missing years. The red ones are excluded because they show a large amount of missing data at the beginning of the time frame and their inclusion in the estimation would lead to a loss of more than one crisis.

Table A28. Baseline estimations – linear variables – Italian data

	Fixed est. no US XR	Fixed est. with US XR	Expanding est. no US XR	Expanding est. with US XR
Number of variables	34	36	34	36
Observations per variables	144	144	144	144
Number crises exploited	9	9	9	9
ME logit	0.19	0.17	0.22	0.22
Precision logit	0	0	0	0.06
Recall logit	0	0	0	0.5
F1score logit	–	–	–	0.11
ME step-wise logit	0.08	0.08	0.17	0.19
Precision step-wise logit	0	0	0	0
Recall step-wise logit	0	0	0	0
F1score step-wise logit	–	–	–	–
ME AdaBoost	0.03	0.03	0.03	0.03
Precision AdaBoost	0	0	0	0
Recall AdaBoost	0	0	0	0
F1score AdaBoost	–	–	–	–

Notes: "–" undefined or not available.

Table A29. Baseline estimations – nonlinear variables included – Italian data

	Fixed est. no US XR	Fixed est. with US XR	Expanding est. no US XR	Expanding est. with US XR
Number of variables	102	108	102	108
Observations per variables	144	144	144	144
Number crises exploited	9	9	9	9
ME logit	0.36	0.39	0.36	0.35
Precision logit	0.07	0.04	0.04	0
Recall logit	1	0.5	0.5	0
F1score logit	0.13	0.07	0.07	–
ME step-wise logit	0.06	0.08	0.13	0.1
Precision step-wise logit	0	0	0	0
Recall step-wise logit	0	0	0	0
F1score step-wise logit	–	–	–	–
ME AdaBoost	0.03	0.03	0.03	0.03
Precision AdaBoost	–	–	–	–
Recall AdaBoost	0	0	0	0
F1score AdaBoost	–	–	–	–

Notes: "–" undefined or not available.

Table A30. Estimations with Casabianca et al. (2019) method of sampling – linear variables – Italian data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	34	36	34	36
Observations per variables	107	107	107	107
Number pre-crises exploited	21	21	21	21
ME logit	0.38	0.36	0.32	0.36
Precision logit	0.06	0.12	0.18	0.16
Recall logit	0.17	0.33	0.5	0.5
F1score logit	0.09	0.17	0.26	0.24
ME step-wise logit	0.28	0.26	0.34	0.38
Precision step-wise logit	0.09	0.1	0.2	0.12
Recall step-wise logit	0.17	0.17	0.67	0.33
F1score step-wise logit	0.18	0.13	0.31	0.17
ME AdaBoost	*0.2*	0.1	*0.1*	*0.1*
Precision AdaBoost	*0*	0	*0*	*0*
Recall AdaBoost	*0*	0	*0*	*0*
F1score AdaBoost	–	–	–	–

Notes: "–" undefined or not available; unstable values in italics (the most frequent result is reported).

Table A31. Estimations with Casabianca et al. (2019) method of sampling – nonlinear variables included – Italian data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	102	108	102	108
Observations per variables	107	107	107	107
Number pre-crises exploited	21	21	21	21
ME logit	0.47	0.42	0.34	0.45
Precision logit	0.15	0.19	0.17	0.18
Recall logit	0.67	0.83	0.5	0.83
F1score logit	0.24	0.31	0.25	0.29
ME step-wise logit	0.17	0.13	0.32	0.17
Precision step-wise logit	0.33	0.43	0.08	0.29
Recall step-wise logit	0.5	0.5	0.17	0.33
F1score step-wise logit	0.4	0.46	0.11	0.31
ME AdaBoost	*0.1*	0.1	0.1	0.1
Precision AdaBoost	*0*	0	0	0
Recall AdaBoost	*0*	0	0	0
F1score AdaBoost	–	–	–	–

Notes: "–" undefined or not available; unstable values in italics (the most frequent result is reported).

Table A32. Baseline estimations – linear variables – Japanese data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	32	34	32	34
Observations per variables	130	130	130	130
Number crises exploited	5	5	5	5
ME logit	0.15	0.15	0.17	0.14
Precision logit	0	0	0	0
Recall logit	0	0	0	0
F1score logit	–	–	–	–
ME step-wise logit	0.06	0.08	0.12	0.14
Precision step-wise logit	0	0	0	0
Recall step-wise logit	0	0	0	0
F1score step-wise logit	–	–	–	–
ME AdaBoost	**	**	0.02	0.02
Precision AdaBoost	–	–	–	–
Recall AdaBoost	–	–	0	0
F1score AdaBoost	–	–	–	–

Notes: "**" not executed; "–" undefined or not available.

Table A33. Baseline estimations – nonlinear variables included – Japanese data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	97	103	97	103
Observations per variables	130	130	130	130
Number crises exploited	5	5	5	5
ME logit	0.49	0.43	0.42	0.52
Precision logit	0	0	0	0
Recall logit	0	0	0	0
F1score logit	–	–	–	–
ME step-wise logit	0.11	0.08	0.08	0.12
Precision step-wise logit	0	0	0	0
Recall step-wise logit	0	0	0	0
F1score step-wise logit	–	–	–	–
ME AdaBoost	**	**	0.02	0.02
Precision AdaBoost	–	–	–	–
Recall AdaBoost	–	–	0	0
F1score AdaBoost	–	–	–	–

Notes: "**" not executed; "–" undefined or not available.

Table A34. Estimations with Casabianca et al. (2019) method of sampling – linear variables – Japanese data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	97	103	97	103
Observations per variables	110	110	110	110
Number pre-crises exploited	15	15	15	15
ME logit	0.33	0.29	0.4	0.4
Precision logit	0.06	0.12	0.05	0.09
Recall logit	0.33	0.67	0.33	0.67
F1score logit	0.1	0.2	0.08	0.15
ME step-wise logit	0.18	0.31	0.22	0.3
Precision step-wise logit	0.18	0.11	0.09	0.67
Recall step-wise logit	0.67	0.67	0.33	0.33
F1score step-wise logit	0.03	0.19	0.14	0.11
ME AdaBoost	**	**	*0.04*	*0.04*
Precision AdaBoost	–	–	*1*	*1*
Recall AdaBoost	–	–	*0.33*	*0.33*
F1score AdaBoost	–	–	*0.5*	*0.5*

Notes: "**" not executed; "–" undefined or not available; unstable values in italics (the most frequent result is reported).

Table A35. Estimations with Casabianca et al. (2019) method of sampling – nonlinear variables included – Japanese data

	Fixed est. no US XR	**Fixed est. with US XR**	**Expanding est. no US XR**	**Expanding est. with US XR**
Number of variables	97	103	97	103
Observations per variables	110	110	110	110
Number pre-crises exploited	15	15	15	15
ME logit	0.42	0.35	0.32	0.4
Precision logit	0.12	0.1	0.06	0.12
Recall logit	1	0.67	0.33	1
F1score logit	0.21	0.17	0.1	0.21
ME step-wise logit	0.18	0.22	0.16	0.2
Precision step-wise logit	0.23	0.15	0.2	0.17
Recall step-wise logit	1	0.67	0.67	67
F1score step-wise logit	0.38	0.25	0.31	0.27
ME AdaBoost	**	**	*0.05*	*0.05*
Precision AdaBoost	–	–	–	–
Recall AdaBoost	–	–	*0*	*0*
F1score AdaBoost	–	–	–	–

Notes: "**" not executed; "–" undefined or not available; unstable values in italics (the most frequent result is reported).

SUMMARY IN ENGLISH

Forecasting business cycles and financial crises with machine learning

Motivation of the research

Policymakers, international organizations, businesses, and financial institutions all depend on accurate economic forecasting (Wieland and Wolters, 2013; Fouliard et al., 2021). For example, central banks, like the European Central Bank (ECB), need data updated more frequently than on the quarterly basis because the latter is insufficient to inform monetary policy decisions that are made every 6 weeks. Financial shocks and economic slowdowns require timely adjustments in policy responses (Sutherland, 2023; Woodford, 2005). This is why these institutions increasingly rely on nowcasting models, predictions on recent past, present for which data is not yet available or future of economic indicators. This is also the reason why they rely on early warning systems, tools for identifying potential financial crises before they escalate, enabling timely interventions by policymakers and market participants.

Financial crises, such as the 2007–2008 crisis, often have severe and long-lasting consequences (Shiller and Shiller, 2011), such as leading to the creation of institutions like the Financial Stability Board (FSB) and the European Systemic Risk Board (ESRB). To lessen financial volatility, these organisations rely on early warning indicators. Economic downturns are difficult to predict, especially those associated with financial crises (Bordo and Haubrich, 2017), therefore the need for reliable forecasting techniques becomes essential in macroeconomics. These forecasts identify potential financial crises before they escalate, enabling timely interventions by policymakers and market participants. But forecasting economic cycles and financial crises using conventional econometric methods can be complex due to their nonlinear patterns (Hamilton, 2005; Akinci and Queralto, 2022).

By utilising Big Data, which comprises unstructured data from text, satellite images, and telecommunications as well as structured economic indicators, advances in machine learning (ML) open up new forecasting possibilities (Hansen et al., 2019; Cerdeiro et al., 2020). This thesis focuses solely on structured data while unstructured data, that needs different methodologies (see for example Alfieri and Gabrielyan, 2024), is left outside the scope of this monograph.

Economic forecasts can be more accurate when ML models are used to capture nonlinear relationships in macroeconomic variables (Mosavi et al., 2020; Athey, 2018). Some studies, such as Casabianca et al. (2022) and Lehmann and Wohlrabe (2016) suggest that ML approaches can outperform traditional models in forecasting financial crises and macroeconomic indices.

Because economic dynamics are becoming more complex and data sources are evolving quickly, forecasting that incorporates machine learning can improve

policymakers' capacity to address economic issues, contributing to financial stability and well-informed decision-making in a world that is incredibly uncertain.

Novelties of the research

The thesis investigates in particular boosting models as relevant examples of machine learning methods. Boosting is a machine learning method that combines multiple base models to create a stronger predictive model. The basic idea behind boosting is to train a sequence of models, where each subsequent model focuses on correcting the mistakes made by the previous models. By assessing the strengths and weaknesses of boosting techniques in regression and classification scenarios, this thesis advances the subject of economic forecasting. It methodically contrasts these approaches with established benchmark models from the literature (Casabianca et al. 2022; Medeiros et al., 2021). The paper presents more intricate benchmark models, such as factor models, which offer a rigorous assessment of the predictive ability of boosting strategies in order to strengthen the robustness of this comparison.

Regression boosting methods have been compared with other machine learning techniques as well in the thesis. For example, they are compared to shrinkage models, which are praised for their ease of use and interpretability, this method provides a more detailed evaluation of boosting models (James et al., 2013).

The use of machine learning techniques on the dataset assembled by Jordà et al. (2017), which offers comprehensive temporal coverage across several nations, is another significant contribution of this study. This dataset, which covers 92 financial crises and 17 advanced economies between 1870 and 2016, provides a more comprehensive view than previous research. The number of crises per country can vary between 9, as for Italy, and 2, like Australia.

For comparison, Casabianca et al. (2022) included 100 countries but, on average, included less than two crises per country, whereas Beutel et al. (2018) looked at 15 countries but just 22 crises over 45 years. This study focuses on systemic financial crises, which enables more accurate cross-country comparisons than Barthélémy et al. (2020), who examined 276 crises across 104 countries but with different crisis characteristics. This study improves our knowledge of financial crisis prediction and the suitability of boosting strategies for economic forecasting by utilising a wealth of data and machine learning techniques.

Aim and Research Tasks

The purpose of this thesis is to assess how well machine learning approaches perform in economic forecasting, with an emphasis on boosting strategies. It looks at how well these models forecast both continuous and discrete economic variables in comparison to conventional econometric techniques. The study investigates how machine learning can improve forecasts in a variety of data contexts, including high- and low-frequency data. The main research tasks include:

1. Predicting economic variables using ML, particularly boosting models, alongside standard econometric methods.
2. Comparing ML predictions with traditional benchmarks to assess their relative effectiveness.
3. Evaluating boosting models for forecasting continuous variables, such as industrial production.
4. Comparing boosting models with other ML techniques in a continuous variable forecasting framework.
5. Assessing boosting models' effectiveness in predicting discrete variables, specifically financial crises, distinguishing between pre-crisis and crisis periods.
6. Analyzing the model performance in different data environments to determine when ML models provide superior forecasting accuracy.

Methodology

The thesis forecasts macroeconomic conditions using machine learning techniques. In addition to a comprehensive literature analysis on the use of machine learning applications on studies of business cycles and financial crises, it starts with a review of the literature on Big Data, ML applications in economics, and macroeconomic forecasting. Following an explanation of the ML models used in the methodological part, two empirical studies using regression and classification techniques are presented.

The empirical evaluation includes comparing machine learning models with traditional econometric methods, assessing regression models and measuring crisis prediction accuracy with standard metrics in the literature. Additionally, the study examines the importance of the variables extracted from the AdaBoost algorithm and its alignment with existing economic research.

More specifically, the machine learning methods used in the first empirical study, aside from boosting models, include other standard ensemble methods such as random forests and bagging, which combine simpler machine learning models to obtain better results. Additionally, shrinkage methods are used, which are similar to linear regression but incorporate a penalty to prevent overfitting. Some example of these methods are lasso and ridge. In the second empirical study, AdaBoost – a boosting algorithm specifically designed for classification tasks – is employed.

First Empirical Study: Forecasting US industrial production using a Big Fat dataset

Using the FRED-MD dataset (McCracken and Ng, 2015) to anticipate US monthly industrial production, the first study examines boosting algorithms in economic regression forecasting. The predictive accuracy of these models is assessed in relation to other machine-learning techniques and traditional econometric approaches. Standard metrics are used to quantify forecast errors and eva-

luate performance. By benchmarking ML models against well-established models, this study provides insights into the added value of boosting models in macroeconomic prediction.

Second Empirical Study: AdaBoost vs logit for identifying periods of crisis

The second study uses the ML classification technique AdaBoost to investigate financial crisis prediction (Freund and Schapire, 1997). In different data settings, AdaBoost's capacity to recognise crisis periods is contrasted with two benchmark models (logistic and stepwise logistic regression models). It also identifies the most pertinent factors for crisis detection at the country-specific level. The research makes use of the Macro-history dataset (Jordà et al., 2017), which comprises 25 annual time series from 17 industrialised economies. This study identifies the circumstances in which boosting models can improve crisis forecasting by examining various economic contexts and predictors.

Contribution

This thesis adds significantly to the literature on machine learning in macroeconomics, especially with regard to the use of boosting approaches. Initially, it broadens the application of boosting techniques in both classification and regression problems, offering a thorough examination of their potential in contrast to conventional econometric models. The research fills in the gaps left by earlier studies (e.g., Medeiros et al., 2021; Yousuf and Ng, 2021) that concentrated on a small number of boosting algorithms. The thesis examines a variety of boosting models in greater detail than earlier research (such as Goulet Coulombe et al., 2022) and emphasises the influence of particular features on enhancing forecasting accuracy such as different loss functions, presence of non linearities as well as different sampling methods.

In classification, the thesis offers a novel application of AdaBoost to forecast country-specific financial crises, in contrast to group-based crisis predictions used in previous studies (e.g., Casabianca et al., 2022; Papadopoulos et al., 2019). By using country-specific data to predict systemic financial crises, it fills a gap in the literature and addresses the challenges of predicting rare events (Beutel et al., 2019; Casabianca et al., 2022). By examining how boosting strategies function with various predictor selection approaches, the work also advances methodological approaches and deepens the body of research on crisis forecasting (Cepni et al., 2019; Jardet and Meunier, 2022).

By using machine learning to the analysis of systemic financial crises (Alessi and Detkon, 2018; Casabianca et al., 2022) and business cycles (Maehashi and Shintani, 2020; Yousuf and Ng, 2021), the thesis develops these strands of the macroeconomic literature. It illustrates how well AdaBoost predicts these occurrences in a specific data environments, especially when nonlinearities are present (Kauppi and Virtanen, 2021). The thesis demonstrates how AdaBoost can out-

perform stepwise logistic regression and conventional logit models when paired with suggested methodological innovations introduced in the thesis.

Moreover, the thesis offers a robust comparison of boosting models with traditional benchmarks, such as factor models (Soybilgen and Yazgan, 2021; Syed and Lee, 2021), and explores the impact of nonlinear predictors on forecasting accuracy, contributing to a better understanding of the effectiveness of machine learning techniques in financial forecasting. The study also looks at monthly and annual time series data from two datasets, which allows it to offer a distinctive viewpoint on how well boosting strategies work in various data frequency and environments.

Limitations of the studies

This thesis has specific boundaries and limitations. It does not focus on more complex ML methods, such as deep learning methods, which are computationally expensive and face challenges with interpretability and the balance between accuracy and efficiency (Zheng et al., 2023). Additionally, the study does not look at unstructured Big Data, like text analysis, which has grown in importance in macroeconomics (Alfieri and Gabrielyan, 2024; Hansen et al., 2019) for gleaning new information from publicly accessible texts. Rather, it emphasizes how to exploit structured data. The thesis does not aim to determine "the best" machine learning methods for all economic applications, given the vast range of methods available (Bluwstein et al., 2020; Chernozhukov et al., 2018) and the multiplicity of possible exercises that can be envisaged. It does not seek to test causal relationships or identify the most effective forecasting model across all contexts. Finally, while the second empirical study includes variables related to systemic financial crises, it does not test any specific macroeconomic theories, focusing instead on evaluating machine learning techniques for forecasting within the context of financial crises. This distinction shapes the study's goals and methodology.

Structure of the study

The thesis is structured into four chapters, following the introduction.

Chapter 1 discusses the challenges of forecasting in macroeconomics, particularly in a nonlinear world. It explores the concept of Big Data and its varied interpretations across different research fields. The chapter also examines the research on machine learning techniques in economics, with a particular emphasis on how they are used to predict financial crises and business cycles.

Chapter 2 introduces the ML models used in the thesis, covering both regression and classification methods. The chapter also explains why certain other ML methods are not utilized. For classification, it focuses on classification trees and AdaBoost.

Chapter 3 presents two empirical macroeconomic exercises. Using the FRED-MD dataset and machine learning regression models, the first study fore-

casts US industrial production in comparison to benchmark models. The second task uses the Macro-history dataset from 1870 to 2016 to detect systemic financial crises as well as pre-crisis and crisis periods. To find important crisis predictors, AdaBoost is used and contrasted with logistic regression models.

Chapter 4 discusses the results of the empirical exercises and the utility of ML methods in various data environments. It concludes with a summary of findings and suggestions for future research.

Summary of the literature

The first subsection of Chapter 1 traces the history of economic forecasting. Early economic forecasting emerged between 1910–1930 but faltered due to the failure to predict the 1929 crisis. Post-WWII, forecasting grew, with Burns and Mitchell (1946) providing a modern definition of the business cycle. The chapter discusses the evolution of economic models, including the Real Business Cycle (RBC) models and the DSGE models, which were refined after the Great Recession.

The subsection highlights the importance of data-driven models, citing Stock and Watson (1998, 2006), who have demonstrated the value of multivariate models in time series forecasting, especially when many predictors are involved. Despite these developments, financial crisis forecasting is still quite difficult because these occurrences frequently involve indicators like household debt and credit cycles.

The subsection also highlights the difference between predicting binary events, such as systemic financial crises, and continuous variables, like industrial production. Key studies that highlight the importance of financial crisis predictors in predicting systemic financial crises are cited in the discussion of these factors, which include public debt and credit growth.

The second subsection delves into the relationship between Big Data and machine learning, with an emphasis on their growing importance in economic research and forecasting. The definition and implications of Big Data are explored in detail, highlighting the exponential increase in data volumes, especially in the post-Google era.

The discussion moves to the structured and unstructured data component of Big Data, where the volume of numerical and non-numerical data like text, audio, and video has grown significantly. Due to its complexity and volume, this data requires specialized ML algorithms for analysis. This leads to an exploration of supervised and unsupervised ML methods, with a focus on their application in a context of a large number of economic predictors. Supervised learning techniques, such as regression and classification, are used to model relationships between predictors and outcomes, while unsupervised learning techniques uncover hidden patterns without predefined labels.

The subsection also covers important ML methods like lasso and ridge regression – both key tools for addressing overfitting and feature selection in complex datasets. It further discusses tree-based methods. The concept of the "curse of

dimensionality," which states that as the number of predictors in a model increases, its complexity raises the chance of overfitting, is discussed.

Traditional econometric models struggle to find causal relationships in Big Data environments, making the search for causal relationships one of the most important potential uses of machine learning in economics. Although the literature on causality matters is currently small in comparison to ML's application in prediction models, the expanding corpus of work in ML-based on causal relationships is recognised.

The subsection concludes by examining the growing interest in ML for forecasting and nowcasting, particularly in the context of economic data. While techniques like AdaBoost and XGBoost have shown promise in predicting economic events (e.g., recessions), challenges remain, especially with the black-box nature of of more complex ML models. It underscores that while these models perform well in non-standard data environments, they often lack interpretability, which limits their broader application in economic theory development.

Given that the discipline of macroeconomics has just lately started to use machine learning techniques, the next part elaborates on the growing interest in ML in macroeconomics as opposed to microeconomics. This delay can be attributed in part to the fact that existing models, such as factor models, have proven so effective at handling large datasets, while macroeconomists often have access to much fewer data points than microeconomists. However, these models frequently fail to recognise the potential of machine learning techniques, particularly their ability to handle complex nonlinearities in data and retain more comprehensive information.

Nonlinearity is an essential aspect of modern macroeconomics, especially in forecasting and analyzing economic data. Scholars like Franses and Van Dijk (2005) emphasize the importance of nonlinear models, especially for long-term forecasting of industrial production. Despite their relevance, these nonlinear models do not always outperform simpler linear models for short-term forecasting, which is why ML methods are increasingly preferred for handling long-term predictions. ML methods, such as boosting methods, provide advantages by accommodating nonlinearities and transformations of variables, as indicated by works like Döpke et al. (2017), who used AdaBoost to predict recessions in Germany.

Traditional econometric methods still hold value in forecasting major economic indicators, such as industrial production, even when many variables are considered. However, the interpretability of models like factor models can be challenging. Scholars are exploring ways to combine ML with econometric models to enhance forecasting accuracy. For example, Li and Chen (2014) combined dynamic factor models with ML-based models to forecast key U.S. macroeconomic indicators, finding that the latter outperformed the former in predicting industrial production.

Other studies reinforce this trend, such as Heij et al. (2011), who found that principal components analysis provided more accurate forecasts than the Conference Board's Composite Leading Index (CLI) for U.S. industrial production.

Similarly, Kock and Teräsvirta (2014) demonstrated that neural networks outperformed traditional linear models (like AR models) in forecasting industrial production, particularly during the Great Recession.

Regression tree methods, such as boosting, have proven highly effective in dealing with nonlinear relationships and outliers in the data. Döpke et al. (2017) highlight how boosting helps mitigate issues that often arise from simple regression trees. Ng (2014) used boosting with over 1,500 predictors to predict U.S. recessions, illustrating how ML methods can manage vast amounts of data effectively. Additional studies, such as Buchen and Wohlrabe (2011), show that component-wise boosting methods also outperform factor models in the long run when forecasting U.S. industrial production.

Boosting methods continue to show utility across various forecasting tasks. For example, Lehmann and Wohlrabe (2015, 2016) used boosting to forecast German regional GDP and German industrial production, respectively, illustrating its versatility for different economic indicators and forecasting horizons. Bai and Ng (2009) demonstrated the power of boosting for predictor selection in factor-augmented regressions, while Carriero et al. (2011) compared more classic macroeconometric methods and multivariate boosting for forecasting economic variables like industrial production growth, inflation, and the federal funds rate. Their results showed that combining simpler ML methods with boosting led to significant improvements in forecasting accuracy.

A thorough assessment of the literature on machine learning techniques and financial crises, as well as business cycles, is presented in the last subsection. Because machine learning approaches can handle data environments with a large amount of variables and identify complicated nonlinear relationships, they have garnered a lot of attention recently for the financial crisis and business cycle predictions. Many researchers have examined the efficacy of machine learning techniques in comparison to conventional methods in the context of the business cycle's predictions. The list includespapers like Exterkate et al. (2016), Maehashi and Shintani (2020), Syed and Lee (2021), Yousuf and Ng (2021) and Zhang et al. (2023).

While machine learning methods have been shown to perform well in capturing nonlinearities and improving forecasting accuracy, there are instances where traditional methods still hold an advantage, particularly when statistical significance is concerned. Scholars contrast ML techniques with factor models, which combine multiple economic variables into a more manageable collection of elements (Cepni et al., 2019; Soybilgen and Yazgan, 2021; Syed and Lee, 2021). Although ML techniques generally perform better than factor models, their effectiveness can vary depending on the time period and data environment.

A recurring theme in the literature is the integration of factor models with machine learning techniques. This hybrid approach helps combine the strengths of both methodologies and has been shown to improve forecasting accuracy. Studies highlight the importance of using a wide range of predictors (Panagiotelis et. al., 2019; Mogliani and Simoni, 2021; Babii et al., 2022), especially when dealing with datasets with a plethora of variables or data with different time

frequencies, and how machine learning can manage such complexity better than traditional models. However, the selection of predictors – whether through pre-selection or automated methods – is another critical consideration (Cepni et al., 2019; Jardet and Meunier, 2022; Borup et al., 2023).

Machine learning methods also excel at detecting and managing nonlinear relationships within economic data. ML based on regression trees have proven effective in capturing these nonlinearities, which are often difficult for traditional models to handle. Another important consideration is how machine learning models perform across different sub-samples or time periods. Studies show that the results can vary significantly based on the chosen time window or country-specific conditions, underlining the importance of tailoring models to specific data environments (Panagiotelis et al., 2019; Soybilgen and Yazgan, 2021; Zhang et al., 2023; Guerrón-Quintana and Zhong, 2023).

In terms of datasets, the FRED-MD dataset – along with its quarterly version, FRED-QD – has been the most frequently used, particularly for U.S. business cycle forecasting. This dataset, which provides a rich collection of macroeconomic variables, has become a standard for comparing ML models. Notable studies, such as those by Exterkate et al. (2016), Guerrón-Quintana and Zhong (2023), and Kauppi and Virtanen (2021), see Table 2 in the thesis, have shown that ML methods can outperform traditional approaches in predicting U.S. economic indicators, though performance can vary depending on the economic conditions during the time period under study.

While the majority of research on machine learning for macroeconomic forecasting focuses on business cycles, there is a growing interest in applying these techniques to predict financial crises (see Table 3 in the thesis). Financial crises are inherently more difficult to predict due to their rarity, unpredictability, and complex nature. However, several studies have found that ML methods can sometimes outperform traditional models in predicting financial crises. For example, Alessi and Detken (2018) showed that random forests slightly outperformed logit models in predicting systemic financial crises in EU countries. Additionally, Liu et al. (2022) demonstrated that ML methods, particularly those using Shapley values, a system of weights derived from game theory, for indicator selection, could improve early crisis prediction. However, the success of these methods in crisis prediction depends heavily on factors such as data characteristics (possible presence of imbalanced data, use of time series instead of panel data), model selection, and country-specific conditions.

In conclusion, machine learning has a lot of promise for financial crisis and business cycle forecasting, but there are drawbacks to its use. The creation of reliable models is hampered by the complexity of the data, the sensitivity to various predictor sets, and the rarity of financial crises. The increasing amount of research and the ongoing development of machine learning techniques, particularly ensemble approaches like random forests and gradient boosting, however, hold promise for improving the precision and reliability of macroeconomic forecasting in the future.

Results and Discussion on the first empirical study

This study investigates the forecasting of the average growth rate of monthly US industrial production from 1963 to 2019 using various boosting models, comparing their performance against other machine learning methods and traditional econometric benchmarks like autoregressive models, random walk, and factor models. The study emphasizes the need for a more comprehensive understanding of how boosting models impact forecasting accuracy and whether they outperform standard econometric models in predicting industrial production.

Boosting models consistently outperform the random walk approach in all estimations, proving to be strong competitors, especially against simpler ML models. When compared to the factor model, AdaBoostRT, component-wise boosting, and XGBoost show superior results. AdaBoostRT performs best in short-term forecasts, while XGBoost is more effective for medium-term predictions. Component-wise boosting delivers the most consistent performance across all forecast horizons. These findings support existing research highlighting the effectiveness of boosting models, particularly in capturing nonlinear patterns in business cycles data.

When comparing boosting models with autoregressive models, the study finds that only component-wise boosting and AdaBoostRT perform better than AR in short-term forecasts, while XGBoost performs better in medium-term forecasts. This is consistent with earlier studies that indicate machine learning models have a hard time beating AR models, even simple ones (Maehashi and Shintani, 2020; Yousuf and Ng, 2021). The examination of boosting techniques, in particular component-wise boosting, is especially relevance since it enhances forecast accuracy by combining factor models and machine learning.

The results also highlight how the specific characteristics of boosting models affect their forecasting power. The study is the first to compare more than two boosting models in the context of industrial production forecasting, providing valuable insights into how boosting algorithms differ. For instance, AdaBoostRT's performance differs significantly from AdaBoostR2. XGBoost, a more complex boosting model, performs better but faces issues that have to be handled.

One important conclusion is that when the period of Great Recession (2007–2009) is included in the data sample, boosting models perform better. Machine learning techniques, such as boosting, appear to manage the data's nonlinearity better in this time of increased economic volatility than in more stable periods. This is consistent with research by Goulet Coulombe et al. (2022), which indicates that machine learning techniques typically outperform other approaches, especially in times of economic crisis.

The study also explores the impact of different testing methods. The expanding window method, which uses more recent data for out-of-sample testing, generally leads to worse performance from XGBoost and AdaBoostRT, suggesting that these models require adjustments for changes in the economic structure over time.

The study concludes by showing that boosting models can perform better than conventional econometric models, particularly for short- and medium-term pro-

jections. The significance of choosing suitable forecasting models based on the data and economic context is shown by the comparison with factor models and the effects of recessionary times. The results highlight the need for additional study on boosting models, especially with regard to their use in predicting macroeconomic variables like industrial production.

Results and Discussion on the second empirical study

The second empirical study focuses on the use of AdaBoost to predict financial crises, particularly systemic ones, while addressing the nonlinearities in financial data. The study uses the dataset from Jordà et al. (2017) and covers sixteen countries from 1870 to 2016, with variables relevant to financial crises. Financial crises differ from regular business cycles and exhibit highly nonlinear relationships, which makes AdaBoost a promising method due to its ability to handle such complexities. AdaBoost's performance is compared to logistic and stepwise logistic regressions, which are common benchmarks in crisis prediction.

AdaBoost shows potential in detecting financial crises, especially when the model is based on a rich set of macroeconomic predictors. However, its effectiveness depends on factors like the estimation method, data environment, and how the dependent variable is sampled. This highlights the importance of appropriate sampling techniques. In the study, the use of country-specific data, rather than pooling countries or using panel data, yields mixed results, though it still shows promise. The results align with previous literature suggesting that more predictors improve machine learning models' forecasting capabilities for financial crises.

The study notes the challenge posed by imbalanced data, which is prevalent in financial crises datasets, where crises are rare events compared to tranquil periods. Common strategies, such as grouping countries or using panel data, aim to address this issue but have shown inconsistent success. A key takeaway is that AdaBoost's simple implementation may not fully capture the potential of machine learning for this task, and more advanced AdaBoost variations could provide better results, as demonstrated by Liu et al. (2022).

The study also explores a different method for crisis prediction, which categorizes different periods of a financial crisis, such as pre-crisis, crisis, and post-crisis phases. This method necessitates modifications to AdaBoost because the model is often intended for binary classification, which may limit its capacity to manage categorical results.

Regarding variable importance, AdaBoost's results suggests that certain macroeconomic indicators, that align with previous research on systemic financial crises, are relevant. For instance, credit variables and asset price fluctuations play a crucial role in forecasting financial crises, especially in advanced economies with well-developed financial systems. For instance, the study confirms findings that excessive financial liberalization and housing price booms can trigger financial crises. Similarly, the debt-to-GDP ratio in Italy, especially in the post-crisis context, indicates the relevance of fiscal space in managing crises.

The study also discusses the significance of international economic variables, such as the current account, which the literature is not unanimous if they are still playing a major role as indicators of financial stability (Davis et al., 2016; Obstfeld, 2012; Barrell et al., 2010) but are found to be relevant in this context. Italy's findings, for instance, suggest a potential connection between trade and systemic financial crises, bolstering the notion that financial ties and trade might influence a nation's susceptibility to crises.

The study reveals considerable heterogeneity across countries in terms of which variables are most predictive of crises. Larger economies with advanced financial structures, like Japan and the US, exhibit unique crisis patterns compared to smaller, more open economies. Market size, international trade openness, and domestic economic productivity also influence a country's susceptibility to both domestic and external financial crises.

In conclusion, while AdaBoost offers promising results in detecting financial crises, its effectiveness is shaped by the chosen data environment and the sampling techniques used. The study suggests that more sophisticated versions of AdaBoost, combined with an appropriate selection of macroeconomic predictors, could improve predictions, particularly in the context of nonlinear relationships inherent in financial crisis data.

General discussion on the thesis

The thesis's highlights how predicting techniques are becoming more sophisticated, especially in the wake of the Great Recession. In an effort to forecast economic cycles and systemic financial crises, researchers are increasingly using novel models and factors. The thesis examines two forecasting schools of thought: one supports a data-driven approach that enables a larger set of variables to capture nonlinearities within both business cycles and financial crises, while the other supports pre-selection of variables based on theoretical macroeconomic foundations.

This research integrates both perspectives, selecting variables informed by macroeconomic theory while incorporating a large number of predictors. Additionally, the study introduces simple nonlinearities in the models, with the second empirical study relying more on variable pre-selection. Notably, the dataset used for the second study, FRED-MD, is grounded in prior studies, such as Stock and Watson's work. This thesis contributes to macroeconomic forecasting by combining traditional econometric methods with modern machine learning, offering a comprehensive framework that improves both forecasting accuracy and augments the economic understanding on financial crises.

The research identifies several implications for future investigations. First, machine learning methods, particularly boosting models, show promise when dealing with Big Data in forecasting business cycles, suggesting that incorporating additional factors can enhance model performance. However, more complex models, such as XGBoost, may not always be effective when training

data is limited or a mixture between strong and weak predictors, emphasizing the importance of careful considerations when it is applied.

Feature selection is also critical in ML, as overly complex models that include irrelevant predictors can degrade performance. For this reason, researchers should prioritize relevant predictors that add value and avoid introducing noise. The study's classification exercise with imbalanced data highlights the importance of properly handling sampling techniques when identifying systemic financial crises.

According to the thesis, more research is required to compare boosting models with more complex factor models that pre-select variables and to examine how nonlinearities are introduced into time series data, as demonstrated in the second study. The AdaBoostRT model's prediction ability during tranquil times might be investigated in more detail. Furthermore, financial data – particularly high-frequency data, such as intraday stock market data – may offer useful insights for business cycle forecasting.

The thesis notes that systemic financial crises in the 19th century are not the same as contemporary crises like the one that occurred in 2007/2008, and it proposes that this difference should be further examined. Reducing the length of the period analyzed, utilising higher-frequency data, or utilising panel data, as observed in previous studies, are all potential remedies. All of these strategies, though, have drawbacks.

The thesis comes to the conclusion that, despite their great potential, machine learning approaches shouldn't take the place of conventional econometric methods. Rather, machine learning techniques can enhance existing procedures, especially when paired with traditional methods that use theoretical knowledge and expert judgment. This well-rounded strategy improves forecasting ability, particularly in complex economic environments. In the end, the study emphasises how critical it is to evaluate the advantages and disadvantages of machine learning techniques, acknowledging that their effectiveness is contingent upon the particular data properties and study goals.

Short summary of the thesis

This thesis examines applications of machine learning techniques to macroeconomic forecasting, with a particular emphasis on business cycles and systemic financial crisis prediction. It looks into two main strategies: using a data-driven approach, the presence of a large number of variables to capture nonlinearities in economic events and pre-selecting variables based on macroeconomic theory. The study applies a combination of theory-based variable selection and a wide range of predictors, integrating both viewpoints.

The findings highlight the importance of handling imbalanced data, feature selection, and nonlinearities in ML models, suggesting that boosting methods can be particularly effective with a large number of variables in the dataset. However, complex models, such as XGBoost, require careful preparations, especially with limited data. The thesis also discusses the challenges of predicting crises, parti-

cularly with historical data, and the need for future research on model optimization.

Overall, the thesis shows how machine learning techniques can improve macroeconomic forecasting by providing a thorough framework that blends contemporary machine learning approaches with conventional econometrics. Additionally, it highlights the necessity of carefully assessing data and methodologies to make sure that machine learning models enhance rather than replace traditional methods.

Indeed, while Big Data has increased the demand for machine learning methods, relying on them uncritically or abandoning traditional approaches is unwise. Machine learning methods can perform better when combined with econometric techniques, as both share statistical foundations. Therefore, many new methods are more closely related to traditional ones than they might seem at first.

SUMMARY IN ESTONIAN

Majandustsüklite ja finantskriiside prognoosimine masinõppe abil

Uurimistöö probleem

Nii poliitikakujundajad, rahvusvahelised organisatsioonid, ettevõtted kui ka finantsasutused vajavad täpseid majandusprognoose (Wieland ja Wolters, 2013; Fouliard jt, 2021). Näiteks on keskpankadel nagu Euroopa Keskpank (EKP) vaja, et prognoose uuendataks sagedamini kui kord kvartalis, et teha teadlikumaid rahapoliitilisi otsuseid, mis avaldatakse iga kuue nädala järel. Finantskriisid ja majanduslangused nõuavad poliitiliste otsuste õigeaegset tegemist (Sutherland, 2023; Woodford, 2005). Seetõttu toetuvad need institutsioonid üha rohkem lähiprognoosimudelitele, mis ennustavad lähiminevikku või -olevikku, mille kohta pole veel andmeid, või tuleviku majandusnäitajaid. Just seepärast kasutatakse ka varajase hoiatamise süsteeme, mis aitavad tuvastada võimalikke finantskriise enne nende eskaleerumist ja annavad poliitikakujundajatele ja turuosalistele võimaluse aegsasti sekkuda.

Finantskriisidel nagu 2007.–2008. aasta kriis on sageli tõsised ja pikaajalised tagajärjed (Shiller ja Shiller, 2011). Teinekord võivad nad muuta lausa institutsionaalset süsteemi, nagu näiteks tõi finantskriis kaasa Finantsstabiilsuse Nõukogu (FSB) ja Euroopa Süsteemsete Riskide Nõukogu (ESRB) loomise. Finantsturgude heitlikkuse ohjeldamisel toetuvad need kriisidele keskendunud organisatsioonid varajastele hoiatustele. Majanduslangusi on raske ennustada, eriti neid, mis on seotud finantskriisidega (Bordo ja Haubrich, 2017), mis teeb usaldusväärsed prognoosimistehnikad makroökonoomilises analüüsis hädavajalikuks. Prognoosid tuvastavad potentsiaalsed finantskriisid enne nende eskaleerumist, nii et poliitikakujundajad ja turuosalised saavad aegsasti reageerida. Tavaliste ökonomeetriliste meetoditega ei saa aga sageli hästi prognoosida majandustsükleid ja finantskriise tulenevalt nende mittelineaarsetest esinemise mustritest (Hamilton, 2005; Akinci ja Queralto, 2022).

Edusammud masinõppe meetodites loovad tänu suurandmetele uusi prognoosivõimalusi. Suurandmete hulka kuuluvad definitsiooni kohaseltnii struktureerimata andmed tekstidest, satelliidipiltidest ja telekommunikatsioonist kui struktureeritud majandusnäitajad (Hansen jt, 2019; Cerdeiro jt, 2020). See väitekiri keskendub ainult struktureeritud andmetele, sest struktureerimata andmed vajavad teistsuguseid uurimismeetodeid (vt näiteks Alfieri ja Gabrielyan, 2024), ja need jäävad selle monograafia haardest välja.

Kui kasutame masinõppemudeleid makromajanduslike muutujate mittelineaarsete seoste väljaselgitamiseks, saame teha täpsemaid majandusennustusi (Mosavi jt, 2020; Athey, 2018). Mõned uuringud, nagu Casabianca jt (2022) ja Lehmann ja Wohlrabe (2016), väidavad, et masinõppemeetodid võivad rahandus-

kriiside ja makromajanduslike indeksite prognoosimisel olla täpsemad kui traditsioonilised ökonomeetrilised mudelid.

Kuna majanduse dünaamika muutub üha keerulisemaks ja andmeallikad arenevad kiiresti, võivad masinõppepõhised prognoosid aidata poliitikakujundajatel majanduslike probleemidega paremini toime tulla. Ühtlasi aitavad need erakordselt ebastabiilses maailmas saavutada finantsstabiilsuse ja tagavad hästi informeeritud otsustusprotsessid.

Uurimistöö uudsus

Siinne doktoritöö uurib eelkõige võimendamisalgoritme (i.k *boosting*) kui asjakohaseid näiteid masinõppemeetoditest. Võimendamine on masinõppemeetod, mis ühendab mitu baasmudelit tõhusamaks prognoosimudeliks. Meetodi põhimõte on õpetada välja mitu mudelit järjest, kusjuures iga järgmine mudel keskendub eelnevate mudelite tehtud vigade parandamisele. Doktoritöö hindab võimendamismeetodi tugevusi ja nõrkusi regressiooni ja klassifitseerimise stsenaariumides, et majandusprognooside valdkonda edendada. Kõrvutame neid lähenemisviise olemasolevate võrdlusmudelitega (Casabianca jt, 2022; Medeiros jt, 2021). Uurimistöö esitleb keerukamaid võrdlusmudeleid nagu faktormudelid, mis pakuvad täpset hinnangut võimendamisstrateegiate ennustamisvõime kohta, et tugevdada võrdluse usaldusväärsust.

Regressiooni võimendamismeetodeid on töös kõrvutatud ka teiste masinõppemeetoditega. Näiteks võrdleme neid kadumismudelitega (i.k *shrinking models*), mida on peetud mugavaks ja lihtsasti tõlgendatavaks, see meetod hindab ka võimendamismudeleid üksikasjalikumalt (James jt, 2013).

Uurimistöö teine oluline panus on masinõppetehnikate kasutamine Jordà jt (2017) andmestikuga, mis pakub andmeid pika ajaperioodi kohta mitmes riigis. Andmestik hõlmab 92 finantskriisi 17 arenenud riigis aastatest 1870–2016 ja pakub põhjalikumat ülevaadet kui varasemad uuringud. Kriiside arv riigi kohta jääb üheksa (Itaalia) ja kahe (Austraalia) vahele.

Võrdluseks, Casabianca jt (2022) uurimus hõlmas 100 riiki, kuid keskmiselt vähem kui kaht kriisi riigi kohta, samas kui Beutel jt (2018) käsitles 15 riiki, kuid vaid 22 kriisi 45 aasta jooksul. Siinne uurimistöö keskendub süsteemsetele finantskriisidele, mis võimaldab teha täpsemaid riikidevahelisi võrdlusi kui Barthélémy jt (2020) töös, kus uuriti 276 kriisi 104 riigis, kuid vaadeldi kriiside erinevaid omadusi. Uurimistöö parandab meie teadmisi finantskriiside ennustamise ja võimendamisstrateegiate sobivuse kohta majandusprognoosides, kasutades rikkalikku andmestikku ja masinõppetehnikaid.

Uurimistöö eesmärk ja ülesanded

Doktoritöö eesmärk on hinnata, kui hästi aitavad majandusprognoose teha masinõppemeetodid, täpsemalt võimendamisstrateegiad. Töös uuritakse, kui hästi prognoosivad need mudelid nii pidevaid kui ka diskreetseid majanduslikke muutujaid, võrreldes traditsiooniliste ökonomeetriliste tehnikatega. Vaatleme masinõppe

võimalusi prognooside tõhustamiseks erinevates andmekontekstides, sealhulgas kõrge ja madala sagedusega andmete puhul.

Peamised uurimishuvid on järgmised:

1. Majanduslike muutujate ennustamine masinõppe abil, eriti võimendamismudelitega, lisaks tavalistele ökonomeetrilistele meetoditele.
2. Masinõppe ennustuste traditsiooniliste mõõdikutega võrdlemine, et hinnata nende suhtelist tõhusust.
3. Võimendamismudelite hindamine pideva muutuja, st tööstustoodangu, prognoosimisel.
4. Võimendamismudelite võrdlemine teiste masinõppetehnikatega pideva muutuja prognoosimise kontekstis.
5. Võimendamismudelite tõhususe hindamine diskreetsee muutuja, st finantskriiside prognoosimisel, eristades kriisieelseid ja -aegseid perioode.
6. Mudeli soorituse analüüs erinevates andmekeskkondades, et selgitada välja, millal pakuvad masinõppemudelid täpsemaid prognoose.

Metoodika

Doktoritöö prognoosib makromajanduslikke tingimusi masinõppetehnikate abil. Lisaks allikatele, mis käsitlevad masinõppe rakendusi majandustsüklite ja finantskriiside uuringutes, tutvustame allikaid, mis puudutavad suurandmeid, masinõppe rakendusi majandusteaduses ja makromajanduslike prognooside tegemisel. Pärast metodoloogia osas kasutatavate masinõppemudelite tutvustamist kirjeldame kaht empiirilist uuringut, milles on kasutatud regressiooni ja klassifitseerimise tehnikaid.

Empiiriline hindamine hõlmab masinõppemudelite võrdlemist traditsiooniliste ökonomeetriliste meetoditega, regressioonimudelite hindamist ja kriisi ennustamise täpsuse mõõtmist kirjandusallikates kasutatud standardsete mõõdikute abil. Lisaks uuritakse AdaBoosti algoritmist saadud muutujate olulisust ja nende suhet senise majandusteadusliku uurimistööga.

Täpsemalt hõlmavad esimeses empiirilises uuringus kasutatud masinõppemeetodid, välja arvatud võimendamismudelid, ka teisi standardseid ansamblimeetodeid nagu juhumetsad (i.k *random forest*) ja *bagging*-algoritmid, mis kombineerivad paremate tulemuste saavutamiseks lihtsamaid masinõppemudeleid. Lisaks kasutatakse kadumismeetodeid, mis sarnanevad lineaarregressioonile, kuid sisaldavad piirangut ülesobituse vältimiseks. Sellised meetodid on näiteks lasso ja *ridge*. Teises empiirilises uuringus kasutatakse AdaBoosti, spetsiaalselt klassifitseerimiseks loodud võimendamisalgoritmi.

Esimene empiiriline uuring: Ameerika Ühendriikide tööstustoodangu prognoosimine suurandmestiku abil

Esimene uuring vaatleb võimendamisalgoritme majandusaktiivsuse prognoosimisel, kasutades FRED-MD andmestikku (McCracken ja Ng, 2015) USA kuise tööstustoodangu ennustamiseks. Nende mudelite ennustustäpsust hinnatakse võrdluses teiste masinõppetehnikate ja traditsiooniliste ökonomeetriliste lähenemistega. Prognoosivigade kvantifitseerimiseks ja soorituse hindamiseks kasutatakse standardseid näitajaid. Uuring annab ülevaate võimendamismudelite lisandväärtusest makroökonoomilistes prognoosides, võrreldes masinõppemudeleid laialdaselt kasutatavate traditsiooniliste ökonomeetriliste mudelitega.

Teine empiiriline uuring: AdaBoost ja logit kriisiperioodide tuvastamisel

Teine uuring analüüsib masinõppe klassifitseerimistehnika AdaBoosti abil finantskriiside ennustamist (Freund ja Schapire, 1997). AdaBoosti võimet kriisiperioode tuvastada võrreldakse erinevates andmekeskkondades kahe võrdlusmudeliga: logistiliste regressioonimudelite ja sammregressioonimudelitega. Uuring tuvastab ka kõige olulisemad tegurid kriisi tuvastamiseks riigi tasandil. Uuring kasutab makroajaloo andmestikku (Jordà jt, 2017), mis sisaldab 25 aasta pikkust aegrida 17 tööstusriikidest. See toob välja olukorrad, kus võimendamismudelid saavad kriisi ennustamist parandada, uurides erinevaid majanduslikke kontekste ja kriisieelseid märke.

Doktoritöö panus

Siinne doktoritöö on oluline lisandus allikatele, mis käsitlevad masinõpet makroökonoomikas, eriti seoses võimendamismeetodite kasutamisega. Esmalt laiendab see võimendamistehnikate rakendusi nii klassifikatsiooni- kui regressiooniprobleemides ja pakub põhjalikku ülevaadet võimendamise potentsiaalist võrreldes traditsiooniliste ökonomeetriliste mudelitega. Uurimistöö täidab varasematest väikesele hulgale võimendamisalgoritmidele keskendunud uuringutest (nt Medeiros jt, 2021; Yousuf ja Ng, 2021) jäänud lünki vaadeldes põhjalikult varasemate uuringute erinevaid võimendamismudeleid (nt Goulet Coulombe jt, 2022) ning tuues välja teatud ennustustäpsust parandavaid mudelite omadusi nagu erinevad kaofunktsioonid, mittelineaarsete elementide olemasolu ja valimimääramise meetodid.

Doktoritöös pakutakse välja, et AdaBoosti võiks rakendada riigispetsiifiliste finantskriiside prognoosimisel, vastandudes varasemates uuringutes (nt Casabianca jt, 2022; Papadopoulos jt, 2019) kasutatud grupipõhistele kriisiprognoosidele. Süsteemsete finantskriiside prognoosimine riigipõhiste andmete põhjal korvab seniste allikate puudujäägi ja käsitleb harvade sündmuste ennustamise väljakutseid (Beutel et al. 2019; Casabianca et al. 2022). Teadustöö uurib, kuidas toimivad võimendamisstrateegiad erinevate kriisi ennustavate muutujate

valimise lähenemiste korral ja arendab seeläbi metodoloogilisi lähenemisviise ning viib edasi kriisiennustamisealast teaduskirjandust (Cepni jt, 2019; Jardet ja Meunier, 2022).

Kasutades masinõpet süsteemsete finantskriiside (Alessi ja Detkon, 2018; Casabianca jt, 2022) ja majandustsüklite (Maehashi ja Shintani, 2020; Yousuf ja Ng,2021) analüüsimiseks, arendab doktoritöö edasi neid makroökonoomika-alase kirjanduse suundi. Töö näitab, kui hästi prognoosib AdaBoost neid juhtumeid spetsiifilistes andmekeskkondades, eriti kui esinevad mittelineaarsused (Kauppi ja Virtanen, 2021). Samuti näitab töö, et AdaBoost võib olla tõhusam kui sammregerssioonilised mudelid ja traditsioonilised logit-mudelid, kui rakendada töös väljapakutud metodoloogilisi uuendusi.

Lisaks võrdleb doktoritöö võimendamismudeleid näiteks traditsioonilise faktoranalüüsga (Soybilgen ja Yazgan, 2021; Syed ja Lee, 2021) ning uurib mittelineaarsete kriisieelsete märkide mõju ennustustäpsusele, aidates paremini mõista masinõppetehnikate tõhusust finantsprognoosides. Kasutades aga kahest andmestikust saadud kuiseid ja aastaseid aegridu, annab töö teistest töödest eristuva vaatepunkti võimendamisstrateegiate tõhususele erineva andmesageduse ja keskkondade korral.

Uuringute piirangud

Nagu igal uurimusel, on ka sel doktoritööl konkreetsed raamid ja piirangud. Töö jätab kõrvale keerukamad masinõppemeetodid nagu süvaõppemeetodid, mis on arvutuslikult kulukad ja milles võib esineda probleeme tõlgendatavuse ning täpsuse ja efektiivsuse tasakaaluga (Zheng jt, 2023). Lisaks ei käsitle uuring struktureerimata suurandmeid nagu tekstianalüüs, mis on makroökonoomikas hiljuti tähtsamaks muutunud (Alfieri ja Gabrielyan, 2024; Hansen jt, 2019),. pigem rõhutades seda, kuidas struktureeritud andmeid masinõppemeetoditega kasutada.

Töö ei püüa kindlaks teha „parimaid“ masinõppemeetodeid kõigi majanduslike rakenduste jaoks, kuna valikus on palju erinevaid meetodeid (Bluwstein jt, 2020; Chernozhukov jt, 2018) ja nende kasutuseks võib luua arvukalt erinevaid ülesandeid. Töö ei sea ka eesmärgiks põhjuslike seoste proovile panemist Viimaks, kuigi teine empiiriline uuring sisaldab süsteemsete finantskriisidega seotud muutujaid, ei kontrolli töö konkreetseid makroökonoomilisi teooriaid, vaid keskendub masinõppe prognoositehnikate hindamisele finantskriiside kontekstis. See eristus kujundab uuringu eesmärke ja metoodikat.

Uuringu ülesehitus

Doktoritöö koosneb neljast sisupeatükist.

Esimene peatükk käsitleb makroökonoomika prognoosimise probleemkohti, eriti mittelineaarses maailmas. Selles selgitatakse suurandmete mõistet ja selle erinevaid tõlgendusi mitmes teadusvaldkonnas. Samuti analüüsitakse uuringuid,

mis puudutavad masinõppetehnikate kasutamist majanduses, pöörates erilist tähelepanu finantskriiside ja majandustsüklite ennustamisele.

Teine peatükk tutvustab väitekirjas kasutatud masinõppemudeleid, nii regressiooni- kui ka klassifitseerimismeetodeid. Põhjendatakse tehtud valikuid ning selgitatakse, miks teatud masinõppemeetodeid töös ei kasutata. Klassifitseerimise osas keskendub töö otsustuspuudele ja AdaBoostile.

Kolmas peatükk tutvustab kaht empiirilist makroökonoomika uuringut. Esimeses uuringus kasutatakse FRED-MD andmestikku ja masinõppe regressioonimudeleid, et prognoosida USA tööstustoodangu mahtu ja hinnata prognoositäpsust suhtes võrdlusmudelitega. Teises uuringus kasutatakse makromajanduse vanema ajaloo andmestikku aastatest 1870–2016, et tuvastada süsteemseid finantskriise ning kriisieelseid ja -aegseid perioode. Oluliste kriisieelsete märkide ehk kriisi ennustavate muutujate leidmiseks kasutatakse AdaBoosti ja saadud tulemusi võrreldakse logistilise regressiooni mudelitega.

Neljas peatükk käsitleb empiiriliste uuringute tulemusi ja masinõppemeetodite kasulikkust erinevates andmekeskkondades. Peatüki lõpus tuuakse välja järeldused ja soovitused tulevasteks uuringuteks.

Doktoritöö kirjanduse kokkuvõte

Esimese peatüki esimene alapeatükk annab ülevaate majandusprognooside ajaloost. Esimesed majandusprognoosid ilmusid aastatel 1910–1930, kuid valdkond takerdus, kuna ei suutdetud ennustada 1929. aasta kriisi. Teise maailmasõja järel arenes prognoosimine valdkond taaskord kiiremini. Selgemaks sai majandustsükli mõiste kui Burns ja Mitchell (1946) pakkusid välja oma määratluse mis on saanud uueks standardiks. Peatükk käsitleb majandusmudelite arengut, sealhulgas reaalsete majandustsüklite (RBC) mudeleid ja DSGE mudeleid, mida täiendati pärast suurt majandus- ja finantskriisi.

Esimene alapeatükk rõhutab andmepõhiste mudelite tähtsust, viidates Stockile ja Watsonile (1998, 2006), kes demonstreerisid mitme muutujaga mudelite väärtust aegridade prognoosimisel, eriti kui kriisieelseid märke ongi mitu. Hoolimata nendest arengutest on finantskriiside ennustamine endiselt üsna keeruline, sealhulgas kuna need hõlmavad sageli näitajaid nagu mis pole otseselt jälgitavaid nagu krediiditsüklid.

Lisaks toob esimene alapeatükk välja, mil moel erineb binaarsete sündmuste, näiteks süsteemsete finantskriiside ja pidevate muutujate, näiteks tööstustoodangu, prognoosimine. Rääkides neist teguritest, mille hulka kuuluvad ka riigivõlg ja krediidi kasv, viitatakse võtmetähtsusega uuringutele, mis rõhutavad finantskriisi kriisieelsete märkide olulisust süsteemsete finantskriiside ennustamisel.

Teine alapeatükk käsitleb suurandmete ja masinõppe suhet, rõhutades nende kasvavat tähtsust majandusuuringutes ja prognoosimises. Peatükis uuritakse üksikasjalikult suurandmete määratlust ja mõju ning toome esile andmemahtude eksponentsiaalse kasvu, eriti Google'i-järgsel ajastul.

Seejärel liigutakse suurandmete struktureeritud ja struktureerimata komponentide juurde, kus on oluliselt kasvanud kvantitatiivsete ja kvalitatiivsete and-

mete, näiteks teksti, heli ja video maht. Andmete mahu ja keerukuse tõttu läheb nende analüüsimiseks tarvis spetsialiseeritud masinõppe algoritme. Töös analüüsitakse juhendatud ja juhendamata masinõppe meetodeid, keskendudes nende rakendamisele suure hulga majanduslike kriisieelsete märkide korral. Juhendatud õppetehnikaid nagu regressioon ja klassifitseerimine kasutatakse kriisieelsete märkide ja tulemuste vaheliste seoste modelleerimiseks, samas kui juhendamata õppetehnikad avavad varjatud mustreid ilma eelnevalt määratletud siltideta.

Alapeatükk käsitleb ka olulisi masinõppemeetodeid nagu lasso ja *ridge* regressioon – mõlemad on olulised tööriistad keerulistes andmestikes ülesobitamise ja tunnuste valimisega tegelemiseks. Samuti tuleb juttu puupõhistest meetoditest. Töö analüüsib antud kontekstis dimensionaalsuse needuse kontseptsiooni, mille järgi prognoosimudelis kasutatavate muutujate arvu kasv suurendab tõusva mudeli keerukuse kaudu ülesobitamise tõenäosust.

Traditsioonilised ökonomeetrilised mudelid ei suuda leida suurandmete keskkondadest põhjuslikke seoseid, muutes nende otsimise üheks masinõppe kõige olulisemaks võimalikuks rakenduseks majanduses. Kuigi põhjuslikkuse küsimustes on hetkel võrreldes masinõppe rakendamisega ennustusmudelites vähe kirjandusallikaid, tuuakse välja masinõppe põhjuslike seostega tegelevate tööde aina kasvava arvu.

Alapeatüki lõpus fookus on kasvav huvi masinõppe abil prognoosimise ja lähiprognoosimise vastu, eriti majandusandmete kontekstis. Kuigi tehnikad nagu AdaBoost ja XGBoost on näidanud majanduslike sündmuste (nt majanduslanguste) prognoosimisel paljulubavaid tulemusi, pole väljakutsed ületatud, eriti keerukamate masinõppemudelite musta kasti sarnase olemuse tõttu. See rõhutab, et kuigi need mudelid toimivad hästi mittestandardses andmekeskkonnas, jätab tõlgendatavus sageli soovida. See piirab laiemaid rakendusi majandusteooria arendamisel.

Kuna makroökonoomikas on masinõppetehnikaid kasutama hakatud alles hiljuti, käsitleme järgmises osas makroökonoomika kasvavat huvi masinõppe vastu, mis mikroökonoomilises analüüs levis juba varem. Osaliselt selgitab hiljem tekkinud huvi asjaolu, et olemasolevad mudelid, näiteks faktormudelid, käsitlevad suuri andmestikke väga tõhusalt, kuid makroökonoomikutel on sageli juurdepääs vähematele andmepunktidele. Need mudelid ei kasuta ära aga tihti masinõppetehnikate potentsiaali, eriti nende võimet hallata andmete keerulisi mittelineaarsusi ja säilitada põhjalikumat teavet.

Mittelineaarsus on kaasaegses makroökonoomikas oluline aspekt, eriti majandusandmete prognoosimisel ja analüüsimisel. Mitmed teadlased, nagu Franses ja Van Dijk (2005), rõhutavad mittelineaarsete mudelite tähtsust, eriti tööstustoodangu pikaajalises prognoosimises. Hoolimata mittelineaarsete mudelite olulisusest pole need lühiajaliste prognooside puhul alati tõhusamad kui lihtsamad lineaarsed mudelid, mistõttu eelistatakse üha enam just pikaajaliste ennustuste tegemiseks masinõppemeetodeid. Masinõppemeetodite, näiteks võimendamismeetodite eeliseks on mittelineaarsustega ja muutujate teisendamisega kohanemine, nagu on näidanud näiteks Döpke jt. (2017), kus AdaBoosti kasutati Saksamaa majanduslanguste ennustamiseks.

Traditsioonilised ökonomeetrilised meetodid on endiselt väärtuslikud peamiste majandusnäitajate, näiteks tööstustoodangu prognoosimisel, isegi kui arvesse tuleb võtta mitut muutujat. Teatud mudelite, näiteks faktormudelite tõlgendamine võib olla aga keeruline. Teadlased uurivad, kuidas kombineerida masinõpet ökonomeetriliste mudelitega, et ennustustäpsust parandada. Näiteks Li ja Chen (2014) ühendasid dünaamilised faktorimudelid masinõppepõhiste mudelitega, et prognoosida USA peamisi makromajanduslikke näitajaid, leides, et masinõppemudelid ületasid tööstustoodangu ennustamisel traditsioonilisi dünaamilisi mudeleid.

Seda suundumust toetavad ka teised uuringud, näiteks Heij jt. (2011), kus leiti, et peakomponentide analüüs andis Ameerika Ühendriikide tööstustoodangu prognoosiks täpsemaid tulemusi kui Conference Boardi liitjuhtindikaator (*Composite Learning Index*, CLI). Sarnaselt näitasid Kock ja Teräsvirta (2014), et närvivõrgud ületasid traditsioonilisi lineaarseid mudeleid (nagu AR mudelid) tööstustoodangu prognoosimisel, eriti suure majandussurutise ajal.

Regressioonipuu meetodid, näiteks võimendamine, on osutunud väga tõhusaks mittelineaarsete seoste ja andmetes esinevate erindite käsitlemisel. Döpke jt (2017) näitab, kuidas võimendamine aitab leevendada lihtsatest regressioonipuudest tekkivaid sagedasi probleeme. Ng (2014) prognoosis USA majandussurutisi võimendamise ja üle 1500 kriisieelse märgi abil, näidates, et masinõppemeetodid suudavad suuri andmehulki tõhusalt hallata. Teised uuringud, nagu Buchen ja Wohlrabe (2011), näitavad, et USA tööstustoodangu prognoosimisel ületavad võimendamismeetodid faktormudeleid ka pikas perspektiivis.

Võimendamismeetodid osutuvad jätkuvalt kasulikuks paljudes erinevates prognoosimisülesannetes. Näiteks kasutasid Lehmann ja Wohlrabe (2015, 2016) võimendamist Saksamaa piirkondliku SKP ja Saksamaa tööstustoodangu prognoosimiseks, näidates selle meetodi mitmekesisust erinevate majandusnäitajate ja prognoosihorisontide kontekstis. Bai ja Ng (2009) näitasid, kui tõhus on võimendamine faktoritega täiendatud regressioonides kriisieelsete märkide valimisel, samas kui Carriero jt (2011) võrdles klassikalisemaid makroökonomeetrilisi meetodeid ja mitme muutujaga võimendamist prognoosides, mis puudutasid majanduslikke muutujaid nagu tööstustoodangu kasv, inflatsioon ja baasintressimäär. Tulemused näitasid, et lihtsamate masinõppemeetodite kombineerimine võimendamisega tõstis oluliselt ennustustäpsust.

Viimane alapeatükk sisaldab põhjalikku ülevaadet masinõppetehnikaid, finantskriise ja majandustsükleid käsitlevast kirjandusest. Kuna masinõppelähenemised suudavad hallata suure hulga muutujatega andmekeskkondi ja tuvastada keerulisi mittelineaarseid seoseid, äratavad nad viimasel ajal palju tähelepanu finantskriiside ja majandustsüklite ennustamise küsimustes. Paljud teadlased on võrrelnud masinõppetehnikate tõhusust traditsiooniliste meetoditega majandustsükli prognoosimise kontekstis. Teiste seas puudutame selliseid teadustöid nagu Exterkate jt. (2016), Maehashi ja Shintani (2020), Syed ja Lee (2021), Yousuf ja Ng (2021) ja Zhang jt. (2023).

Kuigi masinõppemeetodid on näidanud häid tulemusi mittelineaarsuste tabamisel ja ennustustäpsuse parandamisel, on traditsioonilistel meetoditel teatud olu-

kordades endiselt eelis, eriti statistilise olulisuse osas. Akadeemikud võrdlevad masinõppetehnikaid faktormudelitega, mis koondavad mitu majanduslikku muutujat hallatavamaks andmekogumiks (Cepni jt, 2019; Soybilgen ja Yazgan, 2021; Syed ja Lee, 2021). Kuigi masinõppetehnikad toimivad üldiselt paremini kui faktormudelid, võib sõltuvalt ajaperioodist ja andmekeskkonnast nende tõhusus varieeruda.

Mitmed uurimused käsitlevad faktormudelite ja masinõppetehnikate integreerimist. Hübriidlähenemine ühendab mõlema meetodi tugevused ja parandab tõestatult ennustustäpsust. Uuringud rõhutavad, et on oluline kasutada paljusid erinevaid kriisieelseid märke (Panagiotelis jt, 2019; Mogliani ja Simoni, 2021; Babii jt, 2022), eriti andmekogumite puhul, kus on palju muutujaid või erinevate ajasagedustega andmeid, ning et masinõpe suudab sellist keerukust paremini hallata kui traditsioonilised mudelid. Siiski on kriisieelsete märkide valik – nii eelvaliku kui automatiseeritud meetodite kaudu – veel üks kriitiline küsimus (Cepni jt, 2019; Jardet ja Meunier, 2022; Borup jt, 2023).

Masinõppemeetodid paistavad silma ka majandusandmetes mittelineaarsete seoste tuvastamisel ja haldamisel. Regressioonipuudel põhinev masinõpe on osutunud tõhusaks mittelineaarsete seoste tabamisel, millega traditsioonilised mudelid sageli hakkama ei saa. Teine oluline kaalutlus on see, kuidas toimivad masinõppemudelid erinevate alavalimite või ajavahemike lõikes. Uuringud näitavad, et sõltuvalt valitud ajavahemikust või riigispetsiifilistest tingimustest võivad tulemused oluliselt erineda. See rõhutab mudelite konkreetsete andmekeskkondadega kohandamise tähtsust (Panagiotelis jt, 2019; Soybilgen ja Yazgan, 2021; Zhang jt, 2023; Guerrón-Quintana ja Zhong, 2023).

Andmekogudest on FRED-MD andmestikku – koos selle kvartaalse versiooniga FRED-QD – kõige sagedamini kasutatud, eriti USA majandustsüklite prognoosimisel. See andmestik pakub rikkalikku makroökonoomiliste muutujate kogumit ja on saanud masinõppemudelite võrdlemise standardiks. Märgilised uuringud, nagu Exterkate jt. (2016), Guerrón-Quintana ja Zhong (2023) ning Kauppi ja Virtanen (2021), on näidanud, et masinõppemeetodid võivad USA majandusnäitajate ennustamisel traditsioonilisi lähenemisi ületada, kuigi sõltuvalt uuritava ajavahemiku majanduslikest tingimustest võib nende tõhusus varieeruda.

Kuigi enamik senisest masinõppeuuringutest makroökonoomilise prognoosimise valdkonnas keskendub majandustsüklitele, tärkab üha enam huvi nende tehnikate rakendamise vastu finantskriiside ennustamisel. Finantskriisid on oma olemuselt raskemini ennustatavad, kuna need on harvad, ettearvamatud ja keeruka iseloomuga. Mitmed uuringud on aga leidnud, et masinõppemeetodid võivad mõnikord finantskriise traditsioonilistest mudelitest paremini ennustada. Näiteks Alessi ja Detken (2018) näitasid, et EL-i riikide süsteemsete finantskriiside ennustamisel said juhumetsad logit-mudelitest veidi paremini hakkama. Lisaks näitasid Liu jt (2022), et masinõppemeetodid, eriti need, mis kasutavad indikaatorite valimiseks mänguteooriast tulenevaid Shapley väärtuseid, võiksid kriiside varajast prognoosimist tõhustada. Kuid nende meetodite edu kriisi ennustamisel sõltub suuresti andmete omadustest (nt tasakaalustamata andmete

võimalik esinemine, aegridade kasutamine paneelandmete asemel), mudeli valikust ja riigispetsiifilistest tingimustest.

Kokkuvõtteks võib öelda, et masinõppel on finantskriiside ja majandustsüklite prognoosimisel palju potentsiaali, kuid selle kasutamisel on ka puudusi. Usaldusväärsete mudelite loomist takistavad andmete keerukus, tundlikkus erinevate kriisieelsete märkide hulkade suhtes ja finantskriiside haruldus. Teadusuuringute kasvav hulk ja masinõppetehnikate pidev areng, eriti ansamblimeetodid nagu juhumetsad ja gradientide võimendamine, annavad siiski lootust, et makroökonoomiliste prognooside täpsust ja usaldusväärsust saab tulevikus parandada.

Esimese empiirilise uuringu tulemused ja arutelu

Esimene empiiriline töö prognoosib Ameerika Ühendriikide kuist tööstustoodangu kasvumäära aastatel 1963–2019, kasutades erinevaid võimendamismudeleid, võrreldes nende tulemuslikkust teiste masinõppemeetodite ja traditsiooniliste ökonomeetriliste standarditega nagu autoregressiivsed mudelid, juhuslik ekslemine ja faktormudelid. Uuring rõhutab vajadust põhjalikuma arusaamise järele sellest, kuidas võimendamismudelid ennustustäpsust mõjutavad ja kas need ületavad tööstustoodangu prognoosimisel standardseid ökonomeetrilisi mudeleid.

Võimendamisel põhinevad mudelid ületavad juhuslikku ekslemist järjepidevalt kõigis hinnangutes ja näitavad oma konkurentsivõimet, eriti lihtsamate masinõppemudelite vastu. Faktormudeliga võrreldes näitavad AdaBoostRT, komponendipõhine võimendamine ja XGBoost paremaid tulemusi. AdaBoostRT töötab kõige paremini lühiajalistes prognoosides, samas kui XGBoost on tõhusam keskpikkades prognoosides. Komponendipõhine võimendamine tagab kõigi prognoosiperioodide peale ühtlaseima tõhususe. Need tulemused ühtivad varasemate uuringute tulemustega, mis rõhutavad võimendamismudelite tõhusust, eriti majandustsüklite andmetes mittelineaarsete mustrite tuvastamisel.

Võimendamismudeleid autoregressiivsete mudelitega võrreldes leidsime, et ainult komponendipõhine võimendamine ja AdaBoostRT toimivad lühiajalistes prognoosides paremini kui AR, samas kui XGBoost toimib paremini keskpikkades prognoosides. See on kooskõlas varasemate uuringutega, mis näitavad, et masinõppemudelitel on raske isegi lihtsaid AR-mudeleid ületada (Maehashi ja Shintani, 2020; Yousuf ja Ng, 2021). Eriti oluline on võimendamistehnika, iseäranis komponendipõhise võimendamise uurimine, kuna faktormudeleid masinõppega kombineerides parandab see ennustustäpsust.

Lisaks näitavad tulemused, kuidas võimendamismudelite spetsiifilised omadused nende ennustusjõudu mõjutavad. Käesolev uuring on teadaolevalt esimene, mis võrdleb rohkem kui kahte võimendamismudelit tööstustoodangu prognoosimise kontekstis, pakkudes väärtuslikke teadmisi võimendamisalgoritmide erinevuste kohta. Näiteks erineb AdaBoostRT jõudlus oluliselt AdaBoostR2 omast. Keerukam võimendamismudel XGBoost töötab paremini, kuid seisab silmitsi omade tehniliste väljakutsetega.

Ühe olulise järeldusena toome välja, et kui andmevalim sisaldab suure majandussurutise perioodi (2007–2009), toimivad võimendamismudelid paremini.

Näib, et masinõppetehnikad, nagu võimendamine, tulevad suurema majandusliku ebastabiilsuse ajal andmete mittelineaarsusega paremini toime kui stabiilsematel perioodidel. See on kooskõlas Goulet Coulombe jt (2022) uuringutega (2022), mis näitavad, et masinõppetehnikad ületavad tavaliselt teisi lähenemisviise, eriti majanduskriisi ajal.

Uuring käsitles ka erinevate testimismeetodite mõju. Nn laieneva akna meetod (*expanding window method*), mis kasutab uuemaid andmeid väljastpoolt valimit, teeb XGBoosti ja AdaBoostRT tulemused üldjuhul halvemaks, mis viitab sellele, et neid mudeleid tuleb kohandada vastavalt majandusstruktuuri muutustele ajas.

Uuringust võib järeldada, et võimendamismudelid võivad toimida paremini kui traditsioonilised ökonomeetrilised mudelid, eriti lühiajaliste ja keskpikkade prognooside puhul. Võrdlus faktormudelitega ja majandussurutise mõju näitavad, kui tähtis on valida ennustusmudeleid andmete ja majandusliku konteksti põhjal. Tulemused rõhutavad, et võimendamismudeleid tuleks täiendavalt uurida, eriti nende kasutamist makromajanduslike muutujate, näiteks tööstustoodangu, prognoosimisel.

Teise empiirilise uuringu tulemused ja arutelu

Teine empiiriline töö prognoosib AdaBoosti kasutades süsteemseid finantskriise, võttes arvesse finantsandmete mittelineaarsusi. Uuring kasutab Jordà jt (2017) andmestikku, mis hõlmab kuutteist riiki aastatel 1870–2016, koos finantskriisidega seotud muutujatega. Finantskriisid erinevad tavapärastest majandustsüklitest ja sisaldavad tugevalt mittelineaarseid seoseid, mis teeb AdaBoosti meetodi paljutõotavaks, kuna see suudab niisuguseid keerukusi hallata. Töö võrdleb AdaBoosti tulemusi logistiliste regressioonide ja sammregressiooni modelite tulemustega, mis on standardsed kriiside prognoosimise viisid.

AdaBoost näitab potentsiaali finantskriiside tuvastamisel, eriti kui mudel põhineb rikkalikul makroökonoomiliste kriisieelsete märkide kogumil. Selle tõhusus sõltub aga paljudest teguritest, näiteks hindamismeetodist, andmekeskkonnast ja sõltuva muutuja valikust. Õiged valikumeetodid on seega väga olulised. Riigispetsiifiliste andmete kasutamine andmete koondamise või paralleelandmete kasutamise asemel andis uuringus vastakaid tulemusi, kuigi see lähenemine on siiski paljutõotav. Tulemused on kooskõlas varasemate allikatega, mis näitasid, et suurem hulk kriisieelseid märke (kriisi ennustavaid muutujaid) parandaks finantskriiside vallas masinõppemudelite prognoosivõimet.

Uuring toob esile tasakaalustamata andmete väljakutse, mis on finantskriiside andmestikes levinud, kuna kriisid on rahulike perioodidega võrreldes harvad sündmused. Tavaliselt kasutatakse probleemi lahendamiseks riikide rühmitamist või paneelandmeid, kuid need strateegiad pole järjepidevalt edukad olnud. Oluline järeldus on, et AdaBoosti lihtne rakendus ei pruugi näidata kogu masinõppe potentsiaali prognoosiülesande lahendamiseks ja AdaBoosti keerukamad variandid omavad potentsiaali anda paremaid tulemusi, nagu on näidanud ka Liu jt (2022).

Uuring käsitleb ka teistsugust meetodit kriisi ennustamiseks, mis kategoriseerib finantskriisi erinevateks perioodideks nagu kriisieelne, kriisiaegne ja kriisjärgne faas. Selle meetodi jaoks tuleb AdaBoosti kohandada, kuna mudel on loodud tavapäraseks binaarseks klassifitseerimiseks.

Muutujate olulisuse osas näitavad AdaBoosti tulemused selliste makromajanduslike näitajate asjakohasust, mis langevad kokku varasemate süsteemsete finantskriiside uuringutega. Näiteks mängivad muutuvad intressimäärad ja varahindade kõikumised finantskriiside ennustamisel olulist rolli, eriti väljakujunenud finantssüsteemidega arenenud majandustes. Uuring leiab, et liigne finantsliberaliseerimine ja eluasemehindade buumid võivad käivitada finantskriise. Sarnaselt näitab Itaalia võlgade ja SKP suhe, eriti kriisijärgses kontekstis, et fiskaalruum (paindlikkus kulutuste tegemisel)on kriiside haldamisel väga oluline.

Uuring käsitleb ka rahvusvaheliste majandusnäitajate, näiteks jooksevkonto tähtsust. Allikad pole üksmeelel, kas rahvusvahelised muutujad mängivad endiselt finantsstabiilsuse näitajatena suurt rolli (Davis jt, 2016; Obstfeld, 2012; Barrell jt, 2010), kuid selles kontekstis osutusid need siiski asjakohasteks. Itaalia tulemused viitavad võimalikule seostele kaubanduse ja süsteemsete finantskriiside vahel, mis toetaks seisukohta, et finantssidemed ja kaubandus võivad mõjutada riigi vastuvõtlikkust kriisidele.

Uuring toob esile riikide märkimisväärse heterogeensuse selles osas, millised muutujad kriise kõige paremini ennustavad. Arenenud finantsstruktuuridega suuremate majanduste, nagu Jaapani ja USA kriisimustrid erinesid väiksematest, avatuma majandusega riikidest. Turu suurus ja rahvusvahelisele kaubanduse avatus ning kodumaise majanduse tootlikkus mõjutavad samuti riigi vastuvõtlikkust nii kodu- kui välismaistele finantskriisidele.

Kuigi AdaBoosti tulemused finantskriiside tuvastamisel on paljulubavad, mõjutavad selle tõhusust valitud andmekeskkond ja kasutatud valimimeetodid. Uuring näitab, et AdaBoosti keerukamad versioonid koos sobivate makroökonoomiliste kriisieelsete märkidega, võiksid prognoose tõhustada, eriti finantskriisi andmetes leiduvate mittelineaarsete seoste kontekstis.

Tulemuste üldine arutelu

Väitekiri toob esile, kuidas ennustamistehnikad muutuvad üha keerukamaks, eriti pärast suurt majandussurutist. Teadlased kasutavad majandustsüklite ja süsteemsete finantskriiside ennustamiseks üha rohkem uusi mudeleid ja muutujaid. Doktoritöö uurib kaht prognoosimeetodit: üks toetab andmepõhist lähenemist, mis võimaldab suurema muutujahulga abil tuvastada mittelineaarsusi nii majandustsüklites kui ka finantskriisides, teine aga keskendub muutujate eelvalikule teoreetiliste makroökonoomiliste seisukohtade põhjal.

Selles uurimuses põimiti neid meetodeid, valides muutujaid makroökonoomika teooria põhjal ning kasutades suurt hulka kriisieelseid märke. Lisaks on mudelites kasutatud lihtsaid mittelineaarsusi, kusjuures teine empiiriline uuring toetub rohkem muutujate eelvalikule. Tasub märkida, et teises uuringus kasutatud FRED-MD andmestik põhineb varasematel uuringutel, nagu Stocki ja Watsoni

töö. Siinse doktoritöö panus makroökonoomika prognoosimisse on ühendada traditsioonilised ökonomeetrilised meetodid kaasaegse masinõppega ja pakkuda terviklikku raamistikku, mis parandaks ennustustäpsust ja täiendaks ühtlasi majanduslikku arusaama finantskriisidest.

Uuring tuvastas mitu teemat, mida saaks edasi uurida. Esiteks on masinõppe meetodid, eriti võimendamismudelid, suurandmete abil majandustsüklite prognoosimiseks väga paljutõotavad, mis viitab sellele, et rohkemate tegurite kaasamine võib parandada mudelite tulemuslikkust. Siiski ei pruugi keerukamad mudelid nagu XGBoost alati olla tõhusad, kui treeningandmed on piiratud või kasutatakse läbisegi tugevaid ja nõrku kriisieelseid märke, mis näitab, kui tähtis on nende mudelite rakendamist hoolikalt kaaluda.

Ka funktsioonide valik on masinõppes kriitilise tähtsusega, kuna ebaolulisi kriisieelseid märke sisaldavad liialt keerulised mudelid võivad prognoosi kvaliteeti halvendada. Seetõttu peaksid teadlased kasutama eelkõige asjakohaseid ja väärtust lisavaid kriisieelseid märke ja vältima kõrvalisemaid. Tasakaalustamata andmetega klassifitseerimisharjutus rõhutab, kui tähtis on süsteemsete finantskriiside tuvastamisel õigesti koostatud valim.

Doktoritöö leiab, et võimendamismudelite võrdlemiseks neist keerukamate faktormudelitega, kus muutujad eelvalitakse, ning väljaselgitamiseks, kuidas tekib ajaridade andmetesse mittelineaarsus, tuleb teha veel uurimistööd, nagu näitas selgelt teine uuring. AdaBoostRT mudeli ennustamisvõimet rahulikel aegadel võiks põhjalikumalt uurida. Lisaks võivad finantsandmed, eriti kõrge sagedusega andmed nagu päevased aktsiaturu andmed, pakkuda majandustsükli prognoosimiseks kasulikke teadmisi.

Tuleb meeles pidada, et 19. sajandi süsteemsed finantskriisid ei ole päris samasugused kui kaasaegsed kriisid, näiteks 2007–2008. aasta kriis. Seda erinevust tuleks edasi uurida. Lahendusi võivad pakkuda analüüsitava perioodi lühendamine ja kõrgema sagedusega andmete või paneelandmete kasutamine, nagu on tehtud varasemates uuringutes, kuid kõigil strateegiatel on puudusi.

Doktoritöö jõuab järeldusele, et hoolimata potentsiaalikusest ei tohiks masinõppemeetodid asendada traditsioonilisi ökonomeetrilisi meetodeid, mis kasutavad teoreetilisi teadmisi ja eksperthinnanguid, vaid need võivad üksnes täiendada olemasolevaid protseduure. Selline tasakaalustatud strateegia parandab prognoosimisvõimet, eriti keerulistes majanduskeskkondades. Lõpuks rõhutab uuring, kui oluline on hinnata masinõppetehnikate eeliseid ja puudusi, kuna nende tõhusus sõltub konkreetsete andmete omadustest ja uurimise eesmärkidest.

Doktoritöö lühikokkuvõte

Doktoritöö uurib masinõppetehnikate rakendamist makroökonoomilises prognoosimises, pöörates erilist tähelepanu majandustsüklitele ja süsteemsete finantskriiside ennustamisele. Töö käsitleb kaht peamist strateegiat: andmepõhist lähenemist suure hulga muutujatega, et tuvastada majandussündmuste mittelineaarsust, ja muutujate eelvalimist makroökonoomika teooria alusel. Uuringus kasu-

tatakse nii teooriapõhist muutujate valikut kui ka puhtalt andmetest tuvastatud mitmesuguseid kriisieelseid märke, integreerides mõlemat lähenemist.

Tulemused rõhutavad tasakaalustamata andmete käsitlemise, tunnuste valiku ja mittelineaarsuse tähtsust masinõppemudelites, viidates sellele, et võimendamismeetodid võivad olla eriti tõhusad, kui andmestikus on suur hulk muutujaid. Keerulised mudelid nagu XGBoost nõuavad aga hoolikat ettevalmistust, eriti piiratud andmete korral. Töös arutatakse ka kriiside prognoosimise probleemkohtade üle, eriti ajalooliste andmete puhul, ning vajadust tulevaste uuringute järele, et mudeleid optimeerida.

Kokkuvõttes näitab doktoritöö, kuidas masinõppetehnikad võivad makroökonoomilisi prognoose tõhustada, ja pakub põhjalikku raamistikku, mis ühendab kaasaegsed masinõppemeetodid traditsioonilise ökonomeetriaga. Lisaks rõhutab töö andmete ja metoodika hoolsa hindamise hädavajalikkust, veendumaks, et masinõppemudelid üksnes täiustavad, mitte ei asenda traditsioonilisi meetodeid.

Kuigi suurandmed on tõepoolest suurendanud nõudlust masinõppemeetodite järele, poleks mõistlik ainult neile kriitilise mõtlemiseta lootma jääda ja traditsioonilised meetodid hüljata. Masinõppemeetodid võivad toimida paremini koostöös ökonomeetriliste tehnikatega, kuna mõlemad meetodid põhinevad statistikal. Seega on paljud uued meetodid traditsioonilistega rohkem seotud kui esmapilgul tunduda võib.

CURRICULUM VITAE

Name: Luca Alfieri
Date of Birth: 05.01.1986
Work address: Narva mnt 18, Tartu, Estonia
Contact: luca.alfieri@ut.ee

Education

2017–	PhD Studies in Economics, University of Tartu, Estonia
2016–2017	Advanced Studies Program (ASP) in International Economic Policy Research, Kiel Institute for the World Economy, Germany
2013–2015	PhD program in Economics and Finance (exit on September 7, 2015) University of Verona, Italy
2011–2012	Master 2 (M.A.) in Macroeconomics and European Policies, University of Strasbourg, France
2007–2009	M.A. in Local Development, Cooperation and International Markets, University of Parma, Italy (cum laude)
2004–2007	B.A. in Development Economics and International Cooperation (cum laude), University of Parma, Italy

Languages: Italian, English, French

Research and Teaching Experience

Current positions

2023–	Post-doc (Italian assegno di ricerca), Politecnico di Milano – DAStU
2023–	Visiting Lecturer, University of Tartu, Estonia
2017–	Teaching Assistant, University of Tartu, Estonia

Past positions

2024	Lecturer, Politecnico di Milano – DAStU
2021–2023	Lecturer, University of Tartu
2019–2023	Junior Research Fellow, University of Tartu ,Estonia
2022	Visiting Researcher, The Vienna Institute for International Economic Studies, Austria
2020–2021	Visiting Researcher, Bank of Estonia, Estonia

Main research areas: Machine learning, macroeconomics, macroeconometrics, applied economics.

ELULOOKIRJELDUS

Nimi: Luca Alfieri
Sünniaeg: 05.01.1986
Töökoha aadress: Narva mnt 18, Tartu, Eesti
E-post: luca.alfieri@ut.ee

Haridus

2017–	Doktoriõpe majandusteaduses, Tartu Ülikool, Eesti
2016–2017	ASP programm rahvusvahelises majanduspoliitikas, Kieli Maailmamajanduse Instituut, Saksamaa
2013–2015	Doktoriõpe majanduse ja rahanduse erialal (kuni 7. septembril 2015), Verona Ülikool, Itaalia
2011–2012	Magistrikraad (M.A.) makroökonoomikas ja Euroopa poliitikates, Strasbourgi Ülikool, Prantsusmaa
2007–2009	Magistrikraad kohaliku arengu, koostöö ja rahvusvaheliste turgude alal, Parma Ülikool, Itaalia (*cum laude*)
2004–2007	Bakalaureusekraad arenguökonoomikas ja rahvusvahelises koostöös (*cum laude*), Parma Ülikool, Itaalia

Keeled: itaalia, inglise, prantsuse

Teadus- ja õpetamiskogemus

Praegused ametikohad

2023–	Järeldoktorant (Itaalia assegno di ricerca), Milano Polütehniline Ülikool – DAStU, Itaalia
2023–	Külalislektor, Tartu Ülikool, Eesti
2017–	Õppeassistent, Tartu Ülikool, Eesti

Varasemad ametikohad

2024	Lektor, Milano Polütehniline Ülikool – DAStU, Milano, Itaalia
2021– 2023	Lektor, Tartu Ülikool
2019–2023	Nooremteadur, Tartu Ülikool
2022–	Külalisteadur, Viini Rahvusvahelise Majandusuuringute Instituut, Austria
2020–2021	Külalisteadur, Eesti Pank

Peamised uurimisvaldkonnad: Masinõpe, makroökonoomika, makroökonomeetria, rakenduslik majandusteadus.

DISSERTATIONES RERUM OECONOMICARUM UNIVERSITATIS TARTUENSIS

1. **Олев Раю.** Экономическая ответственность и ее использование в хозяйственном механизме. Tartu 1991, 390 с.
2. **Janno Reiljan.** Majanduslike otsustuste analüütiline alus (teooria, metodoloogia, metoodika ja meetodid). Tartu 1991.
3. **Robert W. McGee.** The theory and practice of public finance: some lessons from the USA experience with advice for former socialist countries. Tartu 1994, 123 p.
4. **Maaja Vadi.** Organisatsioonikultuur ja väärtused ning nendevahelised seosed (Eesti näitel). Tartu 2000, 220 lk.
5. **Raul Eamets.** Reallocation of labour during transition. Disequilibrium and policy issues: The case of Estonia. Tartu 2001, 252 p.
6. **Kaia Philips.** The changes in valuation of human capital during the transition process in Estonia. Tartu 2001, 289 p.
7. **Tõnu Roolaht.** The internationalization of Estonian companies: an exploratory study of relationship aspects. Tartu 2002, 193 p.
8. **Tiia Vissak.** The internationalization of foreign-owned enterprises in Estonia: An extended network perspective. Tartu 2003, 221 p.
9. **Anneli Kaasa.** Sissetulekute ebavõrdsuse mõjurite analüüs struktuurse modelleerimise meetodil. Tartu 2004, 260 lk.
10. **Ruth Alas.** Organisational changes during the transition in Estonia: major influencing behavioural factors. Tartu 2004, 210 p.
11. **Ele Reiljan.** Reasons for de-internationalization: An analysis of Estonian manufacturing companies. Tartu 2004, 235 p.
12. **Janek Uiboupin.** Foreign banks in Central and Eastern European markets: their entry and influence on the banking sector. Tartu 2005, 185 p.
13. **Jaan Masso.** Labour reallocation in transition countries: efficiency, restructuring and institutions. Tartu 2005, 241 p.
14. **Katrin Männik.** The impact of the autonomy on the performance in a multinational corporation's subsidiary in transition countries. Tartu 2006, 252 p.
15. **Andres Vesilind.** A methodology for earning excess returns in global debt and currency markets with a diversified portfolio of quantitative active investment models. Tartu 2007, 170 p.
16. **Rebekka Vedina.** The diversity of individual values and its role for organisations in the context of changes. Tartu 2007, 177 p.
17. **Priit Sander.** Essays on factors influencing financing decisions of companies: risk, corporate control and taxation aspects. Tartu 2007, 178 p.
18. **Kadri Ukrainski.** Sources of knowledge used in innovation: an example of Estonian wood industries. Tartu 2008, 263 p.
19. **Kristjan-Olari Leping.** Heterogeneity of Human Capital and its Valuation in the Labour Market. Tartu 2008, 160 p.

20. **Kadri Männasoo.** Essays on financial fragility – evidence from the corporate and banking sectors in Central and Eastern Europe. Tartu 2008, 152 p.
21. **Made Torokoff.** Patterns of learning organisation – Estonian experiences. Tartu 2008, 163 p.
22. **Helena Rozeik.** Changes in ownership structures, their determinants and role in the restructuring of enterprises during transition: evidence from Estonia. Tartu 2008, 268 p.
23. **Jaanika Meriküll.** Technological change and labour demand. Tartu 2009, 148 p.
24. **Anne Aidla.** The impact of individual and organisational factors on academic performance in Estonian general educational schools. Tartu 2009, 219 p.
25. **Alexander Gofman.** Experimentation-Based Product Development in Mature Food Categories: Advancing Conjoint Analysis Approach. Tartu 2009, 273 p.
26. **Anne Reino.** Manifestations of organizational culture based on the example of Estonian organizations. Tartu 2009, 272 p.
27. **Krista Jaakson.** Management by values: the analysis of influencing aspects and its theoretical and practical implications. Tartu 2009, 225 p.
28. **Eve Parts.** Social capital, its determinants and effects on economic growth: comparison of the Western European and Central-Eastern European countries. Tartu 2009, 293 p.
29. **Egle Tafenau.** Welfare effects of regional policy in the constructed capital model. Tartu 2010, 187 p.
30. **Epp Kallaste.** Employee workplace representation: an analysis of selected determinants. Tartu 2010, 173 p.
31. **Danel Tuusis.** Interest rate influence on the behavior of economic subjects. Tartu 2010, 201 p.
32. **Elina Kallas.** Emotional intelligence, organizational culture and their relationship based on the example of Estonian service organizations. Tartu 2010, 304 p.
33. **Dorel Tamm.** Alignment between the factors of the innovation process and public sector innovation support measures: an analysis of Estonian dairy processors and biotechnology enterprises. Tartu 2010, 231 p.
34. **Rasmus Kattai.** The links between private sector indebtedness and banking sector vulnerability: An Estonian case study. Tartu 2011, 164 p.
35. **Kurmet Kivipõld.** Organizational Leadership Capability and its evaluation based on the example of Estonian service organizations. Tartu 2011, 199 p.
36. **Janno Järve.** Downward Nominal Wage Rigidity in the Estonian Private Sector. Tartu 2011, 197 p.
37. **Kristina Toming.** The impact of integration with the European Union on the international competitiveness of the food processing industry in Estonia. Tartu 2011, 241 p.

38. **Andrus Kotri.** Customer experience evoking and management in services. Tartu 2011, 360 p.
39. **Andres Kuusik.** Segmentation of repeat visitors using passive mobile positioning data: customer loyalty based approach. Tartu 2011, 158 p.
40. **Tuuli Pärenson.** Social impact evaluation in social enterprises in Estonia: need, readiness and practices. Tartu 2011, 216 p.
41. **Indrek Saar.** Optimal alcohol taxation in Estonia. Tartu 2011, 185 p.
42. **Kertu Lääts.** Management accounting change in a dynamic economic environment based on examples from business and public sector organizations. Tartu 2011, 250 p.
43. **Reelika Irs.** Teacher performance appraisal and remuneration aspects of performance management on the example of Estonian general educational schools. Tartu 2012, 322 p.
44. **Anne Lauringson.** The impact of the generosity of unemployment benefits on Estonian labour market outcomes in a period of crisis. Tartu 2012, 268 p.
45. **Peeter Peda.** The relationship between governance and performance in water services provision in Estonian municipalities. Tartu 2012, 326 p.
46. **Andres Kuusk.** Financial contagion during times of crisis: a meta-analysis based approach with special emphasis on CEE economies. Tartu 2012, 211 p.
47. **Kerly Espenberg.** Inequalities on the labour market in Estonia during the Great Recession. Tartu 2013, 312 p.
48. **Xiaotian Zhang.** Internationalization processes of Chinese firms: The role of knowledge. Tartu 2013, 274 p.
49. **Helen Poltimäe.** The distributional and behavioural effects of Estonian environmental taxes. Tartu 2014, 141 p.
50. **Eneli Kindsiko.** Organisational Control in University Management: A Multiparadigm Approach on the Example of the University of Tartu. Tartu 2014, 211 p.
51. **Diana Eerma.** A bookkeeping approach to social accounting for a university faculty: the case of the University of Tartu. Tartu 2014, 293 p.
52. **Kaia Kask.** Public sector real estate asset management models and their evaluation. Tartu 2014, 264 p.
53. **Ott Pärna.** Managerial and contextual factors influencing innovation in information technology-based public sector services: an exploratory cross-national study. Tartu 2014, 410 p.
54. **Merle Tambur.** Workplace bullying in Estonian organizations: The prevalence and causes. Tartu 2015, 210 p.
55. **Sten Anspal.** Essays on gender wage inequality in the Estonian labour market. Tartu 2015, 195 p.
56. **Oliver Lukason**. Characteristics of firm failure processes in an international context. Tartu 2016, 180 p.
57. **Marko Viiding.** Role of electricity price in competitiveness of the manufacturing industry in liberalised electricity markets: the case of NordPool. Tartu 2016, 188 p.

58. **Bianka Plüschke-Altof.** Images of the Periphery Impeding Rural Development? Discursive Peripheralization of Rural Areas in Post-Socialist Estonia. Tartu 2017, 237 p.
59. **Tarmo Puolokainen.** Public Agencies' Performance Benchmarking in the Case of Demand Uncertainty with an Application to Estonian, Finnish and Swedish Fire and Rescue Services. Tartu 2018, 247 p.
60. **Karin Sakowski.** The Role of National-Institutional Context in Organisations and in Organisational Innovation: The Case of Western and Central and Eastern European Countries. Tartu 2018, 135 p.
61. **Maryna Tverdostup.** Human capital and labour market disparities. Tartu 2018, 248 p.
62. **Bradley James Loewen.** Towards territorial cohesion? Path dependence and path innovation of regional policy in Central and Eastern Europe. Tartu 2018, 232 p.
63. **Kärt Rõigas.** University-industry cooperation in the context of the national innovation system. Tartu 2018, 208 p.
64. **Tatyana Tsukanova.** Insights into the Export Behavior of SMEs from Emerging Economies: Evidence from Russia and China. Tartu 2019, 251 p.
65. **Gerdien Margreeth Grootens.** Leadership of peripheral places: a comparative study of leadership processes in Estonian and Dutch peripheral places. Tartu 2019, 186 p.
66. **Tõnis Tänav.** Dynamics of firm innovation strategies: relationship with public sector support. Tartu 2020, 280 p.
67. **Gaygysyz Ashyrov.** Essays on firm-level corruption. Tartu 2020, 193 p.
68. **Tõnis Eerme.** Big Science as innovation intermediaries – micro- and meso-level effects from the collaboration with the European Space Agency. Tartu 2020, 173 p.
69. **Isaac Nana Akuffo.** The relationship between authentic leadership competences and nepotism, favouritism, and cronyism – the case of the Ghanaian banking sector. Tartu 2020, 147 p.
70. **Virgo Süsi.** Corporate governance and performance of private SMEs. Tartu 2021, 130 p.
71. **Mariia Chebotareva.** Functional Overlapping Competing Jurisdictions (FOCJs) as a Possible Tool for Inter-municipal Cooperation in the Provision of Russian School Services. Tartu 2021, 343 p.
72. **Aare Värk.** Practice-based exploration of knowledge, knowing and knowledge management. Tartu 2021, 128 p.
73. **Nataliia Ostapenko.** Information, Business cycles and Monetary policy. Tartu 2021, 197 p.
74. **Artur Meerits.** First-level military leaders' leadership competencies and their relationship with unit effectiveness in terms of collectivistic leadership with the example of the Estonian Defence Forces. Tartu 2022, 178 p.
75. **Magnus Piirits.** The Impact of Pension Reforms on Pension Inequality in Estonia: An Analysis with Microsimulation and Typical Agent Models. Tartu 2022, 219 p.

76. **Liina Joller-Vahter.** The government as an enabler and accelerator of diffusion of radical innovations. Tartu 2022, 134 p.
77. **Liis Roosaar.** Essays on labour mobility and labour productivity. Tartu 2022, 182 p.
78. **Sigrid Rajalo.** University-industry collaboration: interaction structure and preconditions. Tartu 2023, 117 p.
79. **Nino Kokashvili.** Public sector evolution under conditions of political business cycle – theory and empirical evidence. Tartu 2023, 194 p.
80. **Alo Lilles.** The relationship between university-industry cooperation and regional capabilities in Europe. Tartu 2023, 142 p.
81. **Mark Kantšukov.** Valuation of companies under the distributed profit taxation system. Tartu 2023, 233 p.
82. **Iulia Trabskaia.** Idea and opportunity identification and implementation within the entrepreneurial process and journey. Tartu 2023, 160 p.
83. **Laura Helena Kivi.** Regional labour markets and assimilation of foreign labour force. Tartu 2023, 157 p.
84. **Vladyslav Soloviov.** The role of culture for innovative processes. Tartu 2024, 138 p.
85. **Diana Gabrielyan.** Essays on inflation, expectations and central bank communication. Tartu 2024, 233 p.
86. **Anastasia Sinitsyna.** Links between segregation processes on the labour and housing markets: evidence from Finland. Tartu 2024, 121 p.
87. **Tanel Hirv.** Internationalisation of the Estonian research system through the lens of bibliometric indicators: criticism and policy recommendations. Tartu 2024, 162 p.
88. **Tran Nguyen Hoang Quan.** Perception of organisational culture in terms of task and relationship orientations in Vietnam. Tartu 2024, 129 p.
89. **Kristjan Pulk.** Influencing personal financial decisions with information provision. Tartu 2024, 168 p.
90. **Liisi Lembinen.** Comprehending Decision-Making in Academic Library Leadership: A Situational Analysis. Tartu 2024, 141 p.