Organisation for Economic Co-operation and Development

ECO/WKP(2018)0

Unclassified **English - Or. English**

ECONOMICS DEPARTMENT

Adaptive Trees: A novel approach to macroeconomic forecasting

Economics Department Working Papers No. 14XX

By Nicolas Woloszko

Authorised for Publication by Luiz de Mello, Director, Policy Studies Branch, Economics Department.

All Economics Department Working Papers are available at www.oecd.org/eco/workingpapers

ABSTRACT/RÉSUMÉ

Adaptive Trees: A novel approach to macroeconomic forecasting

The paper introduces Adaptive Trees, an innovative machine learning algorithm specifically designed for economic forecasting. Economic forecasting is made difficult by economic complexity, which includes non-linearities (multiple interactions and discontinuities) and structural change (i.e. the continuous change in the distribution of economic variables). Our forecast methodology addresses these problems, thus yielding better performance than traditional models. We produced simulations in pseudo-real-time for six major economies (USA, UK, Germany, France, Japan, Italy). Adaptive Trees compare well to both model- and expertise-based benchmark forecasts.

JEL Classification:
Keywords:

****

French title

French abstract.

Classification JEL:
Mots clefs:

Table of contents

ABSTRACT/RÉSUMÉ 3
Adaptive Trees: A novel approach to macroeconomic forecasting 3
French title 3

Adaptive Trees: A novel approach to macroeconomic forecasting........ 6

1. Introduction 6
2. Main findings 8
3. Data sources and benchmarks 11
4. Method 12
   1.4.1. A tree-based approach to tackle non-linearities 12
   1.4.2. From regression trees to adaptive trees: dealing with structural change 17
   1.4.3. Data-driven model selection to address high dimensionality 18
5. Results 20
   1.5.1. Performance 21
   1.5.2. Interpretation 24
6. Conclusion 26

References 28

Annex A. The OECD Indicator Model 30

Annex B. The OECD Economic Outlook 31

Purpose 31
Coverage 31
The forecasting process 31
The role of government economists 32
Risk assessment 32
Medium and longer-term analyses 32

Annex C. Data description 34

Annex D. Full charts 35

Annex E. Full forecast performance tables 41

Annex F. Selected plot of aggregated variable contributions 43

Tables

Table 1. Variables used by the Indicator Models 11
Table 2. Performance of Adaptive Trees for the USA forecast (2007-2017) 22
Table 3. Performance of Adaptive Trees for the Japan forecast (2007-2017) 23

Table A C.1. Data availability per country and number of variables 34
Table A E.1. Performance of Adaptive Trees for the UK forecast (2007-2017).................... 41
Table A E.2. Performance of Adaptive Trees for the USA forecast (2007-2017) 41
Table A E.3. Performance of Adaptive Trees for the France forecast (2007-2017).................... 41
Table A E.4. Performance of Adaptive Trees for the Japan forecast (2007-2017) 42
Table A E.5. Performance of Adaptive Trees for the Germany forecast (2007-2017).................... 42
Table A E.6. Performance of Adaptive Trees for the Italy forecast (2007-2017) 42
Table A E.7. Performance of Adaptive Trees: average overall G6 countries (2007-2017) 42

Figures

Figure 1. Underfitting, right fit, overfitting 10
Figure 2. Bias-variance trade-off 10
Figure 3. Training a regression tree 13
Figure 4. Train and predict 14
Figure 5. Prediction with a regression tree 14
Figure 6. Variable contribution, a simple example 16
Figure 7. Choosing the size of the training sample 18
Figure 8. Cross-validation 20
Figure 9. Incremental learning framework for pseudo-real time simulations 21
Figure 10. Forecast of USA GDP growth at M+3 22
Figure 11. Forecast of Japan GDP growth at M+3 23
Figure 12. Histograms of standardised forecast error for all G6 countries, M+6 forecast 24
Figure 13. Decomposition of the Italy M+3 forecast in aggregate variable contributions, 2007-2009. 25
Figure 14. Aggregated variable contributions, Japan, M+3 26

Figure A D.1. USA 35
Figure A D.2. UK 36
Figure A D.3. France 37
Figure A D.4. Japan 38
Figure A D.5. Germany 39
Figure A D.6. Italy 40
Figure A F.1. Aggregated variable contributions, UK, M+3 43
Figure A F.2. Aggregated variable contributions, USA, M+3 43
Figure A F.3. Aggregated variable contributions, Italy, M+3 44
Figure A F.4. Aggregated variable contributions, Germany, M+3 44
Figure A F.5. Aggregated variable contributions, Japan, M+3 44
Figure A F.6. Aggregated variable contributions, France, M+3 45

Boxes

Box 1. Machine learning and econometrics 9
Box 2. Gradient Boosted Trees (GBT) 16

Adaptive Trees: A novel approach to macroeconomic forecasting

By Nicolas Woloszko[1]

1. Introduction

Machine learning was born in the 1960s as a set of techniques designed to extract information from data. It became a key player around the year 2000 thanks to the advent of Big Data. Big Data provided both the data and the computational means to deal with it, so sophisticated algorithms could be trained and experimented with. Since then, machine learning became ubiquitous in the industry and is at the very core of the artificial intelligence revolution. This paper addresses the following question: how can economic forecasters benefit from such a versatile technology?

That question arises from both the resemblance and asymmetry of machine learning with econometrics (see Box 1). Both disciplines share the purpose of learning from data. Machine learning differs from linear econometrics as it does not require prior domain knowledge (Breiman, 2001), whereas linear econometrics uses an economic model to build an econometric model. As opposed to both linear and non-linear econometrics, machine learning focuses on out-of-sample accuracy, rather than in-sample accuracy and unbiased estimates. As it is distinct from Bayesian econometrics, machine learning does not rely on probabilistic beliefs about the data generating processes and performs model selection on the sole basis of out-of-sample goodness-of-fit.

Linear econometrics is particularly constrained where economic complexity is concerned. Complexity refers among other things to non-linearities. One may specify non-linear relations in a linear model, using polynoms or simple interactions. But things get more complicated when dealing with multiple interactions and multiple discontinuities. Complexity may also refer to the existence of structural change, i.e. the fact that the economy is an ever-changing complex system where the rules may change over time. Linear models suppose stable relations and make a strong hypothesis that the distribution of data remains the same in sometimes long historical samples, as long as structural breaks are not explicitly specified. For instance, it is well documented that the Philipps Curve changed in nature around the 1980s thanks to new frameworks for monetary policy that tamed inflation and inflation expectations. Finally, complexity also refers to the context-dependence of economic relationships. A given policy may have heterogeneous impact across countries or among people across the income distribution.

Multiple interactions, discontinuities and structural breaks are particularly conspicuous around turning points and recessions. A telling example is housing bubbles. Growing

[1] The author is a member of the Economics Department of the OECD and is grateful to Sebastian Barnes, Mohammed Benlaldj, Orsetta Causa, Jean-Marc Fournier, Catherine Mann, Valentine Millot, Annabelle Mourougane, Patrice Ollivaud, Dorothée Rouzet, Nicolas Ruiz, and Alain de Serres for valuable comments and suggestions.

housing prices may signal a high GDP growth up until a given threshold, past which the bubble bursts and the economy decelerates brutally. The economy behaves as it is theoretically supposed to until it stops doing so. The 2008 crisis has called for a renewal of the modelling techniques (Blanchard, 2014; Romer, 2016). Complexity is related to the emergence of crises, and that is why these non-linearities and structural breaks may be considered to be “where the danger lurks”. We suggest that machine learning, as a theory-agnostic methodology is particularly apt at modelling the economy, and in particular should perform better than linear models around turning points and recessions. The paper aims to demonstrate this point.

The more complex an algorithm, the more accurately it can fit a complex reality. But the more complex it is, the less intelligible. There is a trade-off between accuracy and interpretability, often referred to as Occam’s dilemma. In this trade-off, we tend to think that accuracy comes first. Because being able to clearly explain why or how a model works is pointless when the model does not fit the data adequately. The interpretation of the model is relevant only when the model properly fits the phenomenon it is meant to explain. That said, we still need to shed light on the predictive mechanisms and understand the reasons why the algorithm predicts what it predicts. The accuracy-interpretability trade-off can be seen as a continuum, between very interpretable methods (linear models) and very accurate ‘black-box’ methods, such as Support Vector Machines. We believe that tree-based approaches strike a good balance between accuracy and interpretability. This class of algorithms seems to be particularly well-suited to the needs of economists, given that regression trees can capture non-linearities and remain easy to interpret.

The paper introduces an original forecasting method specifically tailored to deal with non-linearities and structural change: Adaptive Trees. This algorithm is based on both existing machine learning methods and an original contribution to the field. The Adaptive Trees algorithm is based upon Gradient Boosted Trees, a widely used machine learning algorithm, which captures non-linear patterns very well. We have modified its functioning in order to tackle structural change. This innovative algorithm is very well-suited to economic forecasting because of its adaptive nature. Our algorithm can be said to be “adaptive” insofar as it adapts to the quantity of structural change it detects in the economy. The more structural changes there are (for instance around major turning points), the more it will focus on the most recent data.

Forecasting is a good exercise for a proof of concept because performance can be measured. We applied Adaptive Trees to the forecasting of GDP growth in G6 countries (US, UK, France, Germany, Japan and Italy). We produced forecast simulations using the same sets of leading indicators as the OECD Indicator Models, a series of model-based forecasts of GDP growth in major economies. The Indicator Models use linear modelling (Vector Autoregressive Models). We build a forecast using the exact same data in order to provide a fair assessment of the characteristics of the new method. The Indicator Models are short-term forecasts. In order to assess the long-term performance of Adaptive Trees, we compare their predictions with the forecast made in the OECD Economic Outlook, which is based on country expertise.

There have been a series of attempts to apply machine learning to macroeconomic forecasting (Biau and D’Elia, 2009; Chakraborty and Joseph, 2017; Tiffin, 2016). Most experiments have applied off-the-shelf machine learning algorithms to economic data and obtained quite good results. Our algorithmic approach was tailored to address the specific challenges of macroeconomic forecasting. Our research not only imported machine learning techniques in economic forecasting but also carved out a new machine learning method in order to meet the needs of economic forecasting.[2]

The paper is organized as follows: we present the data (3), the method (4), the results (5) and then conclude.

2. Main findings

- Adaptive Trees is an innovative algorithm based on regression trees that provides an efficient solution to model non-linearities and structural change in the economy.
- G7 GDP growth forecasts made with Adaptive Trees perform well, displaying high levels of accuracy, directional accuracy and turning point accuracy for all 6 countries at both short (3, 6 months ahead) and long term (9, 12 months ahead).
- At M+6, Adaptive Trees are on average across countries 23% more accurate and 32% more directionally accurate than the Indicator Models.
- The Adaptive Trees forecast at M+6 detect 50% of turning points on average against 25% for the Indicator Models, at the expense of a slightly higher false alert rate.
- At M+12, accuracy and directional accuracy of the Adaptive Trees forecast is comparable to that of the Economic Outlook, with sizeable disparities between countries. Overall, Adaptive Trees' capacity to predict turning points is 74% higher than the Economic Outlook's.
- Forecasts made with Adaptive Trees are 48% less skewed than forecasts made with Vector Autoregressive Models (VAR) at M+6, indicating better performance around recessions.
- A prediction made with Adaptive Trees can be additively decomposed into variable contributions, thus ensuring model interpretability. This feature is not shared by VAR, widely used for macroeconomic forecasts.

[2] It is worth mentioning that there are other research projects meant to introduce machine learning in economics, mostly dealing the with analysis of the heterogeneous impact of policies (Athey and Imbens, 2016; Chernozhukov et al., 2016; Wager and Athey, 2015 among others).

Box 1. Machine learning and econometrics

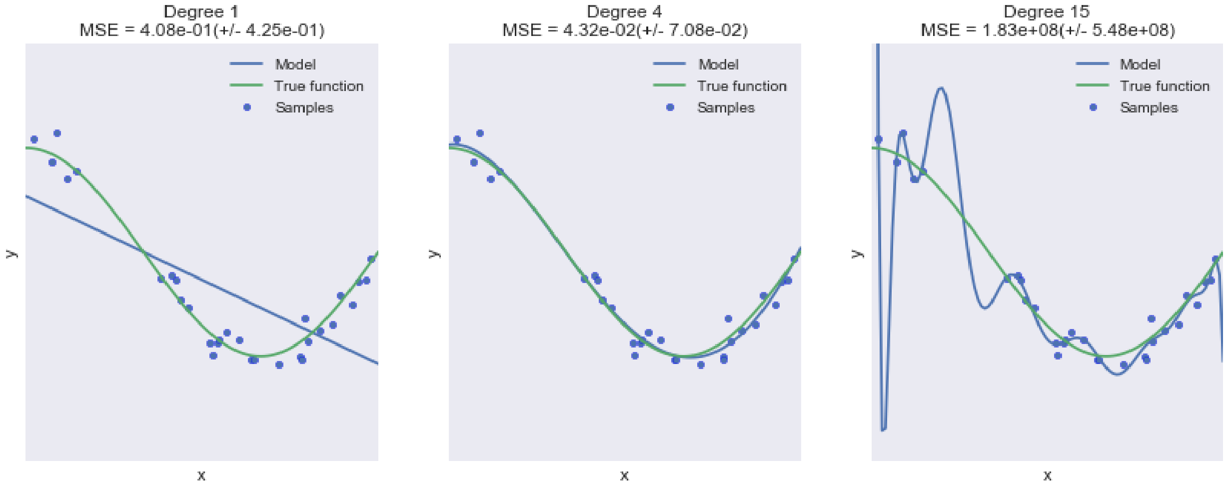
Even though econometrics and machine learning overlap (especially around non-linear methods), the dividing line between the two fields may stem from their distinct objectives. Machine learning differs from linear or non-linear econometrics mostly in that it focuses on maximising out-of-sample accuracy in the prediction of the target variable ($\hat{y}$), whereas econometrics revolves around unbiased estimators of regression coefficients ($\hat{\beta}$). Machine learning relies on techniques meant to prevent overfitting and adjust the level of model complexity in order to maximise out-of-sample goodness-of-fit. Such techniques include cross-validation. Machine learning thus consists in a broader class of estimators than non-linear econometrics, and includes slightly biased estimators that have lower variance and thus better generalisability. While aiming at minimising out-of-sample error, usually measured using a loss function (such as mean square error), machine learning relies heavily on numerical optimization techniques (including gradient descent).

Bayesian econometrics appears to be closer to machine learning in that it relies on computational and algorithmic methods (MCMC) and handles non-linear regressions. Still, Bayesian econometrics supposes that observations are drawn from random variables, and formulates beliefs about their distribution. Model validation is made using the likelihood of a model given the data. Machine learning does not suppose a probabilistic world. Model validation relies only on predictive accuracy. Machine learning thus differs from Bayesian econometrics insofar as it uses algorithmic modelling without the need for supposing the existence of a probabilistic model that generates the data. Both the purely algorithmic approach and focus on out-of-sample accuracy make machine learning a promising pool of resources for economic forecasters.

As machine learning focuses on predictive performance and out-of-sample goodness-of-fit, the bias-variance trade-off is a major issue. In a regression, the out-of-sample mean square error can be decomposed into bias, variance and noise. Bias corresponds to the difference between the average of the predictions and the best possible estimator for the problem. Variance refers to variability of the predictions of the estimator when fit over different samples. And the noise is the irreducible part of the error, due to noise in the data.

There is a trade-off between bias and variance as very simple models will have low variance and high bias (underfitting), whereas very complex models may have a high variance and a low bias (overfitting). In Figure 1, we fit a sinusoid with a polynom. On the left panel, a degree 1 polynom underfits the data. On the right panel, a degree 15 polynom overfits the data. On the centre panel, the degree 4 polynom provides a good fit and low mean square error.

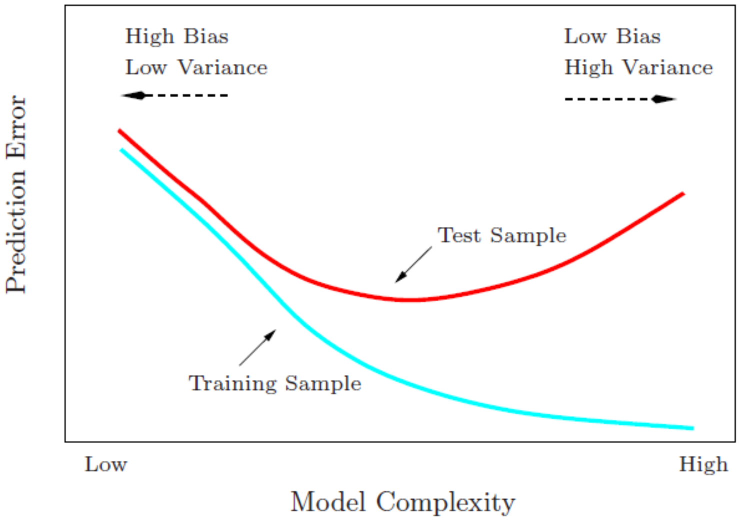
Figure 1. Underfitting, right fit, overfitting



Source: Pedregosa et al., 2011

As shown on Figure 2, the more complex a model, the better the in-sample goodness-of-fit. Past a certain threshold, out-of-sample goodness-of-fit starts decreasing due to overfitting. In order to find the right degree of model complexity and minimise out-of-sample error, we resort to cross-validation.

Figure 2. Bias-variance trade-off



Source: J. Friedman et al., 2001

Cross validation (Efron, 1983; Schneider, 1997) is a model evaluation method that is often used in predictive settings. Some of the data is removed before the training begins. When training is done, the data that was removed can be used to test the performance of the learned model on “new" data. Cross-validation is seen as a watchdog against over-fitting. It is an out-of-sample goodness-of-fit evaluation method.

3. Data sources and benchmarks

The paper deals with forecasting GDP growth in all G7 countries but Canada.[3] For each country, we replicated the setting of the OECD Indicator Models (Ollivaud et al., 2016; Sédillot and Pain, 2003) and used a database including a series of leading indicators, both hard and soft.

Table 1. Variables used by the Indicator Models

Italy	USA	France
Industrial production	Industrial Production	Industrial production
Car registrations	Consumption	Household consumption
PMI (manufacturing)	Employment	Output trend
Houshold confidence	Construction	Business survey
PMI (services)	Inventories	Order book and demand
	Exports	Household confidence
	PMI	
	Housing permits	
	Housing prices	

UK	Germany	Japan
Industrial production	Industrial production	Industrial production
Retail sales	Business surveys expectations	Inventory ratio
Housing prices	Exports	Living expenditure
Business confidence	Manufacturing orders	Job offers to applicants ratio
Economic sentiment indicator	Business survey	Small business sentiment sales
PMI	PMI (manufacturing)	Business sentiment financial position
	PMI (services)	Tankan
	Consumer confidence	PMI
	Vacancies	

All base variables are monthly series, whereas the GDP growth is quarterly. In each country, the target variable is the growth (Q/Q) of the GDP in volume. All variables (including the GDP) come with release delays that were carefully taken into account during the simulations.

The OECD Indicator Models (see Annex A) are a series of short-term forecasts for major economies based upon leading indicators and using VAR. It builds on the work of Sédillot and Pain (2003) and Mourougane (2006) in using short term economic indicators to predict quarterly movements in GDP by efficiently exploiting all available monthly and quarterly information. These models typically combine information from both "soft" indicators, such as business sentiment and consumer surveys, and "hard" indicators, such as industrial production, retail sales, house prices etc. and use is made of different frequencies of data. The forecast was made using the exact same data as the Indicator Models.

The forecast is made in pseudo-real time. For each quarter Q in 2007-2017, a real-time simulation would produce a forecast as if we were standing in Q, using only the data that

[3] Because Canada is a special case (see Mourougane, 2006).

was available to forecasters in Q. The forecast simulation is made in pseudo-real time, as opposed to actual real time, insofar as we use time series that were revised by statistical agencies, not the vintage series that were available in Q that often include measurement errors. In order to ensure comparability, the benchmark Indicator Model was also simulated in pseudo-real time.

The OECD Economic Outlook (see Annex B) forecast is used as a benchmark to assess the performance of the Adaptive Trees forecast on the medium-term (one year ahead). The Economic Outlook projections are produced twice yearly by the OECD Economics Department to provide a consistent view of the world economy, with a specific focus on recent and future macroeconomic developments in current and prospective OECD member countries and the larger non-OECD economies.

The Economic Outlook projections are produced by OECD country experts interacting with topic specialists, taking into account current and prospective developments, officially mandated policies, historical relationships between key variables and new information and indicators related to domestic and global conditions. The forecast process also benefits from detailed consultation and peer review from government economists and policy makers in member and non-member countries and other key international organisations (the International Monetary Fund, World Bank, European Commission, European Central Bank, the Bank for International Settlement).

We thus build upon the variable choices made by the Indicator Models research team, and experiment with a new forecasting algorithm. Comparing forecasts is not straightforward as forecast performance is multi-dimensional. Still, comparing forecast simulations will convey a notion of the characteristics of Adaptive Trees and its functioning. A forecast's best feature, though, is consistency through time. Pseudo-real time simulations can go some way in proving that a method is relevant, but not as far as a forecast track record. Comparing a new algorithm ex post with an existing method that has been in production for years is always tricky, and such a comparison needs to be taken with caution.

4. Method

We developed an original algorithmic approach that draws on latest machine learning research and is an original contribution to the field. The methodology is tailored for economics, and aims at tackling three main challenges in economic forecasting: non-linearities (1), structural change (2) and high dimensionality (3).

1.4.1. A tree-based approach to tackle non-linearities

Given the Xs (the variables), one needs to estimate the form of the relation f between X and Y (the target, in our case, the GDP growth):

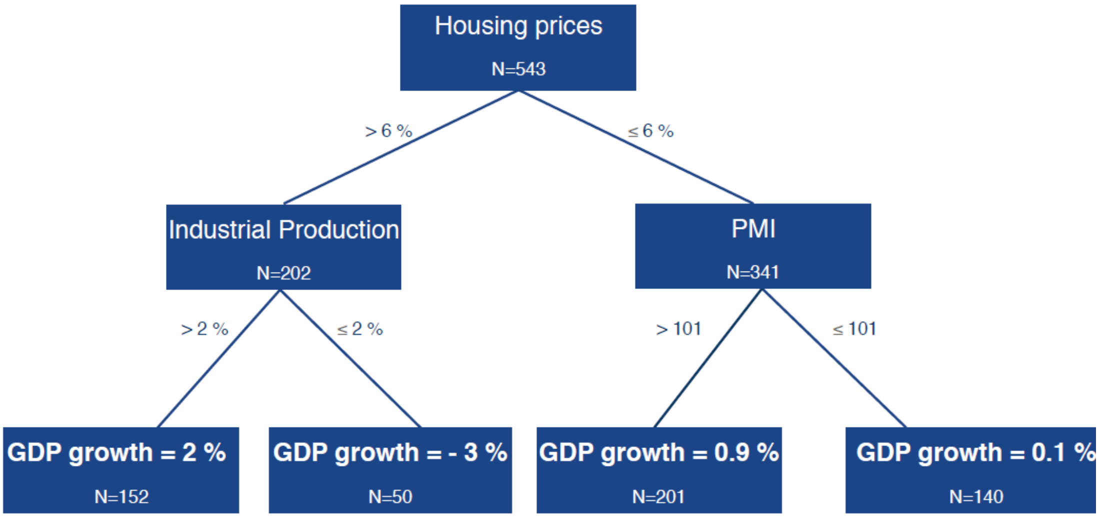
$$Y_{t+1} = f(X_t) + \varepsilon$$

Most models suppose that f is linear, and that the residual ε abides by a certain number of regularity assumptions (null on average, Gaussian, homoscedastic…). Such assumptions may seem too restrictive with regard to economic intuition. A given variable's impact may depend upon its own value (threshold effects) or upon the value of a set of other variables (interactions). For instance, housing price increases growth may signal wealth effects up until a certain threshold, past which they only reflect a housing bubble. The economy is characterized by complex patterns (if… then…) that linear models cannot capture, given that a prediction made with a linear model comes down to a weighted average of its variables.

Tree-based approaches are well suited to dealing with non-linearities and heterogeneity. Regression trees predict the value of a target variable by learning simple "if-then" decision rules from the data.

Figure 5 shows how a regression tree is trained. Regression trees recursively divide the sample of observations into sub-groups so that the within-group variance of the growth of GDP is as low as possible. At first, the algorithm selects the splitting variable (housing prices growth rate) and the splitting point (6%) that best divides the observations (in our case, quarters), i.e. that minimises the variance of the target variable (GDP growth) in the two resulting sub-groups. It iterates this procedure at each node until reaching leaves containing a given minimum number of observations.

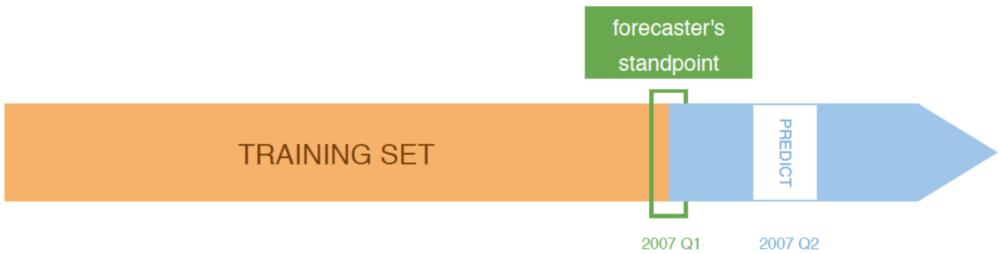
Figure 3. Training a regression tree



Note: At first, the algorithm splits observations between quarters where the growth of housing prices is superior to 6% and observations where it is not. In the first group (left), the algorithms picks industrial production to make a split, because among all variables it is the one that achieves the best split in that group. Among observations where housing price increases were inferior to 6%, observations are split on PMI. In the end, we reach final leaves that are supposed to be homogeneous in terms of the growth of GDP, where the number of observations is sufficient to avoid overfitting.

The training of the regression tree is made using past data (training set in Figure 4). The training phase consists in growing a tree. The next phase is prediction and uses contemporaneous data.

Figure 4. Train and predict

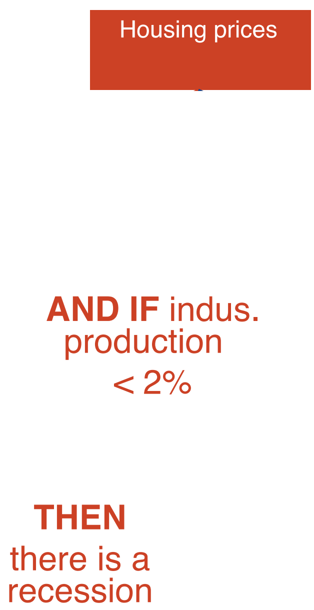


Note: The forecaster who stands in 2007 Q1 uses all past data to train the algorithm, and applies the algorithm to the current situation in order to make a prediction one quarter ahead.

A prediction is made following a path in the tree and predicting the average target value of the past observations that fell in the same leaf (see Figure 5).

Figure 5. Prediction with a regression tree

Type the subtitle here. If you do not need a subtitle, please delete this line.



Intuitively, regression trees may capture multiple interactions and threshold effects. Whereas predictions made with linear regressions are a weighted mean of the covariates (weighted by the regression coefficients), a regression tree introduces a logical structure (if…, and if…, then…). This structure may capture complex non-linear patterns that are to be found in the economy. In this simplistic example, the tree captures a housing price bubble.

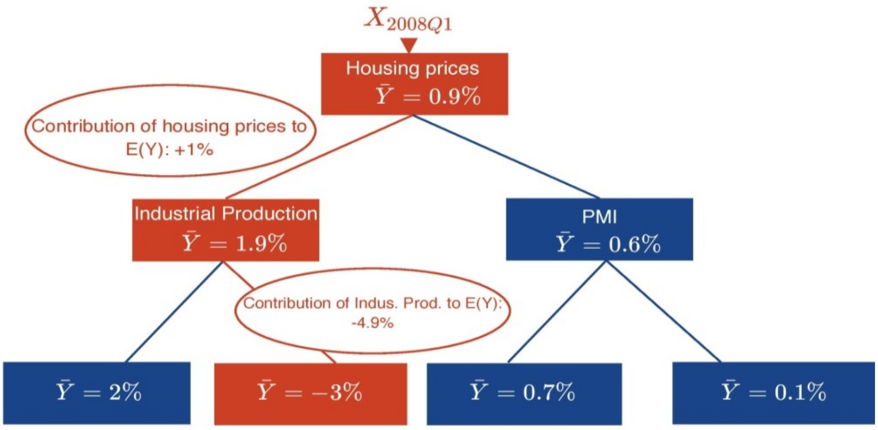
To some extent, regression trees resemble local regressions. A regression tree fits a very simple linear model ($y = \beta + \varepsilon$) in each leaf, as local regressions fit linear models in sub-areas of the data. There are two main differences, though. First, regression trees do not resort to measures of distance, and do not necessarily use all variables, whereas local regressions tend to have bad performance in high dimension. The second difference relates to how parameters are chosen, especially tree depth. A regression tree can be arbitrary deep. The depth of the tree on Figure 5 is equal to 2 as there are two series of splits. The deeper, the more splitting variables and splitting points come into play, the less observations per leaf. Deeper trees are more likely to overfit. In turn, overly shallow trees are likely to miss important patterns in the data. The tree depth is a parameter that is to be optimised using cross-validation. One usually trains a series of regression trees with depth between 2 and 10, and selects the depth that minimises out-of-sample error.

Regression trees can be unstable because small variations in the data might result in a completely different tree being generated. This problem is mitigated by using decision trees within an ensemble, which combines multiple trees by averaging their predictions. We have tested a series of tree-based ensemble methods, including Random Forests and Gradient Boosted Trees (GBT, see Box 2). The latter proved to perform well. Our Adaptive Trees are a transformation of Gradient Boosted Trees meant to address structural change (see below).

One of the main advantages of tree-based methods, including GBT, is their interpretability. Even though ensemble methods such as Gradient Boosted Trees involve a large number of trees, one may compute each variable's contribution to a given prediction. To see how, one simply needs to compute the average of the target variable (the GDP growth) in intermediary nodes (see Figure 6). Let's follow a given observation along the prediction path. At the origin of the tree, the prediction $\hat{y}$ is equal to the population mean, 0.9%. At the next step, the housing prices shunt towards the sub-region where the target mean is equal to 1.9%. The contribution of housing prices so far is thus 1.9% - 0.9% = 1%. And so on. In the end, the final prediction in the final leaf is the sum of the contributions of all the variables that intervened along the prediction path plus the bias, i.e. the population average:

$$\hat{Y} = 0.9\% + \sum Feature\ Contributions$$

Figure 6. Variable contribution, a simple example



Box 2. Gradient Boosted Trees (GBT)

The Gradient Boosted Trees algorithm (Freund et al., 1996) uses boosting, an iterative procedure. In the first step, we train a weak predictor (in our case, a regression tree) on equally weighted observations. At the second step, we train a new regression tree on the residual from the first one and update observation weights in order to give more weight to observations that were inaccurately predicted. We iterate this procedure N times (100, 500, 1000 times…). Observations that are difficult to predict thus receive ever-increasing influence. The final predictor is an average of all regression trees weighted according to their overall accuracy. There are many ways to perform boosting. Most recent techniques use gradient descent to perform numerical optimization.

Here is a mathematical formulation of Gradient Boosted Trees. Gradient Boosted Regression Trees considers additive models of the following form:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x)$$

Where h_m (x) are the simple regression trees (weak predictors). Gradient Boosted Trees builds the additive model in a forward stage-wise fashion:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$h_m(x)$ is a regression tree grown in order to minimize a given loss function, usually, least squares, and γ_m is calculated using numerical optimization.

Two extra variables improve the functioning of GBT: shrinkage (J. H. Friedman, 2001) and subsampling (J. H. Friedman, 2002). Shrinkage is a simple regularization strategy that scales the contribution of each weak learner by a factor ν:

$$F_m(x) = F_{m-1}(x) + \nu\gamma_m h_m(x)$$

The parameter ν is also called the learning rate. It helps reduce the risk of overfitting by reducing the impact of each extra weak predictor. When ν is very low, it takes more time and more predictors to reach sufficient accuracy. When ν is high, the risk of overfitting becomes more important. It is up to the user to choose ν. We discuss optimization below.

Subsampling consists in selecting only a random subsample of available observations when growing each weak predictor h_m (x). Gradient Boosting combined with subsampling becomes Stochastic Gradient Boosting (J. H. Friedman, 2002). At each iteration, a given fraction η of all available data is drawn at random. This randomly selected subsample is used instead of the full sample to fit the weak learner. Introducing some randomness has proved to improve the overall quality of the algorithm. η is another parameter to be optimized by the user.

1.4.2. From regression trees to adaptive trees: dealing with structural change

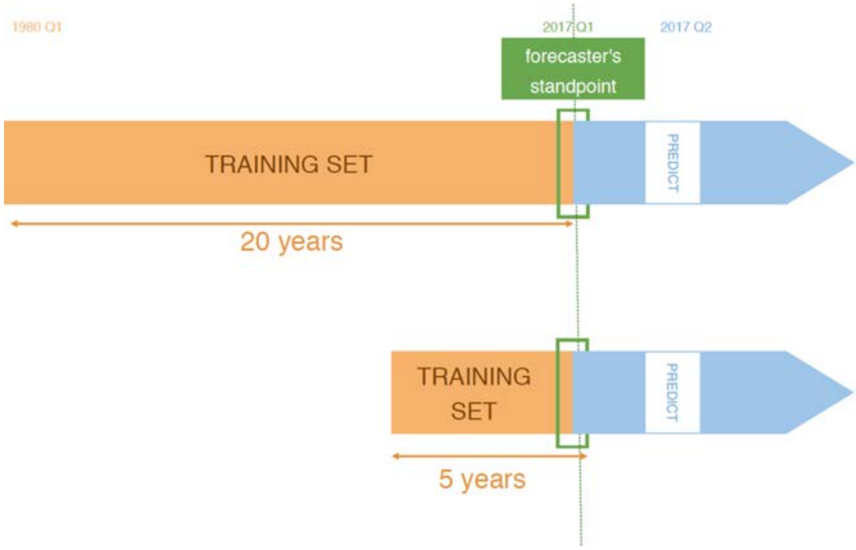
The economy is no stable system. It is ever-changing, and relations that held yesterday may suddenly cease to exist. Most forecasting methods train on a large time window and assume that the economy abides by stable rules. However structural breaks do occur often around crises and change the nature of the relations between the covariates and the target variable. Rigid linear models whose coefficients are supposed to endure along extended periods fail to capture structural changes unless the latter are explicitly introduced in the model. There can be sudden structural breaks, or long-standing structural change, caused for instance by the diffusion of some new house building technology that would change the elasticity of housing prices to wealth.

Structural change is a problem known as “concept drift”[4] in the machine learning literature (see Gama et al., 2014 or Žliobaitė, 2010 for a review). Concept drift refers to the idea that the distribution of data (the distribution of the target Y, the distribution of the variables X, and the joint distribution (X,Y)) may change over time, be it suddenly, incrementally, or through reoccurring contexts.

Figure 7 describes an experiment that sheds light on the importance of structural change to economic forecasts. Using a small training window (bottom panel) yields better results than using a large training window (top panel), even though a larger training window benefits from using more data. The gain from using more data is offset by the gain from using more recent data. The simple experiment described by Figure 7 yields a key insight: the most recent past is more informative about the near future than the more distant past. It follows that there is a trade-off between using only very recent data, and using a large training set.

[4] In the context of machine learning, “concept” refers to the mechanism acting as the source of the data (in our case, the economy).

Figure 7. Choosing the size of the training sample



However, a forecast that would only use very recent history as a guide might be short-sighted. Recognizing a pattern from a distant past may improve accuracy in the case where such pattern would reoccur. For instance, one may want the forecast algorithm to detect a housing bubble had one already occurred in the series 20 years earlier. Training the forecast algorithm on the past five years only would considerably reduce the scope of events the algorithm learns from. Using a small training window is thus a strategy that yields better results than using a large training window but also has major caveats.

In order to overcome this trade-off, we developed an original technique that we named "Adaptive Trees". Adaptive Trees are a transformation of Gradient Boosted Trees. Adaptive Trees adjust to structural change by giving more weight to the recent past when more remote past becomes less informative about the future. Instead of initialising the observation weights to be equal to each other, we gave greater weights to more recent observations. Subsequently, when structural breaks arise, the newest observations that are already heavily weighted will receive ever-increasing weights along the boosting steps given that they will be inaccurately predicted. The adaptive algorithm will thus give a decisively higher importance to newest observations as soon as their distribution differs, signalling concept drift. Adaptive Trees will therefore rely more heavily on the recent past when structural change is at play, while also exploiting information from a more remote past.

1.4.3. Data-driven model selection to address high dimensionality

Dimensionality refers to the large numbers of variables that describe a given phenomenon. The increasing difficulty of modelling phenomena happening in high dimension is commonly referred to as the curse of dimensionality (Bellman, 1957). The more variables there are, the more observations one needs to properly filter out noise variables and train a model. With a fixed number of training samples, the predictive power reduces as the dimensionality increases. High dimension increases the risk of overfitting: in high dimension, it is increasingly difficult to distinguish characteristics of the sample from characteristics of the population.

Macroeconomic data are often characterized by both a high dimension and small samples. The economy is an open and complex system, with endogenous variables and exogenous shocks both at the macro- and micro- levels, from a variety of dimensions, including economic, financial, political, regulatory, technological or natural. Time adds to the dimensionality problem. A given variable, say the number of car registration licences, may be taken contemporaneously, or with lags. There is no telling a priori whether one should look at its contemporaneous values or lagged values, or even maybe complex interactions of the two. The high number of candidate variables along with the multiplicity of lags that one may want to take into account makes it impossible to use all possible variables for a prediction. Failing that, we resort to variable selection methods.

Econometricians would rely on statistical tests, using the hypothesis as to the distribution of observations. The p-values are classically used to tell apart characteristics of the sample from characteristics of the population and ensure generalisability. Significance tests depend on the number of observations in the sample and assume the population is infinite. The machine learning approach is reluctant to making distributional hypotheses and rather resorts to criterion relating to out-of-sample accuracy. Ideally, one would make predictions with all possible subsets of variables and retain the one that performs best. Unfortunately, this method is computationally unfeasible. The machine learning literature contains a very wide array of alternative variable selection techniques. We chose to use Recursive Variable Elimination based upon Variable Importance. In what follows we explain why and how it works.

Let’s first explain variable importance. In a simple regression tree model, a variable is all the more important the higher it is in the tree, and close to the root. During the growing of a tree, the earlier a variable is used as a splitting variable, the more it will impact the shape of the tree. In an ensemble of trees, variable importance measures the likelihood of a given variable being important in each single tree. Variable importance scores are easy to compute. They may be used to perform variable selection. One may want to keep only the n most important variables, in order to remove variables that are not informative enough.

Using variable importance threshold (i.e. keeping only the n most important variables) could be an easy way to proceed, but this method has limitations. Removing the least important 250 in 300 variables would exclude irrelevant variables, but may also remove good predictors that would have collinear counterparts in the remaining 50. Say variable A is a good predictor, and B a noisy version of A:

$$B = A + \varepsilon$$

A and B being substitutes, including B will decrease the variable importance of A. Once B is used as a splitting variable in a regression tree, A becomes useless. Subsequently, selecting only the n most important variables at once will likely remove many good predictors only because they have collinear counterparts in the sample.

Recursive Variable Elimination is meant to prevent this problem. RFE works by creating predictive models, computing variable importance scores, and removing variables with the smallest scores, then repeating the process until a desired number of variables are left.

Using RFE instead of one-shot variable elimination is likely to prevent removing good predictors that are collinear with other variables in the sample. Going back to our previous example, even though A and B will be both underappreciated by the variable importance score, it remains likely that A will rank slightly better than B. Therefore, B will be removed at some stage, thus restoring A’s importance. Moreover, given that variable importance is calculated using an algorithm that captures multiple interactions, the second requisite is satisfied as well.

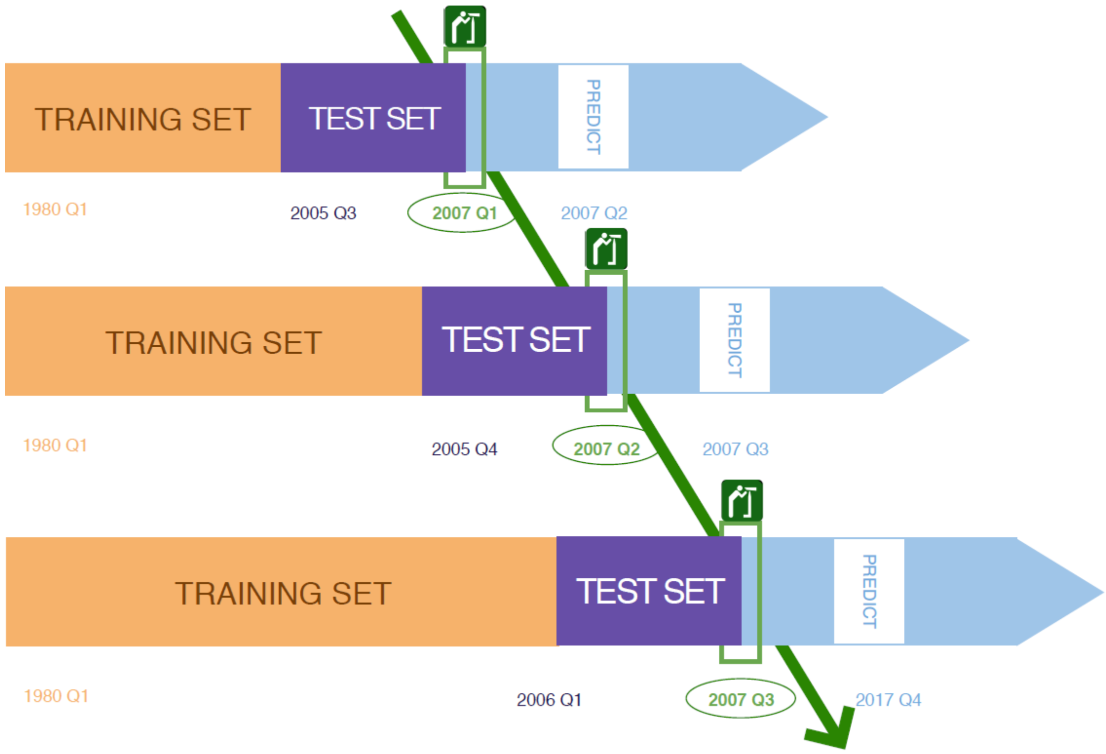
Recursive Variable Elimination provides a ranking of variables. Knowing this ranking does not provide a solution to the variable selection problem. Still, it makes it much easier to solve. One no longer has to test all possible variable subsets, but must simply test subsets along the ranking path: using 10 most relevant variables, 11 most relevant variables, 12 most relevant variables… The number of candidate subsets is reduced to 20, as we consider that we need to use between 10 and 30 variables. Defining the best variable subset from that ranking is done using cross-validation. We perform cross-validation on the test set (see Figure 8). Ultimately, we select the subset of variables that provides the best prediction of the observations in the test set.

Figure 8. Cross-validation

5. Results

In order to assess the performance of the Adaptive Trees forecast, we devised an incremental learning framework that allows us to successively apply the forecast in T for each T in 2007Q1-2017Q1. This framework makes simulations in pseudo-real time: as if we were standing in T, except for the fact that we are using updated series. For each country, we apply the simulation framework. At each T, the algorithm performs some optimization, trains, and makes a prediction. We take care not to use the future when predicting the future, be it during the training or the optimization of algorithm parameters.

Figure 9. Incremental learning framework for pseudo-real time simulations



In the following paragraphs, we present the forecast results and performance for all G7 countries but Canada, at four time horizons: three months ahead (M+3), six months ahead (M+6), nine months ahead (M+9), and twelve months ahead (M+12). We also provide the performance of the OECD Indicator Model for M+3 and M+6, and with the OECD Economic Outlook forecast at M+12 as benchmarks.

We measure forecast performance using three key performance indicators: Root Mean Square Error (RMSE), Forecast Directional Accuracy (FDA) and Turning Point Accuracy (TPA). FDA measures the proportion of forecasts that were in the right direction (up/down) over the simulation period. Turning Point Accuracy measures the proportion of turning points that were accurately predicted (true positives). Performance in terms of true negatives (accurately predicting there will be no turning point: a false alert rate) are shown as well. In order to compare our forecast to the Indicator Model forecast, we display improvement rates for each of the three performance measures. Improvement rates are computed as follows:

$$Gain\ from\ IM = \frac{Perf_{Adaptive\ Trees} - Perf_{Indicator\ Model}}{Perf_{Indicator\ Model}}$$

1.5.1. Performance

The forecasts made with Adaptive Trees generally perform well and compare well to benchmark forecasts. We display key information below and full tables of forecast performance in Annex 5.

Figure 10 displays the forecasts made with Adaptive Trees and by the Indicator Model for the USA between 2007 and 2017, compared to the actual value of the GDP growth. Table 2 shows the performance statistics. The Adaptive Trees forecast has a very good accuracy at all three forecast horizons (3, 6, 9 and 12 months ahead). At M+6, turning point accuracy is 2.3 times that of the Indicator Model. It is worth stressing that the forecast performance slightly decreases as the forecast horizon widens, but Adaptive Trees retain good performance at one year ahead. At M+12, the forecast compares favourably to the Economic Outlook forecast. It is 1.5 times more accurate in forecasting turning points, and has a lower false alert rate.

Figure 10. Forecast of USA GDP growth at M+3

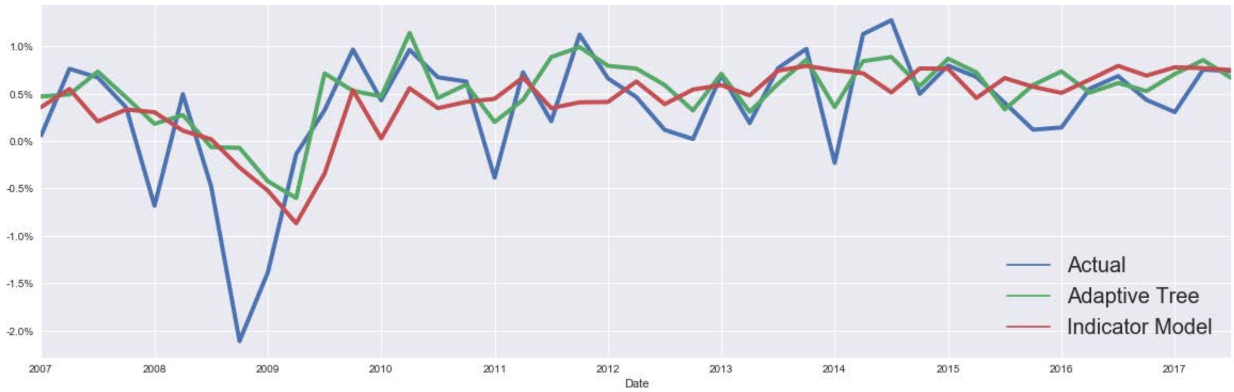


Table 2. Performance of Adaptive Trees for the USA forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.48*	*79%*	*67%*	*53%*
	Indicator Model	0.52	60%	42%	47%
M+6	Adaptive Trees	*0.51*	*71%*	*58%*	*32%*
	Indicator Model	0.64	48%	25%	47%
M+9	Adaptive Trees	*0.54*	*67%*	*50%*	*32%*
M+12	Adaptive Trees	*0.66*	*60%*	*43%*	*43%*
	Economic Outlook	0.62	60%	29%	57%

Note: Performance statistics are computed as an average across all quarters between 2007 and 2017 for the M+3, M+6 and M+9 forecasts. The M+12 forecast is a special case. In order to ensure comparability with the twice yearly EO forecast, performance statistics were computed as an average over Q2 and Q4 of each year between 2007 and 2017.

Figure 11 and Table 3 display plots and performance statistics regarding the Japan forecast. The Japanese forecast is the least accurate of all G7 countries. That is mostly due to wide variations in the GDP growth, linked to the three negative blips after 2010 (in 2010, 2012 and 2014). The Adaptive Trees forecasts accurately predicts two out the three blips (in 2012 and 2014) and captures with a lag the 2010 recession. The three upturns are very well anticipated, although once again with a small lag in 2011. From this point of view, Adaptive Trees would have made a good warning system for the current quarter, even though there is a false alert in 2012Q3 (where the forecast falsely indicates a negative growth). The Japanese forecast at M+12 is satisfying. It accurately predicts the 2010-2011 and 2014 downturns.

Figure 11. Forecast of Japan GDP growth at M+3

Type the subtitle here. If you do not need a subtitle, please delete this line.

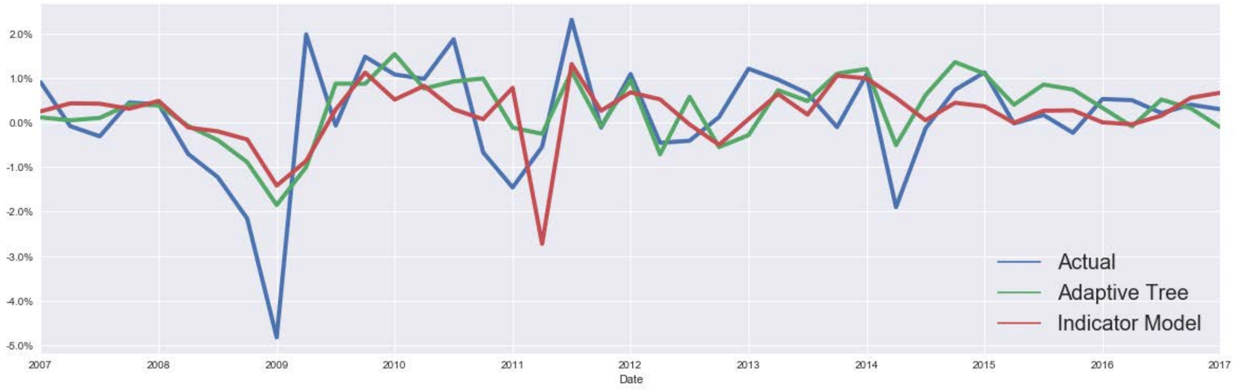


Table 3. Performance of Adaptive Trees for the Japan forecast (2007-2017)

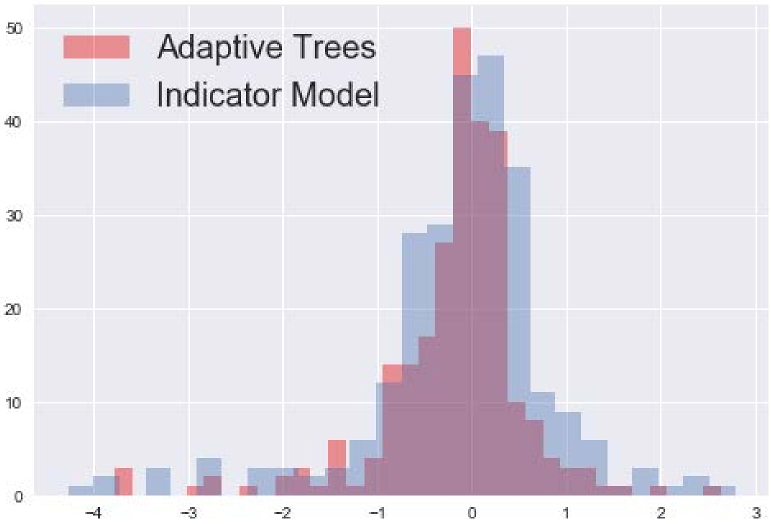
		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.99*	*68%*	*44%*	*38%*
	Indicator Model	1.13	53%	24%	31%
M+6	Adaptive Trees	*0.96*	*66%*	*58%*	*50%*
	Indicator Model	1.35	46%	19%	63%
M+9	Adaptive Trees	*1.04*	*76%*	*65%*	*50%*
M+12	Adaptive Trees	*1.01*	*55%*	*38%*	*38%*
	Economic Outlook	1.15	45%	23%	50%

On average across G6 countries (see Annex 5, Table 7), Adaptive Trees forecasts show better accuracy and directional accuracy than the Indicator Model. For instance, at M+6, Adaptive Trees forecasts are on average 22.5 % more accurate than the Indicator Models, and 32 % more directionally accurate. They are on average twice as capable of accurately predicting turning points, and the false alert rate is lower by 26 %. It is worth noting that at M+12, directional accuracy remains higher than 50% and higher than the Indicator Model's at M+3. It is equal to 63 % on average and equal to 55 % across the period at worst (in France and Japan).

None of the leading indicators we used allow for predicting the financial crisis of 2007. Still, the adaptive nature of our forecasts explains their much better performance during the crisis and around the upturn, especially in the case of the UK (see plots in Annex 4).

Performance during crises can be studied more consistently with Figure 12 that displays histograms of forecast errors at M+6. A forecast error is defined as the predicted value minus the actual values. Forecast errors are standardised: we divide them by the standard deviation of the GDP growth in each country in order to ensure cross-country comparability. Standardised forecast errors of G6 countries for M+6 forecasts are stacked and displayed in Figure 12. Usually, economic forecast errors have a fat left tail because crises are inaccurately predicted. Both forecasts do, except that the Adaptive Trees have a much thinner left tail. The thickness of the left tail is measured by the skewness of the distribution. A large negative skewness signals a fat left tail. The error of the Indicator Model has a skewness equal to -1.47, whereas the Adaptive Trees forecast error skewness is equal to -0.99.

Figure 12. Histograms of standardised forecast error for all G6 countries, M+6 forecast



1.5.2. Interpretation

As explained in section 1.4.1, one may easily compute variable contributions from any ensemble of trees, including Adaptive Trees, in order to explain why the algorithm predicts what it predicts.

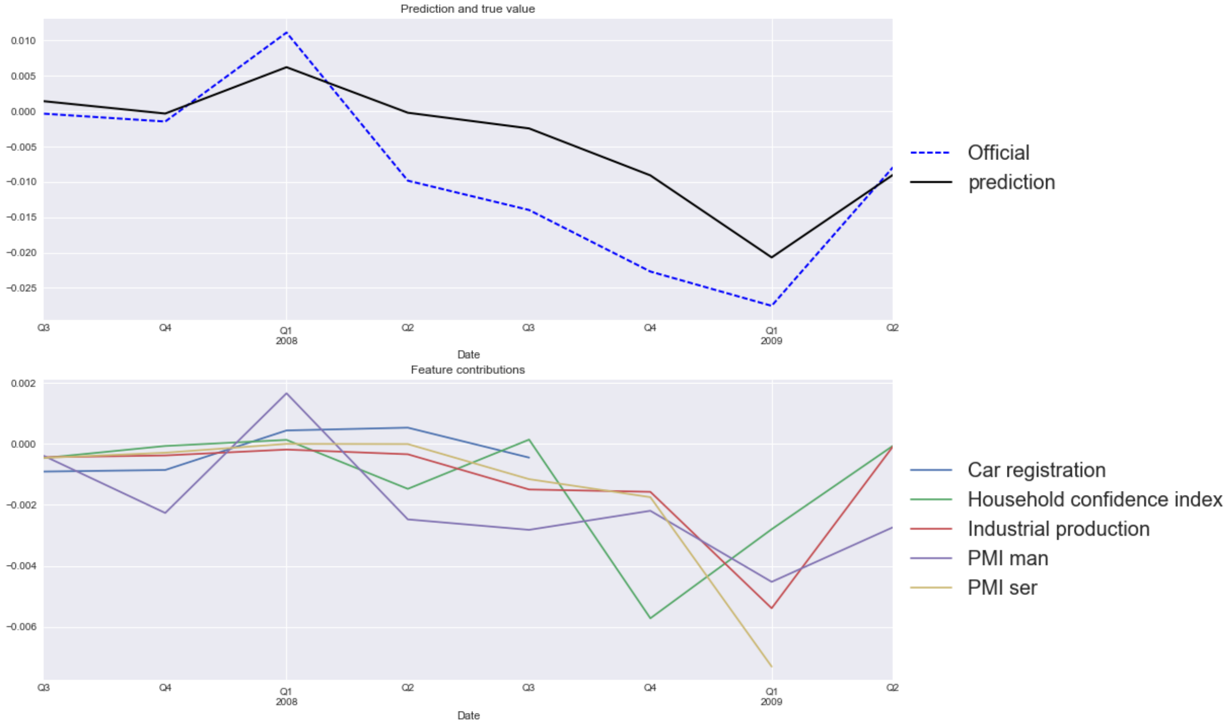
We are interested in knowing which variables most contribute to a prediction. As explained in 1.4.1, a prediction made with Adaptive Trees can be additively decomposed into a sample average and each variable's contribution (that explain the deviation from the sample average). Using variable contributions helps understand the patterns identified by the algorithm and the reasons why it predicts what it predicts. Decomposition into variable contributions provides a window of interpretability on tree-based predictions. Decomposing the prediction into variable contributions is a feature of tree-based approaches that is not shared by Vector Autoregressive models (VAR) used for the Indicator Models.

As the algorithm uses a multiplicity of variables (each variable being introduced at multiple lags), we simplify variable contributions by computing aggregated variable contributions, whereby the contribution of, say, housing prices, will be equal to the sum of the contributions of housing prices at lags M-2, M-3, M-4, … Working with aggregated variable contributions enhances readability.

Figure 13 displays the decomposition of the prediction for the Italy M+3 forecast in aggregate variable contributions during the crisis. It is interesting to note that the first alert comes from the manufacturing PMI, closely followed by the household confidence index. Soft indicators may be better early warnings as they come with smaller release delays.

Figure 13. Decomposition of the Italy M+3 forecast in aggregate variable contributions, 2007-2009

Type the subtitle here. If you do not need a subtitle, please delete this line.

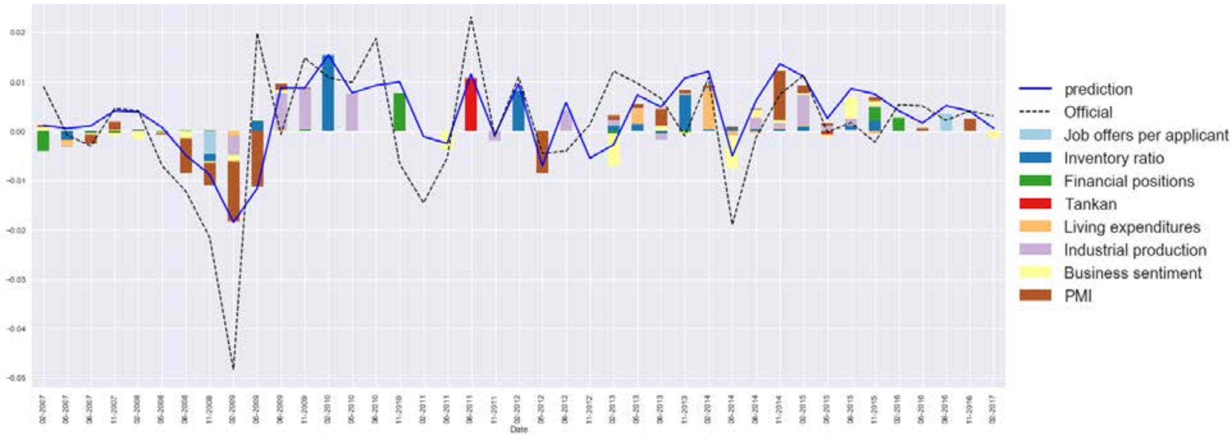


Note: On the top pane, the black line is the prediction made by the Adaptive Trees. It can be decomposed into aggregated variable contributions, as shown on the bottom pane.

Figure 14 displays variable contributions to the Japan M+3 forecast on a larger period. PMI and business sentiment play a critical role in predicting recessions, thus proving again the importance of soft indicators.

Figure 14. Aggregated variable contributions, Japan, M+3

Type the subtitle here. If you do not need a subtitle, please delete this line.



6. Conclusion

The research presented in this paper provides a proof of concept for the use of machine learning in economic forecasting. We have shown that, by using non-linear machine learning algorithms it is possible to increase forecast accuracy, to considerably increase forecast directional and turning point accuracy as well as to produce a consistent modelling framework for both short- and medium-term forecasts.

The methodology has nevertheless some limitations. The simulations are time-consuming and require much more computational power than linear models. And even though one may explain predictions using variable contributions, it is difficult to know why the algorithm learns what it learns. In other words, tree-based approaches allow prediction interpretation, but the complex tree structure underlying predictions comes at the expense of model interpretation.

The Indicator Model and Adaptive Trees are complementary. The Indicator Model provides a stable prediction and aims at the trend. This forecasting strategy yields great accuracy. In turn, Adaptive Trees are a more nervous forecast, with a strong focus on turning points.

The strong performance leaves little doubt about the potential usefulness of machine learning in economic forecasting. This research contributes to open a new horizon for macroeconomic forecasting research. We added to the evidence that it is possible to produce reliable model-based forecasts twelve months ahead. We also proved turning points could be anticipated with a high accuracy long before they occur.

Further research regarding the use of machine learning for economic forecasting could investigate new data sources. In this experiment, we applied a non-linear algorithm to leading indicators that are supposed to be linearly correlated to the growth of GDP. We built from the data choices of the Indicator Models that use linear modelling. In this respect, this can be seen as a constrained experiment. There may be important gains from broadening the scope to variables that are not linearly correlated to the GDP but that remain highly informative about its future growth, such as financial data, policy data, or more granular data (including big data).

It is likely that the reason why our forecast works better than commonly used alternatives is because our algorithm can capture complex non-linear patterns in the economy. That could be the reason why it can make reliable forecasts on longer time horizons. From that point of view, this modelling exercise sheds light on some characteristics of the data. This research provides tentative evidence that the economy is partly driven by non-linear mechanisms, especially in times of crises. This suggests that further investigating the patterns captured by the algorithm may be an interesting area of future work as well.

If Adaptive Trees, or other machine learning algorithms, can capture non-linearities and structural change in the economy, there may be other promising applications than forecasts. In particular, there is a growing body of literature about the estimation of heterogeneous treatment effects of economic policies (Athey and Imbens, 2015a, 2015b; Wager and Athey, 2015). Predictive algorithms may be used to predict counterfactuals in order to perform causal inference. Powerful machine learning algorithms may be very useful in this regard (Hartford et al., 2016; Johansson et al., 2016). Even though linear models facilitate interpretability, the possibility to uncover genuine non-linearities and structural changes provides additional arguments in favour of the use of a more flexible machine learning algorithm for analysing economic policies.

References

Athey, S. and G. Imbens, (2016) “The State of Applied Econometrics-Causality and Policy Evaluation,” arXiv Preprint arXiv:1607.00699.

Athey, S. and G. W. Imbens, (2015a) “Machine Learning Methods for Estimating Heterogeneous Causal Effects,” Stat, Vol. 1050, p. 5.

Athey, S. and G. W. Imbens, (2015b) “Machine Learning Methods for Estimating Heterogeneous Causal Effects,” Stat, Vol. 1050, p. 5.

Bellman, R., (1957) Dynamic Programming.

Biau, O. and A. D’Elia, (2009) “Euro Area Gdp Forecasting Using Large Survey Datasets,” A random forest approach. Availabile via: http://unstats. un.org/unsd/nationalaccount/workshops/2010/moscow/AC223-S73Bk4. PDF.

Blanchard, O., (2014) “Where Danger Lurks,” Finance & Development, Vol. 51(3), pp. 28–31.

Breiman, L., (2001) “Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author),” Statistical Science, Vol. 16(3), pp. 199–231, http://dx.doi.org/10.1214/ss/1009213726.

Chakraborty, C. and A. Joseph, (2017) "Machine Learning at Central Banks", Bank of England working paper No. 674, Bank of England.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and others, (2016) “Double Machine Learning for Treatment and Causal Parameters,” arXiv Preprint arXiv:1608.00060.

Efron, B., (1983) “Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation,” Journal of the American Statistical Association, Vol. 78(382), pp. 316–331.

Freund, Y., R. E. Schapire, and others, (1996) “Experiments with a New Boosting Algorithm,” icml, Vol. 96, pp. 148–156.

Friedman, J. H., (2001) “Greedy Function Approximation: A Gradient Boosting Machine,” Annals of Statistics, pp. 1189–1232.

Friedman, J. H., (2002) “Stochastic Gradient Boosting,” Computational Statistics & Data Analysis, Vol. 38(4), pp. 367–378.

Friedman, J., T. Hastie, and R. Tibshirani, (2001) The Elements of Statistical Learning, Vol. 1, Springer series in statistics Springer, Berlin.

Gama, J., I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, (2014) “A Survey on Concept Drift Adaptation,” ACM Computing Surveys (CSUR), Vol. 46(4), p. 44.

Hartford, J., G. Lewis, K. Leyton-Brown, and M. Taddy, (2016) “Counterfactual Prediction with Deep Instrumental Variables Networks,” arXiv:1612.09596 [Cs, Stat].

Johansson, F. D., U. Shalit, and D. Sontag, (2016) “Learning Representations for Counterfactual Inference,” arXiv Preprint arXiv:1605.03661.

Mourougane, A., (2006) “Forecasting Monthly GDP for Canada.”

Ollivaud, P., P.-A. Pionnier, E. Rusticelli, C. Schwellnus, and S.-H. Koh, (2016) "Forecasting GDP during and after the Great Recession."

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., (2011) "Scikit-Learn: Machine Learning in Python," Journal of Machine Learning Research, Vol. 12, p. 2825−2830.

Romer, P., (2016) "The Trouble with Macroeconomics," The American Economist, Forthcoming.

Schneider, J., (1997) "Cross Validation," A Locally Weighted Learning Tutorial Using Vizier, Vol. 1.

Sédillot, F. and N. Pain, (2003) "Indicator Models of Real GDP Growth in Selected OECD Countries."

Tiffin, A., (2016) "Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon," IMF Working Papers, Vol. 16(56), p. 1, http://dx.doi.org/10.5089/9781513568089.001.

Wager, S. and S. Athey, (2015) "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," arXiv Preprint arXiv:1510.04342.

Žliobaitė, I., (2010) "Learning under Concept Drift: An Overview," arXiv Preprint arXiv:1010.4784.

Annex A. The OECD Indicator Model

For the euro area and individual G7 economies, the near-term assessment also takes particular account of projections from a suite of statistical models using high-frequency indicators to provide estimates of near-term quarterly GDP growth, typically for the current and next quarter or so. This analysis builds on the work of Sédillot and Pain (2003) and Mourougane (2006) in using short term economic indicators to predict quarterly movements in GDP by efficiently exploiting all available monthly and quarterly information. These models typically combine information from both "soft" indicators, such as business sentiment and consumer surveys, and "hard" indicators, such as industrial production, retail sales, house prices etc. and use is made of different frequencies of data and a variety of estimation techniques. The procedures are relatively automated and can be run whenever major monthly data are released, allowing up dating and choice of model according to the information set available.

The most important gains from using the indicator approach are found to be for current-quarter forecasts made at or immediately after the start of the quarter in question, where estimated indicator models appear to outperform autoregressive time series models, both in terms of the size of error and directional accuracy. The main gains from using a monthly approach arise once one month of data is available for the quarter being forecast, typically two to three months before the publication of the first official outturn estimate for GDP. For one-quarter-ahead projections, the performance of the estimated indicator models is only noticeably better than simpler time series models once one or two months of information become available for the quarter preceding that being forecast. Modest gains are nonetheless to be made in terms of directional accuracy from using the indicator models.

Statistical indicator models are nevertheless limited in their ability to forecast quarterly GDP growth. Even with a complete set of monthly indicators for the quarter, the 70 per cent confidence bands around any point estimate for GDP growth in that quarter lie in the range from 0.4 to 0.8 percentage points, depending on the country or region and the degree of uncertainty is found to widen as the forecast horizon lengthens. Forecasting errors can also arise for a variety of reasons, including revisions to the initial published data and inaccuracies in the projections of the incoming monthly data.

Regular indicator model-based estimates of GDP now feed into both routine Economic Outlook assessment exercises and interim analyses and forecast updates released to the press on a routine basis.

Annex B. The OECD Economic Outlook

Purpose

The Economic Outlook projections are produced twice yearly by the OECD Economics Department to provide a consistent view of the world economy, with a specific focus on recent and future macroeconomic developments in current and prospective OECD Member countries and the larger non-OECD economies, most notably Brazil, Russia, India and China (the "BRICs"). Designed to provide a consistent framework for the policy debate in and between Member countries, the OECD forecasts and accompanying analyses are conditional on a consistent set of assumptions about policies and underlying economic and financial conditions, including fiscal and monetary policy settings, exchange rates, oil and non-oil commodity prices and international financial markets (see the Box "Policy and other assumptions underlying the projections" in the "General Assessment of the Macroeconomic Situation" chapter of the relevant Economic Outlook).

Coverage

The projections are made for a range of key macroeconomic variables in quarterly and annual frequencies, over a two to three year future horizon. The variable coverage for Member countries includes the usual range of national accounts demand and production aggregates, supply side and labour market indicators, wage and price inflation measures, monetary conditions, household and public sector accounts, trade volumes and prices and balance of payments accounts. Those for non-Members are considerably less detailed, including summary GDP, inflation, fiscal, trade and current account balances for enhanced engagement and larger economies, and main trade aggregates and balances for other regionally grouped non-OECD economies.

The relevant projections for OECD member countries are summarised in the corresponding Economic Outlook Annex Tables which report developments in key variables by country and broad regional grouping. A fuller coverage of historical data and projections across countries is available in the related Economic Outlook data publications and the OECD's online statistical dissemination system OECD.Stat. For further details of specific variables and coverage see the Economic Outlook Database Inventory and the frequently asked database questions section.

Summary information on the projections, including those for selected non-member economies are also given in the relevant chapters of the Economic Outlook.

The forecasting process

The OECD's projections are produced by its country experts interacting with its topic specialists, taking into account current and prospective developments, officially mandated policies, historical relationships between key variables and new information and indicators related to domestic and global conditions.

The effects of the new elements and revised judgments are typically assessed at the start of each forecasting round taken in conjunction with simulations of the effects of revised assumptions. In making the forecasts, particular attention is paid to consistency at domestic and world levels, to ensure that key accounting identities and relationships are observed, notably with respect to international trade and the balance of payments, a process assisted by the OECD's international trade model (see Pain et al (2005) and Murata et al (2000) and a variety of other estimated relationships between key variables. The overall forecast assessment thereby combines both judgment and a range of econometric based evidence.

The role of government economists

The forecast process also benefits from detailed consultation and peer review from government economists and policy makers in member and non-member countries and other key international organisations (the International Monetary Fund, World Bank, European Commission, European Central Bank, the Bank for International Settlement). These take place through the regular meetings of the OECD's Economic Policy Committee (EPC) and its expert working groups. Thus the main features of the projections and associated policy analyses are discussed by officials from finance or economy ministries and central banks. Country expertise is also drawn from the Economic and Development Review Committee (EDRC) country review process which also contributes to the OECD's regular Country Surveys reports on Member and selected non-member economies. Such discussions are valuable in harnessing Member countries' knowledge and expertise. Although given due consideration, comments and suggestions from Member countries are not automatically reflected in the final version of the Economic Outlook and, overall, the published projections and analyses represent the independent assessment of the OECD Economics Department, published under the responsibility of the Secretary-General.

Risk assessment

Since the OECD's projection are conditional on specific assumptions, an important part of the Economic Outlook analysis typically focuses on associated risks and uncertainties and potential imbalances in the world economy. Such risks are often illustrated by means of alternative scenarios and simulations based on different assumptions about policies, world market conditions or underlying structural factors e.g. different paths of commodity prices, exchange rates, policy mix, business or consumer confidence, using a macro-econometric model and other empirical-based analytical tools. The outcomes of such assessment are typically reported in the chapter General Assessment of the Economic Situation of the Economic Outlook and in separate publications.

Medium and longer-term analyses

In order to elaborate further the underlying nature of possible longer-term build-up or unwinding of specific imbalances and tensions in the world economy, the OECD now routinely constructs longer-term baseline (LTB) scenarios extending the short-term projections numerous years beyond the normal short-term horizon. These also serve as a basis for simulation comparisons with other scenarios based on alternative forecast assumptions. The LTB projections do not however embody a specific view about the nature or timing of future cyclical events but are conditional on number of stylised assumptions about policies and growth, in particular the technical assumption that output gaps close progressively over the projection period bringing the level and growth of actual GDP in each individual country back to estimated potential, within a specified period, see Giorno et al (1995), Beffy et al (2006) and Forecasting tools and analytical methods. Summary details of the OECD's longer-term baselines are typically published in the May/June edition of the Economic Outlook along with related supply-side analyses.

The assessment of forecast accuracy

Given the inherent uncertainties in making economic forecasts, the OECD periodically reviews its projections for predictive accuracy. Typically, such analyses try to distinguish between errors arising from data revisions, changes in underlying assumptions, and errors of judgement about economic conditions and forces shaping the outlook. Typically, larger projection errors appear to occur around major turning points in economic activity. The reasons for this may be due to errors of judgement or a decline in the predictive power of normal economic relationships or the quality of information available in and around cyclical turning points. The most recent published assessments of the forecast accuracy of the OECD projections are given by Vogel (2007), Lenain and Koutsogeorgopoulou (2000).

Annex C. Data description

Table A C.1. Data availability per country and number of variables

	Start	End	Number of observations	Number of variables
Japan	07-2002	01-2017	58	8
Germany	03-1998	01-2017	75	9
USA	10-1985	01-2017	125	9
Italy	10-1998	01-2017	73	5
UK	10-1992	01-2017	97	6
France	10-1987	01-2017	117	6

Annex D. Full charts

Figure A D.1. USA

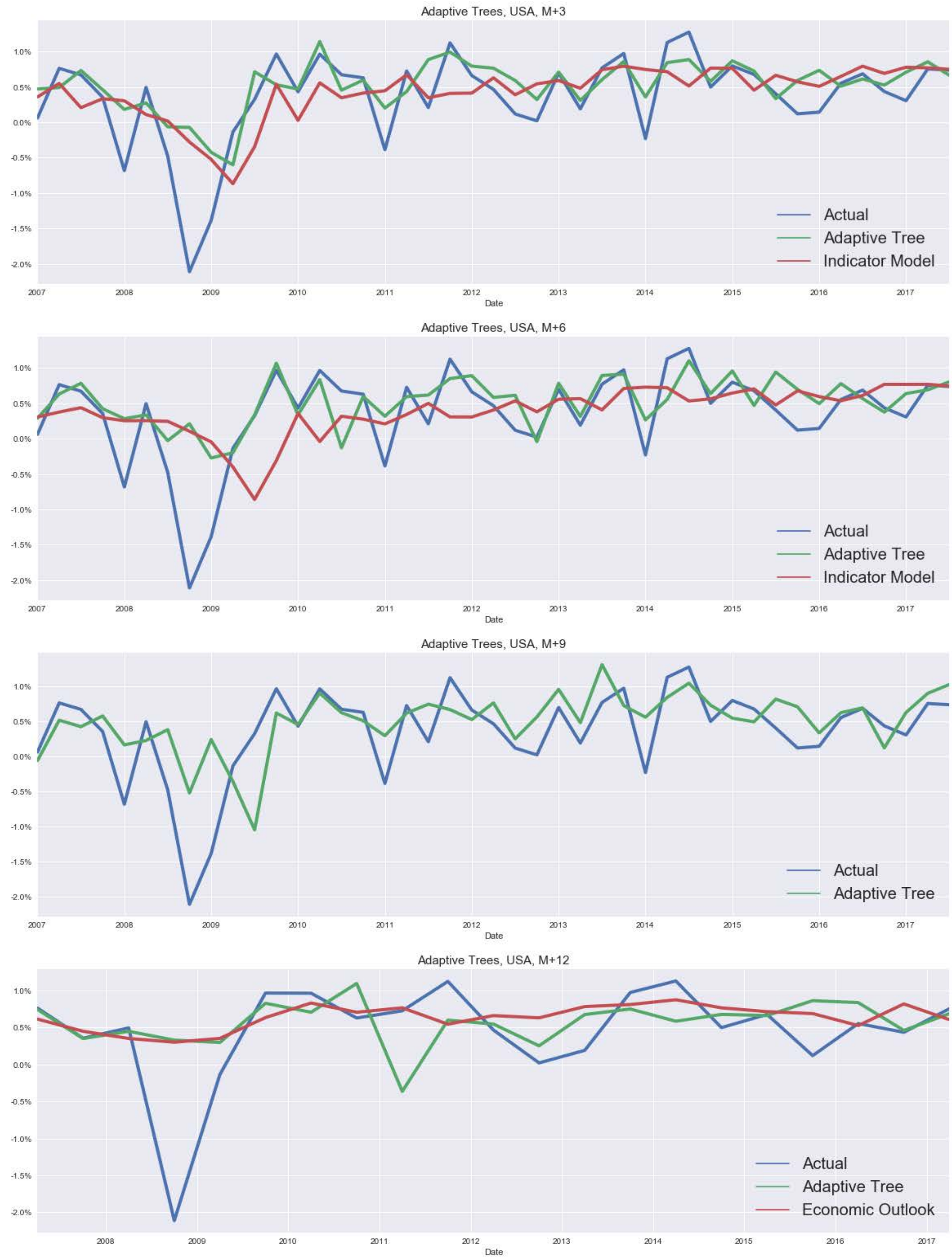


Figure A D.2. UK

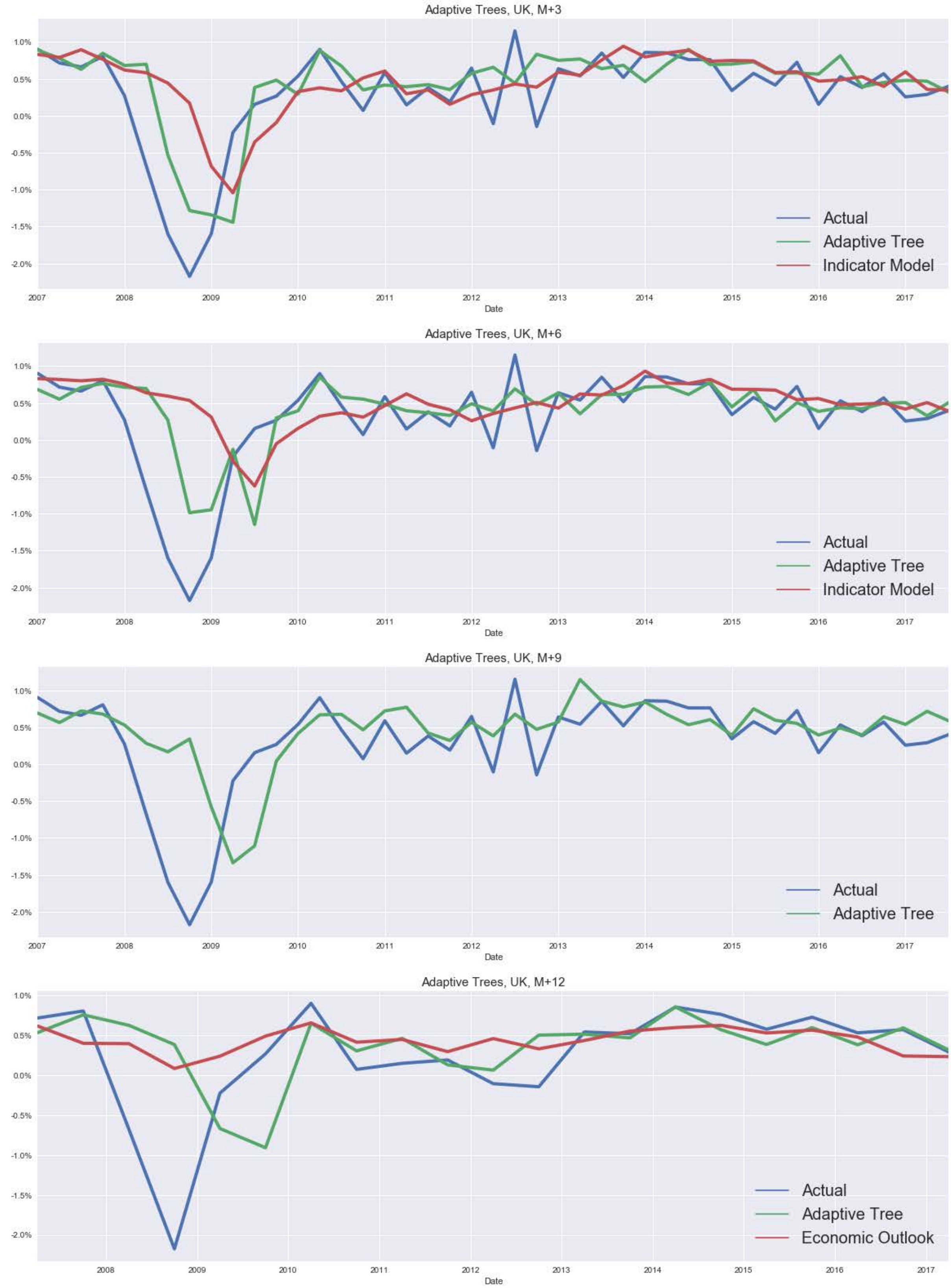


Figure A D.3. France

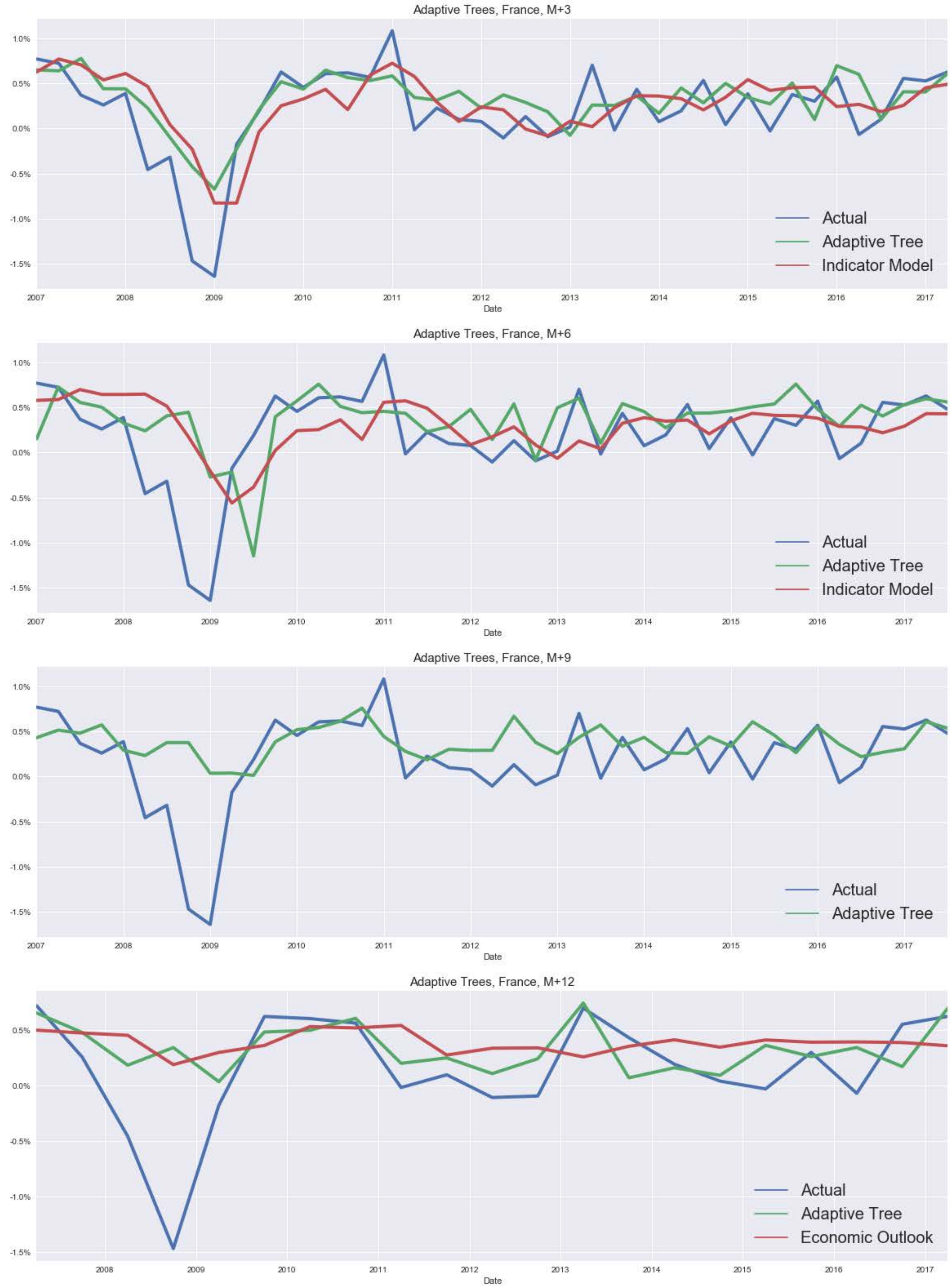
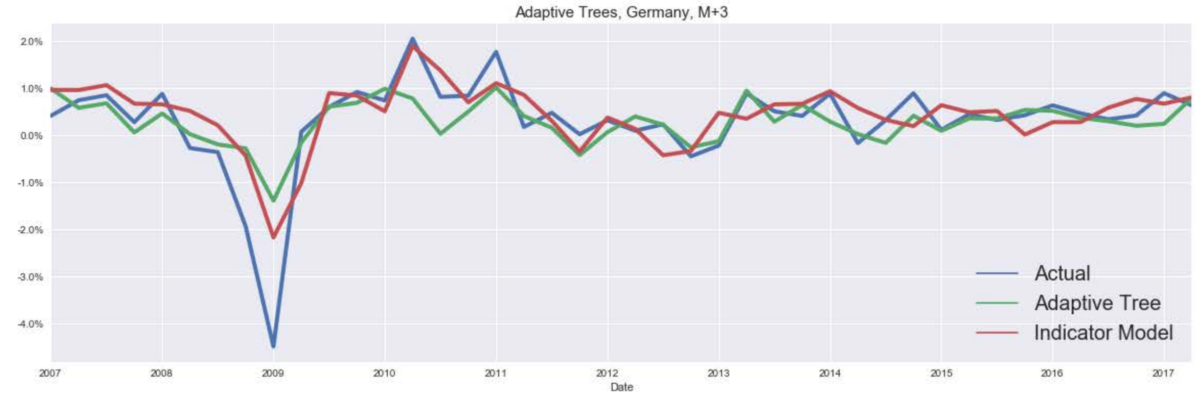
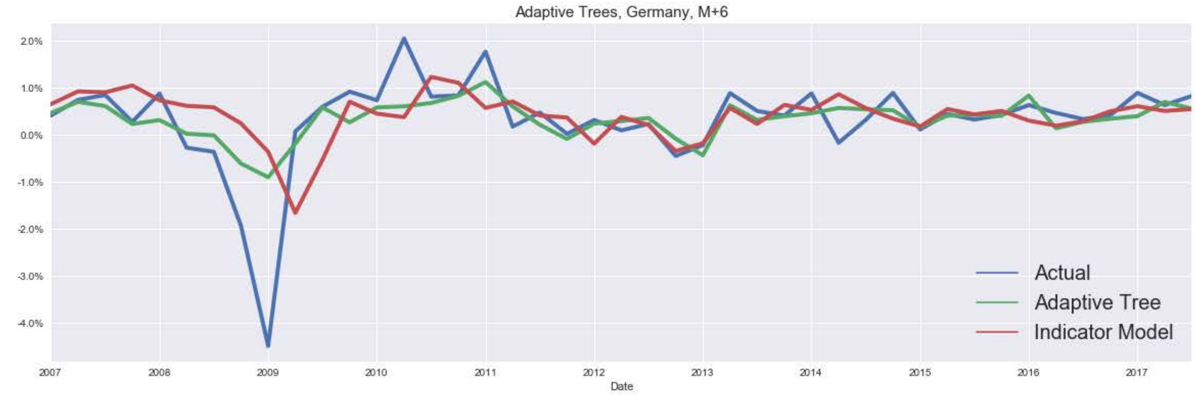
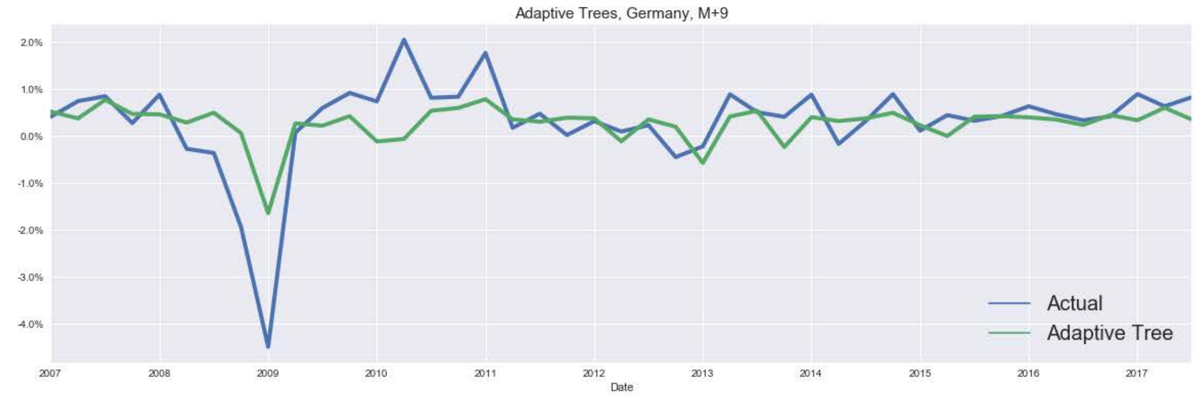


Figure A D.4. Japan

Figure A D.5. Germany







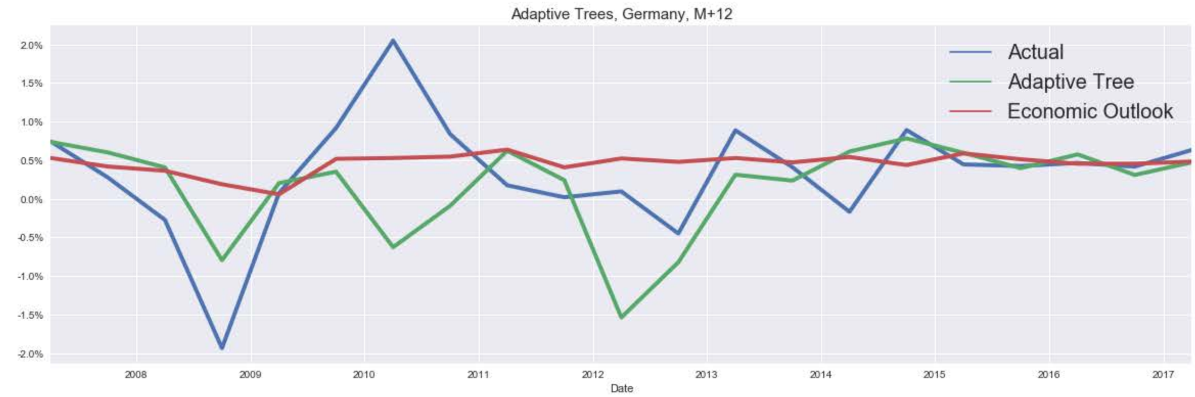
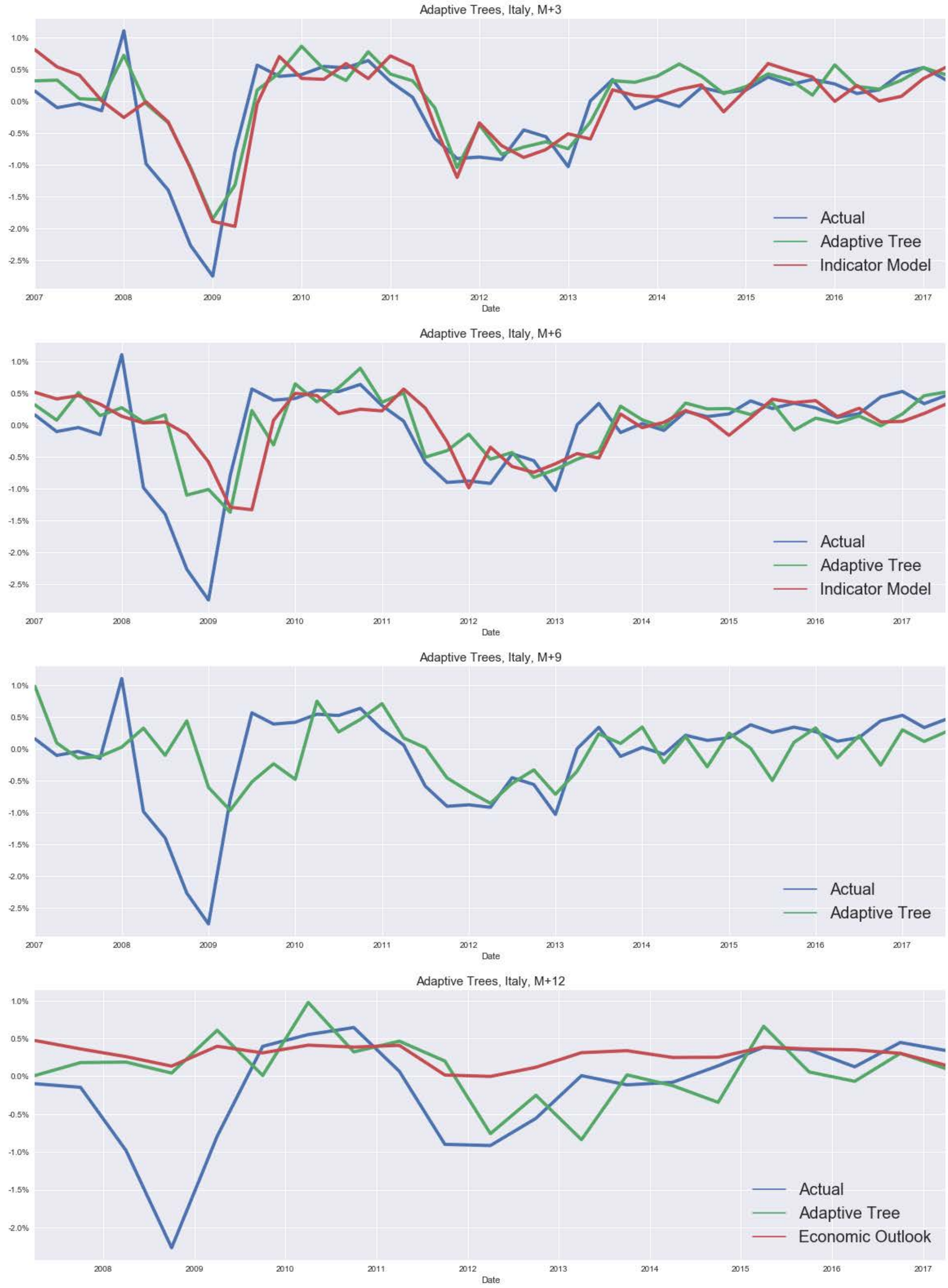


Figure A D.6. Italy



Annex E. Full forecast performance tables

Table A E.1. Performance of Adaptive Trees for the UK forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Treess	*0.45*	*55%*	*47%*	*62%*
	Indicator Model	0.61	57%	40%	77%
M+6	Adaptive Treess	*0.50*	*81%*	*63%*	*62%*
	Indicator Model	0.70	55%	27%	77%
M+9	Adaptive Treess	*0.63*	*71%*	*57%*	*62%*
M+12	Adaptive Treess	*0.71*	*80%*	*71%*	*57%*
	Economic Outlook	0.61	70%	43%	57%

Table A E.2. Performance of Adaptive Trees for the USA forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.48*	*79%*	*67%*	*53%*
	Indicator Model	0.52	60%	42%	47%
M+6	Adaptive Trees	*0.51*	*71%*	*58%*	*32%*
	Indicator Model	0.64	48%	25%	47%
M+9	Adaptive Trees	*0.54*	*67%*	*50%*	*32%*
M+12	Adaptive Trees	*0.66*	*60%*	*43%*	*43%*
	Economic Outlook	0.62	60%	29%	57%

Table A E.3. Performance of Adaptive Trees for the France forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.34*	*68%*	*50%*	*33%*
	Indicator Model	0.39	54%	23%	42%
M+6	Adaptive Trees	*0.51*	*62%*	*39%*	*33%*
	Indicator Model	0.49	55%	29%	58%
M+9	Adaptive Trees	*0.49*	*48%*	*23%*	*33%*
M+12	Adaptive Trees	*0.47*	*55%*	*40%*	*36%*
	Economic Outlook	0.51	40%	10%	55%

Table A E.4. Performance of Adaptive Trees for the Japan forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.99*	*68%*	*44%*	*38%*
	Indicator Model	1.13	53%	24%	31%
M+6	Adaptive Trees	*0.96*	*66%*	*58%*	*50%*
	Indicator Model	1.35	46%	19%	63%
M+9	Adaptive Trees	*1.04*	*76%*	*65%*	*50%*
M+12	Adaptive Trees	*1.01*	*55%*	*38%*	*38%*
	Economic Outlook	1.15	45%	23%	50%

Table A E.5. Performance of Adaptive Trees for the Germany forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.66*	*71%*	*46%*	*64%*
	Indicator Model	0.60	54%	32%	50%
M+6	Adaptive Trees	*0.69*	*62%*	*31%*	*57%*
	Indicator Model	0.91	52%	28%	64%
M+9	Adaptive Trees	*0.73*	*62%*	*41%*	*36%*
M+12	Adaptive Trees	*0.83*	*70%*	*50%*	*67%*
	Economic Outlook	0.68	60%	33%	56%

Table A E.6. Performance of Adaptive Trees for the Italy forecast (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.42*	*76%*	*52%*	*73%*
	Indicator Model	0.53	51%	22%	60%
M+6	Adaptive Trees	*0.55*	*62%*	*50%*	*13%*
	Indicator Model	0.72	50%	21%	27%
M+9	Adaptive Trees	*0.73*	*69%*	*50%*	*27%*
M+12	Adaptive Trees	*0.75*	*55%*	*56%*	*42%*
	Economic Outlook	0.76	70%	33%	58%

Table A E.7. Performance of Adaptive Trees: average overall G6 countries (2007-2017)

		RMSE	FDA	Turning points	Turning points, false alerts
M+3	Adaptive Trees	*0.56*	*69%*	*51%*	*54%*
	Indicator Model	0.63	55%	31%	51%
M+6	Adaptive Trees	*0.62*	*67%*	*50%*	*41%*
	Indicator Model	0.80	51%	25%	56%
M+9	Adaptive Trees	*0.69*	*65%*	*48%*	*40%*
M+12	Adaptive Trees	*0.74*	*63%*	*50%*	*47%*
	Economic Outlook	0.72	58%	29%	55%

Annex F. Selected plot of aggregated variable contributions

Figure A F.1. Aggregated variable contributions, UK, M+3

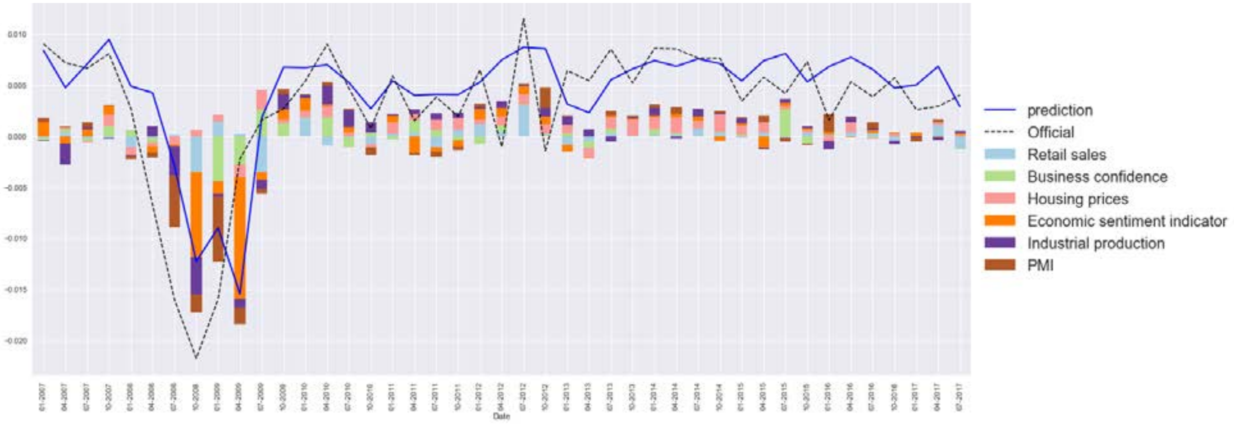


Figure A F.2. Aggregated variable contributions, USA, M+3

Figure A F.3. Aggregated variable contributions, Italy, M+3

Figure A F.4. Aggregated variable contributions, Germany, M+3

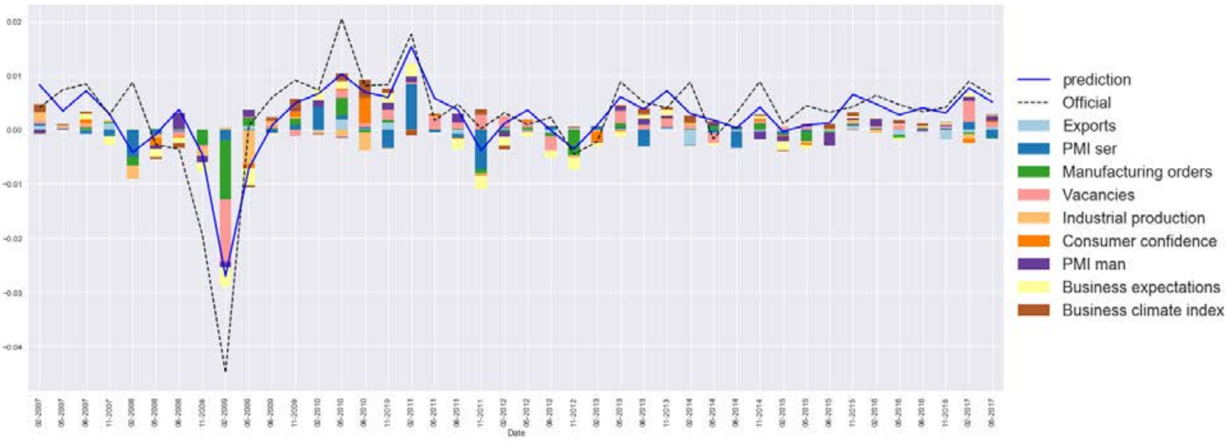


Figure A F.5. Aggregated variable contributions, Japan, M+3

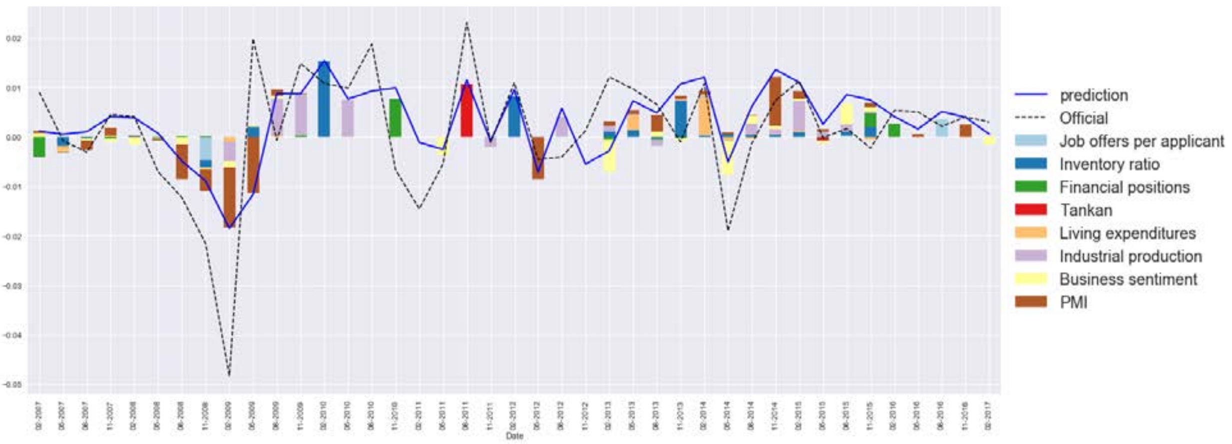


Figure A F.6. Aggregated variable contributions, France, M+3