UNIVERSITY OF TURKU

ESSAYS ON ECONOMIC TIME SERIES FORECASTING

Timo Virtanen

TURUN YLIOPISTON JULKAISUJA ANNALES UNIVERSITATIS TURKUENSIS
SARJA SER. E OSA TOM. 65 | OECONOMICA | TURKU 2020

UNIVERSITY OF TURKU

ESSAYS ON ECONOMIC TIME SERIES FORECASTING

Timo Virtanen

TURUN YLIOPISTON JULKAISUJA ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. E OSA – TOM. 65 | OECONOMICA | TURKU 2020

University of Turku

Turku School of Economics
Department of Economics
Economics
Doctoral programme of Turku School of Economics

Supervised by

Professor Heikki Kauppi
University of Turku
Turku, Finland

Professor Matti Viren
University of Turku
Turku, Finland

Reviewed by

Professor Markku Lanne
University of Helsinki
Helsinki, Finland

Associate Professor Anders Kock
University of Oxford
Oxford, United Kingdom

Opponent

Associate Professor Anders Kock
University of Oxford
Oxford, United Kingdom

ISBN 978-951-29-8196-0 (PRINT)
ISBN 978-951-29-8197-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)
Painosalama, Turku, Finland 2020

UNIVERSITY OF TURKU
Turku School of Economics
Department of Economics
Economics
TIMO VIRTANEN: Essays on Economic Time Series Forecasting
Doctoral Dissertation, 118 pp.
Doctoral Programme of Turku School of Economics
November 2020

ABSTRACT

This dissertation comprises an introductory chapter and three essays on economic forecasting. The first essay addresses the topical question of whether it is possible to predict the occurrence of systemic banking crises, such as the financial crisis of 2008. The question is approached by using a method called exuberance test that can detect the formation of rational asset price bubbles in time series information from various market prices and aggregate debt stocks. A positive test result is interpreted as a potential pre-crisis period. The test can be used as an early warning system that signals increased risk of a systemic banking crisis. The essay studies the usefulness of various time series, such as house prices, aggregate debt stocks and debt service measures for this purpose.

The second essay aims to improve the accuracy of univariate forecasts for macroeconomic and financial time series using a nonlinear model. The estimation method for the nonlinear models, the gradient boosting machine, is obtained from the machine learning literature. The boosting estimator provides a framework for estimating unknown prediction functions and selecting variables from a set of potential predictors. It can be considered as a nonparametric estimator that provides a more favourable bias-variance trade-off than other estimators. In the essay we show that nonlinear modelling can improve the prediction performance for a meaningful share of macroeconomic variables. The result already established in the existing literature, that linear prediction models dominate nonlinear models for the majority of these time series, remains unchallenged. To remedy this issue, a two-stage estimator that combines the linear model and the nonlinear boosting model is proposed.

The third essay continues from the empirical setting of the second essay. Given the result that nonlinear modelling is beneficial for only a fraction of macroeconomic time series, it is necessary to find a practical testing procedure for selecting between the linear and a nonlinear prediction approach. The essay draws on existing literature on tests of expected forecasting accuracy. The alternative test procedures are discussed on a theoretical level. Size and power properties are examined in a simulation study. Finally, the essay concludes in an empirical application, where the effectiveness of the chosen test procedure is demonstrated.

KEYWORDS: Forecasting, forecast evaluation, banking crises, macroeconomy, time series, nonlinearity, machine learning.

TURUN YLIOPISTO
Kauppakorkeakoulu
Taloustieteen laitos
Taloustiede
TIMO VIRTANEN: Essays on Economic Time Series Forecasting
Väitöskirja, 118 s.
Turun Kauppakorkeakoulun Tohtoriohjelma
Marraskuu 2020

TIIVISTELMÄ

Tämä väitöskirja koostuu johdantoluvusta sekä kolmesta esseestä, jotka käsittelevät taloudellisten aikasarjojen ennustamista. Ensimmäisessä esseessä tutkitaan, onko systeemisiä pankkikriisejä mahdollista ennustaa talouden kokonaisvelkamäärien sekä asunto- ja osakemarkkinoiden hinta-aikasarjojen perusteella. Tutkimuksessa käytetään menetelmää, jonka avulla voidaan testata rationaalisen hintakuplan esiintymistä tällaisissa aikasarjoissa. Testin perusteella ajanjaksot luokitellaan mahdollisesti kriisiä edeltäviin periodeihin. Esseessä esitettyjen tulosten perusteella testiä voidaan soveltaa systeemisten pankkikriisien ennakkovaroitusjärjestelmänä.

Toinen essee käsittelee makrotaloudellisten aikasarjojen ennustamista. Esseessä pyritään parantamaan yhden muuttujan mallin ennustetarkkuutta käyttäen epälineaarista ennustemallia. Mallin estimointiin käytetään gradient boosting-menetelmää, joka on lähtöisin koneoppimiskirjallisuudesta. Boosting menetelmän avulla voidaan estimoida annetun aikasarjan tuntematonta ennustefunktiota, joten sitä voidaan pitää epäparametrisena estimaattorina. Menetelmän etuna on sen muihin estimaattoreihin verrattuna edullisempi käyttäytyminen mallin monimutkaisuuden lisääntyessä. Esseessä osoitetaan, että epälineaarisen mallinnuksen avulla voidaan parantaa ennustetarkkuutta merkittävälle osalle makrotaloudellisista aikasarjoista. Samalla kuitenkin vahvistetaan myös aiemmin tunnettu tulos, jonka mukaan lineaarisen mallin ennustetarkkuus on parempi suurimmalle osalle tutkituista aikasarjoista. Esseessä esitetään kaksivaiheinen estimaattori, joka yhdistää lineaarisen mallin epälineaariseen boosting-mallin ja osoitetaan, että lineaarisen mallin yleiskäyttöisyys ja epälineaarisen mallin tarkkuusetu voidaan saavuttaa sen avulla.

Kolmanessa esseessä tutkitaan ennustemallin valintaa. Koska lineaarisen ja epälineaarisen mallin suhteellinen ennustetarkkus vaihtelee eri aikasarjoilla, syntyy tarve löytää käytännöllinen proseduuri, jolla mallien odotettua tarkkuutta voidaan arvioida ennen ennusteen tekemistä. Esseessä käydään läpi kirjallisuuden tarjoamat vaihtoehdot mallien odotetun ennustetarkkuuden testaamiseen. Testausmenetelmien ominaisuuksia, testin kokoa ja voimakkuutta, tutkitaan simulaatioiden avulla. Lopuksi esseessä osoitetaan valitun testausproseduurin toimivuus soveltaen sitä makrotalousaikasarjojen ennustemallin valintaan.

ASIASANAT: Ennustaminen, ennustetarkkuus, pankkikriisit, aikasarjat, makrotalous, epälineaarisuus, koneoppiminen

Acknowledgements

First, I would like to express my gratitude and appreciation to my supervisors, Professors Heikki Kauppi and Matti Viren who have guided me through this long process. Despite your busy schedules, you have found the time to give me the feedback, guidance and encouragement needed to push through the challenges of the research.

I would like to thank the pre-examiners of this thesis, Professor Markku Lanne from the University of Helsinki, and Professor Anders Kock from the University of Oxford, for your valuable feedback. Moreover, I would like to thank Professor Anders Kock for accepting the invitation to act as the opponent for this dissertation.

I would also like to take the opportunity to thank all the co-authors of the first two essays in my dissertation: my supervisors Matti Viren and Heikki Kauppi, Dr. Eero Tölö and Dr. Katja Taipalus from the Bank of Finland. I enjoyed working together with you in the research projects and am sure that as a result of the shared expertise, we were able to cover the research questions in more depth and achieve more meaningful results.

During the dissertation work I have been employed by the department of Economics in the University of Turku and the Bank of Finland. I wish to express my gratitude for these opportunities. I also gratefully acknowledge the financial support from the OP Group Research Foundation and Yrjö Jahnsson Foundation.

Last, but certainly not least, I would like to thank my family for supporting me in this journey. My parents, Kaija and Urpo, and late grandparents, for laying the solid foundations for my life, and my sister Laura, for being my lifelong good friend.

My wonderful wife Emmi, thank you for always being there for me. Without your support and encouragement this project might not have happened. And our dear daughter Taimi, who has brought a lot of love and joy to our lives. Your curious and enthusiastic mind sets a great example for every scientist.

November 19th, 2020

Timo Virtanen

Table of Contents

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Virtanen, Timo; Tölö, Eero; Virén, Matti; Taipalus, Katja. Can bubble theory foresee banking crises? *Journal of Financial Stability*, 2018; 36: 66–81.

II Kauppi, Heikki; Virtanen, Timo. Boosting Nonlinear Predictability of Macroeconomic Time Series. *International Journal of Forecasting*, 2020.

III Virtanen, Timo. Testing for nonlinear predictability.

1 Introduction

This dissertation comprises three essays on economic time series forecasting. The first essay addresses a question that has been topical since the financial crisis of 2008, namely whether it is possible to predict or "foresee" systemic banking crises using financial and macroeconomic time series. Given the huge impact of banking crises to the economy, this question has raised a lot of interest amongst the research community and policy makers. Consequently, the literature on the subject has grown immense. The contribution of this article to the literature is that it investigates a large dataset that covers various macroeconomic and financial variables over 15 countries and introduces the use of a method called exuberance test due to Phillips, Shi and Yu (2015) as an early warning device for detecting price bubbles that may lead to a systemic banking crises.

The second essay considers a more conventional and well-defined forecasting problem; predicting the future values of macroeconomic and financial time series for the time period of one month to one year ahead. The essay does not attempt to answer any specific policy relevant questions but rather focuses on nonlinear modeling and evaluates its benefits compared to the simpler, linear modeling technique. The essay combines elements from the machine learning literature and the econometrics literature. The main contribution is the introduction of a procedure that combines the nonlinear, machine learning based estimator to the traditional linear modeling technique to produce forecasts that are more accurate yet retain the robustness of the linear models.

The third essay looks at the same forecasting problem from a different angle. An issue with nonlinear modeling, as shown in the second essay and many other studies, is that nonlinear modelling may increase the forecasting accuracy for some time series compared to the linear modelling approach but may produce inferior out-of-sample forecasts for others. Thus, a forecaster would like to know in advance which approach is likely to be more accurate given the variable being forecast, specification of the forecasting models and the forecasting horizon. The essay draws on earlier contributions in the literature of forecast accuracy testing, evaluates the alternatives and proposes a testing procedure that can be used for selecting between a linear and a nonlinear forecasting model in this context.

This introductory chapter proceeds as follows. The next section briefly discusses some basic concepts and challenges of economic time series forecasting and provides context to the essays of the dissertation especially in the interest of readers who may not be familiar with the field. Section 1.2 discusses the machine learning approach and its application to forecasting. A summary of the essays is given in Section 2. A brief overview of the key methods that are applied in the research is also given in Section 2.

1.1 Economic forecasting and time series models – a brief introduction

Economic forecasting is usually understood as producing estimates about the expected future development of some variable of interest using information that is currently available. Such forecasts can be based on e.g. aggregated expert opinions or crowdsourced information. More usually, forecasts are made with a mathematical model that takes observed values of some economic quantities as input. The structure of a forecasting model may be based on economic theory or be constructed in a more empirical manner. An example of the former are dynamic stochastic general equilibrium (DSGE) models that are often used in such contexts as forecasting the gross domestic product of a country. A time series model containing explanatory variables that are known to correlate with the forecasted quantity is an example of the latter, although the specification of time series models can also be grounded on theory.

Forecasts can be made for variables that are continuous, such as stock market prices or interest rates, binary, such as a recession occurring during a certain quarter, or in some cases, multinomial. In econometric terms, forecasting a continuous variable is referred to as a regression problem and the binary case as a classification problem. For binary variables, a forecasting model usually predicts the probability of a certain event. For regression problems, the forecasting model may predict the expected value or expected distribution of a variable at some point of time in the future. In the forecasting vocabulary, forecasts for the expected value are called point forecasts and expected distribution are referred to as density forecasts. The first essay in this dissertation examines methods for making binary predictions of systemic banking crisis events. The second and third essay focus on improving univariate point forecasts for continuous variables.

The dissertation focuses on time series models that have been an important workhorse for economic forecasting for several decades. The family of models that can be used for analysis and forecasting of time series includes everything from simple moving average or trend models to complex multivariate models that incorporate information from several explanatory variables in the forecast calculation. A list of different models that have been conceived is likely to be too long to include in this introduction. An interested reader could refer to e.g. Hamilton (1994) for a comprehensive treatment on the subject.

A forecasting process typically starts with the specification of the forecasting model. In almost all forecasting scenarios, multiple alternative models can be considered. It is usually not clear a priori which model will yield the most accurate forecasts out-of-sample. Intuitively, one might think that a more complex model would always be more accurate, since it can incorporate more information in the forecast. In practice, the opposite is often true; Complex models easily overfit the data and a simple model may outperform them in the accuracy of it's out-of-sample predictions. This known finding is also corroborated in this dissertation and it motivates the third essay.

It is also worth noting that the concept of accuracy may be defined in different ways. For continuous variables, the accuracy of a prediction is most often assessed through the mean square error criterion, i.e. the mean of the squared forecasting errors produced by the model for some data. The accuracy of a binary prediction model is usually assessed by the ratio of true (i.e. correctly predicted) prediction signals to all signals. In some cases, a different measure of accuracy, such as the mean absolute error, bias, or some economic measure of the impact of relying on the forecasts of different models may be desirable. For example, in the first essay, a parameterized policy maker loss function is used to compare the usefulness of the predictions of different models from a perspective of an economic policy maker. As another example, when forecasting a financial time series, it may be of interest to compare the economic impact of basing an investment strategy on the predictions of different models rather than comparing the mean square errors of the forecasts.

Most time series models require some parameters to be estimated from historical data before they can be used for forecasting. It is possible to identify three distinctly different approaches to estimating a time series model. The classical, often called frequentist approach assumes that there exist some unknown true values for the parameters of the model. The aim of the estimation process is to use the available data to obtain estimates for the parameters that are as close as possible to the unknown true values. Bayesian analysis, in contrast, assumes that the parameters of the model are random variables. Bayesian estimation uses procedures that can explicitly deal with the uncertainty and produce estimates for the probability distribution of the parameters. The third approach that has emerged more recently is algorithmic estimation that originates from the machine learning literature. The last two essays of this dissertation employ machine learning and attempt to build bridges between the machine learning and classical approaches. We discuss machine learning in more detail in the next section.

1.2 The machine learning approach

In the traditional approach to economic forecasting, it is assumed that the data is generated by an unknown stochastic data model. The forecasting process starts with the specification of the forecasting model, estimation of its parameters and finally, using the model to calculate the forecast. Since the actual data generating model is typically unknown for any empirical data, the aim of the model specification process is to come up with a forecasting model that nests the true data generating model. In practice, any forecasting model constructed in this manner is likely to be misspecified, meaning that the chosen model is unlikely to completely nest the optimal forecasting model.

Machine learning methods take a different approach to this matter. They start with little or no assumptions about the data generating model and use algorithmic tools to “learn” the model from the data. In supervised machine learning, the data that is given to the algorithm includes the dependent variable and a set of potential predictor variables. The algorithm estimates (or trains, as usually called in the machine learning literature) a model that expresses the dependent variable as a function of a subset of the potential predictor variables.

The use of machine learning methods has become increasingly popular in the field of econometrics during the last decade, although it has been argued (Athey and Imbens, 2019) that the adoption has been slower than in many other fields of science. A benefit of machine learning methods is that they can work with large datasets and build, if necessary, complex forecasting models. At least in theory, they provide means to dispense with any prior assumptions of the model structure and to build the model based on correlations that are observed in the data.

This kind of freedom does not come without challenges. Since the model structure is learned from data, successful use of the machine learning methods requires that there is a sufficiently large amount of high quality data available in which these correlations can be reliably estimated – an issue which might become a problem with e.g. macroeconomic data that is published either quarterly or, at best, monthly, is subject to revisions and changes in how the data is collected and aggregated.

In theory, a model constructed by a machine learning algorithm can become arbitrarily complex, if we allow the algorithm to continue training until all dependencies in the training data have been modelled. In this case, the resulting model may not offer a good generalization of the true data generating model. As discussed in the previous section, complexity is not necessarily a desirable feature for a forecasting model. For this reason, controlling for overfitting is a key task when estimating a forecasting model with an algorithmic estimator. The established way of doing this is to use cross validation or backtesting to find the optimal degree of complexity for the model. This entails splitting the estimation data into two parts;

The training dataset is used to estimate multiple candidate models of varying complexity, that is usually controlled by some hyperparameters of the estimation algorithm. The hyperparameters of the model that predicts the test dataset most accurately are then applied for estimating the final model.

Due to these issues, machine learning methods cannot necessarily be considered as "off-the-shelf" tools that can be easily adopted to solve any forecasting problem. Consequently, there is also need for research on how to best apply machine learning in different forecasting scenarios.

One feature of machine learning algorithms that has been critiqued is that they don't easily yield themselves to similar theoretical analysis that has been central to the traditional approach to stochastic data modelling. It has been difficult in many cases to prove even the basic theoretical qualities of some machine learning estimators, such as consistency. Indeed, in the scope of this dissertation, providing a theoretical argument for why the boosting estimator should be expected to perform better in building a time series model for the macroeconomic variables than an artificial neural network turned out to be an elusive goal, so one has to settle for intuitive arguments of why this might be the case and simulation evidence at least for now.

A philosophical idea behind the development of the machine learning algorithms has been that it is not always necessary to be able to prove causality; instead, it usually suffices to know only the correlation. The focus in predicting outcomes has made machine learning tools very successful in building forecasting models. Although the field of causal machine learning has been emerging in the recent years, the limited ability to produce evidence for causal relationships that concerns most machine learning algorithms impairs their use in empirical economics research and policy analysis work, where causality is often an important question.

Most machine learning methods also do not provide means to test for the statistical significance of a given predictor in the model – a practice that is very central to statistics and empirical economics. Additionally, the estimated models may become very complicated and opaque. Thus, interpreting the relationships between different predictors in a machine learning may be difficult.

Nevertheless, there are still conceivable ways how traditional machine learning methods could be utilized in empirical economics work. For instance, one can use algorithms for discovering correlations in large empirical data sets. If the algorithm uncovers correlations that were previously unknown, they can be used as hypotheses in further empirical or theoretical research. If the machine learning method itself cannot produce the necessary evidence on causality, one needs to resort to more traditional research methods to confirm the result.

2 Summary of Essays

2.1 Can bubble theory foresee banking crises?

In the early 2000's, many people thought that financial crises were a thing of the past, at least in the western world. The financial crisis in 2008 and later the Euro crisis in 2010 proved that such events are still a possibility. Among many other things, they spurred interest for research on how such crises could be predicted so that preventive measures could be taken by policy makers in due time.

In this essay, we study the possibilities to predict systemic banking crises[1] using a test procedure that is not, strictly speaking, a forecasting model but an econometric test designed to detect the formation of rational asset price bubbles in financial time series.

Our approach thus differs from the common methodology used for predicting binary events. In most other contributions in this literature, historical data is used to estimate a binary forecasting model, which is then used to predict crisis events. In our case, the forecasting model is fixed, i.e. no parameters are estimated before making a prediction. The hypothesis behind this approach is that systemic banking crises are – at least on many occasions – caused by a bursting asset price bubble in the economy. Thus, if we can detect the formation of such bubbles, we can gain useful information about the risk of a crisis and, to some extent, predict their occurrence in the future.

Forecasting systemic banking crises is a problematic endeavor in many respects. First, it is unclear what constitutes a systemic banking crisis we are trying to forecast. The classification of an event into a systemic crisis or something else is based on expert opinion and there are several differing classifications. To alleviate this issue, we use three different crisis classifications that have been made by different experts in the field. Another problem is that systemic banking crises have historically originated from many sources, although leveraged asset price bubbles are likely to

1 Laeven and Valencia (2012) define a systemic banking crisis as an event where the banking system shows significant signs of financial distress and significant banking policy interventions are taken in response to losses in the banking system

be one of the most common causes. On the other hand, there are several known asset price bubbles that did not result in a banking crisis.

The results are evaluated through a policy maker loss function and a usefulness metric derived from the loss. This practice is customary in the literature and thus also makes our results comparable with those of other studies. The policy maker loss is defined as a function of correctly predicted and unpredicted crises and false alarms. An additional parameter allows adjusting for policy maker preferences, i.e. whether they are more averse towards false alarms or crises that were not predicted. As the name suggests, the best forecasting model is the one that minimizes the loss. In addition to the usefulness measure, we also calculate the Area Under the Receiver Operating Characteristic (AUROC) metric which is a standard way of evaluating the forecasting accuracy of a binary predictor.

Our results show that the proposed method is useful in predicting the occurrence of systemic banking crises and produces more accurate signals than the benchmark signaling method commonly used in the literature. The most useful information is obtained by applying the method to aggregate debt variables, as explosive growth of the aggregate debt stock appears to be associated with the majority of the studied systemic banking crisis events. In many cases, the crises are preceded by a price bubble in the real estate market that could be detected in real time with the test procedure. When the variables are ranked by the usefulness measure, the debt related variables appear better predictors of crises than real estate price measures and other variables.

The results also corroborate many findings in the existing literature. For example, it has been known that many systemic banking crises result from a bursting real estate market bubble while stock market bubbles do not seem to have such clear connection to banking crises. A likely explanation for this is that price bubbles in the real estate market are mainly financed by bank credit are thus more dangerous to the banking sector than stock market bubbles.

2.1.1 Rational bubbles and the PSY test

Asset pricing theory states that the price of any asset should equal the discounted sum of the revenue it is expected to generate in the future, e.g. dividends paid for stockowners or rent for an apartment. Using this notion, we can calculate a fair price for any asset at the time t, the so-called fundamental price, which we denote by P_t^f. Any deviation from this price may be called a bubble. Denoting the bubble component as B_t, we can decompose the current price of any asset as

$$P_t = P_t^f + B_t \tag{1}$$

Assume now that there is no bubble and the price of an asset equals its fundamental price. If the revenue is expected to follow a random walk process with a possible upward trend, the price of the asset also follows a similar process. If a bubble exists, then its expected value must fulfill the following condition:

$$E(B_t) \geq (1 + r)B_t \tag{2}$$

Where r is the risk-free interest rate. If the condition (2) is not met, it is more profitable to sell the asset and invest in a risk-free instrument. This, in turn, would result in the collapse of the bubble.

What (2) implies is that P_t no longer follows a random walk process because the risk-free interest rate is usually above zero. In the presence of a bubble, the price of an asset follows an explosive process, i.e. a process of the form

$$P_t = \alpha + \gamma P_{t-1} + e_t \tag{3}$$

where $\gamma > 1$ while we assumed that $\gamma = 1$ when no bubble is present.

The Generalized Sup ADF (GSADF) test due to Phillips, Shi and Yu (2015) is designed to test the null hypothesis $\gamma = 1$ versus the alternative hypothesis $\gamma > 1$ for time series processes like (3). The test is based on the right-tailed Augmented Dickey-Fuller (ADF) unit root test. The test statistic equals the supremum value of ADF test statistics calculated over a moving and backwards expanding window. The distribution of the test statistic is nonstandard, and the critical values are obtained through a simulation process.

The test procedure produces two test statistics. The GSADF statistic indicates whether the null hypothesis can be rejected for at least some part of the time series (i.e. one or more bubbles are present in the time series). The Backward Sup ADF (BSADF) test is performed for each observation of the time series and it tests whether $\gamma > 1$ at a given time *t*. Thus, the BSADF test statistic can be used to date-stamp the bubble episodes and detect bubbles in real time.

In this essay, the time periods are classified into potential pre-crisis periods based on the BSADF test statistic, which is considered as a proxy classifier. If, for a given time series, the test rejects the null hypothesis at time *t*, that period is classified as belonging to a pre-crisis period, i.e. a period that may be followed by a systemic banking crisis.

To calculate the usefulness values for each individual time series, type I and II error rates are calculated by comparing the predictions to a pre-crisis classification derived from the expert opinions of the timing of past systemic banking crises. A pre-crisis period is defined as the time period of twelve to five quarters before a known systemic banking crisis event.

2.2 Boosting nonlinear predictability of macroeconomic time series

It is often assumed that macroeconomic and financial time series are inherently nonlinear. Several studies have attempted to uncover nonlinear time series models that would allow for more accurate forecasting of these series. However, most studies have concluded that linear models produce the most accurate forecast for most macroeconomic time series. In some cases, however, nonlinear models may significantly improve the forecasting accuracy for some time series so there is continued interest for nonlinear modelling.

These observations arise in many studies including Stock and Watson (1999), Marcellino (2005), Teräsvirta, van Dijk and Medeiros (2005) and Kock and Teräsvirta (2016). Given that the true form of the nonlinearity is often unknown, nonparametric estimators and machine learning have been popular in the macroeconomic forecasting literature. From the machine learning methods, single-hidden-layer feed-forward Artificial Neural Networks (ANN) have most commonly been used for estimating the nonlinear prediction functions.

The first research question of this essay is to determine the share of variables in a database of 128 macroeconomic and financial variables that benefit from nonlinear modeling for 1 to 12 months ahead forecasts and the extent of the potential gain in forecasting accuracy. As a second question we study whether it is possible to retain the robustness of the linear modelling approach while utilizing the nonlinear predictability where it exists. The target of the analysis is the FRED-MD monthly database of macroeconomic variables due to McCracken and Ng (2016).

The boosting estimator from machine learning literature is utilized to estimate the nonlinear forecasting models. To date, the boosting estimator has been used in only few studies in this literature. The only other study, according to our present knowledge, where the boosting procedure is used with component-wise smoothing splines as base learners to forecast macroeconomic variables, is that of Robinzonov, Tutz and Hothorn (2012). Their empirical data, however, includes only variables related to German Industrial Production.

Two different ways to use the boosting procedure for estimating the time series model are examined. The first procedure is based on direct estimation using penalized component-wise regression splines as proposed by Bühlmann and Yu (2003). The other involves estimating the model in two stages. First, a linear autoregressive model is fitted to the data. In the second stage, the boosting procedure is used to fit a nonlinear model to the residuals of the linear model. The purpose of the two-stage procedure is to improve the robustness of estimation for such cases where the optimal prediction function is better represented by a linear function.

The usefulness of the boosting estimator as a method to estimate nonlinear prediction models for stationary time series is evaluated in a simulation study.

Knowing the data generating processes of the artificial data, we can approximate the optimal prediction function and thus the minimum attainable mean square forecasting error (MSFE) and compare the performance of the different estimators to that. The general conclusion from the simulation exercise is that the boosting estimator can, in terms of the MSFE, outperform the linear prediction approach when the data is generated by a nonlinear process. In cases where the optimal prediction function is linear or its linear approximation is nearly equally accurate, the boosting forecasts tend to be slightly less accurate than those of the linear prediction model. This result is intuitive, since any flexible nonparametric estimator will suffer from higher estimation uncertainty compared to a simple parameterized model. In an appendix to the essay, we study the accuracy of the ANN estimator and find that the boosting estimator is generally able to model the nonlinear time series more accurately than the type of ANN we use in the simulations.

In the empirical part of the study we apply the two boosting procedures to the FRED-MD dataset and compare their forecasting accuracy to that of a linear autoregressive model. The forecasting (out-of-sample) period is chosen to be from the beginning of the year 1999 until the end of 2016 and forecasting horizons from one to 12 months ahead. Since the forecasting period includes the financial crisis of 2008, which may be assumed to produce major structural breaks in many of the variables in the data, we also look at the results for the forecasting period of 1999-2007 separately.

We find that 30-40% (depending on the forecasting horizon) of the time series in the dataset can be more accurately modeled by the boosting procedure. The difference is statistically significant for 15-20% of the time series that have at least some predictability in their own history. For 10-25% of the series, the linear prediction procedure is more accurate than the nonlinear procedure and for the rest, the accuracy is essentially the same. Importantly, the list of time series for which the nonlinear prediction is more accurate for the whole out-of-sample time period includes about 75% of the series for which the same is true for the pre financial crisis period, implying that nonlinear predictability may be a persistent feature of certain macroeconomic time series.

One possible explanation for the fact that the linear prediction is often more accurate is the simulation finding that the linear procedure can estimate linear processes more accurately than the boosting procedure. Another finding is that the linear model is in many cases more robust to outliers and large shocks in the forecasted time series. The reason behind this is that the estimated nonlinear model needs to be extrapolated in case the value of some of the predictor values falls outside the range of values in the data used in the estimation of the model. This extrapolation procedure may produce wildly inaccurate results in some cases with the empirical data.

2.2.1 The boosting estimator

Boosting is a supervised machine learning method that can be used for estimating prediction models for classification and regression problems. The model is estimated in a stagewise fashion so that each iteration, "boost", improves the fit of the model. The algorithm can be shown to conduct numerical optimization in function space (Friedman, 2001) and hence, the boosting algorithm can be viewed as a nonparametric estimator.

The boosting algorithm is as follows. Given a sample of observations $(y_i, \boldsymbol{x_i}), i \in 1, \dots, N$, where y denotes the dependent variable and $\boldsymbol{x}$ a (vector of) explanatory variables, the algorithm implicitly estimates a prediction function of the form

$$y_i = \sum_{m=1}^{M} b(\boldsymbol{x_i}, \theta_m) \tag{4}$$

Where $b(x_i, \theta_m)$ is a parametric function called the "base learner" and θ_m parameters estimated at iteration m.

Let F_m denote the estimated prediction function. Given a loss function $L(y, F)$ the base learner for iteration $m = 1,2, \dots, M$ is estimated by minimizing the loss

$$\theta_m = \underset{\theta \in \Theta}{\arg min} \sum_{i=1}^{N} L(y_i, F_{m-1}(\boldsymbol{x_i}) + b(\boldsymbol{x_i}, \theta_m)) \tag{5}$$

At iteration m the prediction function thus becomes

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + b(\boldsymbol{x}, \theta_m) \tag{6}$$

It is essential that the base learner a simple function of the predictor variables for the boosting algorithm to realize it's full potential. Simplicity in this case may mean that the base learner is a function of only a subset of the predictor variables and that it belongs to a narrow class of functions. Regression trees with a limited number of splits are one example of a function that is usually applied in this context.

Component-wise penalized regression splines are a natural choice for base learner when the goal is to estimate smooth nonlinear regression functions. Bühlmann and Yu (2003) establish several important results related to this base learner. First, they show, that regardless of the order of the smoothing splines, the resulting estimator (4) can adapt to any higher order of smoothness. They also show that the resulting estimator achieves the minimax optimal MSE rate of convergence.

Bühlmann and Yu also discuss an important property of the boosting estimator; a different bias-variance trade off, where the bias of the model diminishes exponentially with growing number of iterations, but the variance increases only in an exponentially decaying rate. This property leads to what Bühlmann and Yu call “slow overfitting”, i.e. the out-of-sample loss of the estimated model grows only slowly as a function of the number of iterations, when the optimal number is exceeded. Thus, the accuracy of the forecasts is not likely to collapse even if one fails to determine the optimal number of iterations accurately.

The properties of the boosting estimator discussed above led us to believe that it could suit the purpose of building a nonlinear forecasting model for macroeconomic data while controlling for overfitting better than the artificial neural network. In the simulation study and empirical application, we find evidence to support this claim.

2.3 Testing for nonlinear predictability

In the empirical application of the essay on nonlinear predictability of macroeconomic time series it was discovered that a nonlinear prediction model often has a worse out-of-sample performance in terms of the mean-squared forecasting error than a linear prediction model. There are several possible explanations for this issue. For instance, the nonlinear model may not capture a linear prediction function as accurately as the linear model when it is estimated with a finite data sample. It is also possible that some properties of the time series, such as structural breaks, are estimated as nonlinearities that do not reflect the true prediction function, although the essay does not provide direct evidence to support this hypothesis.

Whatever the reason behind the varying success of the nonlinear models, in the general case it is necessary to perform some kind of a model selection test to select a forecasting model that is expected to yield the most accurate forecast for the time series in question. In this essay, we assume that the forecasting model is either a linear or a nonlinear univariate autoregressive model. As in the previous essay, the nonlinear model is estimated with the boosting procedure.

The literature has developed many tests for comparing the expected predictive ability of alternative forecasting models. Most of them are based on theory derived at the population level, i.e. which forecasting model is expected to be more accurate when the estimation sample size approaches infinity. Some tests address the finite sample hypothesis, i.e. which prediction method is more accurate conditional on the available estimation sample.

The essay provides a theoretical overview of the main test procedures that could be used. In the application considered in this essay, the question of interest is clearly the finite sample hypothesis. Test procedures for it have, however, been much more difficult to develop compared to the population level case. From the existing finite sample tests, the procedure due to Giacomini and White (2006) is found suitable for the application and included in the study. The computational demands of bootstrap-based inference combined with the boosting estimator unfortunately make using approaches such as the one proposed by Clark and McCracken (2015) infeasible.

It is also possible to use a test procedure designed for the population level hypothesis as the basis for model selection. One benefit of population level tests is that the test statistic is usually asymptotically normal. The asymptotic result may however break down in situations where the forecasting models are nested, i.e. the alternative model includes the null model and the models can be made equal with a parametric restriction. If this is the case, the additional parameters of the alternative model approach zero and the estimated models become equal under the null hypothesis when the sample size grows to infinity.

One test procedure where this is a concern is that of Diebold and Mariano (1995). it is a well-known fact that the asymptotic distribution of their test statistic becomes nonstandard if the compared forecasting models are nested. To further complicate the situation, we cannot know a priori if the estimated models will be nested or not, since the boosting procedure selects the predictor variables in the model from a set of candidate variables and may in same cases exclude variables that are included in the null model.

The bigger estimation uncertainty of the nonlinear estimation procedure compared to that of the linear procedure is also an issue that needs to be taken in account. Since the test procedure uses a shorter estimation sample than the actual forecasting process, the test results may be biased in favour of the null model. A test procedure that addresses this concern is proposed by Clark and West (2007). While the test has a nonstandard distribution, Clark and West show that inference based on standard normal critical values may be approximately valid. A bootstrap-based procedure could alternatively be used for inference, but we have omitted this possibility due to the computational difficulties discussed above.

Through the theoretical analysis and the simulation study we reach the conclusion, that the test procedure proposed by Clark and West is best suited for selecting between a linear and nonlinear prediction approach in the empirical application. At least with the sample sizes considered in the simulations, the test procedure of Clark and West has the highest power against the nonlinear alternatives. The size of the test is close to the nominal for different forecasting horizons, while the other test procedures suffer from size distortions that are larger in magnitude. The choice of a windowing method for the simulated out-of-sample procedure is also crucial. Recursive estimation with expanding estimation window appears to produce better estimates of the expected out-of-sample accuracy and more consistent size behaviour than the other options.

In the empirical application, we demonstrate the effectiveness of the test procedure of Clark and West in selecting between forecasting models for the FRED-MD dataset. A "hybrid forecast" constructed from a linear and nonlinear direct boosting prediction based on the model selection test result, is on average clearly more accurate than either of the methods alone. On the other hand, the test does not benefit the two-stage boosting procedure, as the two-stage model already explicitly includes the linear model.

The most favourable results, although only by a small margin, are obtained by using the test procedure to select between the two-stage- and direct boosting approaches. In this setting, the two-stage model can capture linear prediction functions and "mild" nonlinearities, i.e. cases where the prediction function is nonlinear, but the difference to the linear prediction is not statistically significant. The alternative direct boosting model, on the other hand, may be more accurate in cases where the optimal prediction function is nonlinear and the difference between the accuracy of the nonlinear prediction and its linear approximation is significant.

References

Athey, S., and Imbens, G. W. (2019) Machine Learning Methods That Economists Should Know About. Annual Review of Economics 11(1), 1–41.

Bühlmann, P., and Yu, B. (2003) Boosting with the L2 loss: regression and classification,. Journal of American Statistical Association 98(462), 324–339.

Clark, T.E. and McCracken, M.W. (2015) "Nested forecast model comparisons: A new approach to testing equal accuracy," Journal of Econometrics 186, 160–177.

Clark, T. E, and West, K. D. (2007) "Approximately normal tests for equal predictive accuracy in nested models," Journal of Econometrics 138, 291–311.

Diebold, F. X., and Mariano, R. S. (1995) "Comparing predictive accuracy," Journal of Business & Economic Statistics 13(3), 253–263.

Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine," Annals of Statistics, 29 (5), 1189–1232.

Giacomini, R., and White, H. (2006). "Tests of conditional predictive ability," Econometrica, 74 (6), 1545–1578.

Hamilton, J.D. (1994). “Time Series Analysis”, Princeton University Press.

Kock, A. B., and Teräsvirta, T. (2016) "Forecasting macroeconomic variables using neural network models and three automated model selection techniques," Econometric Reviews, 35, 1753–1779.

Laeven, L., Valencia, F., 2012. Systemic Banking Crises Database: An Update. IMF Working Paper n. 163/12.

Marcellino, M. (2005). Instability and non-linearity in the EMU. In C. Milas, P. Rothman, & D. van Dijk (Eds.), Nonlinear Time Series Analysis of Business Cycles. Amsterdam, Elsevier.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006) "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," Journal of Econometrics, 135, 499–526.

McCracken, M. W., and Ng, S.(2016) FRED-MD: A monthly database for macroeconomic research, Journal of Business & Economic Statistics 34(4), 574–589.

Robinzonov, N., Tutz, G., Hothorn, T. (2012) Boosting techniques for nonlinear time series models, Advances in Statistical Analysis, 96, 99–122.

Stock, J. H., and Watson, M. W. (1999) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, In Engle, R. F., and White, H. (eds.) Cointegration, causality and forecasting. Festschrift in Honour of Clive W. J. Granger, Oxford University Press, Oxford, 144

Teräsvirta, T. (2006) Forecasting economic variables with nonlinear models, In: Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.) Handbook of Economic Forecasting, Elsevier, Amsterdam, 459–512.

Teräsvirta, T., van Dijk, D., and Medeiros, M. C. (2005) Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination, International Journal of Forecasting, 21, 755–774.

UNIVERSITY OF TURKU

Painosalama Oy, Turku, Finland 2020

ISBN 978-951-29-8196-0 (PRINT)
ISBN 978-951-29-8197-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)