Piotr Wójcik[1], Bartłomiej Wieczorek

We have just explained convergence with machine learning

P R E L I M I N A R Y   V E R S I O N

Abstract

The analysis of convergence factors has been one of the important research topics in macroeconomics for many years. It also has important implications for economic policy – convergence is one of the most important goals of the European Union, and spending on cohesion policy is an important item in the EU budget. The choice of potential factors of convergence has a key impact on inference about its occurrence. There is no consensus on which set is the best. Conclusions regarding the importance of individual factors may be contradictory. Previous studies have most often focused on determining the initial list of convergence factors and estimating many (often millions) models for various subsets of variables, and then inferring the importance of individual factors on the basis of averaged (also using Bayesian methods) results of all models.

This article takes a novel approach. Its main purpose is to identify the relevant determinants of economic growth and extend the existing empirical research in two ways. First of all, the article attempts to identify potential non-linearities using machine learning methods. Secondly, it uses modern methods that allow to select variables important for the growth and explain the relationship regardless of the specification. Apart from LASSO, previously used in convergence studies, we use support vector regression, random forests and extreme gradient boosting regression. In addition, Explainable Artificial Intelligence (XAI) tools are used measure the importance of convergence factors in a model agnostic way and to interpret what models have learned from the data.

We assume that machine learning tools which allow for non-linearity explain cross-country growth rate with higher accuracy than existing linear models. In addition, these tools still allow for model interpretation and assessing feature importance. Thus, they better identify and explain the factors of convergence and therefore can be helpful in formulating policy recommendations.

keywords: conditional convergence, factors, explainable artificial intelligence

JEL codes: O47, C21, C52

[1] Associate professor, University of Warsaw, Faculty of Economic Sciences, head of Data Science Lab WNE UW, corresponding author, pwojcik@wne.uw.edu.pl.

1. Introduction and literature overview

For over two decades, many researchers were interested in determining the crucial factors of economic growth or income convergence between different countries. The common method used in research is the analysis of *beta* convergence (Barro and Sala-i-Martin, 1995). This approach focuses on the partial correlation between growth in income over time and its initial level. Furthermore, researchers analyze *beta* converge also conditioned on the additional factors. However, they do not agree which factors are really significant and which are not. Durlauf et al. (2009) state that the choice of control variables has an impact on the inference about the convergence and its occurrence.

So far, there were plenty of empirical approaches considered. Levine and Renelt (1992) started with the variant of sensitivity analysis called Extreme Bound Analysis (EBA) (Leamer, 1985) for the cross country growth regressions. Their approach was later challenged by Sala-i-Martin (1997) who showed that EBA had too much power for any variable to pass it. He proposed to „move away from extreme test" and rather to „assign some level of confidence to each of the variables" (Sala-i-Martin, 1997), which was calculated as weighted statistics based on all regressions. In contrary to this approach, Hendry and Krolzig (2004) proposed to estimate just one regression and then apply the general-to-specific procedure. Furthermore, Sala-i-Martin et al. (2004) and Lay and Steel (2007) proposed Bayesian Model Averaging with different priors to address uncertainty of the choice of the correct variables. However, most of those approaches were questioned by Ciccone and Jarociński (2010), both on theoretical and empirical grounds. They stated that only the correct identification of growth factors will allow to form credible recommendations for economic policy.

In case of machine learning tools, LASSO (Tibshirani, 1996) were applied to the growth regressions by some researchers (e.g. Varian, 2014). Hofmacher et al. (2015) showed that LASSO allows to eliminate the problem of collinearity of explanatory variables. Furthermore, the results on out-of-sample predictions of growth rates were better than model averaging techniques.

The drawback of all mentioned methods is that those approaches assume that the relationship is linear. However, Levine and Renelt (1992) predicted that the real model could be non-linear one. The existence of non-linearities of independent variables cannot be ignored in empirical analyses. He and Xu (2019) show that identifying a variable to be a statistically relevant factor of growth in a linear specification might be a result of inappropriately specified model, whereas, if the correct form of the model was linear, this variable might not be significantly correlated with growth.

This article takes a novel approach. Its main purpose is to identify the relevant determinants of economic growth and extend the existing empirical research in two ways. First of all, the article attempts to identify potential non-linearities using machine learning methods. Secondly, it uses modern methods that allow to select variables important for the growth and explain the relationship regardless of the specification. Apart from the above mentioned LASSO, authors used support vector regression (SVR, Vapnik, 1995), modified bagging (bootstrap averaging) approach – random forests (Breiman, 2001) and extreme gradient boosting regression (XGBoost, Chen, 2016), both ensembling multiple decision tree classifiers. All of those techniques allow to exclude potential collinearities among independent variables. Leave One Out cross validation is used for tuning model parameters, based on minimization of the Mean Absolute Error (MAE) metric. In addition to successful prediction, the ability to interpret what a model has learned from the data is of equal importance. This approach is called Interpretable Machine Learning (IML) or Explainable Artificial Intelligence (XAI).

The research hypothesis verified in the article states that machine learning tools which allow for non-linearity explain cross-country growth rate with higher accuracy than existing linear models. In addition, these tools still allow for model interpretation and assessing feature importance. Thus, they better identify and explain the factors of convergence and therefore can be helpful in formulating policy recommendations.

In this article, two datasets are considered, to make results comparable with previous studies. The first dataset was used in Fernandez et al. (2001). The second dataset was introduced in Sala-i-Martin et al. (2004).

The remaining part of the article is structured as follows. In the second part, models methods used in further parts are briefly described. In third part, we present the results of the analysis on two above mentioned datasets and compare them with the findings of other researchers. The last section summarizes our study.

2. Methods

In this section, machine learning tools applied in the empirical part are briefly introduced in a non-technical, intuitive way.

2.1 LASSO (Least Absolute Shrinkage Selector Operator)

LASSO is a regularization technique. It can be viewed as extension of Ordinary Least Squares (OLS) model. It differs from OLS because of its cost function. It not only minimizes the sum of squared residuals, but also takes into account the sum of absolute values of the parameters of the linear model as an additional constraint with a weight λ. Adding a penalty in the optimization results in searching for parameters that fit the data well, but are as small as possible. Parameters at less important (or redundant, e.g. highly correlated) variables will shrink towards zero, some of them will even be set to be equal to zero. At the expense of a certain bias (the estimators of parameters from LASSO model are biased), LASSO often allows to obtain more precise forecasts on the test sample (e.g. Hofmacher et al., 2015). What is crucial also in case of growth regressions, LASSO can be considered as the variable selection method, which can be used even when the initial number of variables exceeds the number of observations. It is often used by researchers as a preliminary stage of analysis, combined with subsequent model estimation on selected variables using OLS (Schneider and Wagner, 2011). No a priori assumptions or selection of subset of variables are needed. One has only to determine the optimal weight for the additional constraint (namely, hyperparameter λ), which can be done via cross-validation.

2.2 SVR (Support Vector Regression)

Similarly to OLS, SVR fits a hyperplane that is positioned as close to all data points as possible. However, while OLS minimizes the sum of squared errors, SVR tries to fit the error within a specified distance from the hyperplame (Smola and Schölkopf, 2004). There is also a regularization parameters C, which controls how much one wants to avoid misclassifying each observation. The most important advantage of SVR over OLS is the ability to model non-linear relationships between variables. It is done with the use of a selected kernel function. Applying a kernel function results in an implicit non-linear mapping into a higher dimensional feature space, where it is guaranteed to find an appropriate hyperplane (Vapnik, 1995). Thus, one can think of SVR as a process of performing linear regression in a transformed and more dimensional space. Two widely used types of kernels are radial basis function and polynomial kernel – both are applied in the empirical part of the article. Radial basis kernel is equivalent to mapping the analyzed data into an infinite dimensional Hilbert space (Ben Ishak, 2016).

2.3 Decision tree regressor

Decision tree regressors are predictive models structured in a tree-like way. The model recursively breaks data into smaller sub-datasets with respect to values of selected variables trying to make these subsets as homogeneous with respect to the explained variable as possible. The decision of how to make splits heavily affects the accuracy of the tree. Decision tree regressors often use MSE metric to decide to split a node in two sub-nodes. Decision trees require some stopping rule, for example the maximum depth of the tree.

In the case of simple regression tree, variable importance can be easily measured. One first needs to calculate the Gini impurity measure for every node. Then, feature importance is calculated as the decrease in node impurity and it is weighted with the probability of reaching that node. The probability of reaching each node can be calculated as the ratio of observations that reach this particular node. The higher the value of decrease in Gini impurity, the more important the feature that creates the node is.

2.4 Bagging (Bootstrap Aggregating) and Random Forests

Bagging is an ensemble technique. It combines multiple models (usually weak learners) trained on different bootstrap samples of the original dataset (selected randomly with replacement). On each of the subsamples, one model of the same type is trained. Then, the prediction from the bagging model is obtained by the majority voting from all models in case of a classification problem, and by averaging the predicted values from all models in case of a regression problem. Bagging is often used to decrease the variance of predictions.

The weak learner is a predictive model, which predictive power is just slightly higher than simple guessing. Combining multiple weak learners can create a strong learner – the model which predictions are much more accurate. The examples of weak learners commonly used in this approach are decision trees (or even decision stumps – trees with just one split of depth equal to 1) or OLS models.

Random forest is a specific case of bagging algorithms. It was first introduced by Breiman (2001) and in simple words, is a combination of decision tree regressors. Each tree is trained on a different bootstrap sample of the original dataset, just like in bagging. In addition, during each split in each tree only a random subset of all predictors is considered. Random forests are robust to the problem of multicollinearity and can be applied on a large number of potential predictors without their initial selection. In addition, they are indifferent to non-linear interlinkages between the data. They require tuning of two parameters – a number of trees and number of a number of predictors tried at each node.

2.5 Boosting

Boosting is another type of ensemble models. However, its principle is different than in bagging. While bagging averages the prediction from multiple independent weak learners, boosting approach combines them iteratively. During each step, a weak learner is trained on the weighted sample. The weights are higher for observations that had high prediction errors in the previous step, which allows to better fit more problematic observations.

There are multiple boosting methods known in literature. One of the most popular is Gradient Boosting. It became popular when Friedman (1999) described the algorithm of gradient descent in the function space, and then applied it to the cost functions of popular algorithms. One of the parameters optimize in the Gradient Boosting approach is the learning rate, which shows how quickly the errors are corrected between the weak learners. Recently, an extension of Gradient Boosting, called eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) is widely used. XGBoost extends Gradient Boosting by the different penalization of trees, adding a shrinkage to the leaf nodes and an extra randomization parameter.

2.6 Cross Validation

Machine learning algorithms require selection of hyperparameter values, i.e. parameters that are not optimized in the model training process (e.g. penalty for too large parameters in LASSO approach, the cost of incorrect classification in SVR, the number of trees in a random forest or the learning rate in Gradient Boosting). Hyperparameters can be chosen arbitrarily, but it is better to choose them consciously. In the empirical part of the paper, we use a Leave-One-Out Cross Validation (LOOCV). For each combination of values of hyperparameters, each model is estimated *n* times on the sample without the 1st, 2nd, 3rd, … observation, respectively. The single observation left aside is used as the validation sample – for assessing the quality of prediction. Based on all predictions for a specific combination of hyperparameters, Mean Absolute Error (MAE) is calculated. Finally, we select and apply the model with the hyperparameters that minimizes the MAE of predictions.

2.7 Variable importance and explainable machine learning

Many machine learning algorithms have their own specific way to measure the importance of each feature. But the lack of model interpretability is the most important limitation of modern machine learning tools. The quality of predictions is important in research, but it is even more important to understand the correct economic mechanism that drives prediction of the particular phenomenon. Recently, a field called Explainable Machine Learning (EML) or eXplainable Artificial Intelligence (XAI) has been developing rapidly (Molnar, 2019). It offers additional tools to overcome the black box dilemma and allow for easy comparability of variable importance across different models. In the empirical part of this paper, we use the measure called *model reliance,* introduced by Fisher et al. (2019), which was inspired by the permutation based approach of Breiman (2001). It describes how much the model's performance *relies* on different covariates. In the permutation based approach, to assess the importance of selected feature, one calculates the error of the prediction from the model on the original dataset, e_{orig} and on the artificial dataset, with the values of that feature randomly permuted, e_{perm}. The higher the ratio e_{perm} / e_{orig}, the more important the feature, as it describes how much the error rises when the feature becomes non-informative. Model reliance generalizes such an approach by taking into consideration not one, but all permutations that permute the values of a selected feature.

After identifying influential variables, one has to understand the relationship between these variables and the response. On the level of each observation, *ceteris paribus* profile can be analyzed. In essence, it shows how a conditional expectation of the dependent variable changes with the values of a particular explanatory variable, while all other variables are kept constant (Goldstein et al, 2015). Averaging *ceteris paribus* profiles over all observations shows how the expected model response behaves as a function of a selected feature. This procedure, called Partial Dependence Profile, was first introduced by Friedman (2000).

2.8 Methods used in the empirical part

In the empirical part of the research, we estimated several machine learning models on two datasets, mentioned earlier. Selected models are LASSO, OLS model with variables selected via LASSO, SVR models with radial basis function and polynomial kernels (individually), random forest, gradient boosting machine with decision trees selected as weak learners and XGBoost models with two different weak learners - OLS models and decision trees. Hyperparameters of the models were selected via Leave-One-Out cross validation, with exception of the number of boosting rounds for boosting algorithms, which was arbitrarily set to 50. After estimation process, we assessed the importance of features using model reliance approach. And finally, we analyzed the Partial Dependence Profiles for initial level of GDP, to verify the possible occurrence of beta convergence.

3. Empirical results

The above mentioned methods – LASSO, SVR with both radial and polynomial kernels, random forest, Gradient Boosting regression with decision trees used as weak learners and XGBoost with both OLS model and decision trees used as weak learners – were applied on two widely studied datasets. The first of the datasets was used by Fernández et al. (2001) and includes 41 explanatory variables for 72 countries – referred to hereafter as FLS. The second one was used by Sala-i-Martin et al. (2004) and includes 67 explanatory variables for 88 countries – referred to hereafter as SDM.

Schneider and Wagner (2011) applied LASSO type regression on these datasets and claim that estimation results were in line with the findings of original papers. However, it is not confirmed below, especially for SDM (2004) data.

In each case, the results of LASSO, OLS with LASSO used as the first selective step, SVR with both kernels, random forest (RF), Gradient Boosting Regression (GBR) and XGBoost with OLS and Decision Trees used as weak learners are reported and compared with the results of the original papers and their follow-ups. Variable importance ranking (based on *model reliance* measure) is provided to show the results in a consistent way and compare them with previous studies.

3.1 Analysis on FLS (2001) data

Table 1 shows the ranking of important variables identified originally in Fernández et al. (2001), compared with Hendry and Krolzig (2004) and previous Sala-i-Martin (1997) results. We show only the first 20 most important variables, according to Fernández et al. (2001) – the lower the number, the more important the variable is.

It appears that all approaches confirm the occurrence of conditional convergence, although in ensemble techniques (random forest and boosting algorithms), the *initial GDP* has a lower rank. In contrary, for LASSO, OLS based on LASSO and SVR models, *initial GDP* level is the most important factor. Most of the approaches also agree on the importance of *fraction Confucian*, *life expectancy* and *equipment investment*. However, machine learning tools do not confirm 5–10 out of 20 most important variables indicated by Fernández et al. (2001). The results of LASSO and SVR models are very consistent with Hendry and Krolzig (2004) – among top 13 variables in their model, 12 are also the most important in LASSO, and among their top 11 variables, 10 are also the most important in both SVR approaches.

Variable	rank of the variable importance										
	FLS (2001)	S (1997)	OLS (HK, 2004)	LASSO	OLS (LASSO)	SVR (poly)	SVR (radial)	RF	GBR	XGBoost (Tree)	XGBoost (linear)
GDP level in 1960	1	1	1	1	1	1	1	10	7	7	13
Fraction Confucian	2	1	11	5	7	5	3	5	15	2	6
Life expectancy	3	7	2	2	3	4	2	7	2	4	21
Equipment investment	4	1	8	8	10	8	6	1	1	1	2
Sub Saharan dummy	5	10	4	4	5	3	4	32	40	41	41
Fraction Muslim	6	1	-	17	20	29	29	22	25	5	18
Rule of Law	7	1	10	11	13	10	10	27	41	28	31
Number of Years open economy	8	1	-	35	31	35	27	4	6	9	10
Degree of Capitalism	9	17	-	18	25	19	17	30	22	34	24
Fraction Protestant	10	22	-	16	23	11	11	14	9	10	7
Fraction GDP in mining	11	13	16	12	17	13	12	25	17	22	23
Non-Equipment Investment	12	19	-	15	19	17	13	3	4	3	3
Latin American dummy	13	8	7	9	8	9	9	26	36	39	39
Primary School Enrollment, 1960	14	15	12	13	12	12	15	20	19	13	22
Fraction Budhist	15	23	-	22	27	21	14	2	3	6	1
Black Market Premium	16	30	-	19	21	18	21	29	27	27	33
Fraction Catholic	17	24	-	-	-	38	40	12	18	23	25
Civil Liberties	18	10	-	20	14	27	33	21	30	21	15
Fraction Hindu	19	35	3	3	2	2	5	34	39	30	30
Primary exports, 1970	20	16	-	28	26	31	30	13	16	19	5
…											
Size labor force	25	28	5	6	4	6	7	9	10	8	8
…											
Etholinguistic fractionalization	28	36	13	10	11	14	16	19	12	17	14
…											
SD of black market premium	30	14	-	33	36	30	22	8	8	11	9
Higher education enrollment	34	39	6	7	6	17	8	7	13	12	29
…											
Public Education share	40	-	-	26	24	22	20	11	11	15	12

Table 1. Importance of growth determinants for FLS data

In addition, most of the machine learning algorithms indicated *Non-Equipment Investment* as an important convergence factor, which is in contrast to original articles. There are also important growth determinants, missed by original approaches, but confirmed by other methods (including Hendry and Krolzig, 2004) – *Size labor force*, *Ethnolinguistic Fractionalization* and *Higher education enrollment*.

We can also identify some similarities between ensemble models. Random forest and boosting algorithms acted quite similarly – all of those models neglected *Sub-Saharan dummy*, *Rule of Law, Latin American dummy* and *Fraction Hindu* factors, which were high in rankings based on the other models. Similarly, for all of above mentioned models, *Number of Years open economy* was listed as an important variable, which is in line with Fernandez et al. (2001) and Sala-i Martin (1997), but was neglected by LASSO, OLS and SVR models.

Based on the estimation results of all machine learning models and replicated HK results, Partial Dependence Profiles for initial income were calculated. One can see them on Figure 1a). The profiles show that the relationship between the average growth rate and initial GDP seems to be linear in case of OLS, LASSO and SVR models (as expected), but it turned out to be non-linear (but partially linear) for the ensemble algorithms (Figure 1b)). The relationship is negative, but, again, much flatter for the ensemble algorithms. When looking closely on ensemble algorithms, one can draw some contrasting conclusions. For instance, if GBM or XGBoost with Decision Trees learners are concerned, one can see that the strongest convergence occurs for the poorest countries, it does not occur for countries with mediocre initial income, and then it occurs again for the group of the richest. However, for XGBoost model with OLS learners, it appears that countries with a wide range of initial GDP have similar growth pace and the only downward trend appears for the low-to-middle income countries. In the case of random forest model, the downward trend is consistent, but rather flat, compared to other ensemble methods.

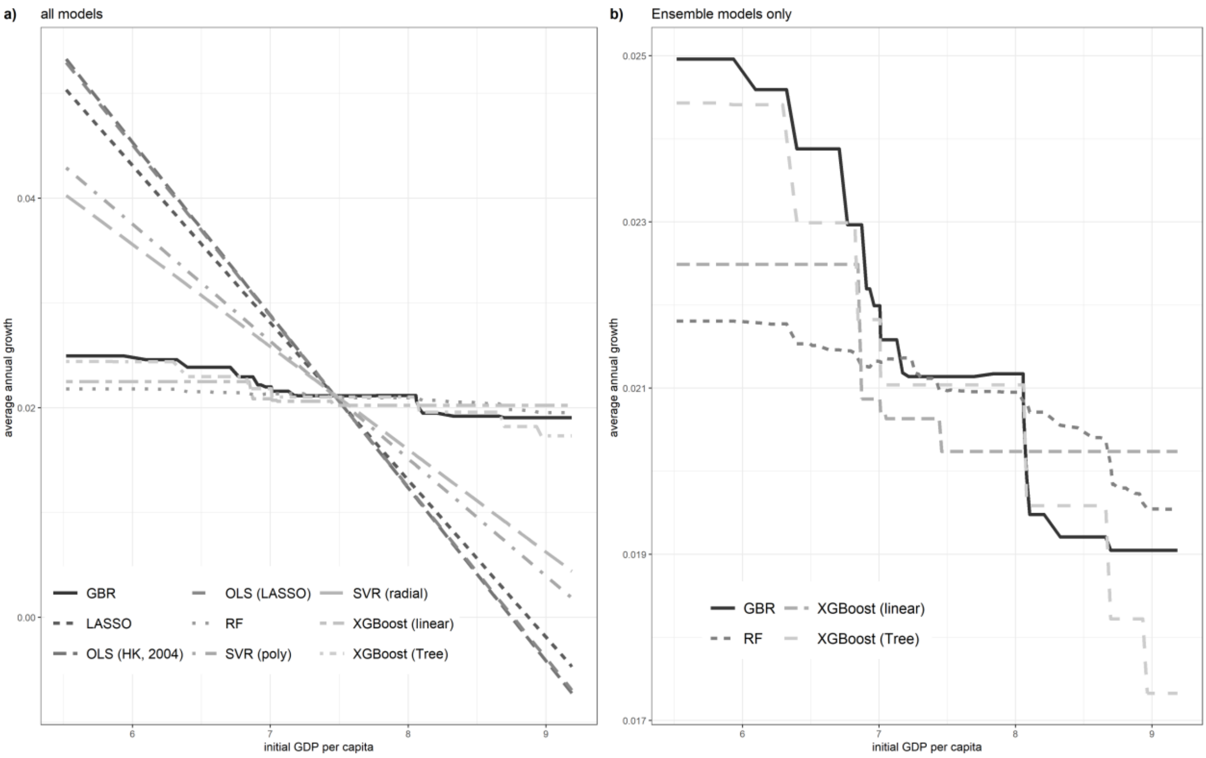

Figure 1. Partial Dependence Profiles for initial income for models estimated on FLS data

In the end, for all estimated models fit to data measures were compared – see Table 2. All models explain more than 90% of the variability of growth on the whole sample. Moreover, each of machine learning algorithms is better than a linear model in terms of every considered measure. Boosting approaches explain the relationship the best among all models, with the lowest prediction errors, which was to be expected looking at their specification. Although the best model in terms of R^2 or the error metrics is XGBoost with OLS learners, we can say that such results confirm our research hypothesis. Generally models that allow for non-linearities explain more variability of growth rates and have lower prediction errors when compared to the linear models like LASSO or OLS.

model	RMSE	MAE	R2
OLS (HK, 2004)	0,0055	0,0041	0,9072
LASSO	0,0040	0,0031	0,9521
OLS / LASSO	0,0037	0,0029	0,9590
SVR (radial)	0,0035	0,0025	0,9621
SVR (poly)	0,0033	0,0024	0,9672
RF	0,0045	0,0033	0,9378
GBR	0,0018	0,0012	0,9897
XGBoost (Tree)	0,0014	0,0011	0,9940
XGBoost (Linear)	0,0006	0,0004	0,9990

Table 2. Measures of fit for all models estimated on FLS data

3.2 Analysis on SDM (2004) data

The differences in conclusions are more striking in case of the second dataset. Table 3 shows the ranking of important variables identified originally by SDM (2004) and in a follow-up study by Doppelhofer and Weeks (2011) who use a „robust" version of BMA, but obtain identical conclusions.

Some of the conclusions based on the ranking are in line with Sala-I-Martin et al. (2004). *East Asian dummy* and *Investment price* – the 1st and 3rd, respectively, the most important factors in SDM – were confirmed by all machine learning tools. Moreover, the 2nd most important factor, *Primary schooling in 1960,* was lower in rankings just for GBR and linear XGBoost with OLS learners. Several other variables seem to have stronger impact on the growth rate, according to most machine learning algorithms: *Malaria prevalence in 1960s, Fraction Buddhist, Fraction Confucian* (it is in line with earlier analysis on FLS data, however, it is far in ranking for GBR), *Life expectancy in 1960* (it is also in line with analysis on FLS, but this time, it was excluded by LASSO).

We can also spot some inconsistencies. For instance, *Population density coastal 1960s* was high in the rankings for LASSO and SVR methods, but much lower for the ensemble models. And, other way around, *Population Density 1960* was considered as important for ensemble models, especially for GBM, but it was neglected by SVR models and excluded from the analysis by LASSO.

variable	rank of variable importance									
	SDM (2004)	DW (2011)	LASSO	OLS (LASSO)	SVR (poly)	SVR (radial)	RF	GBR	XGBoost (Tree)	XGBoost (Linear)
East Asian dummy	1	1	1	2	1	2	3	1	4	5
Primary schooling 1960	2	2	3	4	3	1	4	16	2	18
Investment price	3	3	4	3	2	4	9	7	6	8
GDP in 1960 (log)	4	4	-	-	21	28	38	17	26	33
Fraction of tropical area	5	5	5	1	13	18	22	36	59	49
Population density coastal in 1960s	6	6	9	7	9	9	13	23	16	34
Malaria prevalence in 1960s	7	7	11	12	10	12	1	3	1	1
Life expectancy in 1960	8	8	-	-	11	10	2	6	5	7
Fraction Confucian	9	9	2	5	4	3	10	25	8	16
African dummy	10	10	-	-	8	7	14	48	60	61
Latin American dummy	11	11	-	-	23	34	52	49	30	51
Fraction GDP in mining	12	12	-	-	26	31	42	42	27	50
Spanish colony	13	13	-	-	42	35	48	43	31	59
Years open 1950 - 1994	14	14	6	11	7	8	6	11	13	38
Fraction Muslim	15	15	-	-	41	67	47	56	25	35
Fraction Buddhist	16	16	7	6	5	5	5	2	3	4
Ethnolinguistic fractionalization	17	17	-	-	16	16	25	8	17	12
Government consumption share 1960s	18	18	12	8	18	15	36	44	50	53
Population density 1960	19	19	-	-	47	45	12	5	9	6
Exchange rate distortion	20	20	10	9	6	6	8	9	12	22

Table 3. Importance of growth determinants for SDM data

And finally, the most striking result is the ranking of GDP in 1960 (log). Initial GDP was one of the most important variables for models estimated on FLS data. It is also the key factor for the convergence analysis. However, it was excluded from the analysis by LASSO, and in case of other machine learning approaches, it is in a very far position in the importance ranking. This might suggest the lack of conditional convergence, which was observed in earlier studies. Such conclusion is in line with He and Xu (2019), who suggested that such significance might have been a result of inappropriate model specification. In a correct, non-linear specification, initial GDP can be not (strongly) correlated with growth.

Partial Dependence Profiles for initial income for SDM dataset are plotted on Figure 2a). We discussed that we cannot draw any conclusions about the relation for the LASSO model. Only SVR models show a linear negative relationship between the growth rate and initial GDP, which was to be expected.

In the case of ensemble models (Figure 2b)), we can see some interesting phenomenon. If we look at the GBM model and XGBoost model with Decision Trees learners, we can spot that there are plenty of intervals for initial GDP with the same annual growth rate. However, countries that belong to the "poorer" interval generally grow faster than those which belong to the "richer" interval. We can say that countries with the initial GDP in the given interval, belong to the same convergence club. In the case of XGBoost with OLS learners, we can see only three such intervals, with two among them that do not differ significantly. In the case of random forest, relation is not consistent – we can spot the convergence only among the richest countries.

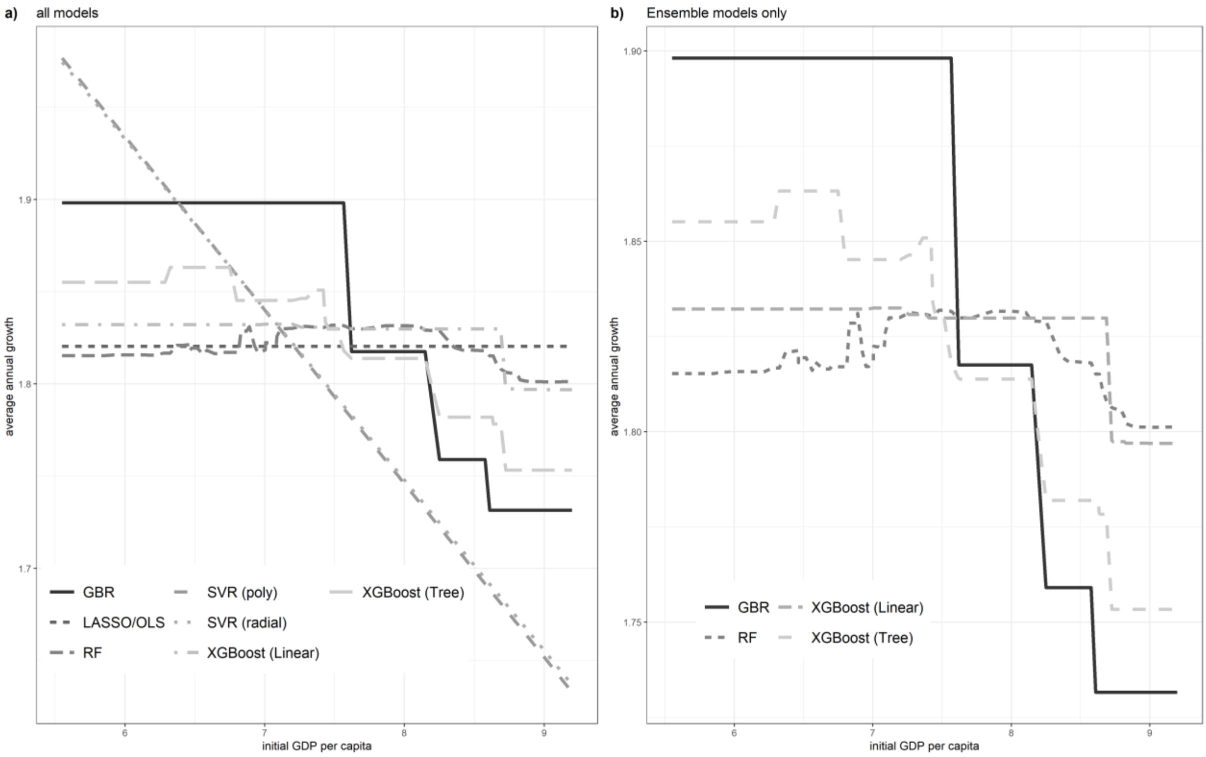

Figure 2. Partial Dependence Profiles for initial income for models estimated on SDM data

In the end, we again show the measures of fit to data for all models (Table 4). Here, only ensemble models explain more than 90% of the variability of growth. Again, the best model is the XGBoost with

OLS learners. However, again, we see the confirmation that models that allow for non-linearities perform better than their linear counterparts (LASSO, OLS or SVR).

Model	RMSE	MAE	R2
LASSO	1,0600	0,8245	0,6859
OLS / LASSO	0,9867	0,7613	0,7278
SVR (radial)	0,9312	0,6573	0,7576
SVR (poly)	0,9313	0,6576	0,7575
RF	0,5052	0,3690	0,9287
GBR	0,4065	0,3119	0,9538
XGBoost (Tree)	0,0670	0,0529	0,9987
XGBoost (Linear)	0,0020	0,0013	0,9999

Table 4. Measures of fit for all models estimated on SDM data

Conclusions

The main purpose of the article was to identify the important factors of economic growth by applying machine learning tools. We applied models which allow to identify non-linearities in the data, namely support vector regression, random forests and boosting algorithms. The models were estimated without any prior assumptions with the use of Leave-One-Out cross validation procedure. Moreover, we used the model reliance measure, which allows to easily assess the importance of features for any model type in a consistent and comparable way. To estimate our models, we used two common datasets – FLS data introduced in Fernandez et al. (2001) and SDM data introduced in Sala-i-Martin et al. (2004). Machine learning tools confirmed the importance of several growth factors, such as life expectancy, investment in the equipment and its price. They also pointed at some factors that were low in the rankings of previous studies using purely linear approach, i.e. ethnolinguistic fractionalization, which measure the ethnic and linguistic diversity in the country. The most striking result from our analysis was the difference between the conclusion about the importance of initial GDP when allowing for nonlinearity of its relationship with the growth rate. For FLS data, this factor was one of the most important, which was consistent with the previous studies. In turn, it dropped down the ranking for SDM data, and was even excluded in case of LASSO approach. This suggests that when using a simplified linear approach one can incorrectly conclude about the occurrence of conditional convergence, while when correctly identifying the non-linear relationship, cross-country convergence is not observed.

When analyzing Partial Dependence Profiles for initial GDP, we could also identify convergence clubs – groups of countries similar in terms of initial GDP per capita for which convergence is observed. We also showed that models that allow for non-linearities generally have higher predictive power and explained more variability of growth rates.

References

Barro R.J. and X. Sala-i Martin. Economic Growth. Prentice Hall of India Private Limited, 2007. ISBN 9788120325517.

Breiman Leo. Random forests. Machine. Learning, 45(1):5–32, October 2001. ISSN 0885-6125. doi: 10.1023/A:1010933404324.

Chen Tianqi and Carlos Guestrin. Xgboost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug 2016. doi: 10.1145/2939672.2939785.

Ciccone Antonio and Marek Jarocinski. Determinants of economic growth: Will data tell? American Economic Journal: Macroeconomics, 2(4):222–46, October 2010. doi: 10.1257/mac.2.4.222.

Doppelhofer Gernot and Melvyn Weeks. Robust Growth Determinants. CESifo Working Paper Series 3354, CESifo, 2011.

Durlauf Steven. The rise and fall of cross-country growth regressions. History of Political Economy - HIST POLIT ECON, 41:315–333, 12 2009. doi: 10.1215/00182702-2009-030.

Fernandez Carmen, Eduardo Ley, and Mark Steel. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics, 16(5):563–576, 2001.

Fisher Aaron, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 20(177):1–81, 2019.

Friedman Jerome H.. Greedy function approximation: A gradient boosting machine. Ann. Statist., 29(5):1189–1232, 10 2001. doi: 10.1214/aos/1013203451.

Goldstein Alex, Adam Kapelner, Justin Bleich, and Emil Pitkin. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24, 09 2013. doi: 10.1080/10618600.2014.907095.

He Qiuqin and Bing Xu. Determinants of economic growth: A varying-coefficient path identification approach. Journal of Business Research, 101:811 – 818, 2019. ISSN 0148-2963. doi: https://doi.org/10.1016/j.jbusres.2018.12.013.

Hendry David F. and Hans-Martin Krolzig. We ran one regression*. Oxford Bulletin of Economics and Statistics, 66(5):799–810, 2004. doi: 10.1111/j.1468-0084.2004.102_1.x.

Hofmarcher Paul, Jesús Crespo Cuaresma, Bettina Grün, and Kurt Hornik. Last night a shrinkage saved my life: Economic growth, model uncertainty and correlated regressors. Journal of Forecasting, 34(2):133–144, 2015. doi: 10.1002/for.2328.

Hoover Kevin D. and Stephen J. Perez. Truth and robustness in cross-country growth regressions*. Oxford Bulletin of Economics and Statistics, 66(5):765–798, 2004. doi: 10.1111/j.1468-0084.2004.101_1.x.

Leamer Edward. Let’s take the con out of econometrics. American Economic Review, 73:31–43, 01 1983.

Levine Ross and David Renelt. A Sensitivity Analysis of Cross-Country Growth Regressions. American Economic Review, 82(4):942–963, September 1992.

Ley Eduardo and Mark F.J. Steel. Jointness in bayesian variable selection with applications to growth regression. Journal of Macroeconomics, 29(3):476 – 493, 2007. ISSN 0164-0704. doi: https://doi.org/10.1016/j.jmacro.2006.12.002. Special Issue on the Empirics of Growth Nonlinearities.

Ley Eduardo and Mark F.J. Steel. On the effect of prior assumptions in bayesian model averaging with applications to growth regression. Journal of Applied Econometrics, 24(4):651–674, 2009. doi: 10.1002/jae.1057.

Molnar Christoph. Interpretable Machine Learning. 2019.

Sala-i Martin Xavier. I Just Ran Two Million Regressions. American Economic Review, 87(2):178–183, May 1997.

Sala-i-Martin Xavier, Gernot Doppelhofer, and Ronald I. Miller. Determinants of long-term growth: A bayesian averaging of classical estimates (BACE) approach. American Economic Review, 94(4): 813–835, September 2004. doi: 10.1257/0002828042002570.

Schneider Ulrike and Martin Wagner. Catching Growth Determinants with the Adaptive LASSO. Economics Series 232, Institute for Advanced Studies, November 2008.

Smola Alex J. and Bernhard Schölkopf. A tutorial on support vector regression. Statistics and Computing, 14(3):199–222, August 2004. ISSN 0960-3174. doi: 10.1023/B:STCO.0000035301.49549.88.

Tibshirani Robert. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996. ISSN 00359246.

Vapnik Vladimir N.. The Nature of Statistical Learning Theory. Springer-Verlag, Berlin, Heidelberg, 1995. ISBN 0387945598.