

SEEJE

South-Eastern Europe Journal of Economics

THE OFFICIAL JOURNAL OF THE ASSOCIATION OF ECONOMIC UNIVERSITIES
OF SOUTH AND EASTERN EUROPE AND THE BLACK SEA REGION

Vol 18 | No 1 | SPRING 2020

web site: <http://www.asecu.gr>
ISSN 1109-8597
ISSN On line: 1792-3115



FUNDING INSTITUTION
UNIVERSITY OF MACEDONIA
THESSALONIKI, GREECE

SEEJE

Editorial Office:

ASECU, University of Macedonia

156 Egnatia Str., 546 36 Thessaloniki, Greece

tel: +30 2310 891 793, fax: +30 2310 891 748

e-mail: asecu@uom.edu.gr

web site: <http://www.asecu.gr>

© ASECU

ISSN 1109-8597

ISSN On line: 1792-3115

Thessaloniki, Spring 2020

contents

Articles

- Revenue Diversification and Bank Performance: Evidence from Turkey**
BURCU BUYURAN, İBRAHİM HALİL EKŞİ 7
- Effect of Foreign Aid on Human Development in Sub-Saharan Africa:
A System GMM Approach**
SAKIRU OLADELE AKINBODE, TOBI MICHAEL BOLARINWA 19
- Forecasting Economic Recessions Using Machine Learning:
An Empirical Study in Six Countries**
ANDREAS PSIMOPOULOS 39
- Re-Examining the Stability of Money Multiplier for the US:
The Nonlinear ARDL Model**
ISMET GOCER, SERDAR ONGAN 101
- Analysis of the Efficiency of Tertiary Education Expenditure
in European Union Member States from Central and Eastern Europe:
An Efficiency Frontier Approach**
KRISTINA STEFANOVA, NIKOLAY VELICHKOV 115

Book Review

- Smart, Sustainable and Inclusive Growth***
Political Entrepreneurship for a Prosperous Europe
edited by Charlie Karlsson, Daniel Silander, Brigitte Pircher
reviewed by Theodore P. Lianos 129

aim and scope of

ASECU was founded in 1996 as Association of South-Eastern Europe Economic Universities with the general aim of promoting the interests of those economic universities in South-Eastern Europe which are public, recognized or financed by the state of origin.

By decision taken in General Assembly of 2007, it was modified *in Association of Economic Universities of South and Eastern Europe and the Black Sea Region*. Presently, Universities and Research Centers are included to the active Full members of ASECU from Albania, Armenia, Bosnia-Herzegovina, Bulgaria, Greece, Montenegro, North Macedonia, Poland, Romania, Russia, Serbia, Slovakia, Turkey and Ukraine. Also, as Associated members, Universities form Egypt, China, Hungary, Kazakhstan, Lebanon and Palestine.

The specific aims of the Association are:

To promote cooperation between Economic Universities, Faculties, Departments; i.e., especially: a) to exchange views and information about syllabi, b) to exchange undergraduate and postgraduate students and c) to exchange teaching and research staff.

To provide members with the opportunity exchange information, opinions etc. by publishing a relevant scientific journal or by cooperation in elaborating scientific studies in relation to the future development of higher education and research as well as to improve their quality in the field of economic studies and business administration.

To undertake initiatives for the protection of the interests of members and their institutions, so as to be supported by international organizations and in particular by the higher education institutions of the European Union.

To encourage cooperation between universities inside and outside the countries referred to in the Association.

To pursue cooperation in the field of higher education with the consolidation of close relations with other organisations having similar aims, e.g. E.U.A.;

To provide opportunities for harmonising the degrees of faculties and departments of the universities participating in the Association;

To promote cooperation between economic universities, faculties, departments in the field of research for the benefit of the economy, the society, peace and the cultural development of the countries referred to the Association.

Presidency and Board Members

The following Board was elected for a four years period, during the 7th General Assembly of ASECU, on the 30th of March 2019.

Grigorios Zarotiadis, President

University of Thessaloniki, Thessaloniki, Greece

Oleg Bodyagin, Vice President

Rostov State University of Economics, Rostov-on-Don, Russia

Leonid Nakov, General Secretary

“Ss Cyril and Methodious” University, Skopje, North Macedonia

Fatmir Memaj, Member

University of Tirana, Tirana, Albania

Dejan Mikerević, Member

University of Banja Luka, Bosnia - Herzegovina

Paskal Zhelev, Member

University of National and World Economy, Sofia, Bulgaria

Vesna Karadžić, Member

University of Montenegro, Podgorica, Montenegro

Bogdan Wierzbinski, Member

University of Rzeszów, Rzeszów, Poland

Zaklina Stojanović, Member,

University of Belgrade, SERBIA

REVENUE DIVERSIFICATION AND BANK PERFORMANCE: EVIDENCE FROM TURKEY

BURCU BUYURAN*
İBRAHİM HALİL EKŞİ

Gaziantep University, Gaziantep, Turkey

Abstract

Over the last three decades, the financial industry in developed as well as in developing countries has experienced major changes. One of these changes is revenue diversification on banking sector. The main purpose of our study is to examine the effects of income diversification on bank performance. Scope of Research is taken as the sample deposit banks operating in Turkey. Using the data of 14 banks between 2010 and 2017, variables were analyzed with dynamic panel data. Because of Herfindahl–Hirschman Index (will be addressed as HHI here after) widely used to measure diversification, we used HHI for analyzing the revenue diversification. In the model, the return on assets (ROA) was taken as the dependent variable representing the bank performance, and the criterion of revenue diversification was HHI (Harfindal Hirsman Index) as the independent variable and other control variables were added. The panel GMM technique was used because of its some features. According to the results; there is a negative significant relationship between HHI Index and bank performance. It means that revenue diversification has a positive effect on bank performance. Results of control variables are also largely consistent with expectations.

JEL Classification: G21

Keywords: Bank Performance, Income Diversification, Panel GMM

*Corresponding Author: Burcu Buyuran, Lecturer, Gaziantep University, Gaziantep, TURKEY.
E-mail: bbuyuran@gantep.edu.tr

1. Introduction

As the most important part of the financial system, banks play an intermediary role in lending surplus funds to deficit units. In this role, households, businesses and governments fall back on banks for credit. Thus, in well-functioning economies, banks tend to act as quality controllers for successful, capital-seeking projects, ensuring higher returns and enhancing growth.

Over the last three decades, deregulation and increased competition have led banks to expanding their activities and developing new lines of businesses besides their traditional interest activities. In relevant literature, such activities are known as diversification.

Banks' income diversification involves banks' activities to gain income not only from conventional interest sources, but also from non-interest sources, such as financial services provided by a bank to its customers, e.g., transfer and trading commissions, credit, e-banking, and so on (Syahyunan *et al.*, 2017).

In banking literature, it is known that revenue diversification, in general, reduces the risks of loan failure. This strategy leads to greater diversification of income sources, which might help banks reduce risks and stabilize profits. However, banking institutions may reach a point of disintermediation by expanding non-interest product activities. Some non-interest generating activities are associated with much higher risks than other income sources and, therefore, they could contribute towards the destabilization of both individual banks and the entire banking system (Brahmana *et al.*, 2018).

The impact of diversification on bank performance is neither theoretically nor empirically certain. According to the portfolio theory, diversified banks benefit from economies of scope that improve performance. Incomes from different sources, which are uncorrelated or imperfectly correlated with each other, result in steady and stable streams of overall bank profits. Otherwise, if the diversified activity is inherently riskier than traditional banking business, the costs of diversification may outweigh its benefits, and banks may become riskier and their overall performance may deteriorate (Nisar *et al.*, 2018). So, in our hypothesis, we will expect a positive relationship between revenue diversification and bank performance.

The essential motivation of this study is to test the effect of diversification on bank performance. In order to assess the effect of revenue diversification on bank profitability in emerging economies, we focused on the case of Turkey. This way, our results can be valid for similar countries, since banking systems in emerging countries have similar characteristics.

This study contributes to relevant literature by examining the strictly regulated Turkish banking sector with a new dataset. This relationship between bank performance and diversification has not been thoroughly examined for the case of Turkish deposit banks. In this aspect, our paper can bridge the gap in existing literature, as

it focuses on the effect of banking diversification in developed markets, yet neglects emerging markets like Turkey. To this end, we conducted our empirical investigation over 8 years using a sample of 14 commercial banks in Turkey.

In Turkey, the financial sector has grown at a tremendous rate over the last 35 years. During the 1990-2003 period, quite a high number of bank failures occurred due to structural problems of the Turkish economy and the fragilities of the Turkish banking sector. As of May 2018, there are a total of 50 banks in the Turkish banking sector, including 32 deposit banks, 13 development and investment banks and 5 Islamic banks (BAT, 2018). However, some bank data are not comprehensive. In addition, since some banks probabilities/profitabilities are negative, HHI is not an applicable measure for these banks. As a result, our sample consists of 14 banks.

We hope that this paper will contribute to relevant literature, especially concerning cases of emerging economies, by identifying the relationship between diversification and bank performance. It is expected to provide useful information about Turkey and similar emerging countries. This study proceeds as follows: literature review, data methodology and results of models.

2. Literature

There are two theories in relation to revenue diversification, namely, the resource-based theory and the risk reduction theory. The diversification decision may be related to the efficiency and risk management of a bank, where joint production of a wide range of financial services should increase the bank's efficiency due to economy of scale (Brahmana *et al.*, 2018).

Previous studies on bank revenue diversification have mainly focused on the benefits of diversification. Three aspects have been observed in literature. The first one concerns the relationship between income diversification and operating performance (Gürbüz *et al.* 2014; Meslier *et al.*, 2014; Alhassan, 2015; Brahmana *et al.*, 2018). The second one concerns the relationship between income diversification and bank risk (Zhou, 2014; Edirisuriya *et al.*, 2019). The third aspect of bank diversification is its effect on bank stability (Amidu and Wolfe, 2013; Nguyen *et al.*, 2012; Syahyunan *et al.*, 2017; Dwumfour, 2017; Abuzayed *et al.*, 2018) . We present a few studies below:

Chiorazzo *et al.* (2008) studied the correlation between non-interest revenue sources and profitability for Italian banks. They found that income diversification increases risk-adjusted returns.

Türkmen and Yiğit (2012) examined the effect of sectoral and geographical diversification on the performance of Turkish banks and tried to show how diversification affects it. The authors used ROA and ROE as measures of performance and Herfindahl Index (HI) as a measure of bank diversification. Results indicated that dependent variables are explained by diversification.

Nisar *et al.* (2018) investigated the impact of revenue diversification on bank profitability and stability in South Asian countries. Overall revenue diversification into non-interest income was found to have a positive impact on the profitability and stability of South Asian commercial banks.

Meslier *et al.* (2014) examined the impact of bank revenue diversification on the performance of banks in an emerging economy, and results indicated that foreign banks benefit more from such a shift than their domestic counterparts.

Alhassan (2015) investigated the non-linear relationship between income diversification and efficiency of Ghanaian banks. His results revealed high levels of efficiency in cost compared with profit to reflect high inefficiencies on the revenue side.

Sissy *et al.* (2017) analyzed the implications of revenue diversification and cross-border banking for risk and return in 29 African countries and results suggested that banks cross borders to diversify across revenue-generating activities. The authors' analyses further showed that banks in Africa derived absolute benefits from diversification if they cross borders while concurrently diversifying their revenue base.

Brahmana *et al.* (2018) investigated the diversification effect on banks' performance using Malaysian banks. In their study, panel regression results showed that income diversification increases a bank's performance confirming the risk reduction hypothesis.

Gürbüz *et al.* (2013) investigated the relationship between non-interest income generating activities and risk-adjusted bank performance; his investigation used GMM for Turkish deposit banks. Authors' results showed that income diversification increases the risk-adjusted financial performance of Turkish deposit banks.

There are a variety of studies that analyzed diversification and bank performance. Deregulating initiatives, which took place in both Europe and the U.S. during the last decades, resulted in an expansion of the scope of bank activities and a shift from traditional to non-traditional sources of income (Meslier *et al.*, 2014). So, a large body of research focuses on the impact of diversification for banks in developed countries, such as the U.S. and Europe (Chiorazzo *et al.*, 2008; Williams and Prather, 2010; Busch and Kick, 2009; Căpraru *et al.*, 2018). While in emerging economies a lot of papers analyze the effect of income diversification on bank performance (such as, bank profitability, risk, stability) in the case of developed countries, during the last years only a few papers have addressed this issue (Grassa, 2012; Nguyen *et al.*, 2012; Amidu and Wolfe, 2013; Adzobu *et al.* 2017, Alhassan, 2015; Sissy *et al.*, 2017; Khalatur *et al.*, 2018; Bapat, 2018; Nisar *et al.*, 2018). However, the empirical relationship between income diversification and bank performance has been found to vary in such studies (Chiorazzo *et al.*, 2008; Molyneux and Yip, 2013; Nisar *et al.*, 2018). Some studies have found evidence of a positive diversification effect on bank performance, like those by Busch and Kick, 2009; Turkmen and Yiğit, 2012; Gürbüz *et al.*, 2013; Gambacorta *et al.*, 2014. In contrast, there are other studies that did not

find any relationship between diversification and bank performance, such as those by Bapat, 2017; Adzobu *et al.*, 2017.

Our study is different from similar studies with regard to the methodology (both difference GMM and system GMM) and sample used.

3. Data and Methodology

Our database consists of 112 reports of annual bank data over the 2010-2017 period. Because of mergers, acquisitions, and banks being closed, it has been impossible to have some of the bank data from 2010-2017. And because of their different working principles, we removed Islamic banks from the sample population. In this manner, the sample finally consisted of 14 banks' annual data. Data came from Istanbul Stock Exchange (ISE) and from the banks' consolidated financial statements posted on their respective web pages for the years studied.

The endogeneity problem has been emphasized in some studies investigating bank performance and diversification, (Acharya *et al.*, 2006; Gürbüz *et al.*, 2013). One ignored variable (e.g., a management skill or the location of the bank) can affect both the income diversification level and the bank's performance. In addition, past and current performance can affect the decision to diversify and vice versa. The endogeneity problem can cause biased estimates in the analysis. In order to overcome a possible endogeneity problem in this study, we used dynamic panel data (Generalized Method of Moments-GMM), for several reasons. Primarily, the main purpose of using dynamic panel data is that the lagged values of dependent variables resulting from the fixed and random effects models and the estimators reached are inconsistent, since the lagged dependent variable is correlated with the error term when lagged dependent variables are used in the fixed and random effect models. This situation has also been observed in studies on this subject in relevant literature (Coşkun and Kök, 2011; Béjaoui and Bouzgarrou, 2014). Furthermore, the financial data used in the model can show highly dynamic effects depending on time (Tunay, 2014).

Differenced and system GMM estimators are considered appropriate for a dynamic panel dataset containing a small t (8 years) and a large N (14 banks), with unobserved fixed-effects and endogeneity between dependent and independent variables (Arellano and Bond 1991; Arellano and Bover 1995; Blundell and Bond 1998).

3.1 Empirical Model

The baseline regression for bank ROA is given by

$$ROA = \beta_0 + \beta_1 ROA_{i,t-1} + \beta_2 DTA_{i,t} + \beta_3 ETA_{i,t} + \beta_4 HHI_{i,t} + \beta_5 LTA_{i,t} + \beta_6 NPL_{i,t} + \varepsilon_{it} \quad (1)$$

Description of the database is presented in the table below ;

Table 1. Description of Database

VARIABLE	SYMBOL	DEFINITION
BANK PERFORMANCE		
Return On Assets	ROA	The Ratio of Net Profit to Total Assets
DIVERSIFICATION		
HHI Index	HHI	The Sum of The Squares of the Share of Net Interest Income and the Share of Non-Interest Income over Net Operating Income
BANK-SPECIFIC CONTROL VARIABLES		
Bank equity	ETA	The Ratio of Equities to Assets
Bank deposit	DTA	The Ratio of Deposits to Assets
Bank size	LTA	The logarithmic for Total Assets
Non Performing Loan	NPL	The Ratio of Non-Performing Loans to Total Loans

We used HHI to measure diversification. This index is widely used for analyzing the diversification (Mercieca *et al.*, 2007; Gürbüz *et al.*, 2013; Amidu and Wolfe, 2013; Sissy *et al.*, 2017).

We used return-on-active (ROA) to test bank effectiveness and performance. ROA and other control variables are also used in various studies, such as those by Turkmen and Yiğit (2012), Beck *et al.* (2013), Acharya *et al.* (2006). In order to show profitability of all assets, we prefer ROA instead of ROE or NIM.

There are differences among sample banks with respect to assets, profitability, and other characteristic differences, which affect empirical results. By including control variables in the models, we tried to ensure that there is no independent variable, such as equity, deposit, size and NPL, excluded.

The asset variable is used to measure bank size. According to literature, larger banks may have better risk management and diversification opportunities; on the other hand, small banks are more flexible in their operations (Gürbüz *et al.*, 2013). There are a lot of papers that use the variable of bank size (Gürbüz *et al.*, 2013; Zhou, 2014).

To measure the financial leverage degree of a bank, the bank equity variable is added, following Gürbüz *et al.* (2013), Zhou (2014), Edirisuriya *et al.* (2019). A higher ratio of equity/total assets reflects risk aversion and protection against bank default risk.

The deposit variable is used to determine a bank's passive structure, following Zhou, 2014, Abuzayed *et al.* 2018. Besides, it is expected that the deposit amount has positive effect on bank performance.

NPL is a standard and widely used statistical value to measure the financial performance of a banking institution, as shown by Nguyen *et al.* 2012, Brahmana *et al.* 2018, Bapat 2018.

3.2 Empirical Results

Summary descriptive statistics of variables are shown in Table 2.

Table 2. Summary Statistics

VARIABLE	MEAN	STD.DEV.	MIN	MAX
ROA	0.012	0.007	-0.016	0.027
ETA	0.11	0.023	0.072	0.193
DTA	0.6	0.088	0.251	0.832
NPL	0.036	0.016	0.008	0.08
LTA	7.677	0.65	5.955	8.559
HHI	0.616	0.093	0.5	0.887

After determining the model the correlation between independent and dependent variables was tested. The matrix of correlation values for the series is shown in Table 3. There is no high correlation between variables.

Table 3. Correlation matrix

VARIABLES	ROA	ETA	DTA	NPL	LTA	HHI
ROA	1					
ETA	0.313	1				
DTA	0.12	0.197	1			
NPL	-0.07	0.199	0.004	1		
LTA	0.491	-0.27	-0.2	-0.17	1	
HHI	-0.24	-0.05	-0.2	-0.13	-0.35	1

First generation panel unit root tests were performed in order to determine the stability of the series.

Panel Unit root tests result are shown in Table 4. According to panel unit root test results, our variables are stable on their level values.

In order to decide whether or not there is cross-sectional dependence in the model, Pesaran's test of cross-sectional independence is performed. According to the results of this test, there is no cross-sectional dependence in either model.

In order to determine whether there is an autocorrelation problem between variables, Wooldridge test for autocorrelation in panel data was performed. According to the results of this test, no autocorrelation problem was found between variables.

Table 4. Panel Unit Root Tests

Variables	<i>Im Pesaran Shin</i>		<i>PP-Fisher ChiSquare Tests</i>	
	<i>Stat</i>	<i>prob</i>	<i>stat</i>	<i>prob</i>
ROA	-2.0411	0.0206	-6.1392	0.000
ETA	-4.4300	0.0000	-6.1724	0.000
DTA	-2.3314	0.0099	-1.9417	0.0275
NPL	-2.5522	0.0054	-6.6056	0.0000
LTA	-2.1133	0.0173	-6.6138	0.0000
HHI	-2.8258	0.0024	-4.0892	0.0001

To determine whether there is a heteroscedasticity problem between variables, Breusch-Pagan/Cook-Weisberg test for heteroskedasticity was performed. According to the results of this test, the heteroscedasticity problem is seen in our model.

Table 5. Heteroscedasticity Autocorrelation Cross Sectional Dependence Problem

	Heteroscedasticity		Autocorrelation		Cross -Sectional Dependence	
	Breusch-Pagan / Cook Weisberg test		Wooldridge test for autocorrelation		Pesaran's test of cross-sectional independence	
	chi2(1)	Prob	F Value	Prob	Value	Prob
Model	0.84	0.3589	4.748	0.0483	6.331	0.0000

Based on all of these test results, it is possible to say our variables are robust. The results of our regression models are presented in table 6;

The first hypothesis of the Arellano-Bond test, which Arellano and Bond developed to test for the presence of autocorrelation of dynamic panel data models, is “no autocorrelation”. In order for GMM estimators to be effective, there should be no second-order autocorrelation in the remains (Tatoğlu, 2012). According to the findings, there is no first and second order autocorrelation in the remains of either model. The Sargan test deals with the validity of instrumental variables and it is a test involving overidentifying restrictions. The calculated values of Sargan test also support the analysis. Both lag values for profitability are statistically significant. It means that the previous year profitability is a factor in estimating the current year performance.

According to GMM results, the HHI variable is statistically significant for bank performance. In addition to HHI, with the exception of the NPL, all control variables are important for bank performance.

According to system GMM results, the HHI variable and all control variables are statistically significant on ROA.

According to both models, banks should make revenue diversification for their profitability. Compared to both models, it is seen that system GMM results are more valid than difference GMM ones.

Table 6. Regression Results

INDEPENDENT VALUES	DEPENDENT VARIABLE:	ROA		ROA	
		DIFF. GMM		SYSTEM GMM	
		Coef	prob	Coef	prob
	ROA (t-1)	0.313	0.025**	0.185	0.010**
ETA	0.216	0.000***	0.190	0.000***	
DTA	(0.013)	0.075*	(0.015)	0.056**	
NPL	(0.064)	0.370	(0.093)	0.058**	
LTA	0.009	0.095*	0.003	0.092*	
HHI	(0.026)	0.006***	(0.013)	0.028**	
Cons	(0.057)	0.156	(0.015)	0.378	
Specification Tests					
Number of Groups	14		14		
Number of Obs	84		84		
Wald x^2	70.57		106.39		
ProbChi2	0.000		0.000		
Sargan Test	26.44		11.33		
P Value	0.1518		0.99		
AR(1)	0.1147		0.1493		
AR(2)	0.6896		0.8451		

All models are estimated using GMM and System GMM, which are significant in all cases AR (1): a test of null of zero first-order serial correlation, distributed $N(0, 1)$ under the null. AR (2) test of null of zero second-order serial correlation, distributed $N(0, 1)$ under the null. Wald statistics: the test is a way of testing the significance of particular explanatory variables in a statistical model. Sargan test for validity of over-identifying restrictions, distributed as indicated under null. This test of over-identifying restrictions is asymptotically distributed as χ^2 under the null of instrument validity.

***Denote significance at 1%, respectively. $p < 0.01$,

**Denote significance at 5%, respectively. $p < 0.05$

*Denote significance at 10%, respectively. $p < 0.1$

4. Conclusion

Especially in recent years, deposit banks have diversified their incomes all around the world, including Turkey. So, there are a lot of papers investigating the effect of diversification on bank performance.

The paper examines the impact of income diversification and some control variables (such as deposit/total asset, equity /total asset, credit/total asset and the log of total asset and non-performance loans) on bank performance. To this end, our study concentrated on the micro bank-level. We used data about 14 deposit banks for the 2010-2017 period. Depending on previous studies in relevant literature, we

use HHI for analyzing diversification. We used ROA as a performance dimension. We used the panel GMM technique due to both its advantages and our data characteristics.

According to results, there is a negative effect of HHI Index on bank performance indicator ROA. It is important that the results of both models are similar. By its formula, the fall of HHI is meant to reflect an increase in diversification. It means that there is a positive relationship between diversification and bank performance. As we expected, in addition to HHI, there is a positive significant effect of the ratio of equity to assets on bank performance. It has been shown that there is a negative relationship between the ratio of deposits-to-total assets and bank performance. There is a negative significant effect of non-performing loans to banks performance on system GMM. Similarly, it has been shown that there is a positive significant effect of the log of total assets on bank performance. Results of control variables are also largely consistent with expectations.

Our analysis findings have one main implication for regulators, bank managers and investors concerning income diversification in Turkish banks. The positive effect of income diversification on banking performance may be a result of increased income of the bank or reduced operating costs of the bank brought about by diversifying operations.

Lastly, it should not be ignored that we have a relatively limited sample period and all sub-categories of non-interest income generating activities as a whole are limited, too. In future studies on the effects of income diversification, a longer sample period can be used and the effects of sub-categories of non-interest income generating activities can also be investigated. In addition, investigations should make a distinction between highly diversified revenue and low diversified revenue.

References

- Abuzayed Bana, Al-Fayoumi , N. Phil Molyneux (2018).Diversification And Bank Stability In The GCC, *Journal Of International Financial Markets, Institutions & Money*, 57 (2018) 17–43.
- Adzobu L. D. Agbloyor E.K. And Aboagye . A. (2017). The Effect Of Loan Portfolio Diversification On Banks' Risks And Return: Evidence From An Emerging Market, *Managerial Finance* Vol. 43 No. 11, 2017 Pp. 1274-1291.
- Acharya V. I. Hasan, A. Saunders. (2006), .Should Banks Be Diversified? Evidence From Individual Bank Loan Portfolios *J. Bus.*, 79 Pp. 1355-1413.
- Alhassan Abdul Latif. (2015) Income Diversification And Bank Efficiency In An Emerging Market, *Managerial Finance*, Vol. 41 No. 12, 2015 Pp. 1318-1335.
- Amidu, M. Wolfe S.(2013) Does Bank Competition And Diversification Lead To Greater Stability?Evidence From Emerging Markets, *Review Of Development Finance* 3 152–166.
- Arellano, M. And Bond, S. (1991). Some Tests Of Specification For Panel Data: Monte Carlo Evidence And An Application To Employment Equations” *The Review Of Economic Studies*, Volume 58, Issue 2, Pages 277–297.
- Arellano, M. And Bover, O. (1995) Another Look At The Instrumental Variable Estimation Of Error-Component Models. *Journal Of Econometrics*, 68, 29-52.

- Bapat Dhananjay (2018) Profitability Drivers For Indian Banks: A Dynamic Panel Data Analysis, *Eurasian Bus Rev* (2018) 8:437–451.
- Beck T., A. Demirguc-Kunt, O. Merrouche (2013), Islamic Vs. Conventional Banking: Business Model Efficiency, And Stability *J. Bank. Finance*, 37 (2) Pp. 433-447.
- Béjaoui, R. - Bouzgarrou H. (2014). Determinants Of Tunisian Bank Profitability. *The International Journal Of Business And Finance Research*, V. 8 (4) P. 121-131, 2014.
- Busch, R. And Kick, T.. (2009). Income Diversification In The German Banking Industry, *Deutsche Bundesbank Eurosystem. Discussion Paper. Series 2: Banking And Financial Studies. No: 09/2009.*
- Brahmana, R. Kontesa M. And Gilbert R. E. (2018) .Income Diversification And Bank Performance: Evidence From Malaysian Banks, *Economics Bulletin*, Volume 38, Issue 2 Volume 38, Issue 2
- Căpraru, B. Ihnatov, I. Pintilie, N.L. (2018). Competition And Diversification In The European Banking Sector Research In *International Business And Finance* Available Online 28 September 2018.
- Chiorazzo, V., Milani, C. And Salvini, F. (2008). Income Diversification And Bank Performance: Evidence From Italian Banks. *Journal Of Financial Services*. 33: 181-203.
- Coşkun Ender And Kök Dünder (2011) Effect Of Working Capital Policies On Profitability: A Dynamic Panel Analysis, *Ege Akademik Bakış*, Cilt:11, 75-85.
- Dwumfour, R.A. (2017) Explaining Banking Stability In Sub-Saharan Africa, *Research In International Business And Finance* 41 (2017) 260–279.
- Edirisuriya, P., Gunasekarage, A. And Shrimal P. (2019) Product Diversification And Bank Risk: Evidence From South Asian Banking Institutions, *APPLIED ECONOMICS* 2019, VOL. 51, NO. 5, 444–464.
- Gambacorta L., Scatigna M., Jing Yang (2014). Diversification And Bank Profitability: A Nonlinear Approach, *Applied Economics Letters*, Vol. 21, No. 6, Pp.438–441.
- Grassa R. (2012) .Islamic Banks' Income Structure And Risk: Evidence From GCC Countries, *Accounting Research Journal*, Vol. 25 Issue: 3, Pp.227-241.
- Gürbüz, A.O, Yanık, Serhat. AYTÜRK, Y. (2013) Income Diversification And Bank Performance: Evidence From Turkish Banking Sector, *BDDK Bankacılık Ve Finansal Piyasalar* Cilt:7, Sayı:1, 2013.
- Zhou Kaiguo (2014) The Effect Of Income Diversification On Bank Risk: Evidence From China, *Emerging Markets Finance And Trade*, Vol. 50, Supplement 3, Pp. 201–213.
- Khalatur, S. Kateryna, Z. Masiuk Y. Velychko, L. Kravchenko M. Assessment Of Bank Lending Diversification In Ukraine, *Banks And Bank Systems*, Volume 13, Issue 3.
- Mercieca, S., Schaeck K. And Wolfe S.. (2007). Small European Banks: Benefits From Diversification?. *Journal Of Banking And Finance*. 31: 1975-1998.
- Meslier, C., R. Tacneng, And A. Tarazi, Bank Diversification, Risk And Profitability In An Emerging Economy With Regulatory Asset Structure Constraints: Evidence From The Philippines. *Archives Ouvertes*, 2014: P. 1-55.
- Molyneux, P. And Yip J. (2013). Income Diversification And Performance Of Islamic Banks, *Journal Of Financial Management, Markets And Institutions*, Vol. 1, N. 1, 47-66.
- Nisar, S., Peng, K., Wang, S., & Ashraf, B. N. (2018). The Impact Of Revenue Diversification On Bank Profitability And Stability: Empirical Evidence From South Asian Countries. *International Journal Of Financial Studies*, 6(2), 40.
- Nguyen, M., Skully, M. Perera. S. (2012). Bank Market Power And Revenue Diversification: Evidence From Selected ASEAN Countries, *Journal Of Asian Economics* 23 Pp.688–700.
- Sissy, A. M., Amidu, M., & Abor, J. Y. (2017). The Effects Of Revenue Diversification And Cross Border Banking On Risk And Return Of Banks In Africa. *Research In International Business And Finance*, 40, 1-18.

- Syahyunan. Muda, S. I., Siregar,H.S., Sadalia,I,, Chandra,G. (2017) . The Effect Of Lerner Index And Income Diversification On The General Bank Stability In Indonesia, Banks And Bank Systems, Volume 12, Issue 4, Pp.56-64.
- Tatođlu, F. Y. (2012). İleri Panel Veri Analizi (Stata Uygulamalı), Beta Publishing, İstanbul.
- The Banks Association Of Turkey (2018) Bankalarımız 2017, https://www.tbb.org.tr/content/upload/dokuman/7519/bankalarimiz_2017.pdf
- Tunay Nejla (2014) Bancassurance Applications In Turkey And Its Dynamic Effects Of Turkish Banking And Insurance Sectors, BDDK Bankacılık Ve Finansal Piyasalar Cilt: 8, Sayı: 1, 2014
- Tunay, K. B., (2014). Türkiye'de Banka Karlılıđının Belirleyicileri: Ölçek Büyüklükleri, Mülkiyet Ve Finansal Krizler Temelinde Yeni Bir Analiz Bankers Journal, Vol. 91, 3-35.
- Turkmen, S.Y. And Yiđit, I. (2012) Diversification In Banking And Its Effect On Banks' Performance: Evidence From Turkey, American International Journal Of Contemporary Research Vol. 2 No. 12;
- Williams B. Prather,L.(2010). Bank Risk And Return: The Impact Of Bank Non-Interest Income, International Journal Of Managerial Finance Vol. 6 No. 3, Pp. 220-244.

EFFECT OF FOREIGN AID ON HUMAN DEVELOPMENT IN SUB-SAHARAN AFRICA: A SYSTEM GMM APPROACH

SAKIRU OLADELE AKINBODE*

TOBI MICHAEL BOLARINWA

Federal University of Agriculture, Abeokuta, Nigeria

Abstract

Using data for 47 SSA countries from 2000 to 2016, the study examined the effect of foreign aid on human development in SSA by employing the System-GMM approach, which is specifically applicable to the present case. Results revealed that aid did not affect human development in SSA, whereas, corruption was found to reduce HDI, while trade openness improved it. Validity of results was confirmed by the Arellano-Bond test for autocorrelation in the disturbance term and the Hansen and Sargan tests for the validity of instrumental variables. The study recommended effective framework for utilization of foreign aid and reduction of corruption.

JEL Classification: O11, O15, O55

Keywords: Human Development, Foreign Aid, Sub-Saharan Africa (SSA)

**Corresponding Author:* Dr. Sakiru Oladele AKINBODE, Department of Economics, Federal University of Agriculture, P.M.B 2240, Abeokuta, Nigeria.
E-mail: deleakinbode@yahoo.com; akinbodeso@funaab.edu.ng

1. Introduction

Human development has to do with expanding people's choices, enhancing people's capabilities and improving the opportunities available to them. It is both a means and an end. Human development is more than considering economic growth as the latter and increases in income are avenues towards achieving human development but not an end in themselves. This is more so as it is the wealth of the people, rather than that of the economies, that is, after all, valuable to people (Human Development Report, 2016). According to UNDP (1997), human development is "the process of enlarging people's choices"; the choices being referred to in this paper allow people to be better educated, to be healthy and have longevity of life, to be able to enjoy a relatively decent standard of living, as well as political freedom, various components of self-respect and other guaranteed human rights.

Sen (1999) posited that "the usefulness of wealth lies in the things that it allows people to do - the substantive freedoms it helps people to achieve". Therefore, the appropriate view of development must indeed go beyond wealth accumulation and continuous increase in GNP. Thus, human development that includes living standard is a far better measure of the quality of life of the people than growth in per capita income, which only measures the wealth of nation that may not be evenly distributed due to inherent inequalities within a country. While some developing economies have made progress in human development since the beginning of the third millennium, a good number are still far behind in most basic development indicators, including education, health, access to clean and safe drinking water, good food and modern sanitation. Sub-Saharan Africa (SSA) is included in the latter category as a developing region, which, according to Human Development Report (2016), since the inception of the Human Development Index (HDI) had been a low human development region (from 1990-2010), but managed to move up the ladder and be categorized as a medium human development region (2011- 2015). In fact, the HDR (2005) reports pointed out that, among the eighteen countries in which human development reversals occurred between 1990 and 2003, twelve were in SSA and the rest were part of the former Soviet Union. The report linked this obvious lack of or slow HDI progress with lack of improvement in education, to economic stagnation and the continuous spread of HIV/AIDS. Other possible causes may also be badly or poorly implemented developmental policies, conflicts, inter-ethnic and civil wars common in Africa, especially in the 1990s.

In Africa, international aid remains one of the powerful weapons available to achieve development goals, particularly in SSA countries over time. For instance, from 2002 to 2009, Africa received the largest part of Official Development Assistance (ODA) from donor countries. Out of approximately \$483 billion received worldwide, about 35 percent was given to African countries (OECD, 2010). Similarly, net disbursements of official development assistance to sub-Saharan Africa rose significantly

between 2000 and 2016 (OECD Statistics). Despite the increase in ODA inflows, the average growth of HDI was 1.67 from 2000 to 2010 and 1.04 from 2010 to 2015 (HDR, 2016).

Mcgillivray and Noorbakhsh (2004) established that aid contributes to aggregate well-being through wage increase as a result of demand for more labour and increases in public and private expenditure on education and health. Aid might be directly used in financing health and education, since most developmental assistance, in the first instance, comes into government accounts in developing countries. This way, aid affects human development indicators in manners that promote human development. The effectiveness of aid can be linked with human development because of its role in supplementing the domestic resource gap, thereby financing public investment in social services, which is directly linked with the welfare of the people. Human capital accumulation is facilitated by channelling foreign aid to education and health facilities, as such, will enhance standards of living if the masses have access to these basic services. However, the pertinent question is the following: 'if aid has the potential to improve human development and growth in the economy, why is the sub-Saharan African region receiving the largest proportion of aid still characterized by low human development, as revealed by the HDI figures?'

The government or its agencies, being the caretakers and users of aid funds, is/are to be directly responsible for using aid efficiently and effectively in order to achieve the human development that is inclusive in nature for the people in SSA countries. However, records and events have shown that, despite the increase in foreign aid to this region, the region is challenged with high ineffectiveness of government and high corruption rates among its government officials. Figures from OECD reports show that foreign aid in the sub-Saharan African region has been increasing over time since the beginning of the century as the growth on average in 2013 reached 6.1 percent in real terms. Meanwhile, a 2002 study by the African Union estimated the cost of corruption in the continent to be around \$150 billion a year; this is quite higher than foreign aid to Sub-Sahara African (SSA) economy, which was around \$134.8 billion dollars from 2000 to 2013. Corruption affects the disbursement of aid to ensure development as many of these countries have highly corrupt government officials whose actions lessen the effectiveness of aid. According to Transparency International (2017), on a scale of 0 to 100, over 21 sub-Saharan countries scored below 30, which shows that the rate of corruption is high in these countries.

Considering the fact that foreign aid inflow into sub-Saharan African countries has been increasing over the years and few or conflicting views are available on the effect of such ODAs on human development in the region; the present study aims at bridging the obvious knowledge gap. Specifically, the study assessed the effect of foreign aid alongside other control variables, such as government effectiveness and corruption, among others, on human development in SSA. The second section of

this paper reviews past literature, while the third section describes the study methodology. The fourth section presents and discusses the results, while the fifth section summarizes and concludes, accordingly.

2. Literature Review

Studies abound on the nature of the effects of foreign aid on development or on some indicators of development in countries and regions around the world, especially in developing and some emerging economies. Such studies have employed several methods to analyse and examine relationships between these two variables to each other or to other macroeconomic variables with varying objectives and types of data. In these studies, some techniques have remained prominent as they have been considerably used in existing literature. These include Ordinary Least Squares (OLS), Two Stage Least Squares (2SLS), Three Stage Least Square (3SLS), Generalized Method of Moment (GMM), Quartile Regression Approach, Vector Autoregression (VAR), Three Stage Least Squares (3SLS), Instrumental Variable (IV) method etc. While some of the studies have found a positive relationship between foreign aid and human development or some of its indicators, others reported a negative relationship. Some studies reported conditional effect, while others found no effect. This section firstly reviews past empirical works on aid effectiveness and, thereafter, reviews several methodologies that have been used in the literature concerning aid effectiveness.

2.1 Empirical Review

Studies concerned with aid effectiveness in the past have mostly concentrated on how foreign aid could affect economic growth. Only few have actually studied the effect of foreign aid on human development, while some have focused on components of human development.

Boone (1996) examined how effective foreign aid programmes were considering political regimes in recipient countries from a panel of ninety-six countries from 1970 to 1990. The study found that aid effectiveness was independent of the government of the countries - whether they were liberal or repressive. It was further reported that poor people did not benefit much from aid, since it failed to impart human development indicators and did not significantly impact growth and investment.

Burnside and Dollar (2000) assessed the relationship between foreign aid, policies that have to do with economy and per capita GDP growth using a panel of fifty-six different countries from 1970 to 1993. Study results showed that there was a positive relationship between aid and economic growth in developing countries that adopted sound fiscal, monetary and trade policies, but had little or no effect where poor policies were adopted. Kosack (2003) assessed the condition necessary for aid effectiveness in improving people's quality of life and found that aid affected the indicators of life quality positively and it is very significant in democratic countries and negative in

autocracies. Gomanee *et al.*, (2003) examined the assertion that “aid can reduce the level of poverty by financing public expenditure”, which is likely to be of great benefit to the poor in a sample of 38 countries studied from 1980-1998. The study also built a pro-poor expenditure index using regression to derive the weighted value of each element in the pro-poor expenditure indicator. It was reported that aid inflows and pro-poor expenditure are associated with higher welfare at all quintiles, i.e. they have greater direct impact on the human development index but an inversely proportional relationship with infant mortality.

Mcgillivray and Noorbakhsh (2004) conducted a study on ninety-four developing countries from 1980 to 2000 to examine the way aid and conflict impact human development. The study reported a significant negative effect of conflict and aid on HDI. Furthermore, it was surprisingly reported that aid effectiveness was not influenced by conflict circumstances. Kumler (2007) used a panel data consisting of 87 countries from 1980 to 2000 to examine the effectiveness of foreign aid in improving the level of human development and aggregate welfare and found that foreign aid has a negative relationship with human development. Asiama and Quartey (2009) examined how development aid affected welfare variables in thirty-nine Sub-Saharan African nations from 1975 to 2003 and reported that aggregate bilateral aid may have a positive effect, but did not show any significant effect on human development; on the other hand, financial sector development aid had a negative and significant effect on the human development indicator (specifically, on infant mortality rate).

Akinkugbe and Yinusa (2009) assessed the effectiveness of technical cooperation in improving state capacity for development for attaining the growth and development desired using a panel of forty-eight Sub-Saharan African countries from 1990 to 2007. Mixed effect was discovered on the link between technical assistance and human development in SSA. Gillanders (2011) examined the likely effect of foreign aid in SSA in a balanced panel of thirty-one SSA countries from 1973 to 2005 and reported that growth of the economy responded more to aid shocks in groups characterized by high level of aid dependency, poor or weak institution and better economic policies. It was further reported that human development responded positively to aid shocks in democracies and in good institutional environments. The finding corroborates that by Burnside and Dollar (2000) on the conditional effectiveness of foreign aid.

Okon (2012), using data from 1960 to 2010, carried out an empirical study on the effectiveness of ODA in determining the level of human development in Nigeria and reported a negative relationship between development aid and human development in this country. David (2017) examined the impact of ODA on poverty reduction within ECOWAS for a panel data from 1980 to 2014 and reported a negative but significant relationship between infant mortality rates (the proxy used for poverty level in the study) and ODA. It was concluded that foreign aid was indeed pro-poor but did not enhance growth in West African countries. Williamson (2018) examined the

likely effect of foreign aid (specifically to the health sector) for countries that received health aid from 1973 to 2004 and reported, after controlling for quality of institution and GDP, that foreign aid specific to the health sector did not significantly improve the overall health outcome among recipient countries. The study also reported that aid came out with the expected sign but was not statistically significant on health development indicators.

2.2 Methodological Review

Studies such as those by Boone (1996), Kosack (2003), Akinkugbe and Yinusa (2009) used the OLS technique involving fixed and random effects estimation. These studies reported different results. While Boone (1996) found aid not to be of benefit to the poor, Kosack (2003) found aid to be beneficial depending on the type of government practiced in such an aid-recipient country: it was more beneficial for democratic governments, without, however, improving the quality of life under authoritarian governments. McGillivray and Noorbakhsh (2004), Kumler (2007), and Okon (2012) examined the effect of foreign aid on human development. These studies used the 2SLS approach and found foreign aid to be negatively affecting human development indicators. A similar approach, namely the three stage least square (3SLS) used by David (2017), found that ODA negatively impacted infant mortality. However, Burnside and Dollar (2000), using the 2SLS approach found aid to be more effective in countries with better economic policies and less effective in countries with unsound economic policies. The result of Burnside and Dollar (2000) was re-examined by Dalgaard and Hansen (2012) using the same procedures but including some variables necessary for the correct specification of the model; the latter found that aid positively affected growth in any policy environment. Asiana and Quartey (2009) used the Generalized Method of Moments (GMM) and found a negative relationship between aid and welfare variables (life expectancy and infant mortality rate in this case). In a similar study, Gillanders (2011) used the Vector autoregressive framework and found that growth of the economy responded more to aid shocks in groups characterized by high level of aid dependency, poor or weak institution and better economic policies. The study carried out robustness check with the GMM technique. It may be concluded that research findings on the effectiveness of aid on human development differ depending on variables, data, the estimation methods adopted, scope (time covered) and countries. However, debates and discussions on the effectiveness of aid are still on-going. The present study employs up-to-date data with the appropriate estimation method and robustness checks and hopes to report more dependable findings regarding the nature of the relationship existing between foreign aid and human development in SSA.

3. Methodology

3.1 Theoretical Background and Model Specification

The focus of several studies on aid effectiveness over the years has been between aid and other macroeconomic variables, such as savings and economic growth. Not until recently have empirical works started studying the nature of the relationship between aid and human development. The first strand of relevant literature on the effectiveness of aid derived their theoretical background from the Harrod-Domar growth model (Rosentein-Rodan, 1961). Thus, the Harrod-Domar model, which became a reference model in the early 1960s, posits that economic growth is directly related to the level of savings and inversely related to the capital-output ratio, which measures the absorptive capacity of capital. In such a model, savings are perceived as an exogenous factor. Since the shortage of savings is the reason for giving foreign aid to underdeveloped countries, it means savings are directly related to foreign aid, i.e., aid bridges the gap where there is obvious shortage of savings. This, in turn, stimulates investment, which leads to economic growth. While some empirical studies found a positive and significant relationship between aid, savings and economic growth (Papanek, 1972; Newlyn, 1973), others reported a negative and insignificant relationship (Rahman, 1968 and Khan *et al*, 1993).

However, the neo-classical model of growth became the reference point for recent empirical findings (Boone, 1996; Burnside and Dollar, 2000; Kosack, 2003; Fasanya and Onakoya, 2012). The leading neoclassical framework that can serve as a reference model for an aid-effectiveness study is the Solow growth model. This model explains the long run growth by focusing on labour or population growth, capital accumulation and growth in productivity, otherwise known as technological progress. According to Burnside and Dollar (2000), a modified neoclassical model can provide an analytical framework for investigation concerning foreign aid and growth. The interpretation can be that foreign aid serves as income transfer to poor countries, which raises the level of savings that can now be invested and ultimately lead to growth. In the same vein, the modified neoclassical model can provide the necessary framework for empirical investigation concerning foreign aid and economic development (human development). This is so because, economic development is the ultimate goal to be reached when an economy continues to enjoy increase in output production (economic growth). In the world today, policy makers are concerned with economic growth that leads to better standards of living for the populace and, thus, economic development, which promotes human development as a priority. Therefore, this study brings its own novelty by using the Solow growth model as a theoretical explanation for the link between foreign aid and human development.

The Solow model exhibits a Harrod neutral production function, i.e., it is labour augmenting, and the basic assumption of the Solow growth model regards the production function that there are constant returns to scale (CRS) and diminishing returns of input factors. The model is given as:

$$Y = f(K, AL) \tag{1}$$

Where; Y = output, K = capital, AL = effective labour. Because of its constant returns to scale assumption, we can rewrite the model in its intensive form, as follows:

$$Y \cdot \frac{1}{AL} = \frac{1}{AL} f(K, AL) \tag{2}$$

$$\frac{Y}{AL} = f\left(\frac{K}{AL}, 1\right) \tag{3}$$

Here, $\frac{Y}{AL} = y$ is the actual output per effective labour and $\frac{K}{AL} = k$ is the capital per effective labour.

$$\text{Thus, } y = f(k) \tag{4}$$

According to the model, k is a function of savings as a proportion of output “ sY ”.

$$\text{Thus, } k = sf(Y) \tag{5}$$

Capital accumulation in the economy depends on savings and the depreciation rate of capital. Hence the capital accumulation equation becomes:

$$\dot{K} = sY - \delta K \tag{6}$$

Where \dot{K} is net capital accumulation or the growth of capital overtime, δ is the depreciation of capital and δK represents the investment necessary to replace worn-out capital, sY is Savings per worker needed to make capital investment.

Given that the growth rate of knowledge overtime $\dot{A}/A = g$ and the growth rate of labour over time $\frac{\dot{L}}{L} = n$, the differentiation of capital per effective labour gives its growth rate overtime:

$$k = K/AL \tag{7}$$

The differentiation of the above leads to:

$$\dot{k} = \dot{K}/AL - K\dot{A}/AALL - K\dot{L}/AALL \tag{8}$$

Given that the growth rate of knowledge overtime $\dot{A}/A = g$ and the growth rate of labour over time $\frac{\dot{L}}{L} = n$ equation (8) can be written as follows:

$$\dot{k} = \dot{K}/AL - K/AL g - K/AL n \tag{9}$$

Furthermore, because $\dot{K} = sY - \delta K$ and $k = K/AL$, the equation (9) can be written as:

$$\dot{k} = (sY - \delta K)/AL - kg - kn \tag{10}$$

$$\dot{k} = (sY)/AL - (\delta K)/AL - kg - kn \tag{11}$$

Since $Y/AL = y = f(k)$ and $k = K/AL$, then equation (11) can be written as follows:

$$\dot{k} = sf(k) - k\delta - kg - kn \tag{12}$$

Collecting the like terms, we would have:

$$\dot{k} = sf(k) - k(\delta + g + n) \tag{13}$$

The equation above is known as the Solow equation. It gives the growth capital per labour ratio, k (also known as capital deepening), and shows that the growth of k depends on savings $sf(k)$ and the cost of capital $(\delta + g + n)$.

3.2 Model Specification

The shortage of savings in underdeveloped/developing economies has, however, led to the need for/reliance on foreign aid. Thus, saving as a proportion of income, in equation 5, is related to aid as a proportion of output or income ($AID/GDP = AIDN$). Therefore, $k = f(AIDN)$ (14)

Equation (3) can be rewritten by replacing k with $AIDN$, $y = f(AIDN)$ (15)

Moving forward to proxy output with human development in the model above, equation (15) can be written as: $HDI = f(AIDN)$ (16)

Considering the perceived influence of corruption and government effectiveness in the ability of aid to improve human development, equation (16) is rewritten to include such variables as control; therefore, the corruption perception index, government effectiveness and trade openness were included as such and the model for the study has become:

$$HDI = f(AIDN, AIDG, AIDC, CORRUP, GOVEFF, OPEN) \tag{17}$$

However, the equation estimated for the purpose of the study is explicitly stated in econometric as follows:

$$HDI_{it} = a_0 + a_1HDI_{it-1} + a_2AIDN_{it} + a_3AIDG_{it} + a_4AIDC_{it} + a_5CORRUP_{it} + a_6GOVEFF_{it} + a_7OPEN_{it} + \eta_i + \mu_t + \epsilon_{it} \tag{18}$$

Where, HDI represents the human development index, AIDN represents aid as a share of the GDP, AIDG is the interaction between aid and government effectiveness, AIDC

represents aid interaction with the corruption index, CORRUP and GOVEFF stand for corruption and government effectiveness, respectively, while OPEN represents trade openness; η represents the country specific effects, μ represents the time effects and ε represents the error term.

The estimation technique preferred in this study is the system-Generalized Method of Moment (System-GMM) regression technique. In relevant literature, it has been established that aid and human development have a bi-causal relationship, which has led to the problem of endogeneity. A well-suited technique to deal with such an endogeneity issue is the GMM methodology, which actually combines the relevant regressors expressed in both their first differences and levels in a system. The GMM technique is divided into two, namely, Differenced GMM and System GMM. The latter is the method preferred in this study, because it has been shown in practice to be capable of correcting for unobserved country heterogeneity, errors due to measurement, omitted variable bias and likely endogeneity problems, which often affect growth estimation (Hoeffler and Tampe, 2001; Blundell and Bond, 1998 and Arellano and Bover, 1995).

There are a number of compelling reasons why the GMM estimation approach is preferred and these are carefully enumerated. The modeling strategy, which is dynamic, enables the control of persistence in human development levels since it has behavioral effects that persist. Persistence can be checked through correlation of the HDI and its corresponding first lag. The number of years is lower than the number of countries, i.e., the value of the time period is lower than that of cross-sections. The method leaves room to account for any likely endogeneity problem by controlling for unobserved heterogeneity with time invariant omitted variables. Variations across countries are controlled in the regressions, and, furthermore, Blundell and Bond (1998) postulate that the system GMM estimator corrects for biases associated with the difference estimator. The GMM approach, in particular, fits well for panel data estimations, when the number of periods T is relatively lower, while the value of cross section units N is relatively higher, there are regressors that are not strictly exogenous (endogenous regressors), and fixed effects exist. It is also useful when heteroskedasticity and autocorrelation exist within each country's data but not across countries.

In the present study, the equation is transformed by orthogonal deviations, which is an alternative to differencing and was proposed by Arellano and Bover (1995). The orthogonal deviation subtracts the average of all future available observations of a variable from the existing data of the variable. "No matter how many gaps, it is computable for all observations except the last for each individual, so it minimizes data loss" (Roodman, 2009). Furthermore, included in all estimations are time dummies that capture time specific effects. Time dummies reflect the assumption of no autocorrelation across countries and help reduce the level of autocorrelation among different countries and the idiosyncratic error term, which will certainly lead to a very robust estimation.

However, to avoid proliferation or over-identification of instruments, which causes bias of the GMM estimator, over-fitting of endogenous variables, and weakening of the Sargan/Hansen test, the rule of thumb is that the number of instruments to be included in the model should not be higher than the number of periods in the cross sections (Asongu and Nwachukwu, 2017). The two-step system GMM estimates, which are robust to heteroscedasticity, and the panel-specific autocorrelation with Windmeijer correction for finite samples, which helps eliminate standard error, are specifically adopted. Arellano and Bover (1995) and Blundell & Bond (1998) emphasize the need to conduct serial correlation tests for the random error term in the GMM estimation. The serial correlation tests are called AR(1) and AR(2) tests. AR(1) test has a null hypothesis that there is no autocorrelation, specifically concerning the first order in the error term series, while the null hypothesis of AR(2) is that there is no serial correlation specifically concerning the second order type in the error term series. For better results, it is important that the null hypothesis of AR(1) test should be rejected, while the null hypothesis of AR(2) test should be accepted. In testing for overall validity of the instrumental variables used, the Sargan and Hansen test, which is a test of over-identifying restrictions, is used. It is good to know that the consistency of the GMM estimator depends on the validity of instruments. The null hypothesis under both Sargan and Hansen tests is that all instrumental variables, as a group, are exogenous. Therefore, a higher p-value (insignificant) is desirable so that the stated null hypothesis may be accepted. According to Bond (2002), the good estimate of the lagged dependent regressor should fall between its OLS and Within-Group (Fixed Effect) estimates. Thus, these estimates provide a useful robustness check on results. To this end, this study carried out the post-estimation exercise in order to establish the validity and correctness of estimates.

3.3 Data Sources and Measurement

This study assessed the effect of foreign aid on human development in sub-Saharan Africa. Given the importance of human development and the role of government in addressing the problem of low human development in the region, data from a total of 47 SSA countries were utilized in the course of the study. Human development, which is the dependent variable, was measured as an index, as given by the human development index of the UNDP database. Foreign aid is measured in Dollars as a ratio of GDP of the countries in the study in constant 2010 US Dollar price. Aid was originally in current US dollar price in the World Development Indicator (WDI) database but was converted to constant 2010 US Dollar price as appropriate. Trade openness measures the degree of openness of the economy to trade, as captured by the addition of import and export as a ratio of the GDP and was sourced from WDI. The interaction of foreign aid and government effectiveness was captured by multiplying both variables. The same was done to capture the interaction of foreign

aid and the corruption index. Government effectiveness and control of corruption estimates, which give a country score ranging from approximately -2.5 to 2.5, were sourced from the World Governance Indicator (WGI).

4. Results and Discussion

4.1 Pre-estimation: Analyses carried out here included descriptive statistics and correlation analyses of study variables.

Descriptive Statistics of Variables

The descriptive statistics presented in Table 1 gives a glimpse of the basic statistics of study variables. It includes measures of central tendencies, dispersion, minimum and maximum values, degree of peakedness (measured by the kurtosis values), asymmetry (measured by the skewness statistics), and the normality test (measured by the Jarque-Bera statistics) of all the series considered in the study. From Table 1, HDI, AIDN, AIDG, AIDC, GOVEFF, CORRUP and OPEN have mean values of 0.471138, 0.083079, -0.06897, -0.05105, -0.72936, -0.61893 and 78.39425, respectively. All variables except AIDG and AIDC were positively skewed, while all variables were leptokurtic, except for GOVEFF and CORRUP, which were both mesokurtic, since their values were close to three. The Jarque-Bera statistics is significant for all variables suggesting that they were all not normally distributed.

Table 1. Descriptive Statistics of Study Variables

	HDI	AIDN	AIDG	AIDC	GOVEFF	CORRUP	OPEN
Mean	0.4711	0.0830	-0.0689	-0.0511	-0.7293	-0.6189	78.3942
Median	0.4540	0.0637	-0.0354	-0.0329	-0.7723	-0.7155	68.6868
Maximum	0.7820	1.0941	0.0377	0.1682	1.0494	1.2167	351.1057
Minimum	0.2550	-0.0020	-1.3973	-0.7970	-2.1632	-1.7727	19.1008
Std. Dev.	0.1065	0.0958	0.1235	0.0845	0.6154	0.6127	42.2426
Skewness	0.68223	4.4740	-5.3305	-2.9502	0.5136	0.7302	2.2811
Kurtosis	3.2372	38.9769	45.9323	18.5009	2.7595	2.9821	11.6891
Jarque-Bera	57.5679	42377.63	60336.17	8482.01	32.2819	61.8655	2825.238
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sum	339.2190	61.4783	-51.0351	-37.7731	-507.64	-430.78	55189.55
Sum Sq. Dev.	8.1585	6.7920	11.2676	5.2842	263.26	260.94	1254459
Observations	720	740	740	740	696	696	704

Source: Authors' computation 2018

Correlation Analysis

Correlation refers to the degree of linear joint movement or relationship between two or more variables and this was computed in the present study as part of the pre-estimation analysis in order to avoid multicollinearity in the model to be estimated. From the Table 2, it can be deduced that the variables in the model did not exhibit high correlation up to 0.95, which, according to Iyoha (2004), can cause serious multicollinearity among variables if they exist together in an econometric model.

Table 2. Correlation Matrix

	HDI	AIDN	AIDG	AIDC	GOVEFF	CORRUP	OPEN
HDI	1						
AIDN	-0.38666	1					
AIDG	0.376004	-0.90855	1				
AIDC	0.484377	-0.79493	0.909374	1			
GOVEFF	0.574897	-0.21225	0.427758	0.471797	1		
CORRUP	0.543456	-0.06056	0.257149	0.440921	0.845457	1	
OPEN	0.440047	-0.01465	-0.02758	0.087155	0.090832	0.173906	1

Source: Author's computation 2018

4.2 Model Results

The estimated results of the two-step system GMM are presented in Table 3. Here, the instrument collapsed and was also set to a lag limit of 2 and limit of (2-1) for level equation, while corruption was removed from the instrument list in order to avoid instrument proliferation. The estimated results revealed that one period lagged HDI has significant and positive effect on present year HDI at 1 percent level of significance. This may reflect a specific trend in human development indicators in the sub-region. It is worthy of note that foreign aid does not significantly affect human development, since the coefficient of the aid variable, albeit positive, was not significant. This corroborates the work of Boone (1996), who also reported that foreign aid does not affect development. Neither the interaction of foreign aid with government effectiveness nor its interaction with the corruption index was significant at any acceptable risk level. However, corruption was revealed to be significant and negative. The coefficient value of -0.0126 implies that a one percent increase in the corruption perception level in the sub-region resulted in a 0.0126 percent decrease in the human development index. The major implication of this is that corruption significantly reduced human development in Sub-Sahara Africa. The reason for this is not farfetched because, if money and other resources needed for developmental projects (such as road, hospital and school constructions, provision of stable power supply and the provision of an enabling environment for businesses to survive) is embezzled because of high level

of corruption, indices of basic human development will decline. Trade openness was also found in this study to be positive and significantly affecting human development in line with *a priori* expectations. Proper management of trade relations with other countries is ordinarily expected to positively impact the livelihood of the people.

The F-statistic indicated statistical significance indicating the overall significance of the model. In addition, the number of instruments in the model was 25, which is below the number of groups, thereby reducing the chance of having the problem of instrument proliferation that weakens the Sargan and Hansen tests.

Table 3. Two-step system-GMM Estimation Results
Dependent Variable: HDI

Two-step estimates with robust standard errors			
Independent variables	Co-efficient	t-statistics	Probability value
LAGGED HDI	0.8962***	3.92	0.000
AIDN	0.1185	0.34	0.732
AIDG	0.0591	0.15	0.884
AIDC	0.0337	0.12	0.901
GOVEFF	0.0153	0.29	0.771
CORRUP	-0.0126**	-2.37	0.011
OPEN	0.0231**	2.12	0.046

F-stat: 1.24e+07

Prob-value (F-stat): 0.000

Number of instruments = 25, **Number of groups** = 47, **Number of observations** = 598

Source: Author's Computation 2018

4.3 Post-estimation Analyses

Serial Correlation Test

The GMM methodology tests for serial correlation using the Arellano-Bond test of autocorrelation are AR(1) and AR(2) tests. The null hypothesis is that there is no autocorrelation. The AR(1) and AR(2) came out with a probability value of 0.031 and 0.237, respectively. The AR(1) test, which tests for serial correlation at first difference, rejects the null hypothesis at 5% level of significance implying the presence of autocorrelation. According to Arellano and Bond (1991), the AR(1) test should be rejected so that the GMM result may be valid. The AR(2) test accepts the null hypothesis, which should be theoretically so. Both the AR(1) and AR(2) tests validate the estimates of the system-GMM result (Table 4).

Table 4. Arellano-Bond test of autocorrelation

	Z-statistics	Probability value
AR(1)	-2.15	0.031
AR(2)	1.18	0.237

Source: Authors' Computation 2018

Sargan and Hansen Tests

The consistency of the GMM estimates depends on the validity of instruments. The Sargan and Hansen tests are both tests of over-identifying restrictions, which test for the overall validity of the instrumental variables used in the estimation process. The null hypothesis was that all instruments as a group were exogenous or, more specifically, that all instruments are valid. The Sargan and Hansen tests statistics in the present study have probability values of 0.224 and 0.317, respectively, both indicating the acceptance of the null hypothesis, i.e., that the instruments were valid.

Table 5. Presentation of Sargan and Hansen tests

	chi-square statistics	Probability value
Sargan test	1.48	0.224
Hansen test	1.00	0.317

Source: Author's computation 2018

OLS and Within-group (Fixed Effect) estimates

Results of the OLS regression in Table 6 reveal that the lagged HDI was significant at 5% level with a coefficient value of 0.9890. All other variables except trade openness (which is significant at 10 percent level) were insignificant at an acceptable level.

Table 6. Summary of Ordinary Least Squares estimates (OLS)
Dependent Variable: HDI

Independent Variables	Regression Coefficients	t- Statistics	Probability Value
LAGGED HDI	0.9890	236.93***	0.000
AIDN	0.0050	0.64	0.521
AIDG	0.0121	1.38	0.167
AIDC	-0.0133	-1.53	0.126
GOVEFF	0.0013	1.37	0.170
CORRUP	0.0004	0.51	0.609
OPEN	0.00001	1.72*	0.087
CONSTANT	0.0089	3.83**	0.000

$R^2 = 0.9973$ Adjusted $R^2 = 0.9972$ F-Stat = 10060.29 Prob (F-stat) = 0.0000

*Significant at 10%, **Sig at 5% and ***sig at 1 %

Source: Author's Computation 2018

Fixed Effect Estimation Results

Table 7 shows the results of the Within Group (Fixed Effect) estimation and it shows that the lagged HDI coefficient was 0.8186, significant at a 5% probability value. The interaction variable between aid and corruption (AIDC) came up with a negative and significant coefficient. This may reveal how corruption slows down or reverses the initial objective foreign aid is meant to achieve. Furthermore, trade openness (OPEN) impacted positively on HDI implying that, if the region opens its economies to beneficial trade, it will certainly translate to positive changes in human development.

Robustness Check of the System GMM Results

According to Bond (2002), the validity of the System GMM results, among other criteria, depends on the ability of the lagged dependent variable to fall in the range of its pooled OLS estimate and its Within-group (Fixed effect) estimates. The lagged HDI of two step system GMM fell in between its values in both pooled OLS (Table 6) and Within-group estimates (Table 7), i.e., $0.8186 < 0.8962 < 0.9890$ in the present study. Hence, the validity of the model is confirmed.

Table 7. Summary of Within-Group (Fixed Effect) Estimation
Dependent Variable: HDI

Independent Variables	Regression Coefficients	t- Statistics	Probability Value
LAGGED HDI	0.8186***	43.52	0.000
AIDN	0.0134	0.97	0.333
AIDG	0.0230	1.79	0.074
AIDC	-0.0251**	-2.15	0.032
GOVEFF	0.0020	1.01	0.313
CORRUP	0.0018	0.95	0.344
OPEN	0.00003***	3.66	0.000
CONSTANT	0.0950***	9.46	0.000

$R^2 = 0.9756$ $F\text{-Stat} = 1012.89$ $Prob(F\text{-stat}) = 0.0000$

*Significant at 10%, **Sig at 5% and ***sig at 1 %

Source: Authors' Computation, 2018

5. Conclusion

There has been influx of Official Development Assistance (ODA), otherwise known as foreign aid, into Sub-Sahara African countries. There is abundant literature on the assessment of the impact of foreign aid on human development, as a whole, or some indices (indicators) of human development (such as education, health, standard of living, etc.) and results have been very diverse. While some authors have reported a positive relationship, others have reported a negative one, while some have found no relationship between the two variables. The system GMM was adopted, because, among other reasons, the number of countries was higher than the number of years considered. Results of the panel analysis show that foreign aid did not significantly affect human development in SSA, and it should be noted that this, actually, corroborates some literature findings, such as those by Boone (1996). The non-significance of foreign aid should be regarded as a serious issue because it implies that the objective of the donors of foreign aid, for which funds were released, were not being achieved. However, some control variables were found to significantly affect HDI in SSA. For instance, it was found that corruption significantly reduced the HDI, while trade openness improved it. This was further corroborated by the result of the OLS and Fixed effect/Within-group estimation, through which it was shown that foreign aid interacted with the corruption index, reduced the HDI, while trade openness improved it. The phenomenon of corruption is a serious 'drag' on development in SSA, as shown in this study, and there is a dire need, now more than

ever, to curb the menace of corruption if the region aspires to develop and level up with other regions around the world. Moreover, no matter what the volume of funds released to SSA in the form of foreign aid may be, the pervasive corruption, especially in the government, will inhibit its effectiveness. The implication is that measures to reduce corruption should be adopted, while an appropriate framework for effective utilization of foreign aid is put in place, so that Sub-Saharan African countries may benefit maximally from aid programmes. Furthermore, SSA countries should be encouraged to open-up their economies to beneficial trade relationships with the outside world, which would ensure human/economic development.

References

- Abouraia, M., 2014, "Impact of foreign aid in economic development of developing countries: A case of the Philippines", *European Journal of Business and Social Sciences*, 33(4), 166-180.
- Akinkugbe, O. and O. Yinusa., 2009, "ODA and Human Development in Sub-Saharan Africa: Evidence from Panel Data", being paper prepared for presentation at the 14th Annual Conference on Econometric Modeling for Africa organized by African Econometrics Society, at the Sheraton Hotel, Abuja-Nigeria, 8-10 July.
- Alemu, A. and J. Lee., 2015, "Foreign aid on economic growth in Africa: A comparison of Low and Middle-Income countries", *SAJEMS NS*, 18(4), 449-462. DOI: <http://dx.doi.org/10.17159/2222-3436/2015/v18n4a1>
- Amartya, S. 1999, "Development for freedom", Oxford Publishing Press
- Arellano, M. and S. Bond., 1991, 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, 58(2), 277-97.
- Arellano, M. and O. Bover., 1995, "Another look at the instrumental variable estimation of error-component models" *Journal of Econometrics*, 68, 29-51.
- Asiama J.P. and P. Quartey., 2009, "Foreign aid and the human development indicators in sub-Saharan Africa", *Journal for Developing Societies*, 25(1), 57-83. DOI: 10.1177/0169796X0902500103
- Asongu S. and J. Nwachukwu., 2017, "Increasing foreign aid for inclusive human development in Africa" *African Governance and Development Institute Working Paper*, WP/17/020. Retrieved from <https://ssrn.com/abstract=2980877>
- Blundell, R. and S. Bond., 1998, 'Initial Conditions and Moment Restrictions in Dynamic Panel Data Models', *Journal of Econometrics*, 87(1), 115-43.
- Boone, P., 1996, "Politics and the Effectiveness of Foreign Aid", *European Economic Review*, 40(2), 289-329.
- Burnside, C. and D. Dollar., 2000, "Aid, policies, and growth", *American Economic Review*, 90(4), 847-868.
- Chenery, H. & A. Strout., 1996, Foreign assistance and economic development. *American Economic Review*, 56(4), 6-19.
- Dalgaard, C. and H. Hansen., 2001, "On Aid, Growth and Good Policies", *Journal of Development Studies*, 37(6), 17-35.
- David, N., 2017, "Contribution of official development assistance to poverty alleviation in the Ecowas" *Journal of Economics and Finance*, 8(2), 37-49.
- Dissa, Y., 2016, "The effectiveness of Official development assistance: A case study of Mali", *International Journal of Research in Social Science*, 6(3), 399-415.
- Fasanya, I.O. and A. Onakoya., 2012, "Does foreign aid accelerate economic growth? An empirical analysis for Nigeria" *International Journal of Economics and Financial Issues*, 2(4), 423-431. Retrieved from www.econjournals.com

- Gillanders, R., 2011, "The effects of foreign aid sub-Saharan Africa", UCD Centre for Economic Research Working Paper Series; WP11/16. Retrieved from http://www.ucd.ie/t4cms/WP11_16.pdf
- Gomanee, K., S. Girma, and O. Morrissey., 2004, 'Aid, Public Spending and Human Welfare: Evidence from Quantile Regressions', *Journal of International Development*, 17(3), 299–309.
- Human Development Report, 2005, United Nations Development Programme. New York. Retrieved from <http://hdr.undp.org>
- Human Development Report., 2016, "Human development for everyone" *Annual Report*. Retrieved from <http://hdr.undp.org>
- Iyoha, M. A., 2004, "Applied econometrics (second edition)", Mindex publishing Benin city Nigeria.
- Kosack, S., 2003, "Effective aid: How democracy allows development aid to improve the quality of life", *World Development*, 31(1), 1-22.
- Kumler T., 2007, "The Impact of Foreign Aid on Development and Aggregate Welfare in Developing Countries". Honors Projects. Paper 19. Retrived from http://digitalcommons.iwu.edu/econ_honproj/19
- McGillivray M. and F. Noorbakhsh (2004), "Aid, conflict and Human Development". A paper presented at conference of Making Peace Work, World Institute For Development Economics Research, United Nations.
- Mohamed B., S. Salahuddin, and M.A. Mohammed., 2015, "Econometric study of official development assistance and the fight against poverty: The case of Mauritania". *International Journal of Development and Economic Sustainability*, 3(2), 66-79.
- Mohamed, M. and S. Mzee. 2017, "Foreign aid and human development: A quantile regression approach" *International Journal of Economics, Management and Accounting*, 25(1), 27-41.
- Naheed, Z., E. R Khan; and Z. Mahmood., 1993, "Foreign aid, Domestic Savings and Economic growth" *The Pakistan Development Review*, 32(4), 1157-1167.
- Newlyn, T. W., 1973. "The Effect of Aid and other Resource Transfers on Savings and Growth in Less Developed Countries- A comment", *Royal Economic Society*, 83(331), 867-869.
- Nourou, M., 2014, "Foreign aid and development: Looking Beyond the growth effect", *Journal of Economic Development*, 39(4), 99-115.
- Okon, E. O., 2012, "Five Decades of Development aid to Nigeria: The impact on human development" *Journal of Economics and Sustainable Development*, 2(1), 32-42. Retrieved from www.iiste.org
- Papanek, G. F., 1972, "The Effect of Aid and other Resource Transfers on Savings and Growth in Less Developed Countries" *Economic Journal*, 82, 934-950.
- Phiri, Y. and B. Tchereni., 2013, "Assessment of the impact of foreign aid on economic growth: Evidence from 26 HIPC sub-Saharan African countries", *Mediterranean Journal of Social Sciences*, 4(13), 201-209. Doi:10.5901/mjss.2013.v4n13p201.
- Rahman, M. A., 1968, "Foreign Capital and Domestic Savings: A test of Haavelmo's Hypothesis with Cross Country Data" *Review of Economics and Statistics*, 30(1), 38-137.
- Roodman, D., 2009a., "A Note on the Theme of Too Many Instruments", *Oxford Bulletin of Economics and Statistics*, 71(1), 135-158.
- Roodman, D., 2009b., "How to do xtabond2: An introduction to difference and system GMM in stata" *The Stata Journal*, 1(9), 86-136.
- Rosentein-Rodan, P. N., 1961, "International Aid for Underdeveloped Countries", *Review of Economics and Statistics*, 43(20), 98-107
- Siraj, T., 2012, "Official development assistance (ODA), public spending and economic growth in Ethiopia", *Journal of Economics and International Finance*, 4(8), 173-191.
- Tezanos, S., A. Quiñones and M. Guijarro., 2012, "Inequality, aid and growth: macroeconomic impact of aid grants and loans in Latin America and the Caribbean", *Journal of Applied Economics*, 16(1), 157-182.

United Nations., 2004, "Millennium Development Goals: Status 2004", UN Department of Public Information, New York.

United Nations Development Programme, 1997, "Human Development Report 1997".

Williamson, C., 2008., "Foreign Aid and Human Development: The Impact of Foreign Aid to the Health Sector." *Southern Economic Journal*, 75(1), 188-207.

FORECASTING ECONOMIC RECESSIONS USING MACHINE LEARNING: AN EMPIRICAL STUDY IN SIX COUNTRIES

ANDREAS PSIMOPOULOS*
ETH Zurich, Switzerland

Abstract

This paper proposes a methodology for forecasting economic recessions using Machine Learning algorithms. Among the methods examined are Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Random Forests. The datasets analysed refer to six countries (Australia, Germany, Japan, Mexico, UK, USA) and cover a time span of more than 40 years. All methods are compared against each other in terms of six evaluation metrics on their out-of-sample performance. In contrast to most similar empirical studies, the methodology developed focuses on the timepoints of the last four quarters before a recession begins rather than on those of a recession *per se*. It has been found that the SVM method tends to outperform the others, as it classified correctly at least 75% of the pre-recessionary periods for half of the countries, with mean overall classification accuracy around 90% in these cases. Moreover, for all the countries under study, the traditional Logit and Probit models are always inferior to at least one Machine Learning-based model. Additionally, it turns out that macroeconomic variables representing a kind of debt – such as, household debt – are most frequently considered as important across the six datasets, in terms of the Mean Decrease Gini measure.

JEL Classification: C18, C45, C53, E37

Keywords: Forecasting recessions, Machine Learning-based Econometrics, Gini importance, Support Vector Machines

Acknowledgements: I would like to thank Prof. Dr. Nicolai Meinshausen for his valuable assistance throughout all phases of this project. I would also like to thank the two anonymous referees for their constructive comments and suggestions.

*Corresponding Address: Andreas Psimopoulos, ETH Zurich, Department of Mathematics, Rämistrasse 101, 8092 Zurich, Switzerland. E-mail: andreas.psimopoulos@alumni.ethz.ch

Introduction

Modern global economy is a highly complex dynamic system. Richard M. Goodwin (1951) highlighted the importance of incorporating nonlinear differential or difference equations in the analysis of business cycles, instead of following oversimplified linear approaches. Wong *et al.* (2011, p. 432) argue that the global financial system is becoming more and more complex, as interconnectivity among financial systems, markets and institutions increases. Moreover, Barnett *et al.* (2015, p. 1750) argue that an alternative explanation for the existence of business cycles is the chaotic nature of economic systems. According to the latter explanation, business cycles are not caused by exogenous shocks, as the popular opinion holds, but they are created endogenously due to the stochastic behaviour of economic systems. Nevertheless, the authors conclude that, currently, it is not feasible to verify whether this chaotic behaviour has its roots in internal factors or not. One of the biggest challenges in Economics is the accurate prediction of some measures of interest, such as the Gross Domestic Product (GDP), for various reasons, e.g., policy making, financial speculation, etc. However, long-term predictions in chaotic systems are impossible, due to the inherent property of systems' sensitive dependence on initial conditions. Since it is not clear whether economic systems are truly chaotic systems or not, the objective of the present paper is to propose a Machine Learning-based methodology, the goal of which is to provide reliable *short-term* predictions of economic recessions. The methodology proposed focuses on the signs that precede significant downturns of economic activity. In other words, our goal is to capture the dynamics of some important macroeconomic factors before a recession occurs, in order to use these signs as indicators for upcoming recessions. The potential benefit is that such predictions can be taken into account by policymakers, giving them the chance to design and apply more effective policies.

According to the International Monetary Fund, there is no official definition of the term *economic recession*. However, a practical definition that seems to be widely accepted is the following: "*Recession is a period of two consecutive quarters of decline in a country's real GDP*" (Claessens & Kose, 2009, p. 52). This definition is also accepted in the framework of this paper. A special case of recessions are the so-called *depressions*. A depression is a severe and long-lasting recession (Hall & Lieberman, 2013, p. 125). Although there is no general consensus regarding the magnitude and the duration that labels a recession as depression, most analysts make this distinction if the decline in real GDP exceeds 10% (Claessens & Kose, 2009, p. 53). Generally speaking, depressions are very rare, and, thus, we do not study them separately in the models presented below.

The capacity to predict economic recessions is not the only open research question regarding them. It is a fact that there has been a two-hundred-year debate in Economics about what causes recessions and depressions – and there has been no general agreement so far (Knoop, 2015, p. 4). Therefore, by looking for macroeconomic signs

before such events, it may be possible to confirm an existing theory about why recessions happen or pave the way for a new one. At this point, it should be clarified that there is no clear distinction between statistical and machine learning methods. They, rather, form a spectrum of different methods with similar goals. Machine learning is a field of Computer Science with sound statistical foundations and Statistics is a branch of Mathematics which is increasingly taking more advantage of algorithms and computational infrastructure. A part of this spectrum is presented in the discussion about methods that can be used for solving our main problem.

The paper commences with a review of the literature about theories and recent findings relevant to economic recessions, focusing on topics related to forecasting. In the next section, the methodology applied is presented, which is followed by the section of corresponding results. The fifth section includes the discussion of results, while the paper is summarised with the main conclusions along with reference to some topics for potential further research. An appendix can be found at the end of the paper, which provides additional details for a variety of topics mentioned in the main text.

Literature Review

A short review of theories relevant to economic recessions

As already mentioned, what causes recessions is an open research question. There is a plethora of schools of economic thought, simply because there is no global consensus on how economies operate. These schools of thought generally build their theories on different axioms and it is likely that two schools may have starkly different opinions about a topic. The existence of economic recessions is such a topic that one can find a lot of different explanations in literature about why they occur. For us to find if any such theory can be empirically verified, we have used several variables in our models, which arise from theories related to economic recessions. In this subsection we briefly present these theories.

It is well known that, during recessionary periods, a characteristic situation in the economy is low profitability of the firms. Adam Smith (1723–1790), a social philosopher considered to be the father of Classical Economics, mentions three reasons that cause low profitability: (a) competition in the labour market, which leads to higher wages, and, therefore, decreased profits; (b) competition in the capital market, which leads to higher prices of capital goods, and (c) competition in the consumer goods market, which forces capitalists to sell at cheaper prices, which also diminishes profits (Smith, 1776 [1977], pp. 129, 469). These reasons are linked to macroeconomic variables like unemployment or inflation, which may be found useful for the models of this paper. Another influential classical economist, David Ricardo (1772–1823), stressed the fact of the negative economic and social consequences that arise due to the endlessly growing population (Ricardo, 1821 [2001], pp. 59-64). Hence, it may

be useful to also include demographic variables in a model intended for predicting recessions, which is what we have done, as presented below. Karl Marx (1818–1883) provides a theoretical framework that specifies the exact point at which an economic crisis erupts, which is the onset of the phenomenon we are interested in. Marx argues that periodical depreciation of existing capital is associated with crises in the production process. The birth of such crises occurs at the point of *absolute over-accumulation of capital* (Marx, 1894 [2010], pp. 176-178). A mathematical explanation for this concept can be found in Tsoulfidis (2010, pp. 119-120). According to his analysis, the absolute over-accumulation of capital happens when the elasticity of profit rate with respect to capital (denoted as $e_{r,c} = \frac{dr}{dc} \frac{c}{r}$) is -1 . This suggests that $e_{r,c}$ is likely to be a good predictor of economic crises and – therefore – recessions.

John Maynard Keynes (1883–1946) – one of the most influential economists of the 20th century – was very critical of the Classical model ideas about business cycles. One of the basic assumptions of the Classical model is that perfect competition exists in all markets, which always leads them in equilibrium. In this model, business cycles do not exist; recessions happen due to government policies and regulations (Knoop, 2015, pp. 40, 44). Keynes' explanation about why recessions happen points to a new – for our analysis – variable: expectations about future earnings (Keynes, 1936 [2013], pp. 46-47). He argues that changes in expectations gradually produce similar oscillations in employment and what mainly determines expectations – especially short-term ones – is the most recent actual results (Keynes, 1936 [2013], pp. 49-51). Hence, qualitative indicators, such as the Business Confidence Index (BCI)¹, might be used for incorporating these aspects in a statistical or a machine learning model. Keynes also referred to the concept of *paradox of thrift*, according to which, every attempt to increase aggregate saving – at the expense of consumption – is necessarily self-defeating (Keynes, 1936 [2013], pp. 83-84). In this framework, higher savings, at the expense of consumption, reduce aggregate demand and, thus, cause production to fall. So, a recession may begin – or may be prolonged – after an increase in aggregate saving, even if such a sign seems good at first sight. Therefore, aggregate saving and consumption are potential predictors of economic recessions, among others. What Keynes proposes for recovering from a recession is that the government should intervene in the economy with expansionary fiscal policies (Knoop, 2015, pp. 56-57).

Milton Friedman (1912–2006) was the founder of the School of *Monetarism*. For reasons that are out of the scope of this paper, Keynes believed that only fiscal policies can be effective². Monetarists are sceptical about such a view. Evidence from the

1. “The business confidence index (BCI) is based on enterprises' assessment of production, orders and stocks, as well as their current position and expectations for the immediate future. Opinions compared to a ‘normal’ state are collected and the difference between positive and negative answers provides a qualitative index on economic conditions.” (OECD, 2018a).

2. A summary of Keynes' justification about this belief can be found in Knoop (2015, pp. 55-57).

economic history of the USA suggests that there is strong correlation between changes in money stock and business cycles. Friedman & Schwartz (1963, pp. 676-695) argue that this correlation can be verified for all U.S. recessions during the 1867–1960 period. Their concluding remark is that changes in money supply play the major role in the formation of business cycles and a less important role in short-term fluctuations of economic activity. Friedman (1968, p. 17) suggests that economic stability can be achieved by setting steady but moderate growth in the quantity of money. According to all these ideas from the Monetarist model, one can say that money supply is, potentially, a good predictor of economic recessions, given that there is a causal relationship from changes in money stock to business cycles.

Irving Fisher (1867–1947) concluded that the two factors with prevailing impact on the evolution of business cycles are over-indebtedness and deflation (Fisher, 1933, p. 341). According to Fisher, debt and price levels are primary variables when studying business cycles, in the sense that other similarly important variables are affected by them. Speaking in a business-related context, firms experience a loss in their profits, due to lower prices (deflation); employment, output and trade are reduced (recession), some of the firms go bankrupt, and pessimism, along with loss of confidence, lead to more money-saving and fewer transactions (Fisher, 1933, p. 342). If this theory is confirmed by data, then what we need to find are those values of deflation and private sector debt that signal the occurrence of an upcoming recession.

Joseph Schumpeter (1883–1950) was one of the most important advocates of *Austrian Economics*. Schumpeter emphasises the evolutionary character of the capitalist process. By using the term *creative destruction*, he puts forward an alternative explanation regarding the existence of business cycles (Schumpeter, 1942 [1994], pp. 81-86): As some firms embody innovation (*creation* of new structures), they become able to produce more and sell at cheaper prices in the long-run, leading their competitors to change or to leave the market (*destruction* of existing structures). It is easily conceivable that this innovation dynamics influences business cycles. Hence, innovation indicators could be useful variables for further analysis, because it may be the case that such an indicator follows a specific pattern before the onset of a – rather prolonged – recession.

Finally, we conclude this subsection with the models of *New Keynesian Economics*. These models were developed in 1980s, as a response to the criticism of Keynesian Economics. Regarding what can cause a recession in the framework of New Keynesian Economics, there are three explanations (Knoop, 2015, p. 136): a) change in expectations (old Keynesian approach); b) contraction in money supply (Monetarist approach); c) increase in default risk perceptions. As we can realise, the new variable for our analysis lies in the third explanation. Speaking in a countrywide context, default risk is one of the factors that influence long-term interest rates. According to the OECD (2018b), these interest rates are determined by the amount charged by the

lender, the risk the borrower undertakes and the fall in capital value. If it is perceived that a country will face difficulties in paying its debt obligations promptly, long-term interest rates imposed on it are increased due to default risk. Consequently, business investment falls. Therefore, one additional potential predictor of economic recessions is long-term interest rates.

Recent studies focusing on prediction of economic recessions

In this part we present a review of recent studies focused on the topic of forecasting recessions. Estrella & Mishkin (1998) investigated several leading indicators for the prediction of U.S. recessions, such as stock prices, interest rates, etc. In order to estimate the probability of the occurrence of a recession, these authors used a Probit model. Their analysis was focused on the out-of-sample performance of their models, up to eight quarters ahead, and they found that the best predictors of U.S. recessions were stock prices and the yield curve spread³. Chauvet & Potter (2005) considered an extended Probit specification. In particular, their work is based on a dynamic Probit framework, where dependent variables are regressed on their lagged values and other exogenous regressors, namely, yield spreads. Their best model allowed for multiple breakpoints across business cycles and autocorrelated errors and it achieved better in-sample fit than the model by Estrella & Mishkin. Christiansen (2013) used a Probit model in order to examine the forecasting ability of yield curve spreads in simultaneous recessions of six countries (Australia, Canada, Germany, Japan, United Kingdom and United States). She considers a recession as 'simultaneous' if it occurs in at least half of the countries studied. She found that, at short horizons, only the German yield spread was significant in explaining future simultaneous recessions, but, at long horizons, both U.S. and German spreads were such.

Dovern & Huber (2015) found that the Global Vector Autoregressive (GVAR) approach produced more accurate predictions of recessions than country-specific time series models. The authors defined a period of at least two consecutive quarters with declining GDP as recession, and they used a dummy variable (binary indicator) to encode this in their data. They investigated 36 countries over a period of ten years⁴ and their goal was to provide good probability forecasts for the occurrence of a recession, within a Bayesian framework. Using the real GDP, the change of CPI, the real equity prices, the real exchange rate and the short/long-term interest rates as variables, they successfully constructed a GVAR model that outperformed both of the two benchmark models⁵ in forecasting accuracy for the majority of the countries.

3. A yield curve with a negative slope is often considered as a sign of an upcoming recession. The negative slope of the yield curve means that the interest rate spread is negative.

4. They used quarterly data. The dataset contained data from 1979.Q2 (i.e., the second quarter of 1979) to 2013.Q4. For the verification period, they used 40 observations from 2004.Q1 to 2013.Q4.

5. These models were the Bayesian VAR (BVAR) and the Bayesian univariate autoregressions (AR).

Kauppi & Saikkonen (2008) found that dynamic Probit models outperform static ones in terms of both in-sample and out-of-sample predictions. The authors' goal was to build different forecasting models to predict U.S. recessions and to compare them to each other. They used the definition of recessions⁶ by the National Bureau of Economic Research (NBER) and they also encoded recessionary periods with a binary variable (1: recession, 0: otherwise). They also used quarterly data referring to the 1955.Q4 – 2005.Q4 period and the best model of their analysis was only based on the interest rate spread and the binary variable.

We proceed with the paper by Gogas *et al.* (2015). To the best of that paper authors' knowledge, it was the first attempt to forecast GDP cycles using a Support Vector Machines (SVM) classifier⁷ on data relevant to the yield curve. Gogas *et al.* found that both the short-term and the long-term interest rates had an important role in forecasting future recessions. They used quarterly data of the U.S. GDP and interest rates, from 1967.Q3 to 2011.Q4. Moreover, they used a definition for recessions that differed from the two mentioned previously: They considered every deviation of GDP *under* the long-run trend as a recessionary period. The best performing model was a radial kernel SVM, which achieved in-sample test accuracy of 73.3% and out-of-sample overall accuracy of 66.7%. Döpke *et al.* (2017) applied a machine learning approach known as *Boosted Regression Trees* (BRT). Their goal was to find the predictive value of several leading indicators for forecasting recessions in Germany. They used 35 leading indicators related to the German economy, which were collected on a monthly basis. Some of them were data about money supply, unemployment, price levels, exchange rates and data about interest rates. Their sample period was from 1973.M1 (i.e., January 1973) to 2014.M12. They did not use a specific definition for recessions; for their context, recessionary periods are those characterized as *troughs* by the Economic Cycle Research Institute (ECRI, 2013) and they encoded them in a binary variable. Regarding their findings, they provide evidence that the BRT approach has better out-of-sample performance in comparison to variants of the alternative Probit approach, and that the most influential leading indicators were: a) the short-term interest rate of money market instruments⁸, and b) the spread *yield on ten-year government bonds minus money market rate*.

6. "The NBER does *not* define a recession in terms of two consecutive quarters of decline in real GDP. Rather, a recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales." (NBER, 2010).

7. More details about the Machine Learning methods mentioned in this paper can be found in Appendix, [A.1](#). A short description of the methods used is cited at the end of the next section.

8. Also known as *money market rate*.

Plakandaras *et al.* (2017) compared the performance of SVM models with that of dynamic Probit models in forecasting U.S. recessions. They used data from 1871.M1 up to 2016.M6 and, with regard to how the recessionary periods were specified, they used the definition of NBER, as Kauppi & Saikkonen (2008) did. Plakandaras *et al.* included in their analysis explanatory variables about stock prices, oil prices, financial indicators, money supply and the yield curve. They found that, for short-term predictions, Probit models outperformed SVM, but, for longer horizons, the latter provided more accurate forecasts. To be more specific, Probit models showed better out-of-sample performance in forecasting horizons of 1 and 3 months, while SVM performed better in forecasting horizons of 6, 12, 18, 24 and 36 months. Lastly, we complete this section with the paper by Kiani (2008). The goal of this paper was to apply Artificial Neural Networks (ANN) on forecasting recessions for a set of countries. These countries were Canada, France, Germany, Italy, Japan, the UK and the USA. Kiani's paper provides some first insights about the capabilities of ANN in forecasting recessions. The data used for this research were quarterly for all countries, from 1965.Q1 to 2004.Q4. Variables referred to money supply, stock price indices, several interest rates and others. The author investigated ten models: one for each of the seven variables used and three models based on variable combinations. Regarding the results, each country had a different set of candidate predictors of recessions, according to the out-of-sample forecasting performance of the models employed. The most common ones were the stock price indices and the spread between bank rates and risk-free (i.e., T-Bill equivalent for all countries) rates. Regarding forecasting accuracy, the author defined a new metric that considers both Type I and Type II errors; according to his definition, missed recessionary periods and other missed periods, respectively. In this metric (Kiani, 2008, p. 4), forecasting accuracy exceeded 80% for most of the countries, which is noteworthy. It seems that the flexibility of ANN makes it possible to sufficiently capture the nonlinear features of business cycles, regardless of the country. It would be interesting to see, though, how well these models score in other metrics, too.

Methodology

In this section, we present all methodological issues concerning the models developed. In order to better evaluate the performance of the presented methodology, the author's decision was to apply it for the datasets of six countries, namely: Australia (AUS), Germany (GER), Japan (JAP), Mexico (MEX), the United Kingdom (UK), and the United States of America (USA). There is no objective, strictly defined criterion for this selection, but the selection was not random, either. The main rationale was to choose countries from different continents that are likely to differ from each other in economic terms. If results from heterogeneous countries converge, this is an indication that the pre-recessionary conditions are common across countries and the corresponding model tends to generalise well. If they do not, one could associate pre-recessionary conditions with country-specific characteristics. Additionally, since

the Organization for Economic Co-operation and Development (OECD) provides a comprehensive database of economic and other factors, a prerequisite condition was that the countries to be selected should be OECD's members, in order to take full advantage of the database of the Organisation for the purposes of this paper⁹.

Table 1 contains all the main variables of the datasets constructed. These variables are called 'main', in the sense that all other variables of the final datasets are some transformations of the former. The initial goal was to have each variable in quarterly frequency, from 1969.Q1 up to 2017.Q4 (196 possible observations). As GDP is at the centre of our attention in the context of this paper, this decision was made because the highest available frequency of GDP-related data was on a quarterly basis. However, some variables were provided only in yearly frequency and, for some countries, many variables had their first observation some time during 1990s. The consequence of these circumstances was that there were a lot of missing values in the six datasets, which, in turn, added an extra challenge during the pre-processing steps¹⁰.

At this point it is important to make some remarks regarding the data downloaded. First, variable *GFCF* refers only to domestic investment. The OECD provides a variable for Foreign Direct Investment (FDI), but it is not included in the datasets because *no more* than thirteen observations were available for any country (i.e., a large number of values were missing). A similar situation came up during the collection of Mexico's *BankRate* data. The central bank of Mexico provides relevant data only from 2008 and on. Therefore, Mexico's dataset does not contain the *BankRate* variable. Regarding variable *M1*, the decision was to exclude it from Germany's dataset, because, since the establishment of the Eurozone, *M1* has been the same for all its member-states.

With regard to variable *PPP*, as already mentioned in Table 1, it is measured in *national currency units/US dollar*. This means that for the USA this variable is always equal to one, which is a problematic situation if we want to keep it in the USA dataset. By looking at *PPP*, one essentially compares a country's cost of living with that of the USA. This can be considered as a price ratio between the two countries, where each price refers to a basket of goods and services. So, the question was what to do if we want to measure *PPP* for the USA. The idea here was to reverse the concept. Apart from the time series of separate countries, the OECD also provides *PPP* data for EU28; this means one additional time series, weighted across the 28 countries of the European Union. This time series was also expressed in relation to the U.S. dollar. Therefore, the idea here was to invert the PPP_{EU28} in order to express USA *PPP* in a hypothetical common currency of EU28 (i.e., *US dollars/1 unit of "EU28"*). The ideal situation would be to have a time series weighted across *all* countries except for the USA, but inverting that of EU28 also seems a good approximation¹¹.

9. References regarding the variables downloaded can be found in Appendix, [A.2](#).

10. A detailed presentation of the pre-processing steps can be found in Appendix, [A.3](#).

11. In OECD's database, no other weighted *PPP* time series consisted of more than 28 countries. In addition to that, the fact that the EU is in aggregate of the largest economies worldwide makes the choice of these 28 countries even more suitable for our purpose.

Variable name	Short description	Unit	Linked to a specific theoretical concept or paper
BankRate	Bank rate; interest rate at which the national central bank lends money to domestic banks.	% of principal	It was introduced in order to find if it has similar predictive importance to other types of interest rates.
BCI	Business Confidence Index; enterprises' expectations for the immediate future.	—	Expectations about future earnings, <i>John M. Keynes</i> .
CPI	Consumer Price Index; an index of the price level of consumer goods.	—	3 rd argument about firms' low profitability, <i>A. Smith</i> . Also, part of <i>Fisher's</i> debt-deflation theory.
Fdebt	Financial corporations' debt to equity ratio; a measure of corporations' debt.	—	Debt-deflation theory, <i>Irving Fisher</i> .
Fprof	Net operating surplus of financial corporations; an indicator of the financial sector profitability.	% of net value added	Necessary for the construction of variable 'elasticity of profit rate with respect to capital', <i>K. Marx</i> ; <i>L. Tsoulfidis</i> (2010).
Gdebt	General government debt-to-GDP ratio.	—	It was introduced because variables related to private sector's debt are also included.
GFCF	Gross Fixed Capital Formation; i.e. investment (domestic).	Growth rate	(generally used concept)
Gov	General government spending.	% of GDP	One of two main fiscal policy instruments, <i>J. M. Keynes</i> .
GrowGDP	Percentage change of real GDP.	—	(generally used concept)
HHC	Households' consumption expenditure.	% of GDP	Consumption shrinkage in the concept of 'paradox of thrift', <i>J. M. Keynes</i> .
HHdebt	Household debt.	% of net disposable income	Debt-deflation theory, <i>J. Fisher</i> .
LJR	Long-term interest rate; this can be considered a measure of default risk.	% of principal	Default risk is one of the potential causes of recession in the framework of <i>New Keynesian Economics</i> .
M1	An index of money supply (coins, banknotes and overnight deposits).	—	The core concept of Monetarism, <i>Milton Friedman</i> .
Mports	Trade in goods and services: Imports.	% of GDP	(generally used concept)
NFdebt	Non-Financial corporations' debt to surplus ratio; a measure of corporations' debt.	—	Debt-deflation theory, <i>J. Fisher</i> .
NFprof	Net operating surplus of non-financial corporations; an indicator of non-financial sector profitability.	% of net value added	Necessary for the construction of variable 'elasticity of profit rate with respect to capital', <i>K. Marx</i> ; <i>L. Tsoulfidis</i> (2010).
Pop	Population.	10 ⁶ persons	Threats by endlessly growing population, <i>David Ricardo</i> .
PPI	Producer Price Index; an index of the price level of producer goods.	—	2 nd argument about firms' low profitability, <i>A. Smith</i> .
PPP	Purchasing Power Parities: the rates of currency conversion that equalize the purchasing power of different currencies by eliminating differences in price levels between countries.	National currency units / US dollar	Currency conversion (i.e., exchange rate) variables have been used in the papers by <i>Dovern & Huber</i> (2015) and <i>Döpke et al.</i> (2017).
Rnd	Gross domestic spending on Research and Development; a measure of innovation.	% of GDP	Role of innovation in the concept of 'creative destruction', <i>Joseph Schumpeter</i> .
Sav	Saving rate.	% of GDP	Increase of savings in the 'paradox of thrift', <i>J. M. Keynes</i> .
SIR	Short-term interest rate. Monetary policy can be conducted by using it.	% of principal	Variable mentioned in the paper by <i>Gogas et al.</i> (2015).
SPI	Share Price Index; another name for the already mentioned Stock Price Index.	—	A similar variable was mentioned in the paper by <i>Kiani</i> (2008).
Tax	Total tax revenue.	% of GDP	One of two main fiscal policy instruments, <i>J. M. Keynes</i> .
Unemp	Unemployment rate.	% of labor force	1 st argument about firms' low profitability, <i>A. Smith</i> .
Wage	Average wage in total economy.	US dollars	1 st argument about firms' low profitability, <i>A. Smith</i> .
Xports	Trade in goods and services: Exports.	% of GDP	(generally used concept)

Table 1. The main variables. References are presented in Appendix, A.2.

We close this section's introduction by mentioning the fact that certain variables were not at all available for some countries at the time of data collection, namely:

Table 2. List of missing variables per country.

Country	Missing variables
Australia	<i>Fprof, NFprof</i>
Japan	<i>PPI, Fprof, NFprof</i>
Mexico	<i>PPI, HHdebt, Gov</i>

The variable selection procedure

After the main pre-processing steps, we had six datasets with over a hundred variables in each one, and very few missing values at their top and/or bottom parts. The first question that came up before feeding the models with data was how to choose the explanatory variables. The goal was to have a relatively small number of well-chosen predictors, in order to build parsimonious models and provide reliable results. Thus, the idea was to exploit some beneficial outcomes of the Random Forests modelling. As seen in James *et al.* (2013, pp. 319-321), random forests are based on many different decision trees and, in turn, each decision tree is generally based on a different set of variables¹². This fact gives us the possibility to measure a variable's importance in terms of how much an impurity measure – like the Residual Sum of Squares (RSS) or the Gini index – is decreased on average, for all the times this variable is selected as a predictor. This measurement can easily be made using the function `varImpPlot()` from package `randomForest`. The output of this function is a plot that one can use in order to assess importance of variables in a Random Forests model. For our case, since the target variable is binary, the measure that `varImpPlot()` uses is the Mean Decrease Gini (MDG). Consequently, the decision here was to fit a Random Forests model¹³ and look at the output of `varImpPlot()` for the group of the most important variables in terms of MDG. This means that the number of initially selected variables varies from country to country and the selection procedure is mainly based on a soft (visual) rule. In other words, we are looking for a small group of variables that lie far from others in terms of MDG. This procedure is, in some sense, a competition among different theories and hypotheses about economic recessions based on real world data. If some variables are systematically characterised as important, this implies that the hypotheses they represent are close to reality. To the author's best knowledge, this is the first empirical study that compares the strength of all these hypotheses, which are briefly presented in the last column of Table 1.

12. Details regarding the methods mentioned can also be found in Algorithms 3 and 4 of the present paper.

13. Regarding the value of parameter `mtry` (number of variables sampled before each split) in `randomForest()`, the overall out-of-bag (OOB) error was taken into account by looking at the output of function `plot.randomForest()` (black line). To be specific, a value of `mtry` that quickly minimizes the OOB error was selected for each country, following a trial and error approach. Parameter `ntree` (number of trees to grow) was set at 10,000 in order to estimate each variable's importance to the best possible degree.

The Average Trees algorithm

A new algorithm was developed in order to identify macroeconomic conditions that precede recessions. The Average Trees algorithm can be considered as a robust version of Decision Trees and, specifically, a robust version of classification trees. The latter means that the target variable is a binary one, i.e., the *PreRecess*, which represents a pre-recessionary period¹⁴. The main idea of the methodology developed is to build one classification tree for each country, the decision rules of which at every splitting point is an average of the corresponding rules from other similar classification trees. The latter trees are fitted on slightly different samples. To put it more simply, the idea is to exclude some observations from the dataset, fit a classification tree on the remaining ones and repeat; finally, extract one classification tree the decision rules of which are an average of all previously fitted trees' rules. The average rule is calculated because, for interpretability reasons, the main goal is to have found *a single* rule at the end of the day (at least, one per country). The motivation for this idea was the need for this paper to apply a method which, on the one hand, is as easily interpretable as a Decision Trees model and, on the other, is not too sensitive for overfitting as a Random Forests model (ideally). The kind of decision trees proposed is expected to be less prone to overfitting than simple ones, in the sense that the former are not based on a single dataset but built through a resampling procedure. This is why they are called robust in the context of this paper.

For this purpose, two functions were written in R, which differ only in the way they select the observations excluded. The first function does this without replacement, in a way identical to the sampling procedure of *K*-fold cross-validation. The second, instead of *K*-fold sampling, uses a random sampling procedure with replacement. Algorithm 1 describes in more detail the method proposed:

Algorithm 1 The Average Trees algorithm

INPUT: *dt*, *imp* = {*v*₁, *v*₂, ...}, *nt*, *s* // Dataset, important variables, number of trees to fit¹⁵ and size of excluded sample per iteration, respectively. //

Step 1: For each *i* = 1, 2, ..., *nt*:

- 1) Exclude *s* observations from *dt*. // Sampling method: depending on which function has been selected. //
 - 2) Fit a decision tree on remaining data.
 - 3) Extract the decision rules from the model fitted and save them in a table named *RS*.
-

14. More details about how a pre-recessionary period was defined can be found in Appendix, [A.3](#), in reference "b)" about function `def.recess(dt)`.

15. This parameter is available only in the function which performs random sampling with replacement. In the other one, the maximum number of trees is restricted by the number of excluded observations per iteration (here, parameter *s*). Thus, in that case, *nt* is calculated by the algorithm.

Step 2: Search in *RS* for the most frequent tree structure (i.e., for the moment, ignore the numerical part of each rule and only care about the splitting variable and the inequality direction) and select it.

Step 3: Take into account only the trees that have the most frequent structure and, at each split, compute the average rule from the numerical parts previously excluded. These average rules, which are based on the most frequent tree structure, form the *average tree*.

OUTPUT: *average tree*

For example, let us assume that 90% of the trees fitted have the simple rule that, if variable $V_1 > a$, then $PreRecess = 1$; otherwise, $PreRecess = 0$, where, in general, number a differs from tree to tree. Additionally, the remaining 10% are trees that have the rule: if $V_2 > b$ then $PreRecess = 1$; otherwise $PreRecess = 0$ (again, the values of b are generally different for each tree of this structure). In this case, Algorithm 1 drops out the second tree structure, because only 10% of the trees fitted have it. Consequently, the majority rule has been applied here in order to find the most prevalent tree structure. After that, Algorithm 1 keeps only the first structure and, in place of a , it substitutes every a with the average value. This is the *average tree*. The rationale behind these steps is that, since a tree structure appears more frequently than any other in different samples of the same population, it is more likely that this is the structure that best represents reality. As these trees differ only in the numerical values of their decision rules, we take the average rule by averaging these values.

The evaluation metrics

A plethora of statistical and machine learning models were compared to each other in terms of six evaluation metrics for each country. These evaluation metrics are based on the so-called *confusion matrix* for discrete classifiers of binary classification problems. The confusion matrix has the following form (Aggarwal, 2015, pp. 637-638):

Table 3. The confusion matrix in binary classification problems.

		Predicted class	
		Class-1	Class-0
Actual Class	Class-1	True Positives (TP)	False Negatives (FN)
	Class-0	False Positives (FP)	True Negatives (TN)

At this point we form the convention that instances of class “1” are called *positives* and instances of class “0” are called *negatives*. The sum of true positives and true negatives is the number of correctly classified observations; the rest are wrongly classified. Let CM be the squared confusion matrix that contains numbers TP , FN , FP and TN from Table 3. Then, we can define the following evaluation metrics for a binary discrete classifier c :

$$acc(c) = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$recall(c) = \frac{TP}{TP + FN} \quad (2)$$

$$precision(c) = \frac{TP}{TP + FP} \quad (3)$$

$$specificity(c) = \frac{TN}{FP + TN} \quad (4)$$

$$falarm(c) = \frac{FP}{TN + FP} \quad (5)$$

Eq. 1 is called *classification accuracy* and it is a metric which simply denotes the overall proportion of well classified instances. Eq. 2, also known as *sensitivity* or *true positive rate* (TPR), is the proportion of instances correctly classified as ‘positives’ from among all truly positive instances. Eq. 3 is the proportion of instances correctly classified as ‘positives’ from among all instances classified as ‘positives.’ Eq. 4, also called *true negative rate* (TNR), is analogous to the sensitivity for negative instances, and *false alarm* of Eq. 5, also known as *false positive rate* (FPR), is the proportion of instances wrongly classified as ‘positives’ from among all truly negative instances. Another evaluation metric, based on *recall* and *precision*, is the F_β -score:

$$F_\beta(c) = (1 + \beta^2) \frac{recall(c) * precision(c)}{recall(c) + [\beta^2 * precision(c)]} \quad (6)$$

It may be the case that we care about both *recall* and *precision* and we want a single metric for these two. With F_β -score we take both into account and we can adjust their weights by choosing the appropriate β . A higher β means that more emphasis is placed on *recall*, while a lower β attributes more weight to *precision* (i.e. $F_\beta(c) \rightarrow recall(c)$ and $F_\beta(c) \rightarrow precision(c)$, respectively). For example, if we use classifier c for predicting recessions, it is not obvious – without further investigation – whether the cost of *not predicting* a recession or that of *wrongly preparing* for a recession is

greater. A policymaker may acquire this knowledge after some investigation and (s) he can adjust β accordingly, in order to incorporate this cost-centred perspective into an evaluation metric. If $\beta = 1$ we have the so-called F_1 -score, which is simply the harmonic mean of *recall* and *precision*.

A short presentation of the methods used

In this subsection we briefly present the basic components of the methods used to produce this study's results. References for detailed descriptions of all of them are presented in Appendix, A.1.

- Logit: A Logit model is described by the following equation:

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (7)$$

where p_i is the probability that the response (binary) variable equals 1, \mathbf{x}_i^T is the $1 \times (K+1)$ vector of i -th observation values on K independent variables plus a constant, and $\boldsymbol{\beta}$ is the $(K+1) \times 1$ vector of K parameter values plus a constant.

- Probit: With regard to Probit models, according to Baltagi (2002, pp. 332-333), these differ from Logit only in the tails of their CDFs¹⁶. Both have the CDF of a t-distribution; Probit has the one with infinite degrees of freedom, while Logit has that of seven. More specifically, Probit has the following CDF:

$$p_i = \Phi(\mathbf{x}_i^T \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (8)$$

- Support Vector Machines (SVM): In a general context, we can split a K -dimensional space into two regions using a $(K-1)$ -dimensional *hyperplane*. The mathematical definition of a $(K-1)$ -dimensional hyperplane is given by the following equation¹⁷:

$$\beta_0 + \sum_{k=1}^K \beta_k X_k = 0. \quad (9)$$

16. Cumulative Distribution Function.

17. Using words, hyperplane is a flat affine (i.e., not necessarily passing through the origin) subspace of dimension *one less* than its surrounding space (James *et al.*, 2013, p. 338).

Any point \mathbf{x} , the coordinates of which satisfy Eq. 9, lies on the hyperplane. But it may be the case that when substituting a point's coordinates in the LHS of Eq. 9, the result is either > 0 or < 0 , instead of being equal to zero. This simply means that this point is either 'above' or 'below' the hyperplane in the K -dimensional space. This idea can be used for constructing classifiers by using proper hyperplanes as decision boundaries, which divide the K -dimensional space into two regions, one for each class of observations. In the context of SVM, this hyperplane is the result of the following optimisation problem:

$$\underset{\beta_0, \beta_1, \dots, \beta_K, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n}{\text{maximize}} \quad W \quad (10)$$

$$\text{subject to} \quad \sum_{k=1}^K \beta_k^2 = 1, \quad (11)$$

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_K X_{iK}) \geq W(1 - \varepsilon_i), \quad (12)$$

$$\varepsilon_i \geq 0, \quad \sum_{i=1}^n \varepsilon_i \leq C, \quad (13)$$

where W is the width of the margin¹⁸, β_k s are the parameters which define the hyperplane along with X_k s, ε_i s, which are slack variables¹⁹, each $y_i \in \{-1, 1\}$ represents the class of the i -th observation, and C is a tuning parameter for the tolerance of margin violations. If $C = 0$ no margin violations are allowed. It can be shown that the solution of the above optimisation problem involves only the inner products²⁰ of observations, which leads to the conclusion that the linear classifier can be represented as:

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^n \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle, \quad (14)$$

18. In this context, margin is the distance between a hyperplane and the closest points (i.e., observations) on either side. However, for SVM to also be used in cases where two classes cannot be perfectly separated by a hyperplane, a margin can be violated to some extent by correctly classified observations and/or by misclassifications.

19. If $\varepsilon_i > 0$, this means that the i -th observation has violated the margin and if $\varepsilon_i > 1$, this means that the i -th observation is on the wrong side of the hyperplane.

20. For two observations $\mathbf{x}_1, \mathbf{x}_2$, their inner product is defined as: $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \sum_{k=1}^K x_{1k} x_{2k}$.

where α_i s are parameters to be estimated²¹. Equivalently, one can write Eq. 14 as:

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^n \alpha_i K(\mathbf{x}, \mathbf{x}_i), \quad (15)$$

where function $K(\mathbf{x}, \mathbf{x}_i)$ is called *linear kernel* and it simply computes the inner product of two vectors. The advantage of this approach is that one may very well use another kernel function in Eq. 15 – probably a nonlinear one – to produce a much more flexible decision boundary. Such an example is the *radial kernel*:

$$K(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp\left(-\gamma \sum_{k=1}^K (x_{ik} - x_{i'k})^2\right). \quad (16)$$

The use of a radial kernel in SVM produces nonlinear margins and decision boundaries and, more specifically, boundaries of ‘circular shape’. However, we should not overlook the fact that SVM is always a *linear* approach. Regardless of the kernel used, SVM’s solution is always a linear decision boundary at a higher-dimensional space (larger than K). In the original feature space though, it turns out that the decision boundary is generally nonlinear.

- The k -Nearest Neighbours (k -NN) algorithm:

Algorithm 2 k -nearest neighbours (k -NN)

INPUT: $\mathbf{X}_{n \times J}$, $\mathbf{x}_{new} \in \mathbb{R}^J$, $C = \{c_1, \dots, c_M\}$, $\mathbf{y} \in C^n$, $k \in \mathbb{N}^*$

/// A data matrix of n observations and J predictors, the observation to be classified, a set of M classes, the class vector and the number of k nearest neighbours, respectively. ///

```

1:  for each  $\mathbf{x}_i$  in  $\mathbf{X}_{n \times J}$            //  $i$  from 1 to  $n$ .
2:     $d_i \leftarrow \text{Distance}(\mathbf{x}_i, \mathbf{x}_{new})$ 
3:     $D = \{(d_1, 1), \dots, (d_n, n)\}$  // Set of ‘distance-index’ tuples.
4:     $D_s \leftarrow \text{Sort}(D, 1)$        // Sort by ascending order, by distance.
5:     $I \leftarrow \text{Untuple}(D_s, 2)$    // Extract indices.
6:     $c_{nn} = \{y_{I[1]}, \dots, y_{I[k]}\}$  // Classes of  $k$ -nearest neighbours.
7:     $y_{new} \leftarrow \text{Mode}(c_{nn})$    // Most frequent class of  $k$ -NN.

```

OUTPUT: y_{new}

21. In this formulation, parameters α_i have substituted the original β_k , $k > 0$.

By looking at Algorithm 2, one realises that the class of a new observation is determined *only* by the classes of its k -nearest neighbours. Using the majority rule, \mathbf{x}_{new} is assigned to the class which predominates. Function *Distance* in line 2 computes the distance between two vectors and it may use any distance metric, depending on their domain. For example, such a metric is the *Euclidean distance*²². The underlying assumption for the k -NN classifier is that observations of the same class are similar to each other, which means that – given a distance metric – they are close to each other. This simple assumption provides the classifier with great flexibility: unlike the previously presented methods, the true shapes of decision boundaries do *not* need to be taken into consideration before solving the classification problem. In fact, decision boundaries may be of any shape and the classifier can still be reliable if number k is properly chosen. Moreover, no distributional assumptions are made.

- Decision Trees: Decision Trees is a method based on a procedure called *recursive binary splitting*:

Algorithm 3 Recursive Binary Splitting

INPUT: $\mathbf{X}_{n \times J}$, $\mathbf{x}_i \in \mathbb{R}^J$, n_{min} , $\mathcal{C} = \{c_1, \dots, c_M\}$, $\mathbf{y} \in \mathbb{R}^n$ or $\mathbf{y} \in \mathcal{C}^n$

// domain of \mathbf{y} depends on the problem (regression or classification, respectively). //

Step 1: Choose splitting variable j and splitting point s so that quantity 17 (regression problem) or 18 (classification problem) is *minimised*:

$$\sum_{i: \mathbf{x}_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: \mathbf{x}_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \quad (17)$$

$$\sum_{c \in \mathcal{C}} \hat{p}_{R_1,c} (1 - \hat{p}_{R_1,c}) + \sum_{c \in \mathcal{C}} \hat{p}_{R_2,c} (1 - \hat{p}_{R_2,c}) \quad (18)$$

Step 2: For each of the resulting regions repeat Step 1, until no terminal node (i.e., tree leaf) R_g contains more than n_{min} observation(s).

OUTPUT: $T_0 = \{R_1, R_2, \dots, R_G\}$ // Tree consisted of G regions.

22. It is defined as follows: $d_E(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^K (x_k - y_k)^2}$ (Kubat, 2017, p. 46).

The procedure above shows how a decision tree is built. In regression problems, Algorithm 3 chooses the split (i.e., variable j and point s) which minimises the RSS of the two resulting regions $R_1(j, s) = \{X|X_j < s\}$ and $R_2(j, s) = \{X|X_j \geq s\}$.

Here, $\hat{y}_{R_g} = \frac{1}{n_g} \sum_{i: x_i \in R_g} y_i$, where n_g is the number of observations which lie in R_g . In classification problems, Algorithm 3 minimises the *Gini index*²³ instead of RSS, as the latter cannot be used in a classification setting.

Here, $\hat{p}_{R_g, c} = \frac{1}{n_g} \sum_{i: x_i \in R_g} \{1_{y_i=c}\}$, where notation $1_{y_i=c}$ means that this quantity equals one, if $y_i = c$ or zero otherwise. Quantity $\hat{p}_{R_g, c}$ is also an estimate of the probability that an observation of class c lies in region R_g . As recursive binary splitting algorithm builds a large tree, which is very well fitted to the training data, it becomes necessary to properly 'prune' it in order to achieve better generalisability. By and large, this is what the next steps of Decision Trees algorithm do in order to produce more accurate predictions.

- **Random Forests:** Random Forests is an algorithm which is based on decision tree estimators and it aims to reduce their high variance. Its general form is the following:

Algorithm 4 Random Forests

INPUT: $X_{n \times J}$, $x_i \in \mathbb{R}^J$, $C = \{c_1, \dots, c_M\}$, $y \in \mathbb{R}^n$ or $y \in C^n$, n_{min} , B , n_b , h

Step 1: For $b = 1, 2, \dots, B$: Draw with replacement a random subsample²⁴ of size n_b from observations given.

Step 2: For each subsample build a decision tree T_b as follows:

While there is a region containing more than n_{min} observations:

- 1) Choose randomly h out of the J variables.
- 2) From these h variables, select the variable j and the point s which minimise quantity 17 or 18, depending on the problem.
- 3) Split variable j at point s .

OUTPUT: $RF = \{T_1, \dots, T_B\}$

23. In general, the Gini index is defined as follows: $G = \sum_{c \in C} \hat{p}_{gc} (1 - \hat{p}_{gc})$, where \hat{p}_{gc} is the proportion of observations from class c in region g .

24. A so-called *bootstrap sample*.

We observe that the output of Algorithm 4 is a set of B decision trees. In order to make a single prediction, we need to perform an action called *bootstrap aggregation* – also known as *bagging*:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}) \quad (19)$$

$$\hat{y} = \text{mode}(\{T_1(\mathbf{x}), \dots, T_B(\mathbf{x})\}), \quad (20)$$

where Eq. 19 is used for predictions in regression problems and Eq. 20 for predictions in classification problems. Notation $T_b(\mathbf{x})$ indicates the prediction of tree T_b , given observation \mathbf{x} .

- **Boosted Regression Trees (BRT):** BRT algorithm is an extension of the Decision Trees algorithm and, specifically, of regression trees. It has the following form:

Algorithm 5 Boosted Regression Trees (BRT)

INPUT: $\mathbf{X}_{n \times J}$, $\mathbf{x}_i \in \mathbb{R}^J$, $\mathbf{y} \in \mathbb{R}^n$, B , λ , d // Data and tuning parameters.

Step 1: Set $\hat{f}(\mathbf{x}) = 0$ and $r_i = y_i \forall i \in \{1, \dots, n\}$, where r_i represents the residual of the i -th observation. // \mathbf{r} is a vector of residuals.

Step 2: For $b = 1, 2, \dots, B$ repeat:

- Fit a tree \hat{f}^b with d splits (i.e., $d + 1$ terminal nodes) to the training data (\mathbf{X}, \mathbf{r}) .
- Update \hat{f} by adding in a shrunken version of the new tree:

$$\hat{f}(\mathbf{x}) \leftarrow \hat{f}(\mathbf{x}) + \lambda \hat{f}^b(\mathbf{x}) \quad (21)$$

- Update residuals:

$$r_i \leftarrow r_i - \lambda \hat{f}^b(\mathbf{x}_i) \quad (22)$$

Step 3: The boosted model $\hat{f}(\mathbf{x})$ is:

$$\hat{f}(\mathbf{x}) = \sum_{b=1}^B \lambda \hat{f}^b(\mathbf{x}) \quad (23)$$

OUTPUT: $\hat{f}(\mathbf{x})$

Boosting is a technique that builds an estimator sequentially (James *et al.*, 2013, p. 321). This means that, at each step, information from the previously constructed estimator(s) is used, which is what Step 2 of Algorithm 5 does.

- Artificial Neural Networks (ANN): We focus on the simplest class of ANN, the *feed-forward neural network*, also known as *multilayer perceptron* (MLP). We still use notation X_1, \dots, X_K for independent (explanatory) variables and Y for the dependent variable. In the ANN context, we call the former *input variables*, because they are the input of an ANN system. Similarly, the result related to Y (either a real value or a class probability, for regression or classification problems, respectively) is the *output* of an ANN system. What we want to do with an ANN model is to approximate a nonlinear function f , for which:

$$y(\mathbf{x}, \mathbf{w}) = f\left(\sum_{j=1}^M w_j \varphi_j(\mathbf{x})\right), \quad (24)$$

where y corresponds to the output, \mathbf{x} is an observation vector of length K , \mathbf{w} is a vector of some weight parameters, f is called *activation function*²⁵ and φ is a nonlinear *basis function* (i.e., a nonlinear transformation of vector \mathbf{x}). We observe that output y is a nonlinear transformation of M linear combinations. The following equations compose the mathematical representation of a feed-forward neural network model:

$$a_j = \sum_{k=1}^K w_{jk}^{(1)} x_k + w_{j0}^{(1)}, \quad j \in \{1, \dots, M\}. \quad (25)$$

Parameters w_{jk} are called *weights*, parameters w_{j0} are the *biases*²⁶, a_j s are called *activations* and superscript '(1)' denotes that the corresponding quantities belong to the first layer of the network. For each a_j there is a differentiable nonlinear activation function h :

$$z_j = h(a_j) \quad (26)$$

Quantities z_j are called *hidden units*. In general, h is chosen to be a sigmoidal function. The following holds for the second layer:

$$a_l = \sum_{j=1}^M w_{lj}^{(2)} z_j + w_{l0}^{(2)}, \quad l \in \{1, \dots, L\}. \quad (27)$$

25. Which means that, given some input, it is a function that determines the output.

26. Here, the term *bias* is used to describe a parameter w_0 that allows any fixed offset in the data (Bishop, 2006, p. 138). It is often useful to define $\varphi_0(\mathbf{x}) = 1$. In the linear regression context, the bias parameter is the intercept.

If there are only two layers, then L is the number of outputs and quantities a_l are called *output unit activations*. Finally, the output units are calculated as follows:

$$y_l = \sigma(a_l), \quad (28)$$

where activation function σ is the identity function for regression problems, the logistic sigmoid function for (multiple) binary classification problems (Eq. 29) or the softmax function (Bishop, 2006, p. 198) for multiclass classification problems (Eq. 30):

$$\sigma(a_l) = \frac{1}{1 + \exp(-a_l)} \quad (29)$$

$$\sigma(\alpha_l) = \frac{\exp(a_l)}{\sum_{r=1}^L \exp(a_r)}. \quad (30)$$

Eqs. 29 and 30 represent the conditional probability for an observation to belong to class l given \mathbf{x} , i.e. they give values in $[0,1]$. Putting it all together, the MLP model takes the following functional form:

$$y_l(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=1}^M w_{lj}^{(2)} h \left(\sum_{k=1}^K w_{jk}^{(1)} x_k + w_{j0}^{(1)} \right) + w_{l0}^{(2)} \right), \quad (31)$$

$$\forall l \in \{1, \dots, L\}.$$

For more than two layers, the generalisation is straightforward.

Results

In the previous section, we presented the important methodological aspects of this paper. In this section, we focus on result evaluation. The first question that needs to be answered is whether the Average Trees algorithm can provide more reliable results than classic Decision Trees and Random Forests. Subsequently, another question is about which method(s) perform(s) better in our datasets. And, apart from the questions related to methods, we need to provide an answer to the question about which economic theory seems more plausible according to the evidence – if any of them stands out. The results presented in this section contribute towards finding answers to these questions.

The evaluation procedure was based on K -fold cross-validation. Parameter K was chosen for each country to ensure a sufficient number of observations is included in test sets²⁷ but, in parallel, that the training sets also are of a sufficient size for building

27. In general, the target was fifteen observations.

the most reliable models possible. The following metrics were used for evaluating each method: 1) Classification Accuracy, 2) Sensitivity, 3) Precision, 4) Specificity, 5) False Alarm and 6) F_1 -Score. For each method, the results from the K -fold cross-validation are summarized by calculating a *weighted average* on each metric. The weights are proportional to test set sizes. It is expected that the K -th test set is usually of smaller size than the previous $K-1$ test sets, because the division of observation number by K is likely to give a non-zero remainder. The last test set may show, for example, 100% Classification Accuracy, since it consists of only one (correctly classified) observation. But it is obvious that such results do not have the same significance as those from the other $K-1$ test sets. Thus, the impact of the last test set should be shrunk proportionally to the number of cases used for evaluating a classifier on this set. This detail is taken care of by using the weighted average mentioned above.

The ten methods evaluated are the following²⁸: Average Trees (both variants of the algorithm; i.e., K -fold sampling and random sampling with replacement), Decision Trees, Random Forests, Logit, Probit, k -NN, Boosted Regression Trees (Logistic Regression model), Support Vector Machines, and Artificial Neural Networks (single-layer, feed-forward).

It has been mentioned that the important variables were selected through a procedure based on Random Forests. This was a first step to reasonably reduce the number of variables from ~ 150 to ~ 10 , according to their MDG. However, depending on the method, an additional approach was followed in order to further decrease model complexity, in cases where such a decrease improves the generalisability of a model. To be more specific, for the Logit and Probit models a stepwise forward selection was applied, using AIC as the model selection criterion. For the k -NN algorithm, Principal Component Analysis (PCA)²⁹ was applied before feeding the model with data, in order to avoid the ‘curse of dimensionality’³⁰. The number of principal components selected was defined manually for each country, by looking at the output of R function `plot.prcomp()`, which shows the proportion of variance explained by each principal component. The rationale behind the choice of this number is subjective and similar to that of the initial variable selection based on MDG. As for recursive partitioning methods (including the Average Trees), their model complexity was left

28. The corresponding R packages used are the following (explanation is given if package and function have different names): `rpart`, `randomForest`, `stats` (for `glm()`), `class` (for `knn()`), `gbm`, `e1071` (for `svm()`) and `nnet`.

29. More details about this method can be found in James *et al.* (2013, pp. 230-237).

30. According to Kubat (2017, p. 54), this term describes a situation where, as the number of explanatory variables increases, it becomes less likely that two observations are close to each other in the high-dimensional space. Therefore, it is hard to distinguish whether the large distance between them indicates class differentiation or not. Nevertheless, one may overcome this problem by increasing the number of observations or by applying a dimension reduction technique, such as the PCA.

to be controlled by the internal R procedures; i.e., the relevant parameters remained at their default values. For the BRT, parameter shrinkage³¹ was chosen such that the out-of-sample Classification Accuracy would be the largest. Regarding the SVM, a range of values were tried for each country regarding parameters C and γ ³². A three-dimensional plot was proved very helpful for finding ‘parameter areas’ of high Classification Accuracy (an example is presented in Appendix, A.5). The values selected were those that gave the largest mean out-of-sample Classification Accuracy. Finally, regarding the ANN, a *weight decay regularization*³³ was applied by searching over a set of values that provided better out-of-sample performance. Note that for any method used there is no guarantee for a globally optimal parameter setting, since the latter was selected by trial and error.

Australia

The presentation of evaluation results begins with Australia. In this dataset, data are from 1972.Q1 to 2014.Q4 (length: 43 years – 172 observations). Figure 1 shows the results of `varImpPlot()` for the initial variable selection.

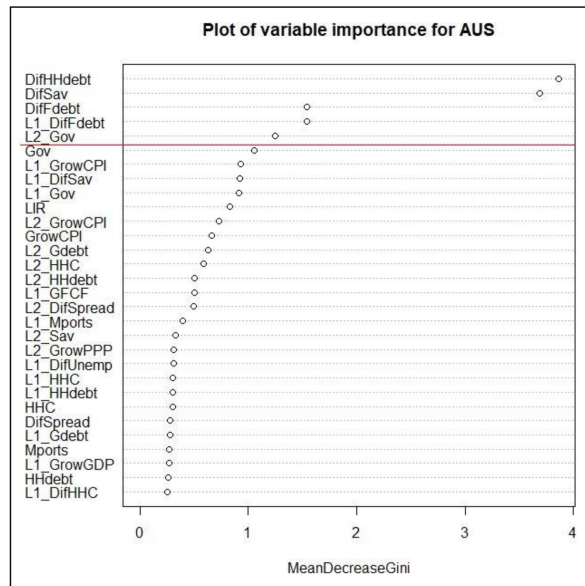


Figure 1. The five – out of 131 – variables chosen for Australia (those above the red line).

31. This is parameter λ from Eq. 23.

32. See also optimization problem 10-13 and Eq. 16. Apart from the radial kernel, the sigmoid was also tried. It is defined as: $K(\mathbf{x}_i, \mathbf{x}_{i'}) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_{i'} + r)$. For reasons related to computation time, parameter r was left at default value 0.

33. Weight decay regularization is a method for reducing potential overfitting of an ANN model. This is done by decaying some weight parameters towards zero. More details can be found in Bishop (2006, pp. 256-257).

Table 4 contains the out-of-sample performance of the methods described above in six evaluation metrics. In Appendix, A.8, one can also find the corresponding tables for the in-sample performance of the methods presented. Their parameterisation is the same as that emerged from the out-of-sample evaluation procedure. Moreover, test sets are also of the same size; the difference is that the training sets consist of all available observations. Lastly, in Figure 8 one can visually compare the performance of all methods presented.

There was some doubt about including *L2_Gov* in the set of important variables. As their total number is relatively small, the final decision was to include it.

Table 4. Evaluation results – Australia. Each cell is the average of the results of each test set. Regarding the *K*-fold cross-validation procedure: $K = 12 \rightarrow$ test set size = 15.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (<i>K</i> -fold)	81.98%	45%	20.54%	88.59%	11.41%	37.96%	ex.size = 0.12
Average Trees (random)	86.63%	53.57%	42.71%	90.92%	9.08%	54.86%	exsize = 0.45, tree nr = 200
Decision Trees	81.98%	45%	20.54%	88.59%	11.41%	37.96%	probability threshold: 0.5
Random Forests	88.37%	55%	47.78%	94.39%	5.61%	61.11%	probability threshold: 0.6
Logit	88.95%	45%	43.33%	96.28%	3.72%	53.33%	probability threshold: 0.75
Probit	89.53%	35%	54.17%	97.45%	2.55%	45%	probability threshold: 0.75
<i>k</i> -NN	90.7%	60%	68.89%	96.4%	3.6%	79.05%	<i>k</i> = 6, # of principal components: 2
BRT	90.7%	45%	77.78%	98.41%	1.59%	76.19%	shrinkage = 0.018, probability threshold: 0.45
SVM	92.44%	86.43%	63%	92.76%	7.24%	70.74%	$\gamma=0.0005$, $C=1.3$, sigmoid kernel, probability threshold: 0.15
ANN	90.12%	88.57%	52.06%	89.92%	10.08%	69.43%	probability threshold: 0.54, # of hidden nodes: 5

The probability thresholds mentioned in some cells of “Comments” column refer to the *least* estimated probability for which the predicted class is “1 – Recession³⁴”; i.e., any estimated probability under the threshold gives the prediction “0 – No Recession”. These thresholds were chosen so that the value of Classification Accuracy is the maximum possible³⁵. Methods with no reference to probability thresholds are those

34. Actually, class “1” refers to a pre-recessionary period.

35. Thresholds that produced Sensitivity=0 or False Alarm=1 were not considered in any dataset.

that do not provide predictions in probabilistic form. Regarding the parameters of Average Trees algorithm, *ex.size* refers to the percentage of observations to be excluded³⁶ and *tree.nr* refers to parameter *nt* from Algorithm 1.

Germany

We move on to the German dataset. The data in this case are from 1973.Q1 to 2014.Q4 (length: 42 years – 168 observations).

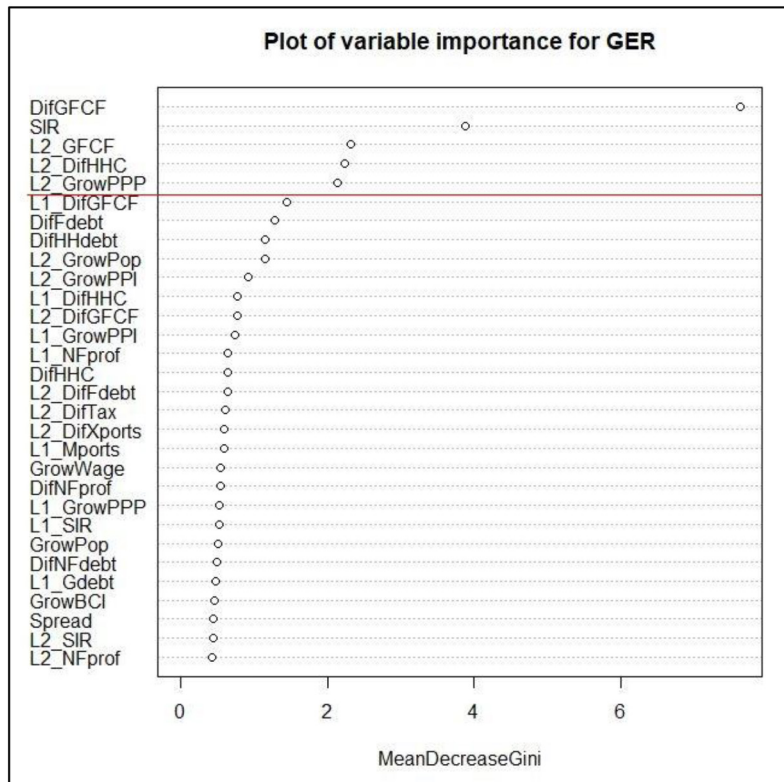


Figure 2. The five – out of 146 – variables chosen for Germany.

36. In Algorithm 1, this quantity is expressed in integer form (parameter *s*).

Table 5. Evaluation results – Germany³⁷.Regarding the K -fold cross-validation procedure: $K=11 \rightarrow$ test set size=16.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F_1 -Score	Comments
Average Trees (K -fold)	82.74%	43.75%	67.86%	94.44%	5.56%	59.34%	ex.size = 0.13
Average Trees (random)	84.52%	50%	72.22%	95.24%	4.76%	65.56%	ex.size=0.225, tree.nr=200
Decision Trees	80.95%	18.75%	75%	98.41%	1.59%	50%	probability threshold: 0.8
Random Forests	86.9%	71.88%	71.88%	91.27%	8.73%	68.54%	probability threshold: 0.37
Logit	83.93%	31.25%	88.89%	99.21%	0.79%	85.71%	probability threshold: 0.85
Probit	83.93%	31.25%	88.89%	99.21%	0.79%	85.71%	probability threshold: 0.85
k -NN	87.5%	53.13%	71.43%	95.63%	4.37%	72.38%	$k = 3$, # of principal components: 3
BRT	85.71%	68.75%	72.08%	90.48%	9.52%	66.96%	shrinkage = 0.01, probability threshold: 0.3
SVM	85.12%	59.38%	57.21%	90.48%	9.52%	60.93%	$\gamma=0.0014$, $C=2.9$, radial kernel, probability threshold: 0.35
ANN	79.76%	54.69%	52.5%	84.52%	15.48%	55.26%	probability threshold: 0.67, # of hidden nodes: 5

37. Regarding the number of principal components in the case of k -NN algorithm, the initial choice based on the visual output of `plot.pcomp()` was to choose two principal components. However, it was discovered that choosing three improves mean Classification Accuracy.

Japan

We proceed to the dataset of Japan. In this case, the data are from 1973.Q1 to 2015.Q4 (length: 43 years – 172 observations).

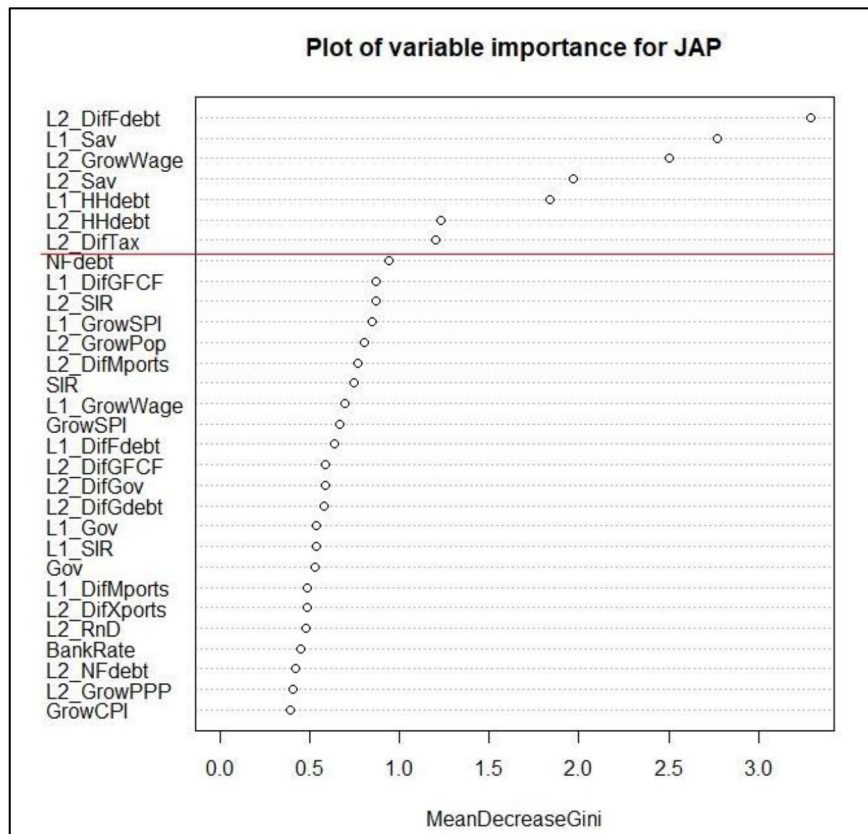


Figure 3. The seven – out of 128 – variables chosen for Japan.

Table 6. Evaluation results – Japan³⁸.Regarding the K -fold cross-validation procedure: $K=12 \rightarrow$ test set size=15.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	81.4%	32.93%	38.16%	90.63%	9.37%	48.96%	ex.size = 0.09
Average Trees (random)	81.4%	32.93%	38.16%	90.63%	9.37%	48.96%	ex.size = 0.1, tree.nr = 200
Decision Trees	83.14%	8.54%	33.33%	98.64%	1.36%	50%	probability threshold: 0.8
Random Forests	86.63%	24.8%	77.27%	94.19%	5.81%	83.77%	probability threshold: 0.45
Logit	83.72%	42.07%	31.98%	86.38%	13.62%	49.66%	probability threshold: 0.35
Probit	84.3%	42.07%	31.98%	86.96%	13.04%	49.66%	probability threshold: 0.35
k-NN	86.05%	39.02%	61.19%	95.2%	4.8%	60.9%	k = 2, # of principal components: 3
BRT	81.98%	13.11%	27.65%	96.26%	3.74%	32.95%	shrinkage = 0.035, probability threshold: 0.3
SVM	89.53%	81.71%	61.35%	87.1%	12.9%	75%	$\gamma=0.0049$, $C=1.2$, radial kernel, probability threshold: 0.3
ANN	70.93%	63.41%	33.41%	66.33%	33.67%	57.87%	probability threshold: 0.5, # of hidden nodes: 6

38. Similarly, in this dataset the decision based on the output of `plot.pcomp()` was to select the first two principal components, but the addition of a third one improved Classification Accuracy.

Mexico

The next country to be presented is Mexico. Data in the Mexican dataset are from 1973.Q2 to 2015.Q4 (length: 42.75 years – 171 observations). In Figure 4 we see, for the first time so far, two trade-related variables at the top of the table.

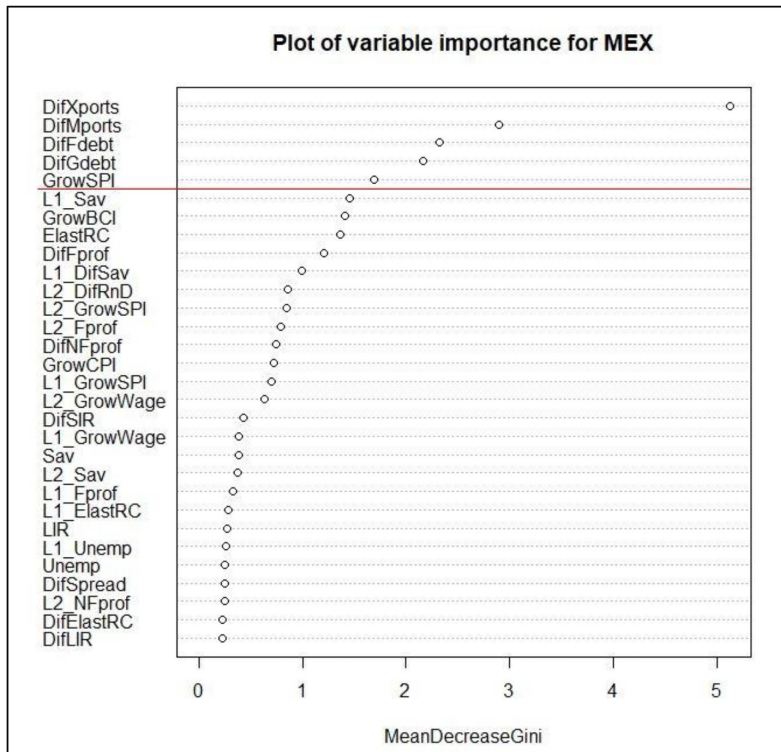


Figure 4. The five – out of 128 – variables chosen for Mexico.

The following Table 7 is the first table that contains results from an ensemble model. As no single model provided satisfactory results for the case of Mexico, an ensemble model was built in order to combine the strengths of the best three models into one and, hopefully, mitigate their weaknesses³⁹. It makes a prediction by applying the majority rule on the predictions of the three models selected. Lastly, it may seem somewhat strange that the Average Trees algorithm with random sampling has mean F_1 -Score 73.33%, while the corresponding values of Sensitivity and Precision are 30% and 41.67%, respectively. Such peculiarities may be found in other datasets, too, since the F_1 -Score is calculated *only if* both Sensitivity and Precision values exist. Particularly for this dataset, only two values were calculated for the F_1 -Score of Average Trees (random), because both Sensitivity and Precision were simultaneously real numbers for only two test sets.

39. Explanations about this topic are presented in [Discussion](#).

Table 7. Evaluation results – Mexico.Regarding the K -fold cross-validation procedure: $K=12 \rightarrow$ test set size=15.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F_1 -Score	Comments
Average Trees (K -fold)	81.29%	50%	35.56%	87.77%	12.23%	66.23%	ex.size = 0.12
Average Trees (random)	82.46%	30%	41.67%	92.56%	7.44%	73.33%	ex.size = 0.1, tree.nr = 200
Decision Trees	80.7%	50%	34.67%	86.98%	13.02%	64.76%	probability threshold: 0.5
Random Forests	84.21%	25%	50%	95.11%	4.89%	46.67%	probability threshold: 0.6
Logit	84.21%	5%	20%	96.81%	3.19%	22.22%	probability threshold: 0.5
Probit	81.29%	5%	5.56%	93.04%	6.96%	10%	probability threshold: 0.4
k -NN	82.46%	22.5%	25.71%	92.72%	7.28%	27.78%	$k = 1$, # of principal components: 4
BRT	85.96%	20%	43.75%	97.24%	2.76%	38.33%	shrinkage = 0.01 probability threshold: 0.3
SVM	84.8%	22.5%	68.18%	94.42%	5.58%	37.78%	$\gamma=0.006$, $C=0.08$, radial kernel, probability threshold: 0.16
ANN	85.38%	42.5%	48.81%	92.66%	7.34%	46.21%	probability threshold: 0.55, # of hidden nodes: 3
Ensemble	85.96%	20%	50%	97.45%	2.55%	41.9%	Majority rule: {Average Trees (K -fold), BRT, ANN}

United Kingdom (UK)

We move on to the dataset for the United Kingdom. In this case, the data are from 1973.Q2 to 2014.Q4 (length: 41.75 years – 167 observations).

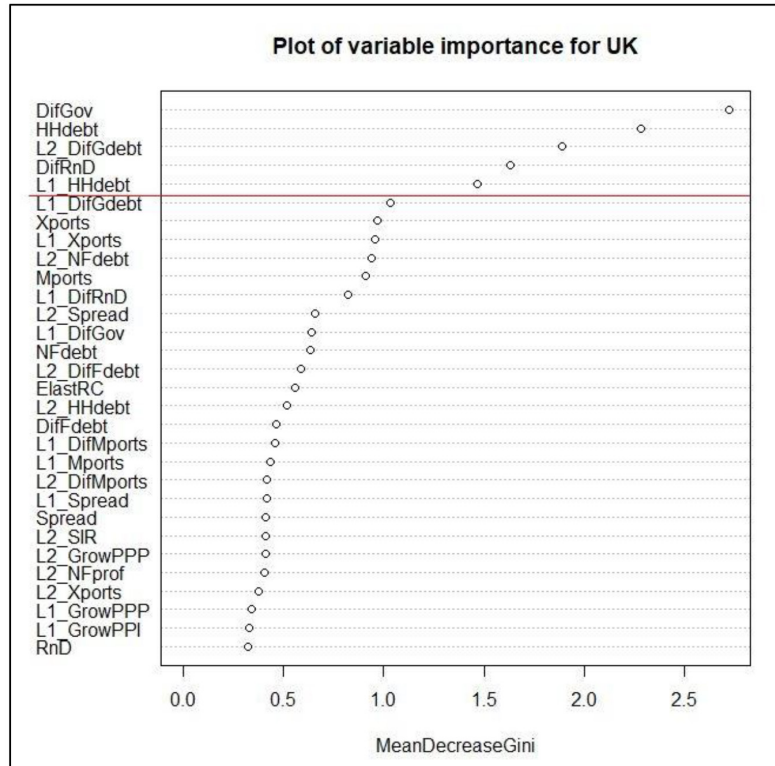


Figure 5. The five – out of 149 – variables chosen for the UK.

Table 8. Evaluation results – United Kingdom.Regarding the K -fold cross-validation procedure: $K=11 \rightarrow$ test set size=16.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F_1 -Score	Comments
Average Trees (K-fold)	88.62%	50%	41.27%	94.13%	5.87%	76.36%	ex.size = 0.25
Average Trees (random)	88.62%	50%	41.27%	94.13%	5.87%	76.36%	ex.size = 0.15, tree.nr = 200
Decision Trees	88.62%	43.75%	35.42%	95.13%	4.87%	77.5%	probability threshold: 0.5
Random Forests	88.62%	33.33%	64.58%	96.33%	3.67%	64.29%	probability threshold: 0.7
Logit	88.02%	20.83%	62.5%	97.13%	2.87%	71.43%	probability threshold: 0.5
Probit	88.02%	20.83%	62.5%	97.13%	2.87%	71.43%	probability threshold: 0.5
k -NN	89.82%	37.5%	60%	98.6%	1.4%	77.78%	$k = 10$, # of principal components: 2
BRT	89.22%	43.75%	63.33%	95.53%	4.47%	73.33%	shrinkage = 0.01 probability threshold: 0.3
SVM	85.63%	12.5%	22.22%	93.41%	6.59%	47.06%	$\gamma=0.0003$, $C=0.009$, radial kernel, probability threshold: 0.95
ANN	86.83%	12.5%	22.22%	96.41%	3.59%	57.14%	probability threshold: 0.51, # of hidden nodes: 4
Ensemble	88.62%	50%	41.27%	94.13%	5.87%	76.36%	Majority rule: {Average Trees (random), BRT, k -NN}

United States of America (USA)

The dataset for the USA consists of data from 1973.Q1 to 2014.Q4 (length: 42 years – 168 observations).

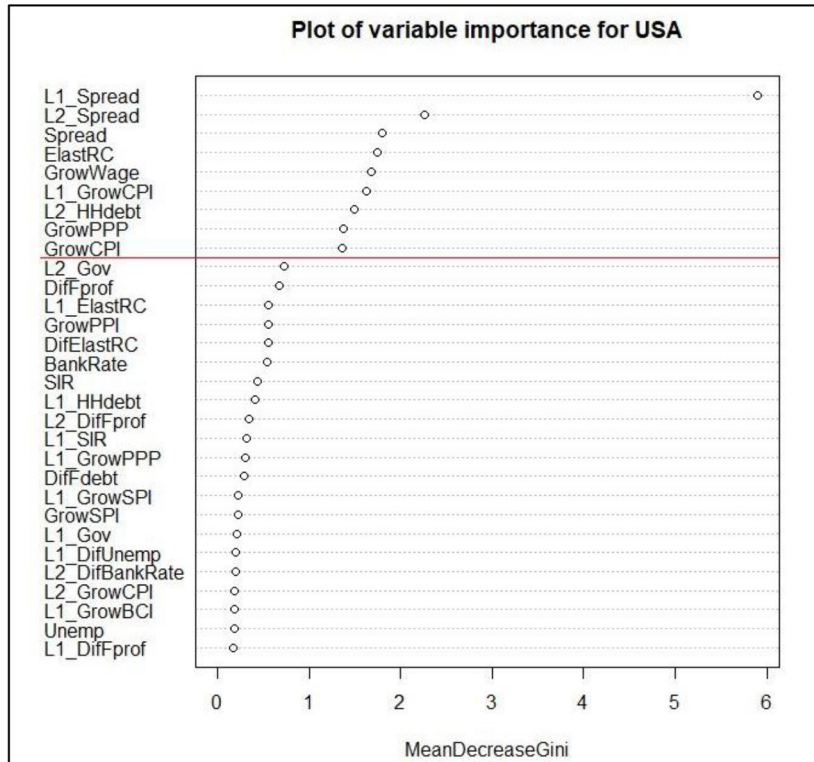


Figure 6. The nine – out of 149 – variables chosen for the USA.

Table 9. Evaluation results – USA.Regarding the K -fold cross-validation procedure: $K=11 \rightarrow$ test set size=16.

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F_1 -Score	Comments
Average Trees (K -fold)	86.9%	50%	46.94%	90.48%	9.52%	50.38%	ex.size = 0.1
Average Trees (random)	87.5%	55%	48.61%	90.48%	9.52%	53.72%	ex.size = 0.1, tree.nr = 200
Decision Trees	86.9%	50%	46.94%	90.48%	9.52%	50.38%	probability threshold: 0.5
Random Forests	90.48%	50%	63.49%	95.24%	4.76%	69.96%	probability threshold: 0.4
Logit	91.07%	45%	65.56%	96.83%	3.17%	67.41%	probability threshold: 0.8
Probit	92.26%	45%	76.67%	98.41%	1.59%	74.07%	probability threshold: 0.92
k -NN	92.26%	55%	77.78%	96.83%	3.17%	81.9%	$k = 10$, # of principal components: 2
BRT	91.07%	40%	73.33%	97.62%	2.38%	84.44%	shrinkage = 0.009, probability threshold: 0.5
SVM	93.45%	75%	73.33%	95.24%	4.76%	71.43%	$\gamma=0.00025$, $C=0.92$, radial kernel, probability threshold: 0.25
ANN	88.69%	65%	42.98%	90.67%	9.33%	62.93%	probability threshold: 0.52, # of hidden nodes: 5

Discussion

In order to specify the conditions that precede economic recessions, the author's choice was to develop a method based on Decision Trees. The goal of the so-called 'Average Trees algorithm' is to provide results with the straightforward interpretability of the classic Decision Trees and the robustness of Random Forests. The choice of developing a method based on Decision Trees was made because our central topic is to specify the macroeconomic conditions before a recession commences. Decision trees provide results in a form that best suits this purpose. The question is whether Average Trees have better out-of-sample performance than classic Decision Trees and/or Random Forests.

Before answering this question, one can realise from Tables 4 – 9 that the K -fold variant of Average Trees *never* provided better mean Classification Accuracy than the variant of random sampling with replacement. Therefore, we can argue that random sampling improves the generalisability of the method, probably because, at each iteration, the observations excluded are not from a very specific period of the dataset, but they may be from *any* period with the same probability. So, for this vari-

ant, the information that is considered during the training phase is generally from the entire available time span. Thus, the main question is whether the performance of the random sampling variant of Average Trees exceeds that of Decision Trees and/or Random Forests. With the single exception of the Japanese dataset, classic Decision Trees never showed better performance than the Average Trees random sampling variant in terms of mean Classification Accuracy. Indeed, in four out of the remaining five datasets, Decision Trees performance was worse than that of Average Trees (random sampling). However, in no dataset did the random sampling variant of Average Trees have better performance than Random Forests. Thus, after this first application of the Average Trees algorithm on real datasets, we can say that its models tend to generalise better than classic Decision Trees without losing their straightforward interpretability. However, it seems that Average Trees cannot achieve better performance than Random Forests.

The Average Trees algorithm was developed for this paper in order to identify rules that lead to recessions. However, before looking at any data, it was not known whether such a concept existed. In other words, it may be the case that true classes cannot be efficiently separated in the high-dimensional space by a recursive partitioning method. So, even if it was not possible to identify such global rules accurately predicting recessions, the goal of predicting the latter using some other methods would still exist. For this reason, many statistical and machine learning methods were examined in the framework of this paper. In the previous section we reported the performance of ten different methods based on six datasets and the question is whether some of these methods are consistently better than others.

Before proceeding to answer this last question, it is important to define what the term 'better method' means. Classification Accuracy is an important metric for evaluating different classification methods, but this number alone is not always the only thing we care about. Especially in our problem, which holds that class "0" appears much more often than class "1" due to the rarity of economic recessions, a classifier could possibly achieve more than 80% Classification Accuracy just by correctly predicting only class "0". Such an example is the Logit classifier from Table 7, which achieved Classification Accuracy 84.2%, with Specificity 96.8%, but only 5% in Sensitivity. Apparently, this model should not be used by policymakers for predicting recessions in Mexico, despite its indisputably sound Classification Accuracy. This example makes it clear that, in order to decide which method is better, we must also take into consideration metrics other than Classification Accuracy. In the author's opinion, a model can be considered good in the framework of this paper, if its Classification Accuracy is at least 85%, its Sensitivity and Precision at least 70%, and its False Alarm at most 10%. Of course, this is a reasonable, albeit subjective, choice. In practice, the method to be chosen depends a lot on how much FPR can be tolerated.

Before discussing which methods tend to prevail, it is important to shed some

light on the interpretation of predictions. Taking into account how variable *PreRecess* was constructed, we realise that if a classifier predicts class “1” for a quarter Q_t , it essentially predicts that a recession is going to begin within that year; i.e., in one of the four quarters from Q_t to Q_t+3 . This holds because we have focused on pre-recessionary periods rather than on recessions *per se*.

Tables 4 – 9 show the out-of-sample predictive performance of ten methods in six datasets and, at this point, the best ones for each country are presented. We begin our analysis with Table 9 and the USA. For this dataset, the method to be selected seems to be an easy decision. SVM achieved the highest Classification Accuracy (93.45%), having the highest Sensitivity (75%) and one of the lowest values for False Alarm (4.76%). Also, the fact that it has the third highest Precision (73.33%) among all tested models makes SVM a very reasonable choice for predicting recessions in the USA. We move on to Table 4 and the dataset for Australia. In this case, the decision is not as obvious as the previous one. The BRT model achieved the highest Precision (77.78%) with the lowest False Alarm (1.59%), but, despite its good Classification Accuracy (90.7%), its Sensitivity is not adequate (45%). The two models distinguished here are those of SVM and ANN. The latter achieved the highest Sensitivity of the table (88.57%), with Classification Accuracy at 90.12%, but it has the third highest False Alarm (10.08%). On the other hand, SVM has the highest Classification Accuracy of the table (92.44%), with the second highest Sensitivity (86.43%) and its False Alarm is 7.24%. The fact that SVM achieved better Precision than ANN (the third highest of the table: 63% versus 52.06%) is an additional argument favouring the opinion that SVM is the most appropriate method for the case of Australia as well.

Regarding the dataset for Japan (Table 6), the most appropriate model seems to be, once again, that of SVM. It has the highest Classification Accuracy (89.53%), the highest Sensitivity (81.71%) and the second largest Precision (61.35%). Its only weakness seems to be the False Alarm of 12.9%. However, the methods that follow in terms of Classification Accuracy (Random Forests and k -NN) – which, additionally, have lower False Alarm – achieved Sensitivity under 40%. So, in order to avoid the cost of missing a lot of true positives, we would rather tolerate some more false positives than those corresponding to our acceptable percentage and, consequently, select the SVM. In the dataset for Germany (Table 5), k -NN algorithm achieved the highest Classification Accuracy of the table: 87.5%. However, its moderate Sensitivity (53.13%) discourages us from stating that it can be used for predicting recessions in Germany. The most suitable model here seems to be that of Random Forests. It achieved Classification Accuracy 86.9%, with Sensitivity 71.88% (the highest of the table), Precision 71.88% and False Alarm 8.73%.

Regarding the dataset for Mexico (Table 7), it is not so easy to reach a decision about which method performs best. For a moment, let us ignore the performance of the Ensemble model. In such a table, Decision Trees and Average Trees algorithm

(*K*-fold) show the highest Sensitivity (50%), but they exhibit the two highest False Alarm values (13.02% and 12.23%, respectively). Note the fact that the highest Sensitivity in the dataset for Mexico is only 50%, which disputes its data quality. The BRT model holds the highest Classification Accuracy (85.96%) and the lowest False Alarm (2.76%), but its Sensitivity is only 20%, which makes it seem useless. So, if we ignore the Ensemble model, the best choice for the case of Mexico appears to be the ANN. It has the second largest Classification Accuracy of the table (85.38%), Sensitivity equal to 42.5% (it is the next value after the 50% value mentioned above), the third largest Precision (48.81%) and False Alarm 7.34%. However, since none of these methods alone provided satisfactory results, the idea was to build an ensemble model that makes best use of all three of them. More specifically, the Ensemble model applies the majority rule on the predictions of the Average Trees (*K*-fold), BRT and ANN models⁴⁰ in order to make a rather more accurate prediction. Although it achieved the highest Classification Accuracy (85.96%; same as BRT) and the lowest False Alarm of the table (2.55%; even better than BRT), it exhibited only 20% Sensitivity. Therefore, after considering the entire table, it seems that ANN is probably the best choice.

A similar situation exists in the results of the UK dataset (Table 8). The highest value for Sensitivity is 50% and it is achieved only for both variants by the Average Trees algorithm. They provided identical results: Classification Accuracy 88.62%, Precision 41.27% and False Alarm 5.87%. The highest Classification Accuracy (89.82%) is achieved by the *k*-NN algorithm, which also has the lowest False Alarm value (1.4%) and Sensitivity equal to 37.5%. However, the author's decision here was to choose the BRT method. Its model achieved the second largest Classification Accuracy (89.22%), the second highest Sensitivity (43.75%) and the second largest Precision of the table (63.33%). Its False Alarm is 4.47%, which seems quite acceptable. However, we should not neglect the fact that it was not possible for the UK dataset, either, to find a model that achieved above 50% Sensitivity with good Classification Accuracy. For this reason, an ensemble model was constructed for the UK, too, but its results were identical to those of Average Trees. Thus, this country is another case for which it was not possible to find a satisfactory model. If this did not happen because of poor data quality or because of omitting some important variables, then it may be the case that, for some methods, the correct set of parameters cannot not be found. As it is computationally infeasible to search through all possible combinations among model parameters (e.g. *C* and γ in the SVM framework), it is likely that we missed the opportunity to fit a better model for Mexico and the UK due to computational restrictions. Table 10 shows which methods were selected by the author for each country:

40. Their parameterisation remained unchanged.

Table 10. The methods selected per country.

Country	Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score
AUS	SVM	92.44%	86.43%	63%	92.76%	7.24%	70.74%
GER	Random Forests	86.9%	71.88%	71.88%	91.27%	8.73%	68.54%
JAP	SVM	89.53%	81.71%	61.35%	87.1%	12.9%	75%
MEX	ANN	85.38%	42.5%	48.81%	92.66%	7.34%	46.21%
UK	BRT	89.22%	43.75%	63.33%	95.53%	4.47%	73.33%
USA	SVM	93.45%	75%	73.33%	95.24%	4.76%	71.43%

The mean performance of each method is presented in summary in the two tables of Figure 7:

Out-of-sample				In-sample			
Method	Accuracy	Sensitivity	False Alarm	Method	Accuracy	Sensitivity	False Alarm
Average Trees (K-fold)	83.82%	45.28%	8.99%	Average Trees (K-fold)	92.81%	68.73%	3.74%
Average Trees (random)	85.19%	45.25%	7.67%	Average Trees (random)	91.05%	72.20%	6.30%
Decision Trees	83.72%	36.01%	6.96%	Decision Trees	92.23%	65.61%	2.84%
Random Forests	87.54%	43.34%	5.58%	Random Forests	99.60%	97.74%	0%
Logit	86.65%	31.53%	4.56%	Logit	88.50%	36.19%	3.19%
Probit	86.56%	29.86%	4.63%	Probit	88.30%	33.96%	3.07%
k-NN	88.13%	44.53%	4.10%	k-NN	95.07%	76.08%	2.09%
BRT	87.44%	38.44%	4.08%	BRT	93.50%	70.98%	3.66%
SVM	88.50%	56.25%	7.77%	SVM	81.39%	58.12%	16.24%
ANN	83.62%	54.45%	13.25%	ANN	85.11%	59.92%	12.86%

Figure 7. Mean performance of each method in terms of Classification Accuracy, Sensitivity and False Alarm.

The same information is presented in the bar chart of Figure 8. From these two figures we realise that SVM had the best out-of-sample performance on average, in terms of Classification Accuracy and Sensitivity. Moreover, we see that the Logit and Probit models showed very poor performance in terms of Sensitivity. The Average Trees algorithm (especially the variant that performs random sampling) provided better results than classic Decision Trees and, on average, it outperformed ANN in terms of Classification Accuracy. As for the in-sample performance, we observe that the Random Forests models tend to perfectly overfit their training data.

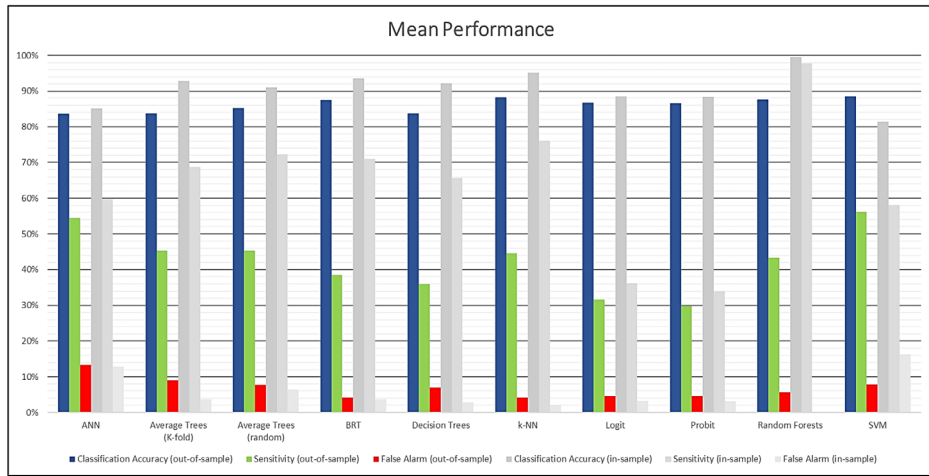


Figure 8. Mean performance of each method.

Australia Method: SVM							Mexico Method: ANN						
Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score	Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score
3 months	87.50%	100.00%	21.05%	87.07%	12.93%	34.78%	3 months	83.19%	12.50%	25.00%	94.17%	5.83%	16.67%
6 months	88.24%	100.00%	22.22%	87.83%	12.17%	36.36%	6 months	83.05%	18.75%	30.00%	93.14%	6.86%	23.08%
1 year	88.89%	100.00%	23.53%	88.50%	11.50%	38.10%	1 year	81.90%	12.50%	22.22%	93.00%	7.00%	16.00%
2 years	91.15%	100.00%	28.57%	90.83%	9.17%	44.44%	2 years	83.04%	15.38%	20.00%	91.92%	8.08%	17.39%

Germany Method: Random Forests							UK Method: BRT						
Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score	Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score
3 months	91.45%	75.00%	81.82%	95.70%	4.30%	78.26%	3 months	77.59%	33.33%	18.18%	82.69%	17.31%	23.53%
6 months	84.48%	45.83%	68.75%	94.57%	5.43%	55.00%	6 months	80.00%	33.33%	21.05%	85.44%	14.56%	25.81%
1 year	80.70%	25.00%	60.00%	95.56%	4.44%	35.29%	1 year	76.99%	41.67%	20.83%	81.19%	18.81%	27.78%
2 years	75.45%	8.33%	28.57%	94.19%	5.81%	12.90%	2 years	67.89%	41.67%	15.15%	71.13%	28.87%	22.22%

Japan Method: SVM							USA Method: SVM						
Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score	Horizon	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F1-Score
3 months	80.88%	80.00%	64.00%	81.25%	18.75%	71.11%	3 months	88.89%	75.00%	35.29%	89.91%	10.09%	48.00%
6 months	73.13%	60.00%	54.55%	78.72%	21.28%	57.14%	6 months	88.79%	62.50%	33.33%	90.74%	9.26%	43.48%
1 year	64.62%	50.00%	43.48%	71.11%	28.89%	46.51%	1 year	82.46%	62.50%	22.73%	83.96%	16.04%	33.33%
2 years	68.85%	63.16%	50.00%	71.43%	28.57%	55.81%	2 years	87.27%	62.50%	31.25%	89.22%	10.78%	41.67%

Figure 9. The forecasting performance of the methods selected per country, at four horizons.

In Figure 9, we can assess the performance of the methods selected for different forecasting horizons. Parameterisation remained the same for each model. SVM performed very well for Australia and the USA, while it showed moderate increase in False Alarm for larger forecasting horizons in the case of Japan. Moreover, we observe that the Random Forests model for Germany loses its ability to correctly predict an upcoming recession for horizons longer than six months. Instead, the SVM models seem more robust in increasing the forecasting horizon order. For the case of Australia, we observe perfect performance of SVM in terms of Sensitivity. Since the parameterisation of the SVM model (γ and C) was defined from the same data during

the training phase, it is very likely that an effect of overfitting exists here, despite the fact that forecasting at horizons comprises, by definition, a type of out-of-sample performance evaluation. In general, we could avoid this situation by leaving a specific set of observations out of any training procedure. However, in our framework this would not be a wise decision because recessions are rare events and such a test set probably would barely contain even one pre-recessionary period. In our analysis, every pre-recessionary period appeared at least once in a test set for every model. Lastly, for the cases of Mexico and the UK, the situation remained, more or less, the same.

The second question was whether any method prevails in predicting recessions. From the analysis above it seems that SVM tends to do so. It was chosen for three out of six countries, while no other method was chosen more than once. It probably has the potential to perform better for the other three countries, too, if a more suitable combination of C and γ is found. Nevertheless, the trial and error procedure on the SVM parameter setting could not provide better out-of-sample performance for those countries. As for the Average Trees algorithm, it is true that its results are simple and easily interpretable. However, given that a recession is going to start within next year, it seems that it is not possible to predict the upcoming recession in half of such cases by using this method⁴¹. Despite the fact that such simple rules about macroeconomic variables cannot predict economic recessions in an accurate way⁴², after this analysis we can argue that methods providing fewer insights from an economic perspective, such as SVM or Random Forests, can be proven useful for policymakers because of giving correct early warning signs in the majority of the cases.

Another question was whether a set of general rules that lead to recessions exists, at least for the countries studied. From the findings of this paper the answer is *no*. Important variables differ among countries, so pre-recessionary conditions are necessarily different according to our findings. We can argue that there are some *country-specific* macroeconomic conditions often preceding recessions – this is the output the Average Trees algorithm provides – but we cannot support the opinion that national economies operate in a way that consistently follows such simple rules. At a second glance, we could say that these rules – if they exist – need to be represented by a more complex concept than a tree-like one.

Closing this section, we must also answer the question whether there is any economic theory verified through the findings of this paper. In order to give an answer to this question, we must look at the set of variables selected for each one of the six countries. Each economic theory selected from relevant literature is represented in our analysis through certain variables. Assuming our data contain no systematic errors,

41. This conclusion arises from the method's performance in Sensitivity.

42. An example of Average Trees algorithm visual output is presented in Appendix, [A.6](#).

and they are what they are supposed to be⁴³, we expect that, if a theory is consistent with reality, its variables tend to be characterised as important by the procedure applied. So, by looking at the relevant plots (Figures 1–6), we realise that in all countries except Germany there is always a debt-related variable in the set of the ones selected. For the case of Germany, the first such variable is two positions under the red line – not so far from being chosen, though. The remaining variables appear less frequently in the sets of the ones selected. Debt-related variables were introduced in the datasets due to I. Fisher's debt-deflation theory. Can we say that this is the theory which best corresponds to reality? The answer is *maybe*. While we could agree with Fisher that factors related to debt have an important impact on the evolution of business cycles⁴⁴, we completely miss the 'deflation' part in our analysis. In the context of this paper, deflation means $GrowCPI < 0$. Apart from the USA, there is no other country for which an inflation-related variable is characterised as important. Therefore, we cannot say that the debt-deflation theory is truly verified through this paper. Alternatively, we could say that it will be a rather beneficial decision to include debt-related variables in similar future works.

Concluding remarks and topics for further research

This paper was an opportunity for investigating the topic of economic recession forecasting from a new basis. Instead of incorporating in our analysis only the variables most papers in literature suggest, we started our investigation from point zero. The rationale for choosing the main variables presented stems from theories established very early on in economic science. Through this work it was made possible to test these theories against each other and discover which of their hypotheses are confirmed by data. Moreover, we reviewed some recent papers from relevant literature in order to present the kind of methodologies applied today and to find which additional variables may potentially make good predictors.

One innovation of this paper is focusing on the short period before a recession begins and not on the recession *per se*. The advantage of this choice is that the predictions resulting from it refer to potentially pre-recessionary periods. This means it is very likely that if a policymaker takes them into account, they have the time to design a proper policy and, ultimately, intervene in the economy. If predictions referred to recessionary periods, it would probably be too late for a policymaker to take precautionary measures. On the other hand, though, many explanatory variables behave in a known much clearer manner during recessionary periods than in the last quarters

43. For example, *BCI* is a variable that indeed sufficiently captures the expectations businesspeople have.

44. We found that models based on such variables can predict pre-recessionary periods with adequate accuracy, at least for three out of the six countries (Australia, Japan and the USA).

prior to a recession. For example, unemployment may rise even from the onset of a recession. Thus, following the conventional approach translates to less uncertain predictions. However, in this paper, the choice was to detect early recessionary signs, even allowing for some false alarms. An additional issue that refers to the whole methodology is that of data availability. For us to be able to make a timely prediction, it is important to have updated necessary data, as far as this is possible. However, there are many variables – at least in the OECD database – that are published once a year, which makes such a goal more challenging.

Probably the most important innovation of this paper is the Average Trees algorithm. It generally achieved better out-of-sample performance than classic Decision Trees, while its good interpretability remained unchanged. It is characterised as most important because it provides us with an alternative way to extract decision tree rules, which are very likely more generalisable than classic ones. However, we should not neglect the contribution of all methodological steps before building average trees. Initially, for every dataset more than 100 variables existed, while there were roughly 170 observations. The number of variables had to be significantly smaller, in order to build parsimonious, yet well-generalisable, models in the steps to follow. The variable selection procedure through fitting an initial Random Forests model with 10,000 trees was proved efficient; models of different methods based on only ‘important variables’ showed very good out-of-sample performance in evaluation metrics. Even the inclusion of lagged variables and variables of percentage changes and differences has proved helpful, as numerous such variables were above the red line in variable importance plots. Furthermore, the construction of some new variables seemed to be a correct decision in the attempt to build well-generalisable models. And even if it was somehow expected for variable *Spread*, as we know that many researchers take it into account in similar works, it was probably not expected for variable *ElastRC* from the Marxian analysis⁴⁵. For the case of the USA, these two variables were quite enough to predict more than half of the recessions using the Average Trees algorithm (see Appendix, A.7). Therefore, although Average Trees did not show better out-of-sample performance than Random Forests, it seems that the overall dataset preparation procedure had a positive impact on evaluation results in general.

We saw that for four out of six countries it was possible to find models with satisfactory performance. This is a good sign for the methodology developed, but six datasets cannot probably be considered a sufficiently large number. A suggestion for further research is to repeat the same methodological steps for many more countries and evaluate them in the same manner. Thus, it could probably be possible to draw more certain conclusions regarding the effectiveness of the methodology applied.

45. This is said because it was not possible for the author to find a quantitative study that included a variable identical – or similar – to *ElastRC* in models aiming at the prediction of economic recessions.

Additionally, it would probably be possible to give more certain answers about which theory seems more plausible. Moreover, testing some additional methods may lead us to find satisfactory models for the two countries this was not achieved (i.e., Mexico and the UK). Lastly, probably the most interesting aspect of investigating this topic in a future research project is through a complex network framework. In this paper we completely missed the dimension of interconnectivity among countries. Each country was studied as if it were isolated from the rest of the world. However, economies are interconnected and a recession in country *A* may cause a recession in country *B* after some time. In order to build a realistic model for predicting economic recessions, one should look at this aspect, too, because the global economy does operate as a single system. In fact, countries are not isolated from each other. In this paper it was not feasible to follow such an approach, since it essentially requires training and testing complex network models that encompass dozens of countries. In a more extensive future research, though, it is very likely that such an approach will provide us with even better models, because it may incorporate a very important aspect of business cycles evolution; namely, that of interconnectivity among national economies.

Appendix

A.1 Details about the methods mentioned

With regard to the Logit and Probit models, details can be found in Baltagi (2002, pp. 332-333). Regarding Support Vector Machines (SVM), a very good presentation of the method can be found in James *et al.* (2013, pp. 341-353). The k -NN algorithm can be found in Kubat (2017, p. 44). Decision Trees and Random Forests are presented in James *et al.* (2013, pp. 304-313), and (2013, pp. 319-321), respectively. Artificial Neural Networks (ANN) are extensively presented in Bishop (2006, pp. 225-231). Finally, the Boosted Regression Trees (BRT) algorithm can be found in James *et al.* (2013, p. 323).

A.2 Data references

1) *BankRate*

Bank of England (2018), Official Bank Rate History Data from 1694. <https://www.bankofengland.co.uk/monetary-policy/the-interest-rate-bank-rate> (Accessed on 10 June 2018)

Bank of Japan (2018), The Basic Discount Rates and Basic Loan Rates. <https://www.boj.or.jp/en/statistics/boj/other/discount/index.htm/> (Accessed on 10 June 2018)

Bank of Mexico (2018), Representative interest rates. <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?accion=consultarCuadroAnalitico&idCuadro=CA51§orDescripcion=Precios&locale=en> (Accessed on 14 June 2018)

Deutsche Bundesbank (2018), Discount rate, Lombard rate and base rate. https://www.bundesbank.de/Navigation/EN/Statistics/Money_and_capital_markets/Interest_rates_and_yields/Tables/table.html (Accessed on 10 June 2018)

FRED Economic Data (2017), Interest Rates, Discount Rate for United States. <https://fred.stlouisfed.org/series/INTDSRUSM193N> (Accessed on 10 June 2018)

Reserve Bank of Australia (2018), Cash rate. <https://www.rba.gov.au/statistics/cash-rate/> (Accessed on 10 June 2018)

2) *BCI*

OECD (2018), Business confidence index (BCI) (indicator). doi: 10.1787/3092dc4f-en (Accessed on 05 June 2018)

3) *CPI*

OECD (2018), Inflation (CPI) (indicator). doi: 10.1787/eee82e6e-en (Accessed on 04 June 2018) (Accessed on 05 June 2018)

- 4) *Fdebt*
OECD (2018), Financial corporations debt to equity ratio (indicator). doi: 10.1787/a3108a99-en (Accessed on 08 June 2018)
- 5) *Fprof*
OECD (2018), Value-added in financial corporations (indicator). doi: 10.1787/f891bfeb-en (Accessed on 09 June 2018)
- 6) *Gdebt*
OECD (2018), General government debt (indicator). doi: 10.1787/a0528cc2-en (Accessed on 08 June 2018)
- 7) *GFCF*
OECD (2018), Investment (GFCF) (indicator). doi: 10.1787/b6793677-en (Accessed on 05 June 2018)
- 8) *Gov*
OECD (2018), General government spending (indicator). doi: 10.1787/a31cbf4d-en (Accessed on 08 June 2018)
- 9) *GrowGDP*
OECD (2018), Quarterly GDP (indicator). doi: 10.1787/b86d1fc8-en (Accessed on 05 June 2018)
- 10) *HHC*
OECD (2018), Household spending (indicator). doi: 10.1787/b5f46047-en (Accessed on 06 June 2018)
- 11) *HHdebt*
OECD (2018), Household debt (indicator). doi: 10.1787/f03b6469-en (Accessed on 08 June 2018)
- 12) *LIR*
OECD (2018), Long-term interest rates (indicator). doi: 10.1787/662d712c-en (Accessed on 08 June 2018)
- 13) *M1*
OECD (2018), Narrow money (M1) (indicator). doi: 10.1787/7a23d68b-en (Accessed on 06 June 2018)
- 14) *Mports*
OECD (2018), Trade in goods and services (indicator). doi: 10.1787/0fe445d9-en (Accessed on 06 June 2018)
- 15) *NFdebt*
OECD (2018), Non-Financial corporations debt to surplus ratio (indicator). doi: 10.1787/dc95ffa7-en (Accessed on 08 June 2018)

- 16) *NFprof*
OECD (2018), Value-added in non-financial corporations (indicator). doi: 10.1787/731f0874-en (Accessed on 09 June 2018)
- 17) *Pop*
OECD (2018), Population (indicator). doi: 10.1787/d434f82b-en (Accessed on 08 June 2018)
- 18) *PPI*
OECD (2018), Producer price indices (PPI) (indicator). doi: 10.1787/a24f6fa9-en (Accessed on 05 June 2018)
- 19) *PPP*
OECD (2018), Purchasing power parities (PPP) (indicator). doi: 10.1787/1290ee5a-en (Accessed on 08 June 2018)
- 20) *RnD*
OECD (2018), Gross domestic spending on R&D (indicator). doi: 10.1787/d8b068b4-en (Accessed on 08 June 2018)
- 21) *Sav*
OECD (2018), Saving rate (indicator). doi: 10.1787/ff2e64d4-en (Accessed on 06 June 2018)
- 22) *SIR*
OECD (2018), Short-term interest rates (indicator). doi: 10.1787/2cc37d77-en (Accessed on 08 June 2018)
- 23) *SPI*
OECD (2018), Share prices (indicator). doi: 10.1787/6ad82f42-en (Accessed on 05 June 2018)
- 24) *Tax*
OECD (2018), Tax revenue (indicator). doi: 10.1787/d98b8cf5-en (Accessed on 08 June 2018)
- 25) *Unemp*
OECD (2018), Unemployment rate (indicator). doi: 10.1787/997c8750-en (Accessed on 06 June 2018)
- 26) *Wage*
OECD (2018), Average wages (indicator). doi: 10.1787/cc3e1387-en (Accessed on 07 June 2018)
- 27) *Xports*
OECD (2018), Trade in goods and services (indicator). doi: 10.1787/0fe445d9-en (Accessed on 06 June 2018)

A.3 Pre-processing steps

To facilitate reproducibility of this paper's results, in this section we present all steps followed to build the final dataset for each country from the raw data downloaded. The data in OECD's database are grouped per variable. This means that each file downloaded referred to one variable⁴⁶ and contained data about *all* countries available. Therefore, the first step was to create six .CSV files – one per country – for each main variable (except for *BankRate*). This simple step was executed in a spreadsheet. Regarding the *BankRate* variable, which is the only one not available in OECD's database, a different procedure was followed. Data about *BankRate* are officially provided by central banks. Each country's central bank makes an announcement when a change in its bank rate is to take place, but these announcements have no specific frequency. Moreover, there are differences in file format⁴⁷ and/or data structure among the data published by each central bank. This means that, while for the rest of the variables it was possible to import values into an initial data frame in an automated way (as indeed happened), for the *BankRate* this was not possible; or, at least, it was not worth the effort to build such a complicated procedure just for one variable. Consequently, the data concerning *BankRate* were entered into the initial data frames manually⁴⁸.

After creating the .CSV files, everything was ready for the construction of the initial data frames, i.e., this very first version of the six datasets before the main pre-processing steps. At this point, there was a folder for each country, which contained as many .CSV files as the number of that country's main variables. Initial data frames were built using the script⁴⁹ Building Datasets.R, in RStudio. This script was written in order to transfer all data from the .CSV files into the six data frames quickly and accurately. Data from variables of quarterly frequency were just copied one-by-one, because they already had the frequency desired. Data from yearly variables were entered into the first quarter (Q1) of each year, without filling the empty cells yet. Variable *BCI* was the only one of monthly frequency. For this variable, a three-month average had already been computed in the spreadsheet for every available quarter. These averages were entered into the corresponding quarters of the initial data frames using the Building Datasets.R script, as well. Finally, *BankRate* was the only variable that was entered into the initial data frames manually, as already mentioned.

46. An exception was the "Trade in goods and services" indicator, which contained data about both imports and exports – two different variables in our datasets.

47. For example, Deutsche Bundesbank publishes these data in .PDF format, while the other five central banks provide them in either .XLS or .CSV format.

48. As these data were not provided in quarterly format, a weighted average of the corresponding bank rates was calculated for each quarter. For example, if a bank rate changed from a to b in the middle of a quarter, a simple average of these two bank rates was entered into the initial data frame for that quarter. In fact, the weights varied a lot. As this procedure was accomplished in a spreadsheet by hand, it is expected that there are some small deviations from the actual weighted quarterly averages.

49. All the script files mentioned, along with the data downloaded, are available to anyone interested by request via email.

At that point we had six datasets with raw data and many missing values. The goal of the next steps was to fully prepare the datasets for model fitting. These are the main pre-processing steps. For this purpose, several functions were written in R by the author, in order to make these pre-processing steps a fully automated procedure. These functions can be found in the `Preprocessing.R` script, which is the same file for all countries. The main function of this script is the `preprocess(dt)`, where `dt` is an argument that represents a data frame. The input of this function is a dataset which contains only raw data, as described above, and the output is the final version of that dataset. The `preprocess()` function is based on other functions that perform the main pre-processing steps, i.e., its only purpose is to coordinate the entire process. These functions are the following:

- a) `q1.q4(dt)`: This function selects the variables of yearly frequency and transfers their observations from Q1 to Q4. This is a preparatory step for the transformation of yearly variables into quarterly ones. The rationale of this step lies behind the fact that an observation of yearly frequency contains information about all four quarters. In other words, information about Q4 cannot, in fact, be known in Q1–Q3 and, therefore, the best probable choice is to transfer the yearly observation into Q4. This choice is still a convention that stems from our lack of further information.
- b) `def.recess(dt)`: This function captures the pre-recessionary periods. In order to find them, it checks only the values of *GrowGDP*. The definition used is that a recession is *a period of two consecutive quarters of decline in real GDP*. Each time the function finds such a period, it labels the first recessionary quarter and the previous three with number “1”. In contrast to the common practice of labeling the recessionary periods with number “1”, in this paper we label the pre-recessionary periods as such, because what concerns us is the conditions *before* a recession. The only reason we mark the first recessionary quarter as well is because a recession might, in fact, have begun some time during the quarter, and, therefore, part of that quarter is also a pre-recessionary period. The remaining periods are marked with number “0” and a new binary variable is imported into the dataset, namely, the *PreRecess*. It now becomes clear how a classification method could capture the conditions before the beginning of a recession.
- c) `interpolate(dt)`: This is the first function that deals with the issue of missing values. It ignores the missing values before the first real observation (there is another function concerning them) and it completes the remaining missing values by applying linear interpolation, where this is feasible. Essentially, `interpolate()` is applied to variables of yearly frequency⁵⁰. If, for example, a variable has the

50. The function does *not* make a distinction between yearly and quarterly variables, but, in practice, it intervened only in variables of yearly frequency due to the way these were entered into the dataset.

values 1990.Q4: 100, 1991.Q1: NA, 1991.Q2: NA, 1991.Q3: NA and 1991.Q4: 500, the function fills the three missing values with the numbers 200, 300 and 400, respectively. The main problem of all yearly variables is the fact that their missing values appear in a systematic way: after every observation, three missing values necessarily follow. For these variables, linear interpolation is a simple and a rather reliable solution to this problem. The only implicit assumption is that the dots (i.e., the actual observations) are connected by a straight line; in the face of uncertainty, there is no reason to assume something more complicated.

- d) `new.vars(dt)`: This function creates two new variables in the dataset. One is *Spread*, which refers to the interest rate spread, and the other is *ElastRC*, which refers to the elasticity of profit rate w.r.t. capital ($e_{r,c}$). The transformations were the following: $Spread_t = LIR_t - SIR_t$ and $ElastRC_t = \frac{(Fprof_t + NFprof_t) - (Fprof_{t-1} + NFprof_{t-1})}{GFCF_t * (Fprof_{t-1} + NFprof_{t-1})}$, where t indicates time (quarter of year). With *ElastRC* we want to approximate quantity $\frac{dr}{dc} \frac{c}{r}$ from the analysis in Tsoulfidis (2010, pp. 119-120). By definition, $GFCF = \frac{dc}{c}$, it is the growth rate of capital formation. For the calculation of $\frac{dr}{r}$, we approximate⁵¹ total profit rate with quantity $r \cong Fprof + NFprof$. Therefore, we approximate quantity $\frac{dr}{r}$, by calculating: $\frac{r_t - r_{t-1}}{r_{t-1}} = \frac{\Delta r}{r}$. Putting everything together, we end up with the transformation given above for *ElastRC*.
- e) `imput.NAs(dt)`: This is the function that handles the remaining missing values. With function `interpolate()` we have filled all missing values between the first and the last actual observation of every variable. However, interpolation is not feasible for the missing values before the first actual observation and after the last one. This is the task of `imput.NAs()` function. The following steps are a summary of its main steps on each variable with many missing values⁵²:

51. Variables *Fprof* and *NFprof* are indicators of profitability. However, according to the full definitions provided by OECD (references in Appendix, A.2), they are not constructed as profit-to-capital ratios and this is why the verb 'approximate' is used here.

52. We allow up to sixteen missing values per variable, in the total of the 196 possible observations. If there are more than sixteen, then the function proceeds to missing data imputation. Even without this threshold, `imput.NAs()` cannot complete all missing values in general, because the imputation procedure is based on observations of other variables. In other words, the output of the function is always restricted by the number of other variables' missing values. Therefore, in order to save computational time (i.e., to avoid running the whole procedure just to replace "NA" with "NA"), we allow a relatively small number of missing values to exist in datasets.

- 1) The first step is to find the twenty variables that are the ones most correlated with the selected one. A correlation matrix for all variables of the dataset has already been computed. At this point it is important to mention that, where it was feasible, lagged variables and variables of first-order differences or percentage changes have been calculated for every main variable before running the `imput.NAs()` function. This means that, at the time `imput.NAs()` runs, the total number of variables is by far larger than 100 for all countries. This is the reason why only twenty variables were selected. The functions that created those extra variables are `grows.diffs(dt)` and `L1.L2(dt)`, which are presented at a later point in this discussion.
- 2) Variables with more than sixteen missing values are excluded from the set of twenty. The reason becomes apparent in the next step.
- 3) A linear model is fitted, using as a response variable the one that needs imputation, and as predictors the remaining variables from the set of twenty. Since these remaining variables are going to be used as predictors in a linear model, they should not have many missing values⁵³; this is the explanation for the previous step. Function `step()` is used to build a parsimonious linear model, using BIC as model selection criterion. Argument `direction` is set to “forward”, so a variable is added only if it reduces the BIC value.
- 4) Using the fitted model, `imput.NAs()` estimates the missing values of the variable selected, where this is possible. In order to reduce a potential high variance of these estimates, a smoother is applied on them before completing the missing values in the dataset. Again, in the face of uncertainty, the preference here is to capture the (more certain) trend-like movements rather than the (uncertain) data noise. Additionally, sometimes it may be the case that this procedure produces negative values for a variable that takes only non-negative values; e.g. variable *Wage*. Function `imput.NAs()` addresses this issue by properly “squeezing” all estimates into the positive range. To be more specific, when such a violation is verified, the minimum estimate (which is negative) takes the value of 0.001 (this is a convention) and all others are proportionally moved towards the first actual observation of the variable⁵⁴. Thus, estimates close to the first *actual* observation are subject to minor changes, while further estimates do change more. This seems a good approach for coping with the problem of wrongly negative estimates, because, on the one hand, the correlation between the response variable and its predictors is maintained (as the estimates’ transformation is linear)

53. The number of missing values that `imput.NAs()` will not finally complete, equals that of the predictor with the most missing values.

54. In mathematical terms, the transformation is the following: $v'_i = v_i + \alpha \frac{\beta - v_i}{\gamma} + 0.001$, where v_i is the i -th estimate, $\alpha = \left| \min_{i:\{1,2,\dots\}} v_i \right|$, β is the first actual observation of the selected variable, and $\gamma = \alpha + \beta$.

and, on the other hand, the non-negative nature of this variable⁵⁵ is respected. An evaluation of the `imput.NAs()` method through an example can be found in Appendix, A.4.

- f) `grows.diffs(dt)`: This function constructs the variables of percentage changes and the variables of first-order differences. Including such variables in our analysis allows us to examine the main problem of this paper from a dynamical perspective as well. Each variable, which is *not* a percentage or a ratio, has a counterpart variable with the prefix *Grow* before its main name. This indicates a variable that refers to percentage changes. For example, *BCI* is an index, i.e., not a percentage or ratio. Thus, a variable *GrowBCI* is also included in the six datasets, and is equal to $\frac{BCI_t - BCI_{t-1}}{BCI_{t-1}} \times 100$. The reason for using percentage changes in these cases, and not merely differences, is because we want to provide models with variables that have comparable values over time. For example, an increase in average income by €1,000 today does not create the same buying power, as it could have done 50 years ago, due to inflation. The use of percentages for describing such changes makes comparison much more meaningful. For variables that are already percentages or ratios, their first-order differences are calculated in their counterpart variables with the prefix *Dif*. For example, variable *Tax* is expressed as a percentage of GDP. Therefore, in this case the variable $DifTax = Tax_t - Tax_{t-1}$ was constructed.
- g) `L1.L2(dt)`: This function constructs variables of first and second lag order for every variable of the dataset (i.e., also for *Grow*-s and *Dif*-s). Exceptions are the variables *PreRecess*, *BCI*, *CPI*, *M1*, *Pop*, *PPI*, *SPI* and *Wage*. The first one is a binary variable. The reason for excluding the rest is that these variables are of no interest at their levels, because their values are not comparable over time. Instead, their *Grow*-s and *Dif*-s variables are entered on to the function `L1.L2()` for constructing the corresponding lagged variables.
- h) `remain.Lags(dt)`: This function recalculates all lagged variables. After imputing the missing data, it is necessary to update the lagged variables with the new values.

55. An alternative approach applied in order to cope with the problem of wrongly negative estimates was to use `glm()` instead of `lm()` for the linear model, with argument `family` set to `Gamma`. This approach succeeded in giving only positive estimates. However, regardless of the variable selected, almost all estimates produced were slightly above zero. Thus, this approach was not adopted, since it failed to produce realistic estimates.

To put everything together, the function `preprocess()` constructs a dataset as follows:

Algorithm A.1 The main pre-processing steps

INPUT: *dt* // A data frame with raw data.

- 1: *dt* \leftarrow `q1.q4(dt)`
- 2: *dt* \leftarrow `def.recess(dt)`
- 3: *dt* \leftarrow `interpolate(dt)`
- 4: *dt* \leftarrow `grows.diffs(dt)`
- 5: *dt* \leftarrow `L1.L2(dt)`
- 6: *dt* \leftarrow `imput.NAs(dt)`
- 7: *dt* \leftarrow `new.vars(dt)`
- 8: *dt* \leftarrow `grows.diffs(dt)`
- 9: *dt* \leftarrow `remain.Lags(dt)`

OUTPUT: *dt*

A.4 Evaluating the `imput.NAs()` method

There are many R packages the objective of which is reliable missing data imputation. Before creating the function `imput.NAs()`, such a package was tested on the datasets of this paper in order to decide whether it was worth the effort to create a new procedure for this purpose or to let an existent method complete the remaining missing values. This package was the `missForest`. At first glance, it seemed that function `missForest()` could not capture time trends in any variable. Thus, the author's decision was to create function `imput.NAs()` in order to produce more realistic estimates of missing values. In order to justify this choice, a comparison was made between the two functions in terms of Mean Squared Error (MSE) on known data⁵⁶. The two functions were tested on the initial dataset of the USA. Eight variables with fewer than sixteen missing values were randomly selected, and a range of observations was defined, within which no value was missing from the dataset⁵⁷. Then, 50 missing values were artificially created on these eight variables, in order to simulate a real situation, when some variables present a lot of missing values. The goal was to see how well each function 'predicted' the 50 missing values.

56. The relevant code is provided in the script `Miscellaneous.R`.

57. This range was 15th observation – 184th observation.

	missForest() MSE	imput.NAs() MSE	Difference
HHC	2.269	0.192	2.077
Mports	1.830	0.385	1.445
Tax	0.662	0.649	0.013
M1	200.651	24.221	176.430
PPI	188.876	10.189	178.668
NFprof	0.084	1.959	-1.875
SPI	125.146	5.928	119.218
Unemp	1.231	1.062	0.169

Figure A.1 Comparison between missForest() and imput.NAs() in terms of MSE.

In Figure A.1 we see that, in seven out of eight variables, function imput.NAs() performed better. In Figure A.2, we can visually assess the performance of the two functions on the eight variables selected. Black lines represent real data, red lines represent estimates of missForest() and green lines represent estimates of imput.NAs().

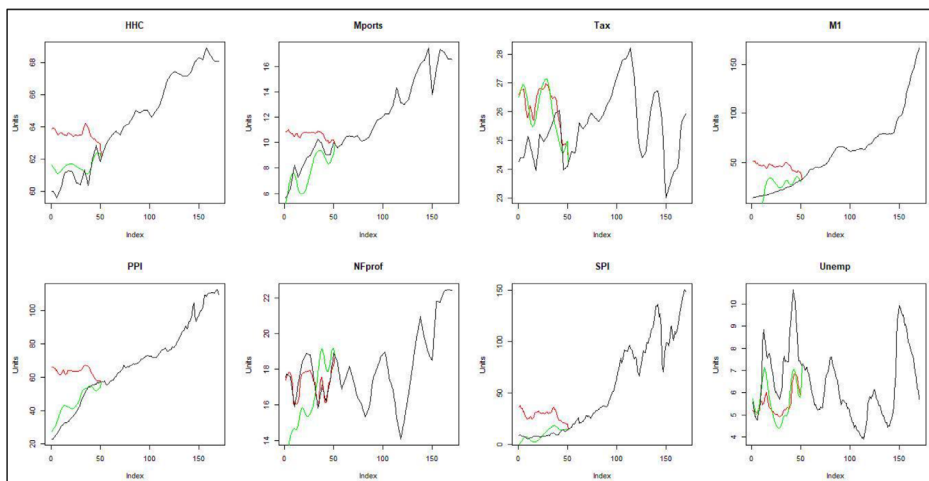


Figure A.2 Optical evaluation of missForest() and imput.NAs() performance (red and green lines, respectively).

Consequently, function imput.NAs() was selected for completing the remaining missing values of the six datasets, since it seems to produce more reliable estimates. This does not imply that imput.NAs() is a generally better imputation method than missForest(); we simply have evidence that it performs better in this specific kind of datasets and, therefore, it was finally used. At this point we must note that the performance of imput.NAs() was dramatically improved when the lagged variables and

the variables of first-order differences or percentage changes were calculated before the missing values imputation. Initially, the plan was to calculate these variables as a last step; however, having calculated all of them before running the `imput.NAs()` function improved the latter's performance to a large extent.

A.5 Searching for optimal parameter setting in SVM (example)

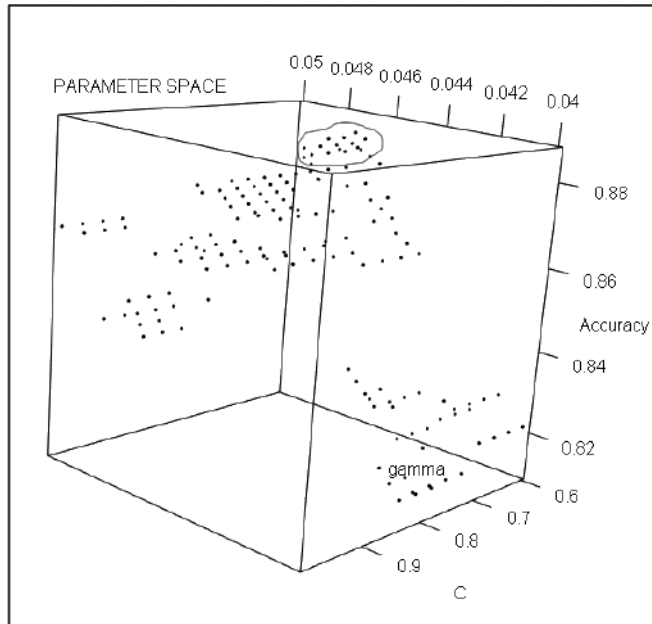


Figure A.3 An example of the three-dimensional plot used for finding the combination of C and γ that gives the best out-of-sample Classification Accuracy.

A.6 The visual output of the Average Trees algorithm (example)

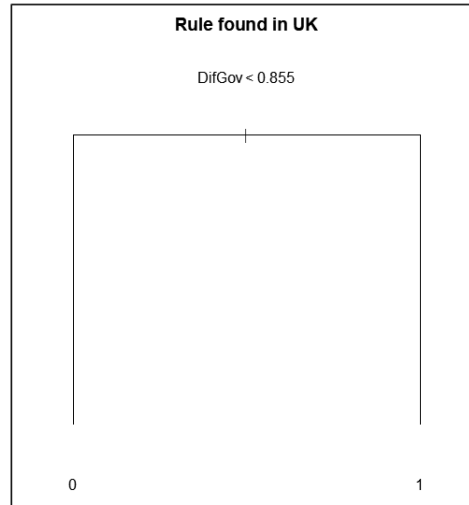


Figure A.4 The result of Average Trees (random sampling) using all 167 observations of the UK dataset (the model parameterisation is presented in Table 8). In this example, a recession is expected within next year provided the current quarter holds that $DifGov \geq 0.855$.

A.7 The result of Average Trees algorithm for the case of the USA

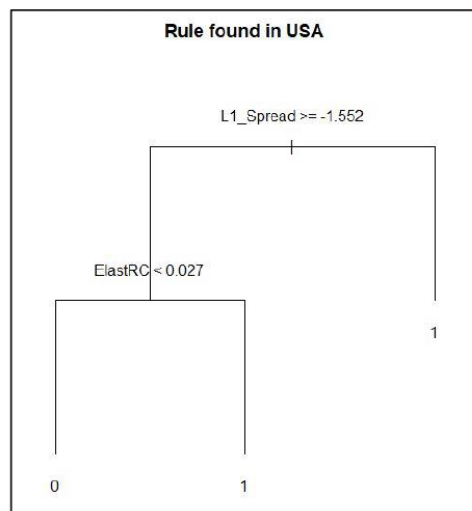


Figure A.5 The output of Average Trees (random sampling) for the dataset of the USA. Regarding the splitting point (inequality), answer “Yes” is on the left-hand side, as usual. The parameterisation of this model is presented in Table 9.

A.8 Evaluation results: In-sample performance**Table A.1** Australia

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	96.51%	77.14%	91.67%	99.33%	0.67%	93.08%	ex.size = 0.12
Average Trees (random)	88.95%	78.57%	50.87%	90.13%	9.87%	64.89%	ex.size = 0.45, tree.nr = 200
Decision Trees	96.51%	77.14%	91.67%	99.33%	0.67%	93.08%	probability threshold: 0.5
Random Forests	100%	100%	100%	100%	0%	100%	probability threshold: 0.6
Logit	91.86%	43.57%	85.42%	98.24%	1.76%	60.3%	probability threshold: 0.75
Probit	92.44%	43.57%	93.75%	98.91%	1.09%	65.3%	probability threshold: 0.75
k-NN	91.86%	68.57%	50%	95.7%	4.3%	60%	k = 6, # of principal components: 2
BRT	94.19%	73.57%	88.33%	97.32%	2.68%	74.91%	shrinkage = 0.018, probability threshold: 0.45
SVM	88.37%	86.43%	45.63%	88.18%	11.82%	64.95%	$\gamma=0.0005$, $C=1.3$, sigmoid kernel, probability threshold: 0.15
ANN	90.7%	83.57%	54.76%	91.51%	8.49%	70%	probability threshold: 0.54, # of hidden nodes: 5

Table A.2 Germany

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	89.88%	53.13%	94.44%	98.41%	1.59%	77.62%	ex.size = 0.13
Average Trees (random)	86.9%	62.5%	80.54%	93.65%	6.35%	71.02%	ex.size=0.225, tree.nr=200
Decision Trees	89.29%	56.25%	94.44%	98.41%	1.59%	80%	probability threshold: 0.8
Random Forests	100%	100%	100%	100%	0%	100%	probability threshold: 0.37
Logit	84.52%	31.25%	100%	100%	0%	88.89%	probability threshold: 0.85
Probit	84.52%	31.25%	100%	100%	0%	88.89%	probability threshold: 0.85
k-NN	97.62%	87.5%	100%	100%	0%	91.67%	k = 3, # of principal components: 3
BRT	92.26%	87.5%	81.11%	92.46%	7.54%	81.65%	shrinkage = 0.01, probability threshold: 0.3
SVM	85.12%	59.38%	56.19%	90.48%	9.52%	60.95%	$\gamma=0.0014$, $C=2.9$, radial kernel, probability threshold: 0.35
ANN	88.1%	84.38%	68.52%	88.29%	11.71%	73.33%	probability threshold: 0.67, # of hidden nodes: 5

Table A.3 Japan

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	93.6%	77.13%	84.32%	92.42%	7.58%	88.61%	ex.size = 0.09
Average Trees (random)	93.6%	77.13%	84.32%	92.42%	7.58%	88.61%	ex.size = 0.1, tree.nr = 200
Decision Trees	89.53%	39.63%	100%	100%	0%	82.22%	probability threshold: 0.8
Random Forests	100%	100%	100%	100%	0%	100%	probability threshold: 0.45
Logit	87.79%	66.46%	56.23%	87.9%	12.1%	59.06%	probability threshold: 0.35
Probit	87.79%	66.46%	56.23%	87.9%	12.1%	59.06%	probability threshold: 0.35
k-NN	98.84%	95.43%	95.43%	99.21%	0.79%	95.43%	k = 2, # of principal components: 3
BRT	97.67%	95.43%	93.02%	96.3%	3.7%	93.52%	shrinkage = 0.035, probability threshold: 0.3
SVM	87.21%	95.43%	48.56%	81.62%	18.38%	70.73%	$\gamma=0.0049$, $C=1.2$, radial kernel, probability threshold: 0.3
ANN	69.77%	63.41%	31.34%	63.42%	36.58%	55.91%	probability threshold: 0.5, # of hidden nodes: 6

Table A.4 Mexico

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	92.4%	62.5%	73.33%	97.45%	2.55%	83.52%	ex.size = 0.12
Average Trees (random)	92.4%	72.5%	66.67%	96.07%	3.93%	79%	ex.size = 0.1, tree.nr = 200
Decision Trees	92.4%	62.5%	73.33%	97.45%	2.55%	83.52%	probability threshold: 0.5
Random Forests	100%	100%	100%	100%	0%	100%	probability threshold: 0.6
Logit	86.55%	15%	60%	98.41%	1.59%	66.67%	probability threshold: 0.5
Probit	87.72%	25%	86.67%	98.41%	1.59%	48.89%	probability threshold: 0.4
k-NN	100%	100%	100%	100%	0%	100%	k = 1, # of principal components: 4
BRT	93.57%	70%	92%	98.41%	1.59%	74.03%	shrinkage = 0.01, probability threshold: 0.3
SVM	49.71%	42.5%	14.38%	48.62%	51.38%	32.78%	$\gamma=0.006$, $C=0.08$, radial kernel, probability threshold: 0.16
ANN	85.96%	47.5%	45.24%	92.45%	7.55%	47.87%	probability threshold: 0.55, # of hidden nodes: 3

Table A.5 United Kingdom

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	89.82%	62.5%	61.11%	93.93%	6.07%	70%	ex.size = 0.25
Average Trees (random)	89.82%	62.5%	61.11%	93.93%	6.07%	70%	ex.size = 0.15, tree.nr = 200
Decision Trees	90.42%	78.13%	51.76%	90.94%	9.06%	68.47%	probability threshold: 0.5
Random Forests	97.6%	86.46%	100%	100%	0%	91.96%	probability threshold: 0.7
Logit	88.02%	20.83%	62.5%	97.13%	2.87%	71.43%	probability threshold: 0.5
Probit	86.83%	12.5%	50%	97.13%	2.87%	50%	probability threshold: 0.5
k-NN	89.82%	50%	48.89%	95.73%	4.27%	84.44%	k = 10, # of principal components: 2
BRT	91.02%	59.38%	75.42%	94.33%	5.67%	59.17%	shrinkage = 0.01 probability threshold: 0.3
SVM	86.83%	0%	0%	100%	0%	0%	$\gamma=0.0003$, $C=0.009$, radial kernel, probability threshold: 0.95
ANN	86.83%	15.63%	50%	95.93%	4.07%	35%	probability threshold: 0.51, # of hidden nodes: 4

Table A.6 United States of America

Method	Classification Accuracy	Sensitivity	Precision	Specificity	False Alarm	F ₁ -Score	Comments
Average Trees (K-fold)	94.64%	80%	76.67%	96.03%	3.97%	75.81%	ex.size = 0.1
Average Trees (random)	94.64%	80%	76.67%	96.03%	3.97%	75.81%	ex.size = 0.1, tree.nr = 200
Decision Trees	95.24%	80%	79.33%	96.83%	3.17%	77.59%	probability threshold: 0.5
Random Forests	100%	100%	100%	100%	0%	100%	probability threshold: 0.4
Logit	92.26%	40%	93.33%	99.21%	0.79%	71.53%	probability threshold: 0.8
Probit	90.48%	25%	83.33%	99.21%	0.79%	71.43%	probability threshold: 0.92
k-NN	92.26%	55%	77.78%	96.83%	3.17%	81.9%	k = 10, # of principal components: 2
BRT	92.26%	40%	90%	99.21%	0.79%	94.44%	shrinkage = 0.009, probability threshold: 0.5
SVM	91.07%	65%	71.79%	93.65%	6.35%	67.07%	$\gamma=0.00025$, $C=0.92$, radial kernel, probability threshold: 0.25
ANN	89.29%	65%	53.73%	91.27%	8.73%	62.93%	probability threshold: 0.52, # of hidden nodes: 5

References

- Aggarwal, C. C. (Ed.). (2015). *Data Classification - Algorithms and Applications*. Boca Raton, FL (USA): CRC Press, Taylor & Francis Group.
- Baltagi, B. H. (2002). *Econometrics* (3rd ed.). Berlin; Heidelberg (Germany): Springer-Verlag. doi:10.1007/978-3-662-04693-7
- Barnett, W., Serletis, A., & Serletis, D. (2015). Nonlinear and Complex Dynamics in Economics. *Macroeconomic Dynamics*, 19(8), pp. 1749-1779. doi:10.1017/S1365100514000091
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY (USA): Springer Science and Business Media.
- Chauvet, M., & Potter, S. (2005, Mar.). Forecasting recessions using the yield curve. *Journal of Forecasting*, 24(2), pp. 77-103. doi:10.1002/for.932
- Christiansen, C. (2013). Predicting severe simultaneous recessions using yield spreads as leading indicators. *Journal of International Money and Finance*, 32, pp. 1032-1043. doi:10.1016/j.jimonfin.2012.08.005
- Claessens, S., & Kose, A. M. (2009, Mar. IMF Publications.). What Is a Recession? *Finance & Development*, 46(1), pp. 52-53. Retrieved from <http://www.imf.org/external/pubs/ft/fandd/2009/03/pdf/basics.pdf>
- Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting*, 33(4), pp. 745-759. doi:10.1016/j.ijforecast.2017.02.003
- Dovern, J., & Huber, F. (2015). Global prediction of recessions. *Economics Letters*, 133, pp. 81-84. doi:10.1016/j.econlet.2015.05.022
- ECRI. (2013). Business cycle peak and trough dates, 1948–2011. Retrieved (by the authors) Sep. 19, 2013, from <https://www.businesscycle.com/ecri-business-cycles/international-business-cycle-dates-chronologies>
- Estrella, A., & Mishkin, F. S. (1998, Feb.). Predicting U.S. Recessions: Financial Variables as Leading Indicators. *The Review of Economics and Statistics*, 80(1), pp. 45-61. doi:10.1162/003465398557320
- Fisher, I. (1933, Oct.). The Debt-Deflation Theory of Great Depressions. *Econometrica*, 1(4), pp. 337-357. doi:10.2307/1907327
- Friedman, M. (1968, Mar.). The Role of Monetary Policy. *The American Economic Review*, 58(1), pp. 1-17. Retrieved Apr. 2, 2018, from <http://www.jstor.org/stable/1831652>
- Friedman, M., & Schwartz, A. J. (1963). *A Monetary History of the United States, 1867-1960*. Princeton, NJ (USA): Princeton University Press.
- Gogas, P., Papadimitriou, T., Matthaiou, M., & Chrysanthidou, E. (2015, Apr.). Yield Curve and Recession Forecasting in a Machine Learning Framework. *Computational Economics*, 45(4), pp. 635-645. doi:10.1007/s10614-014-9432-0
- Goodwin, R. M. (1951, Jan.). The Nonlinear Accelerator and the Persistence of Business Cycles. *Econometrica*, 19(1), pp. 1-17. doi:10.2307/1907905
- Hall, R. E., & Lieberman, M. (2013). *Macroeconomics: Principles & Applications* (6th ed.). Mason, OH (USA): South-Western, Cengage Learning.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning - with Applications in R*. New York, NY (USA): Springer Science and Business Media. doi:10.1007/978-1-4614-7138-7
- Kauppi, H., & Saikkonen, P. (2008, Nov.). Predicting U.S. Recessions with Dynamic Binary Response Models. *The Review of Economics and Statistics*, 90(4), pp. 777-791. doi:10.1162/rest.90.4.777
- Keynes, J. M. (1936 [2013]). *The Collected Writings of John Maynard Keynes: The General Theory of Employment, Interest and Money* (4th ed., Vol. VII). New York, NY (USA): Cambridge University Press.

- Kiani, K. (2008). On Forecasting Recessions via Neural Nets. *Economics Bulletin*, 3(13), pp. 1-15. Retrieved May 27, 2018, from <https://pdfs.semanticscholar.org/af11/30f6a52e1e41057d001a365f78968a2bbfe7.pdf>
- Knoop, T. A. (2015). *Business Cycle Economics: Understanding Recessions and Depressions from Boom to Bust*. Santa Barbara, CA (USA): ABC-CLIO, LLC.
- Kubat, M. (2017). *An Introduction to Machine Learning* (2nd ed.). Cham, ZG (Switzerland): Springer International Publishing AG. doi:10.1007/978-3-319-63913-0
- Marx, K. (1894 [2010]). *Capital: A Critique of Political Economy* (Vol. III). (F. Engels, Ed.) New York, NY (USA): International Publishers.
- NBER. (2010, Sep. 20). *US Business Cycle Expansions and Contractions*. Retrieved May 25, 2018, from <http://www.nber.org/cycles/>
- OECD. (2018a). "Business confidence index (BCI)" (indicator). Retrieved Mar. 27, 2018, from <http://dx.doi.org/10.1787/3092dc4f-en>
- OECD. (2018b). "Long-term interest rates" (indicator). Retrieved Apr. 6, 2018, from <http://dx.doi.org/10.1787/662d712c-en>
- Plakandaras, V., Cunado, J., Gupta, R., & Wohar, M. E. (2017). Do leading indicators forecast U.S. recessions? A nonlinear re-evaluation using historical data. *International Finance*, 20(3), pp. 289-316. doi:10.1111/infi.12111
- Ricardo, D. (1821 [2001]). *On the Principles of Political Economy and Taxation* (3rd ed.). (reprint), Kitchener, Ontario (Canada): Batoche Books.
- Schumpeter, J. A. (1942 [1994]). *Capitalism, Socialism and Democracy* (5th ed.). London (UK): Routledge.
- Smith, A. (1776 [1977]). *An Inquiry into the Nature and Causes of the Wealth of Nations* (5th ed.). (E. Cannan, Ed.) (reprint), Chicago, IL (USA): University Of Chicago Press.
- Tsoufidis, L. (2010). *Competing Schools of Economic Thought*. Berlin; Heidelberg (Germany): Springer-Verlag. doi:10.1007/978-3-540-92693-1
- Wong, A. Y.-T., Pak, T., & Fong, W. (2011, Dec.). Analysing interconnectivity among economies. *Emerging Markets Review*, 12(4), pp. 432-442. doi:10.1016/j.ememar.2011.06.004

RE-EXAMINING THE STABILITY OF MONEY MULTIPLIER FOR THE US: THE NONLINEAR ARDL MODEL

ISMET GOCER^{a*}
SERDAR ONGAN^b

^aAydin Adnan Menderes University, Turkey

^bSt. Mary's College of Maryland, St. Mary's City, MD, USA

Abstract

The rising uncertainties in the economy and the entwined global financial markets can easily cause nonlinear (asymmetric) behaviors among economic actors. Accordingly, this study re-considers the stability of the money multiplier from a different methodological perspective from that of prior studies, which assumed a linear relationship between money supply and monetary base. To this aim, the nonlinear ARDL model is applied for the US for the 2000M1-2018M9 period. Empirical findings of the nonlinear model indicate that only increases in positive monetary base shocks have a proportional relation with money supply. Additionally, the nonlinear ARDL detects proportional relationships between money supply and monetary base lower degree than the linear model.

JEL Classification: E510, E520

Keywords: Money Multiplier, Asymmetry, Linear and Nonlinear ARDL

**Corresponding Author:* Ismet Gocer, Aydin Adnan Menderes University, Department of Econometrics, Turkey. E-mail: ismetgocer@gmail.com

Introduction

Controllable money supply may help central banks achieve their monetary policy objectives effectively and as desired. On the other hand, controllable money supply may also require a stable money multiplier and a controllable monetary base. The long-run proportional relationship between money supply and monetary base implies and requires a stable money multiplier. There are two main approaches to the determination of money supply: the Money Multiplier Approach and the Portfolio Approach. The two approaches differ in their assumption of whether the money multiplier is stable or not, and the monetary base is controllable. However, there is no academic consensus on the matter. According to the Money Multiplier Approach (Friedman and Schwartz, 1963; Brunner and Meltzer, 1964), variations (changes) in the money multiplier, caused by the amount of currency in circulation, time-demand deposits and bank reserves, may dominate money supply in the short-run and become stable and predictable in the long-run (Brunner, 1961). According to the Portfolio Approach, the components of the money multiplier are determined by the portfolio choices of economic agents via cash demand, time deposits and excess reserves. These portfolio choices are sensitive to changes in relative return rates, risk levels, financial innovations, and the structure of financial markets. Thus, there is little reason to assume the validity of a stable money multiplier, since all these market forces may cause a money multiplier to be unstable (Goodhart, 1989).

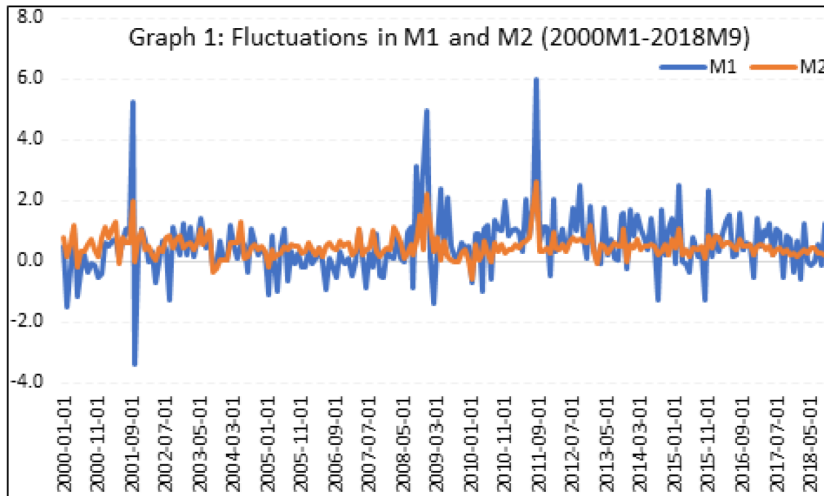
There are some empirical studies supporting the Money Multiplier Approach. For instance, Darbha (2002) applied the residual-based cointegration test by Gregory-Hansen (1996) for India and found a stable, but time-varying, long-run relation between money supply and monetary base. Rusuhuzwa (2015) used the Engle-Granger (1987) cointegration test and the Gregory-Hansen (1996) method for Rwanda and found a stable money multiplier. Tule and Ajilore (2016) applied the cointegration test for Nigeria and found a stable money multiplier. Lastly, Bhatti and Khawaja (2018) applied the cointegration tests of Engle-Granger (1987), Phillips-Ouliaris (1990) and Johansen and Juselius (1990) for Kazakhstan and found that the money multiplier was stable for the country.

However, the growing body of empirical studies is in favor of the latter approach, supporting an unstable money multiplier and uncontrollable monetary base. For instance, Nachnae (1992) for India, Ford and Morris (1996) for the UK, Baghestani and Mott (1997) for the US, and Sen and Vaidya (1997) for India applied the cointegration test and found that the money multiplier was not stable and the monetary base was uncontrollable for their respective countries. Furthermore, Uchendu (1995) applied the moving average regression technique for Nigeria and found an unstable money multiplier. Sahinbeyoglu (1995) used the unit root test for Turkey and found that the money multiplier was unstable. Moosa and Bhatti (1997) applied a battery of econometric tests for Kuwait and found an unstable money multiplier.

Howells and Hussein (1998) applied the causality tests based on cointegration and the Error Correction representation and found an unstable money multiplier for all G7 countries. Khan and Khan (2007) applied the cointegration test for Pakistan and found that the money multiplier was not stable. White (2006) applied the residual based cointegration test for Jamaica and found an unstable money multiplier. Downes *et al.* (2006) applied the descriptive statistics and unit root tests approach for six African countries and found that the money multiplier was not stable. Panagopoulos and Spiliotis (2008) used the Error Correction Vector Autoregressive (VAR) causality for the G7 countries and found an unstable money multiplier for most of the countries except France and Japan. Badarudin *et al.* (2013) used the Trivariate VAR and Granger causality for the G7 countries and found that the money multiplier was unstable for all countries except the UK and the US for two short time periods. Similarly, Odior (2013) applied the Generalized Method of Moments (GMM) model for Nigeria and found that the money multiplier was unstable. The reasons for this inconsistency among monetary economists and their empirical studies may lie in the different time horizons, economic sizes, and structures of financial markets of sample countries, as well as the different methodologies applied. However, it is also important to note that all these studies assume an expected long-run proportional relationship (change) on the money supply is determined by a linearly distributed monetary base series. Yet, this distribution may potentially be nonlinear (asymmetric) because rising uncertainties in the economy can easily cause nonlinear behaviors among economic actors. This means that increases and decreases in the monetary base may have different impacts on money supply. While increases (positive variations) in monetary base may have proportional impacts on money supply, decreases (negative variations) may not. Additionally, even if increases and decreases have the same effect, they may have different scale impacts on money supply. Accordingly, the stability of the money multiplier can also be determined via these increases and decreases, separately, in a nonlinear context. The nonlinear autoregressive distributed lag (ARDL) model by Shin *et al.*, (2014) allows for this decomposition in the monetary base series.

Therefore, this study, differently from previous studies, aims to re-consider and test the stability of the money multiplier from this new methodological perspective for the US. Meanwhile, before starting methodological analyses, it would be better to look at the fluctuations in M1 and M2 for the US, which may correspond to some economic-financial crises and, thereby, change the stability of the money multiplier.

Graph 1 shows that money aggregates M1 and M2 continuously fluctuate. However, sharp fluctuations correspond to some structural break dates, such as the 2001 recession, the 2008 financial crisis and FED's 2012 quantitative easing (QE) policy. All may cause potential nonlinear behaviour of US money multipliers.



Source: Graph 1 was created from the data of the Federal Reserve Bank of St. Louis (FED)

The rest of this study is structured as follows: Section 2 explains the empirical methodology and data set, Section 3 provides empirical results, and Section 4 concludes the study.

2. Empirical Methodology and Data Set

The *money multiplier model (money multiplier relationship)* by Brunner (1961) and Brunner and Meltzer (1964) is the most frequently used model in empirical studies and it is usually written in the following proportional form:

$$M^S = k(.) * H \quad (1)$$

According to this model, money supply can be determined in two ways. First, a change in monetary base (H) proportionately changes money supply (M^S), if the money multiplier (k) is stable. This means that the money supply is exogenously determined by the central bank, in this case the U.S. Federal Reserve (FED). Second, both k and H change the money supply (M^S), if k is not stable, and it is a function of several endogenous factors, mentioned above in parenthesis. This means that the money supply process is endogenous. Eqn. 1 can be written in the following logarithmic form:

$$\text{Log}M^S = \text{Log}k + \text{Log}H \quad (2)$$

Under the assumption of a stable k (denotes $k = M_t^S/H_t=1$), Eqn. 3 is obtained in the following regression form:

$$\text{Log}M_t^S = \beta_0 + \beta_1 \text{Log}H_t + e_t \quad (3)$$

where e_t is the error term and β_0 is the logarithm of k . For a controllable or exogenous money supply process, k must be stable (stationary) and M_t^S and H_t must be stationary or cointegrated, if the series are not stationary at the same order of integration (Thenuwara and Morgan, 2017; Bhatti and Khawaja, 2018). This means that β_0 must be zero, implying a logarithmic k , and β_1 must be 1, implying a proportional relationship between M_t^S and H_t . We test this relationship for both M1 and M2, separately. The monthly data of money supply and monetary base were provided from the database of the Federal Reserve Bank of St. Louis (FRED). The sample period of this study was between 2000M1-2018M9, because the sharp fluctuations in M1 and M2, caused by different economic crises and the FED's monetary policy changes, were mostly seen after 2000. Therefore, this monthly period (223 observations) may give us more and clearer information about the stability of the US money multiplier through monetary aggregates.

In order to estimate both short-run and long-run impacts of changes in H_t on the M_t^S we apply bounds testing to cointegration and the error correction model (ECM) within the ARDL model developed by Pesaran *et al.* (2001). The ECM indicates the speed of adjustment toward the long-run equilibrium from the short-run. Thus, we obtain the following ECM in the linear form of the ARDL model in Eqn.4:

$$\Delta \text{Log}M_t^S = \beta_0 + \sum_{j=1}^p \beta_{1j} \Delta \text{Log}M_{t-j}^S + \sum_{j=0}^q \beta_{2j} \Delta \text{Log}H_{t-j} + \beta_3 \text{Log}M_{t-1}^S + \beta_4 \text{Log}H_{t-1} + e_t \quad (4)$$

In Eq.4, Δ is the difference operator. Short-run and long-run impacts of changes in H_t on M_t^S are determined by the scale and significances of β_{2j} and β_4 respectively. For the validity of potential nonlinear relationships, we follow Shin *et al.*, (2014) and decompose H_t as H_t^+ (increases) and H_t^- (decreases). This decomposition of H_t^+ and H_t^- is constructed with the concept of the partial sum process in the following form:

$$H_t^+ = \sum_{j=1}^t \Delta H_j^+ = \sum_{j=1}^t \max(\Delta H_j, 0) \quad (5)$$

$$H_t^- = \sum_{j=1}^t \Delta H_j^- = \sum_{j=1}^t \min(\Delta H_j, 0) \quad (6)$$

Hence, we obtain the following form of the nonlinear ARDL model in Eq.7.

$$\Delta \text{Log} M_t^S = \beta_0 + \sum_{j=1}^p \beta_{1j} \Delta \text{Log} M_{t-j}^S + \sum_{j=0}^q \beta_{2j} \Delta \text{Log} H_{t-j}^+ + \sum_{j=0}^r \beta_{3j} \Delta \text{Log} H_{t-j}^- + \beta_4 \text{Log} M_{t-j}^S + \beta_5 \text{Log} H_{t-j}^+ + \beta_6 \text{Log} H_{t-j}^- + e_t \quad (7)$$

The impacts of long-run increases and decreases in H_t on M_t^S are determined by the signs and significances of normalized long run coefficients, such as β_5/β_4 and β_6/β_4 . Thus, with decomposed variables, we will be able to understand how M_t^S responds to the H_t^+ and H_t^- separately, in terms of stability of the money multiplier. Consequently, this model will also reveal whether changes in H_t^+ and H_t^- have symmetric or asymmetric effects on M_t^S . If the coefficient values of H_t^+ and H_t^- are of different scale or one of them is significant and the other one is not, this will imply asymmetric impacts. However, for a formal decision of asymmetry, we apply the Wald test for both the short-run (W_{SR}) and the long-run (W_{LR}). $\sum_{j=0}^q \beta_{2j} \neq \sum_{j=0}^r \beta_{3j}$ and normalized long run coefficients, such as $(-\beta_5/\beta_4) \neq (-\beta_6/\beta_4)$ will confirm short-run and long-run asymmetries, respectively.

3. Empirical Results

Before running the ARDL models, we must make sure that the series are stationary. To this aim, we apply Augmented Dickey Fuller (1981) (ADF) and Phillips-Perron (1988) (PP) Unit Root Tests. The results of these two tests are reported in Table 1.

Table 1. ADF and PP Unit Root Test Results

	ADF		PP	
	Level	First Difference	Level	First Difference
	Prob.	Prob.	Prob.	Prob.
<i>LogM1</i>	0.74	0.05*	0.67	0.00***
<i>LogM2</i>	0.35	0.00***	0.45	0.00***
<i>LogH</i>	0.77	0.00***	0.97	0.00***
<i>LogH⁺</i>	0.53	0.02**	0.64	0.00***
<i>LogH⁻</i>	0.99	0.30	0.99	0.00***

Note: ***, ** and * denote statistical significances at 1%, 5% and 10% level respectively. Optimal lags were automatically selected by using the Modified Akaike Information Criterion. For the levels and first differences, trend-intercept and intercept models were used, respectively.

The results of the unit root tests indicate that all series are $I(1)$. Hence, M_t^S and H_t must be cointegrated for a long-run stable money multiplier. To test the cointegration relationship between these two variables, we applied bounds testing. Results of the bounds testing for the linear and nonlinear models are reported in Table 2.

Table 2. Test Results of Bounds Testing

Dependent Variable	k	F stat.	Critical Values					
			I0 Bound			I1 Bound		
			10%	5%	1%	10%	5%	1%
<i>Linear</i>								
LogM1	1	8.96***	3.02	3.62	4.94	3.51	4.16	5.58
LogM2	1	10.33***	3.02	3.62	4.94	3.51	4.16	5.58
<i>Nonlinear</i>								
LogM1	2	7.87***	2.63	3.1	4.13	3.35	3.87	5.00
LogM2	2	7.84***	2.63	3.1	4.13	3.35	3.87	5.00

Note: k is number of regressors. ***; denotes cointegration at 1% significance level.

The critical bounds have been tabulated by Pesaran *et al.* (2001). The test results of the bounds testing indicate that the series are cointegrated at the 1% level for both linear and nonlinear models, since the F-statistics of M1 and M2 exceed the critical values. The results of the linear ARDL model for the short-run and the long-run, as well as diagnostic tests are reported in Table 3.

The long-run estimates of the linear model in Table 3 indicate that changes in the monetary base (H_t) have a cointegrated proportional relationship with money supply (for both M1 and M2, since their coefficients are significantly positive. However, the coefficient values of for both M1 and M2 are less than 1 (*one-to-one relation*), implying that the money multiplier is unstable for both of them. On the other hand, M1 responds to changes in the monetary base (H_t) more than M2 (0.70 and 0.39). Short-run estimates indicate that there is no considerable proportional relationship between and in the short-run. Furthermore, the Error Correction Term (ECT) mechanisms for M1 and M2 work, since their coefficients are significantly negative. This means that short-run variations converge (adjust) with long-run values. The speed of M1 adjustment is higher than that of M2. The results of the nonlinear ARDL model, both in the short-run and the long-run, as well as the diagnostic tests, are reported in Table 4.

Table 3. Linear ARDL Model Estimation Results

<i>Variable</i>	<i>Dependent Variable: (M1)</i>		<i>Dependent Variable: (M2)</i>	
	<i>Coef.</i>	<i>Prob.</i>	<i>Coef.</i>	<i>Prob.</i>
Short-Run				
$\Delta \text{Log}M_{t-1}^S$	-0.19***	0.00	0.17**	0.01
$\Delta \text{Log}M_{t-2}^S$	0.03	0.56	0.03	0.60
$\Delta \text{Log}M_{t-3}^S$	0.19***	0.00	0.19***	0.00
$\Delta \text{Log}M_{t-4}^S$	-	-	-0.18***	0.00
$\Delta \text{Log}M_{t-5}^S$	-	-	0.08	0.22
$\Delta \text{Log}H_t$	0.06***	0.00	0.01*	0.09
$\Delta \text{Log}H_{t-1}$	0.03	0.20	0.001	0.88
$\Delta \text{Log}H_{t-2}$	0.01	0.57	0.01*	0.07
$\Delta \text{Log}H_{t-3}$	-0.02	0.32	-0.01	0.11
$\Delta \text{Log}H_{t-4}$	-0.03	0.20	0.01	0.17
$\Delta \text{Log}H_{t-5}$	0.007	0.80	-0.01	0.29
$\Delta \text{Log}H_{t-6}$	0.03	0.17	0.01	0.22
$\Delta \text{Log}H_{t-7}$	-	-	-0.01	0.18
$\Delta \text{Log}H_{t-8}$	-	-	-0.0007	0.94
$\Delta \text{Log}H_{t-9}$	-	-	-0.01*	0.08
ECT_{t-1}	-0.01***	0.00	-0.005***	0.00
Long-Run				
$\text{Log}H_t$	0.70***	0.00	0.39***	0.00
<i>Constant</i>	2.65***	0.00	6.79***	0.00
Diagnostic Tests				
	<i>Test Stat.</i>	<i>Prob.</i>		
R^2	0.99	-	0.99	-
\bar{R}^2	0.99	-	0.99	-
F	35272.10	0.00	104.13	0.00
DW	1.98	-	2.01	-
χ_{SC}^2	0.11	0.94	1.11	0.57
χ_{HET}^2	19.58	0.07	32.88	0.11
χ_{FF}^2	0.68	0.40	0.31	0.57

Note: ***, ** and * denote statistical significances at 1%, 5% and 10% levels, respectively. CUSUM and CUSUM of Squares test graphs are reported in Appendix 1. High R^2 and \bar{R}^2 values show that explanatory powers of the models are high. DW , χ_{SC}^2 and χ_{HET}^2 statistics indicate that there is no autocorrelation or heteroscedasticity problems in the models. χ_{FF}^2 statistics indicates there is no model misspecification error. Models are stable, since CUSUM and CUSUMQ graphs in Appendix 1 are within confidence intervals.

Table 4. Nonlinear ARDL Model Estimation Results

<i>Dependent Variable: (M1)</i>			<i>Dependent Variable: (M2)</i>		
<i>Variable</i>	<i>Coef.</i>	<i>t-stat.</i>	<i>Variable</i>	<i>Coef.</i>	<i>t-stat.</i>
$LogM_{t-1}$	-0.06***	-5.04	$LogM_{t-1}$	-0.021**	-2.35
$LogH_{t-1}^+$	0.02***	5.98	$LogH_{t-1}^+$	0.006***	2.76
$LogH_{t-1}^-$	-0.008	-0.89	$LogH_{t-1}^-$	-0.0002	-0.08
$\Delta LogM_{t-1}$	-0.22***	-3.54	$\Delta LogM_{t-1}$	0.14**	2.34
$\Delta LogM_{t-3}$	0.16**	2.01	$\Delta LogM_{t-5}$	0.11**	1.96
$\Delta LogH_t^+$	0.11***	5.27	$\Delta LogH_{t-6}^+$	0.02**	2.35
$\Delta LogH_{t-4}^+$	-0.08***	-3.39	$\Delta LogH_{t-7}^+$	-0.02**	2.35
$\Delta LogH_{t-10}^+$	-0.05**	-2.18	$\Delta LogH_{t-9}^+$	-0.02***	-2.68
$\Delta LogH_{t-12}^+$	-0.07***	-3.03	$\Delta LogH_{t-11}^+$	-0.02***	-2.88
-	-	-	$\Delta LogH_{t-1}^-$	-0.02**	-2.25
-	-	-	$\Delta LogH_{t-3}^-$	-0.02**	-2.32
$\Delta LogH_{t-11}^-$	0.16***	2.70	$\Delta LogH_{t-4}^-$	0.03***	2.58
<i>Constant</i>	0.43***	5.07	<i>Constant</i>	0.19**	2.40
ECT_{t-1}	-0.25**	-2.18	ECT_{t-1}	-0.17**	-2.25
Normalized Long-Run Coefficients					
$LogH_t^+$	0.39***	8.91	$LogH_t^+$	0.29***	6.11
$LogH_t^-$	-0.14	0.97	$LogH_t^-$	0.013	0.91
Diagnostic Tests					
	<i>Test Stat.</i>	<i>Prob.</i>		<i>Test Stat.</i>	<i>Prob.</i>
R^2	0.29	-	R^2	0.28	-
\bar{R}^2	0.25	-	\bar{R}^2	0.24	-
F	8.26	0.00	F	6.59	0.00
DW	2.00	-	DW	1.97	-
χ_{SC}^2	0.19	0.90	χ_{SC}^2	4.87	0.08
χ_{HET}^2	12.32	0.26	χ_{HET}^2	14.76	0.25
χ_{FF}^2	0.79	0.37	χ_{FF}^2	11.75	0.17
W_{LR}	28.17	0.00	W_{LR}	-0.30	0.00
W_{SR}	0.12	0.07	W_{SR}	0.042	0.05
EG_{MAX}	-9.80	0.00	EG_{MAX}	-10.83	0.00

Note: ***, ** and * denote statistical significances at 1%, 5% and 10% levels respectively. W_{LR} and W_{SR} are long and short-run Wald tests. Normalized long-run coefficients are obtained with $LogH_t^+ = -\beta_5/\beta_4$, $LogH_t^- = -\beta_6/\beta_4$. EG_{MAX} : Engle and Granger cointegration test statistics. Critical t -table values are 2.57, 1.96 and 1.64 at 1%, 5% and 10%. CUSUM and CUSUM of Squares test graphs are reported in Appendix 2. High R^2 and \bar{R}^2 values show that the explanatory powers of the models are high. DW , χ_{SC}^2 and χ_{HET}^2 statistics indicates that there is no autocorrelation or heteroscedasticity problem in the models. χ_{FF}^2 statistics indicates the there is no model misspecification error. Models are stable, since CUSUM and CUSUMQ graphs in Appendix 2 are within confidence intervals. The long-run (W_{LR}) and short-run (W_{SR}) Wald tests confirm asymmetry both in the long-run and the short-run. EG_{MAX} statistics confirm cointegration relationships between variables.

The normalized long-run estimates of the nonlinear model in Table 4 indicate that only increases in the monetary base (H_t^+) have cointegrated proportional relationships with money supply (M_t^S) for both M1 and M2, since their coefficients are significantly positive. This means that the FED's only expansionary monetary policy, which increases the monetary base (H_t^+), leads to an increase in money supply (both in M1 and M2). However, the coefficient values of (H_t^+) for M1 and M2 are lower than 1, thereby implying that the money multiplier is not stable in terms of increases in the monetary base (in expansionary monetary policy). On the other hand, decreases in monetary base (H_t^-) (contractionary monetary policy) do not have a cointegrated proportional relationship with the money supply (M_t^S) for both M1 and M2 in the long-run, since their coefficients are insignificant. Thus, it can be concluded that the money supply process for both M1 and M2 is not *exogenous* for the US *because the coefficient values of H_t^+ and H_t^- are lower than 1 or insignificant, respectively.*

The comparative result of the linear and nonlinear ARDL models is that the nonlinear model detects a weaker proportional relationship between money supply and monetary base than the linear model for both M1 (0.39) and M2 (0.29) in the long run. These coefficients were found to be 0.70 and 0.39 by the linear model. This may be interpreted to mean that the money supply determination process (mechanism) in the US exhibits a nonlinear character. Furthermore, the nonlinear model indicates that increases (H_t^+) and decreases (H_t^-) in the monetary base have asymmetric impacts on money supply (M_t^S) for M1 and M2, since $(-\beta_5/\beta_4) \neq (-\beta_6/\beta_4)$ in the long-run and $\sum_{j=0}^q \beta_{2j} \neq \sum_{j=0}^r \beta_{3j}$ in the short-run.

4. Conclusion

Understanding the money supply determination process in detail is extremely important for a country. If a nation's central bank (such as the FED in the US) increases or decreases the monetary base more or less than the amounts required, this can easily cause undesirable inflation or deflation rates and, thereby, lead to unwanted results concerning all other macroeconomics variables. Therefore, this study reconsiders the stability of the money multiplier via the traditional *money multiplier model* from a different methodological perspective. Contrary to previous empirical studies, this study assumes that there may be a potentially asymmetric (nonlinear) proportional relationship between money supply and monetary base, because of potentially asymmetric (nonlinear) behaviors of financial actors in response to rising uncertainties in the economies and entwined financial markets of countries. To this aim, the nonlinear ARDL model is applied for the US alongside the linear ARDL

model. The empirical findings of this study are three-fold. First, both linear and nonlinear models indicate that the money multiplier in the US is not stable for both M1 and M2, therefore, implying that the money supply process is not exogenous for the FED. Second, the nonlinear model detects a weaker proportional relationship between money supply and the monetary base than the linear model. Third, the nonlinear model indicates that the FED's only expansionary monetary policy has impacts on money supply, while the contractionary monetary policy of the FED has no impact on the money supply determination process.

In conclusion, new methodological approaches in the nonlinear context, such as the nonlinear ARDL model, may help central banks understand how money supply responds separately to increases and decreases in the monetary base. This study also shows the need for more empirical studies conducted using different methodologies in a nonlinear context for other countries as well.

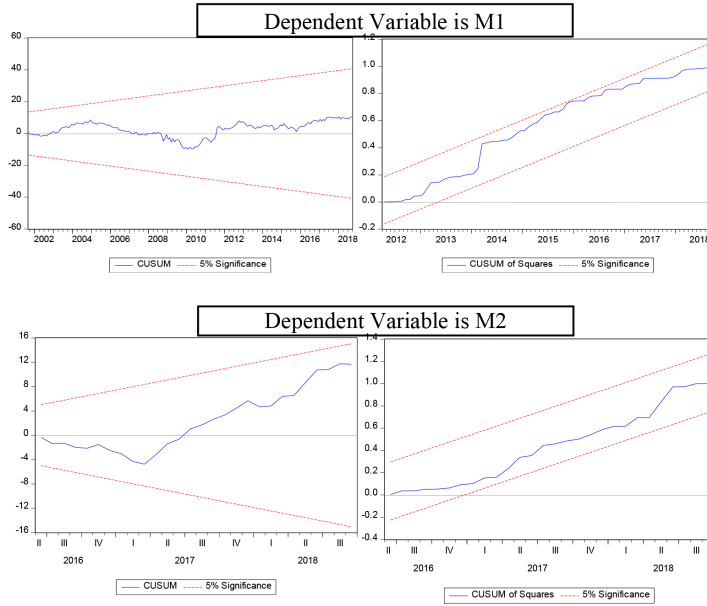
References

- Badarudin, Z.E., M. Ariff and A.M. Khalid, 2013, Post-Keynesian Money Endogeneity: Evidence in G-7 Economies, *Journal of International Money and Finance*, 3, 146-162.
- Baghestani, H, and T. Mott, 1997, A Cointegration Analysis of US Money Supply Process, *Journal of Macro Economics*, 19, 269-283.
- Bhatti, R.H, and M.J. Khawaja, 2018, The Existence of a Stable Money Multiplier in The Small Open Economy of Kazakhstan, *Journal of Economic Studies*, 45(6), 1211-1223.
- Brunner, K. 1961, A Schema for The Supply Theory of Money, *International Economic Review*, 2(1), 79-109.
- Brunner, K. and A.H. Meltzer, 1964, Some Further Investigations of Demand and Supply of Money, *Journal of Finance*, 19, 240-283.
- Darbha, G. 2002, Testing for Long-run Stability- An Application to Money Multiplier in India, *Applied Economics Letters*, 9(1), 33-37.
- Dickey, D.A. and W.A. Fuller 1981, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Econometrica*, 49, 1057-1072.
- Downes, D., W. Moore and Jackson, 2006, Financial Liberalization and The Stationarity of Money Multiplier. *International Economic Journal*, 20, 227-260.
- Engle, R.F. and C.W.J. Granger, 1987, Co-integration and Error-correction: Representation, Estimation, and Testing, *Econometrica*, 55(2), 251-276.
- FRED, (Federal Reserve Bank of St. Louis). (2015). Economic Research.
- Ford, J.L. and J.L. Morris, 1996, The Money Multiplier, Simple Sum, Divisia and Innovation-Divisia Monetary Aggregates: Cointegration Tests for the UK, *Applied Economics*, 28(6), 705-714.
- Friedman, M. and A. Schwartz, 1963, *A Monetary History of United States, 1867-1960*. Princeton University Press, Princeton.
- Goodhart, C.A.E. 1989, The Conduct of Monetary Policy, *Economic Journal*, 99, 293-346.
- Gregory, A.W. and B.E. Hansen, 1996, Residual-based Tests for Cointegration in Models with Regime Shifts, *Journal of Econometrics*, 70(1), 99-126.
- Howells, P.G.A. and K. Hussein, 1998, The Endogeneity of Money: Evidence from the G7, *Scottish Journal of Political Economy*, 45(3), 329-340.
- Khan, M.A. and S. Khan, 2007, Restructuring of Financial Sector in Pakistan, *The Lahore Journal of Economics*, (Special Edition), 97-124.

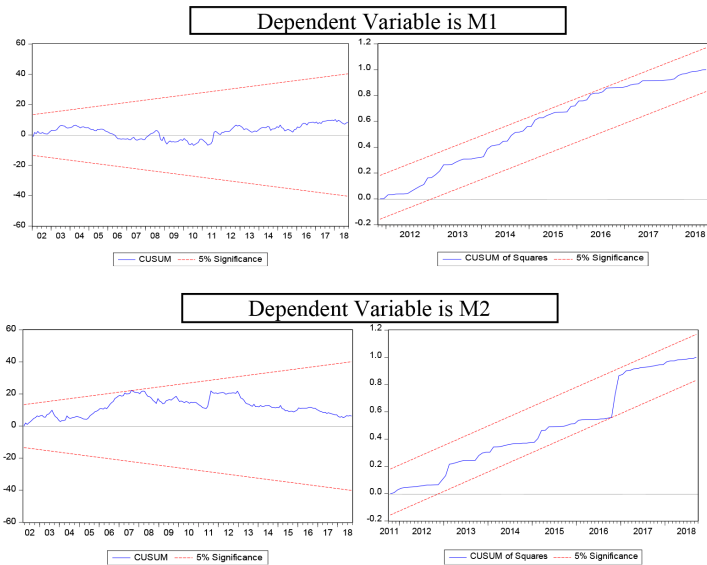
- Johansen, S. and K. Juselius, 1990, Maximum Likelihood Estimation and Inference on Cointegration with Applications to The Demand for Money, *Oxford Bulletin of Economics and Statistics*, 52(2), 169-210.
- Moosa, I.A. and R.H. Bhatti, 1997, International Parity Conditions: Theory, Econometric Testing and Empirical Evidence. Published simultaneously by Macmillan Press Ltd (UK) and St Martin's Press (USA), New York, NY and London.
- Nachnae, D.M. 1992, Money Multiplier in India: Short-run and Long-run Aspects, *Journal of Quantitative Economics*, 8, 51-66.
- Odior, E.S.O. 2013, Pure Portfolio Approach to Money Supply Determination in Nigeria: A Generalized Method of Moments Approach, *Journal of Finance and Economics*, 1(1), 10-26.
- Panagopoulos, Y. and A. Spiliotis, 2008, Alternative Theories of Money, *Journal of Post Keynesian Economics*, 30(4), 601-622.
- Pesaran, M. H., Y. Shin and R. J. Smith, 2001, Bounds Testing Approaches to the Analysis of Level Relationships, *Journal of Applied Econometrics*, 16, 289-326.
- Phillips, P.C. B and P. Perron, 1988, Testing for a Unit Root in Time Series Regression, *Biometrika*, 75(2), 335-346.
- Phillips, P.C.B. and S. Ouliaris, 1990, Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, 58(1), 165-193.
- Rusuhuzwa, T.K. 2015, Monetary and Financial Innovations and Stability of Multiplier in Rwanda, *Issues in Business Management and Economics*, 3(1), 1-8.
- Sahinbeyoglu, G. 1995, The Stability of Money Multiplier: A Test for Cointegration, Central Bank of the Republic of Turkey, DP No. 9603.
- Sen, K. and R.R. Vaidya. 1997, *The Process of Financial Liberalization in India*. Delhi: Oxford University Press.
- Shin, Y., B. Yu and M. Greenwood-Nimmo, 2014, Modelling Asymmetric Cointegration and Dynamic Multipliers in A Nonlinear ARDL Framework. *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*, eds. by R. Sickels and W. Horrace: Springer: 281-314.
- Thenuwara, W and B. Morgan, 2017, Monetary Targeting in Sri Lanka: How Much Control does the Central Bank have Over the Money Supply? *Journal of Economics of Finance*, 41(2), 276-296.
- Tule, K.M. and O.T. Ajilore, 2016, On the Stability of Money Multiplier in Nigeria: Cointegration Analyses with Regime Shifts in Banking System Liquidity, *Cogent Economics and Finance*, 4, 1-16.
- Uchendu, O.A. 1995, Can Money Supply in Nigeria be controlled through Base Money? An Empirical Analysis, *CBN Economic and Financial Review*, 33, 284-316.
- White, O. 2006, Forecasting the Jamaican Money Multiplier, Working Paper, Monetary Analysis and Programming Department, Research and Economic Programming Division, Bank of Jamaica.

Appendix

Appendix 1. Cusum and CusumQ Figures of Linear ARDL Models



Appendix 2. Cusum and CusumQ Figures of Nonlinear ARDL Model



ANALYSIS OF THE EFFICIENCY OF TERTIARY EDUCATION
EXPENDITURE IN EUROPEAN UNION MEMBER STATES
FROM CENTRAL AND EASTERN EUROPE:
AN EFFICIENCY FRONTIER APPROACH

KRISTINA STEFANOVA*

NIKOLAY VELICHKOV

University of National and World Economy, Sofia, Bulgaria

Abstract

The main purpose of this paper is to test the efficiency of tertiary education expenditure in European Union Member States from Central and Eastern Europe, in comparative terms, through the application of an efficiency frontier approach (Data Envelopment Analysis). The results from the study conducted show that the most efficient country with respect to tertiary education expenditure is Romania, followed by the Czech Republic, Lithuania and Slovenia. Estonia and Bulgaria are classified as the most inefficient countries in terms of tertiary education expenditure, with the largest deviation from the efficiency frontier, even though investment in the field is relatively high.

JEL Classification: C14, H52, I21, I23, O52

Keywords: Efficiency frontier approach, Data Envelopment Analysis, Tertiary education, Expenditure efficiency, CEE countries, EU, Comparative analysis

**Corresponding Author:* Kristina Stefanova, Department of Economics, University of National and World Economy, Studentski grad (Students' town), 1700 Sofia, Bulgaria, Office 4027.
E-mail: k.petrova@unwe.bg

Introduction

The development of the scope and quality of tertiary education is important for every country, as it is considered to be an investment in human capital, leading to many positive effects for both individual recipients of the service and the economy as a whole. On the one hand, tertiary education has a positive effect on labour market placement and the welfare of the individual. On the other hand, according to endogenous growth theories (e.g. Romer, 1986), human capital is a factor that has a positive effect on long-term economic growth. In view of the above, it is important to increase investments in the field from both public and private sources and to improve the efficiency of the expenditure incurred.

Unlike secondary education, which advocates the principle of equal access and equal opportunity to a greater extent, tertiary education is not compulsory, and, in most European countries, it is funded from mixed sources.

Table 1. Differences between secondary and tertiary education in EU member states from Central and Eastern Europe (CEE) on average

	Population with secondary education	Population with tertiary education
Unemployment rates - age group 25-29 (%), 2018¹	6.7	6.3
Employment rates - age group 25-29 (%), 2018	77.7	82.9
Monthly earnings - age group 20-64 (euro), 2014	731	1066
People at risk of poverty or social exclusion -,age group 25-49 (%)², 2018	18.8	6.4

Source: Authors' calculations based on Eurostat data.

-
1. The value of the indicator is calculated as an average value for CEE countries from the EU except for Estonia and Lithuania, due to the lack of data on Eurostat.
 2. Except for Slovakia and Lithuania, due to the lack of data for 2018 on Eurostat.

The acquisition of a tertiary education degree provides an advantage and implies a more successful labour market realization. Table 1 shows that the employment rate (age group 25-29) among the population with secondary education in the member states of the European Union from Central and Eastern Europe (CEE), in 2018, is on average 77.7 percent, while for the population with tertiary education it is 82.9 percent. The same trend is observed for youth unemployment (6.7 percent for the population with secondary education vs. 6.3 percent for the population with tertiary education in CEE Member States, in 2018) and income (the average monthly earnings of a person with secondary education in the CEE member states, in 2014, was 731 euro, while for a person with tertiary education it was 1066 euro). In addition, in 2018, in CEE, a significantly smaller proportion of the population with tertiary education was at risk of poverty or social exclusion.

The positive effects on individuals with tertiary education, as well as the external effects that tertiary education generates, are an incentive for public policy in this field. One goal is for most of the population to acquire a tertiary education degree. In this way, more people will benefit from better opportunities to enter the labour market, thereby also affecting inequality in society. Increasing tertiary educational attainment is precisely one of the goals of the Europe 2020 Strategy to achieve smart growth. In 2018, according to Eurostat data, the EU average goal was reached (tertiary education attainment for age group 30-34 was 40.7 percent with a target of 40 percent by 2020).

The data in Figure 1 show that the achievement of the national target is also characteristic of the Czech Republic, Estonia, Latvia, Lithuania, Poland and Slovenia. The other CEE countries studied are close but have not yet achieved the national target set in the Europe 2020 Strategy in this area. The highest value for tertiary education attainment (age group 30-34) is characteristic of Lithuania (57.6 percent for 2018). Other countries in Central and Eastern Europe (Estonia, Poland, Latvia and Slovenia) also present higher values in tertiary education than the 2018 EU-28 average.

Obtaining a tertiary education degree influences labour market integration and welfare, but what is more important is the skills acquired during training. Exploring the relationship between education and economic growth, Barro (2013) points out that "quality and quantity of schooling both matter for growth but that quality is much more important" (Barro, 2013, p. 228). For this reason, the second area in which efforts in tertiary education should focus on is achieving quality of service.

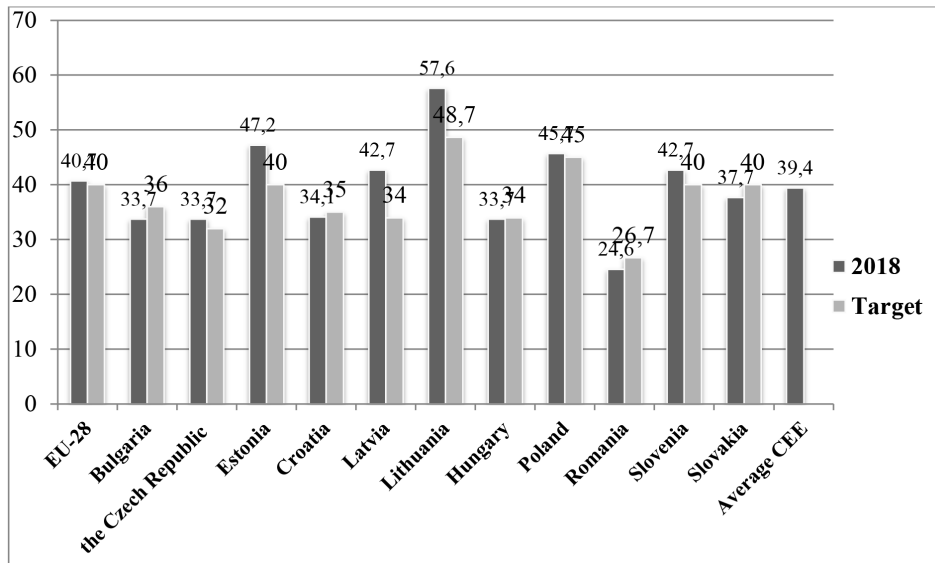


Figure 1. Tertiary educational attainment age group 30-34 (%)

Source: Authors' calculations based on Eurostat data.

Increasing the positive effects of tertiary education in both directions can be achieved by increasing investment in the field or improving the efficiency of the expenditure incurred. Interestingly, against the background of an increase in the proportion of the population with tertiary education, in the period 2008 - 2017, public expenditure on tertiary education, as a share of total public expenditure, decreased, on average, in the EU-28, on average in the CEE Member States, and, also, in all CEE countries except Hungary (see Figure 2). Tertiary education, however, can be a quasi-public good (training at public universities), with a way of excluding consumers, due to the fee charged, or a pure private good (training at private universities). In this regard, tertiary education expenditure has a public as well as a private source, even though in all CEE Member States public exceeds private expenditure³.

3. For a more detailed analysis of higher education funding in CEE Member States, see Yotova and Stefanova (2017).

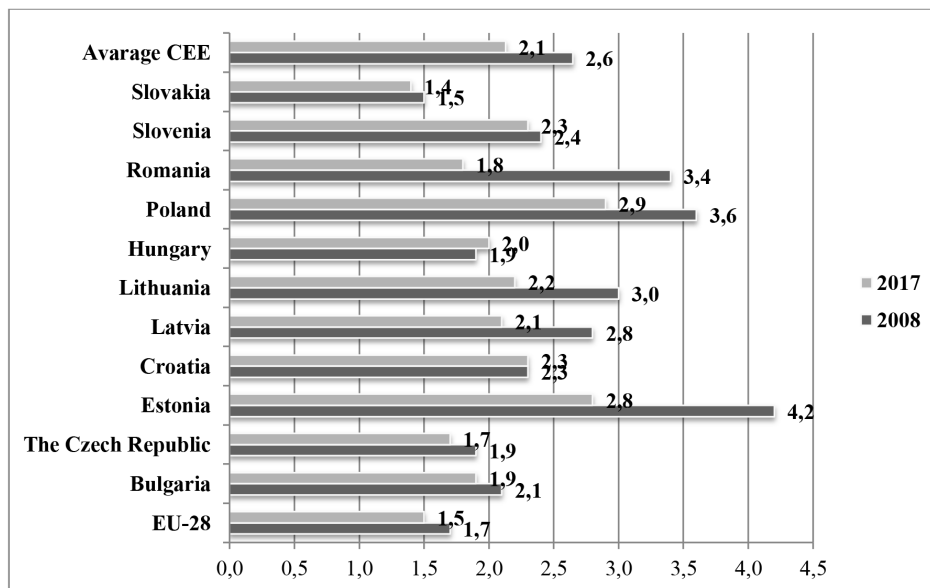


Figure 2. Share of tertiary education public expenditure in total public expenditure (%)

Source: Authors' calculations based on Eurostat data.

Increasing expenditure in tertiary education is important but ensuring that the investment is spent efficiently is even more important. The concept of efficiency is associated with the theory of the firm, but is also increasingly being applied to public policy evaluations. Expenditure efficiency for tertiary education is achieved when the resources given produce the maximum possible results or when the results given are achieved with minimal resources.

The purpose of this study is to carry out a comparative analysis of the expenditure efficiency of tertiary education in the EU Member States of Central and Eastern Europe. To achieve this, the approach adopted is one that aims to increase the validity of conclusions. It differs from those used in relevant literature in the field of combining several methodological decisions: First, it accounts for mixed higher education funding when choosing an input resource indicator (the sum of public and private expenditure on tertiary education per student as a percentage of GDP per capita). Second, it takes account of the time lag between costs spent on higher education and results manifested. Third, it uses three separate models with a common indicator for input and different indicators for output, which is a kind of robustness test concerning the results. Fourth, it takes account of both the direct quantitative effects of expenditure spent on tertiary education and some of the indirect costs related to the labour market realization and welfare.

The paper is structured as follows: The next section provides a brief summary of existing relevant literature in the field. The second part describes the methodology adopted. The third part presents the main results of the study conducted, identifying the most efficient and comparatively inefficient countries, in terms of expenditure on tertiary education. The last part presents the main conclusions drawn from the analysis.

Literature Review

Studies on efficiency of educational expenditure of countries, in comparative terms, predominantly apply non-parametric methods. The most used method of this type is DEA (Data Envelopment Analysis). Existing research studies in the field adopt various methodological approaches that determine the scope, nature and validity of results and conclusions, as well as the contributions made to relevant literature.

Despite the important role of tertiary education in building human capital and placement on the labour market, many of the research studies conducted have not independently examined this important issue. They consider the efficiency of education expenditure as a whole or at different levels (e.g., Afonso and S. Aubyn, 2005; Herrera and Pang, 2005; Jafarov and Gunnarsson, 2008; Aristovnik, 2013; Fonchamnyo and Sama, 2016; Dutu and Sicari, 2016). Fewer studies (e.g., St. Aubyn et al., 2009; Toth, 2009; Yotova and Stefanova, 2017; Jelic and Kedzo, 2018; Stefanova, 2019) trace the specifics of tertiary education, in particular.

Another major feature of much of the research in this field is the use of public tertiary education expenditure as an input resource indicator (e.g., Jafarov and Gunnarsson, 2008; Herrera and Pang, 2005, Fonchamnyo and Sama, 2016; Dutu and Sicari, 2016; Ahec Sonje *et al.*, 2018; St. Aubyn *et al.*, 2009), without taking into account the fact that both public and private funding is available in the sphere in most European countries, including Central and Eastern European countries. Fewer studies (e.g., Toth, 2009; Yotova and Stefanova, 2017; Stefanova, 2019), including this one, consider the mixed nature of higher education funding and use the sum of public and private expenditure as an input resource indicator.

The diversity of methodological approaches in the specialized literature examining the efficiency of higher education expenditure in comparative terms is also largely determined by the choice of output indicators. Existing approaches include indicators for the direct effects of education, such as, for example, labour force with tertiary education (% of total), school enrolment, tertiary educational attainment, ratio of people with diploma to total population (e.g., Aristovnik, 2013; Yotova and Stefanova, 2017; Toth, 2009) and indicators reflecting the quality of education received but having an indirect effect, such as unemployment in the tertiary education population (e.g., Aristovnik, 2013; Ahec Sonje *et al.*, 2018; Jelic and Kedzo, 2018; Stefanova, 2019), the employment rate of tertiary education population (e.g., Toth, 2009; Yotova and Stefanova, 2017; Stefanova, 2019), etc.

In addition to the use of output indicators, the range of countries included in the study is the other major parameter that has a significant impact on the results obtained for the efficiency of using DEA. This is because the method evaluates comparative efficiency and the inclusion or exclusion of a country affects the efficiency frontier calculated and the classification of individual countries as efficient or inefficient. There are also different approaches in this area. Some studies (Herrera and Pang, 2005; Afonso and S. Aubyn, 2005; St. Aubyn *et al.*, 2009; Toth, 2009; Aristovnik 2013; Dutu and Sicari, 2016; Jelic and Kedzo, 2018) examine a broader range of countries that are, however, not homogeneous in terms of economic development, historical features, etc. Others (Jafarov and Gunnarsson, 2008; Yotova and Stefanova, 2017; Ahec Sonje *et al.*, 2018), as well as the current study, focus on a smaller and relatively homogeneous group, such as the EU member states from Central and Eastern Europe.

Due to different methodological approaches, and, especially, due to different country choices, the conclusions from the studies are not identical, making it difficult to define a common conclusion about the countries that show the highest efficiency of tertiary education expenditure. However, some common features of the results of studies on the countries of Central and Eastern Europe can be indicated. For example, Aristovnik (2013) and AhecSonje *et al.* (2018) classify the Czech Republic, Latvia and Lithuania as efficient countries. These countries are also efficient in at least one of the models implemented by Yotova and Stefanova (2017). Lithuania and the Czech Republic are among the most efficient countries in the EU, according to Jelic and Kedzo (2018). As to Romania's place in terms of efficiency of expenditure on tertiary education, there are also some similarities in the results of the studies, with the country being in second place according to Ahec Sonje *et al.* (2018), Aristovnik (2013) and Yotova and Stefanova (2017). Slovakia, on the other hand, is classified as an efficient country by Jelic and Kedzo (2018) and Toth (2009), and Bulgaria is among the most inefficient countries, according to Jafarov and Gunnarsson, Yotova and Stefanova (2017) and Jelic and Kedzo (2018).

Methodology

The evaluation of the efficiency of tertiary education expenditure in this study is done through the application of an efficiency frontier approach. In particular, the method applied is DEA (Data Envelopment Analysis)⁴. The method has been increasingly applied in research on public sector efficiency, and especially for comparative analysis of the efficiency of education and health expenditure. It can also be used in more extensive analyses (e.g., Alfonso *et al.*, 2006).

4. For a more detailed analysis of DEA (Data Envelopment Analysis), see Cooper *et al.* (2011).

The widespread application of the method is a consequence of its advantages as a non-parametric method, in which the form of the efficiency frontier or the functional relationship between inputs and outputs need not be defined in advance, but determined on the basis of specific empirical data on inputs and outputs, through mathematical programming. It is precisely for this reason that the method is particularly suitable for use in the field of tertiary education expenditure efficiency, since it is difficult to determine in advance a specific relationship between input resource and output result, because the effects of higher education are predominantly indirect. Furthermore, according to Mihaiu (2010), the DEA recognises a complex nonlinear relationship between results and inputs, while parametric methods typically limit this relationship, based on a linear relationship or simple forms of a nonlinear one.

The DEA uses linear programming and other forms of mathematical programming methods in order to calculate the efficiency frontier and to derive efficiency coefficients. The DEA classifies countries as efficient (with efficiency coefficient one and situated on the efficiency frontier) and inefficient (with efficiency coefficient under one and situated below the efficiency frontier).

Applying DEA assesses the efficiency of certain units in a comparative way. This means that the method does not provide a theoretical criterion for efficiency, but, rather, indicates which countries are more efficient than others included in the study. For this reason, the choice of countries is essential. In this regard, the current study covers a relatively homogeneous group of countries with similar characteristics, in terms of historical features and economic development, that are relevant to the area under study.

DEA can be used to analyse the efficiency of input resources or output results. If there is inefficiency, with respect to input resources in one country, this means that such input resources must be reduced until the efficiency frontier is reached. In the case of inefficiency with respect to the output result, it must be increased in order to achieve efficiency. In this study, DEA is used to analyse the efficiency of input resources, since these are easier to model and can be directly affected. The influence on output results is more complicated since it cannot be directly addressed and influenced. For this reason, the input-resource DEA model provides a better opportunity to make recommendations to policy makers. The model is applied at variable returns of scale, as this takes account of the different scales of the individual units and allows different input-output ratios to be defined as efficient. In other words, the choice of the variable returns of scale removes the scale effect if some units are not functioning at optimal scale. When studying efficiency in a comparative aspect, under constant returns of scale, only one correlation between input resources – output result is assumed as efficient, and all other units have to be compared against it without taking into account the scale at which individual units, subject to classification, act.

This study employs three models that use one input resource indicator and different output indicators drawing data from the Eurostat database. This approach aims at increasing the validity of results, while also serving as a robustness test of results.

The input indicator selected is Total expenditure on tertiary education per student, as a percentage of GDP per capita. Despite the preponderance of studies on the efficiency of public expenditure on tertiary education in relevant literature, the methodological approach here is different. Because of the mixed funding system, it is impossible to clearly distinguish what part of the results is due to public and what to private sources. In this regard it is more appropriate to use the total tertiary education expenditure, which, in this study, is calculated as the sum of public expenditure and private household expenditure and presented in relative terms.

A critical question of the methodology of this study is how the outputs of tertiary education expenditure incurred should be defined and measured. The first area in which results can be explored reflects the quality of the education received, which, in turn, affects labour market realisation and the welfare of the population with tertiary education. The indicators employment rate of population with tertiary education and population with tertiary education not at risk of poverty and social exclusion were selected to reflect the qualitative aspect in this study⁵. Although the indirect effects of higher education are more important, they can clearly be also influenced by other factors not necessarily related to educational attainment, such as IQ, personal qualities, talents, etc. For this reason, in order to carry out the robustness test of results, as already indicated, an additional third model has been applied. It reflects the direct quantitative effects of the degree obtained using tertiary education attainment as an output indicator.

In particular, the following indicators are used as outputs for the three models. The first model uses Tertiary education attainment (age group 25-34 years). The second model applies Employment rate of population with tertiary education (age group 25-29 years), and the third model uses Population with tertiary education not at risk of poverty and social exclusion (age group 25-49 years). The choice of output indicators is predetermined by the existence of a strong direct theoretical relationship between the input resource and the output that is required for the application of DEA, since the purpose of the method is not to calculate the coefficient of significance of the relationship and verify that it exists, but to determine, through comparative analysis, which countries achieve the greatest resource efficiency (achieving the highest result with a given resource or achieving a given result using the least amount of resources).

5. The indicator population with tertiary education not at risk of poverty and social exclusion is obtained by subtracting the percentage of the population with higher education at risk of poverty and social exclusion from 100%.

DEA works with data for a given year or averaged data for a specific period. The study uses averaged data over two years, the purpose being to prevent any extreme values to affect results. Due to the time lag between the time when the expenditure is incurred and the effects it has (about four years, given the average duration of a Bachelor's degree course), input and output data are taken over different periods. Input resource data for the year 2013-2014 on average is used. Averaged data for the 2017-2018 period⁶ is used for output results since these are the latest data available for the indicators.

The three indicators are defined for a specific age group. The lowest limit is determined by the age at which it is generally considered that genuine integration into the labour market has begun, and at least a Bachelor's degree has been attained. The highest limit is the lowest possible, according to available Eurostat data. Data on tertiary education population indicators are at levels 5 to 8 according to ISCED 2011. All methodological decisions described aim at increasing the reliability of study results.

Results

The results of applying Data Envelopment Analysis to the group of ten EU Member States⁶ from Central and Eastern Europe show that the only country classified as efficient in all three models applied is Romania. The Czech Republic has an efficiency coefficient equal to one, according to the first and third models, and according to the second model, it is the closest to the efficiency frontier among the countries studied. Lithuania is defined as efficient, according to the first and second models, but according to the third model, it is ranked eighth in terms of efficiency coefficient. In this regard, the Czech Republic and Lithuania can be described as relatively efficient, according to this study. Slovenia shows an efficiency coefficient equal to one, according to the first model, while according to the second and third ones, it is ranked fourth and third, respectively. All other countries are classified as inefficient with different deviations from the efficiency frontier according to the three models since their efficiency coefficients are under one (see Table 2).

It is important to note that, according to the three models, the ranking of countries (except Lithuania) is relatively similar, which increases the validity of the conclusions, since the application of the three models also serves as a robustness test. In addition to the similar results of the three models, the validity of the findings of this study is also supported by the validation of their more important part, not only

6. For the indicator Population with higher education not at risk of poverty and social exclusion for Lithuania and Slovakia, only data for 2017 are used, as data for 2018 were missing on Eurostat at the time of the study.

by other authors applying similar methodologies, but also in studies using different output indicators, different time intervals and a much more diverse and comprehensive choice of countries studied. For example, Lithuania and the Czech Republic are classified as efficient countries by Aristovnik (2013), Yotova and Stefanova (2017), Jelic and Kedzo (2018) and Ahec Sonje *et al.* (2018). Romania is classified as efficient by Stefanova (2019) and it is in second place in studies by Ahec Sonje *et al.* (2018), Aristovnik (2013) and Yotova and Stefanova (2017).

Table 2. Efficiency Coefficients

	First Model		Second Model		Third Model	
	Efficiency Coefficient	Rank	Efficiency Coefficient	Rank	Efficiency Coefficient	Rank
Bulgaria	0.7823	6	0.7760	8	0.7760	7
Czech Republic	1.0000	1	0.9938	2	1.0000	1
Estonia	0.7638	7	0.6741	9	0.6741	9
Latvia	0.8626	5	0.7791	7	0.7791	6
Lithuania	1.0000	1	1.0000	1	0.7487	8
Hungary	0.9826	2	0.9788	3	0.9788	2
Poland	0.9659	3	0.8507	5	0.8507	4
Romania	1.0000	1	1.0000	1	1.0000	1
Slovenia	1.0000	1	0.8931	4	0.8931	3
Slovakia	0.8776	4	0.8448	6	0.8448	5
Average	0.9235		0.8790		0.8545	

Source: Authors' calculations through applying DEA on Eurostat data.

The most inefficient country, according to the efficiency coefficients obtained from the three models, is Estonia. It should be noted that the highest percentage of expenditure per student, as a percentage of GDP per capita (see Figure 3), is observed in Estonia, but the results from costs incurred in the areas studied are not satisfactory. At the same time, in Romania and the Czech Republic there is a relatively low value for the input resource indicator. From these observations, a relationship between the expenditure on tertiary education and the efficiency coefficient can be assumed. However, there are also exceptions, since one of the most efficient coun-

tries (Lithuania) ranks second according to the input indicator used. It cannot, therefore, be determined unequivocally that there is an inverse relationship between the magnitude of tertiary education expenditure and the efficiency coefficient. Both countries that spend less and invest more in the field can be efficient. The process of providing the service is more important.

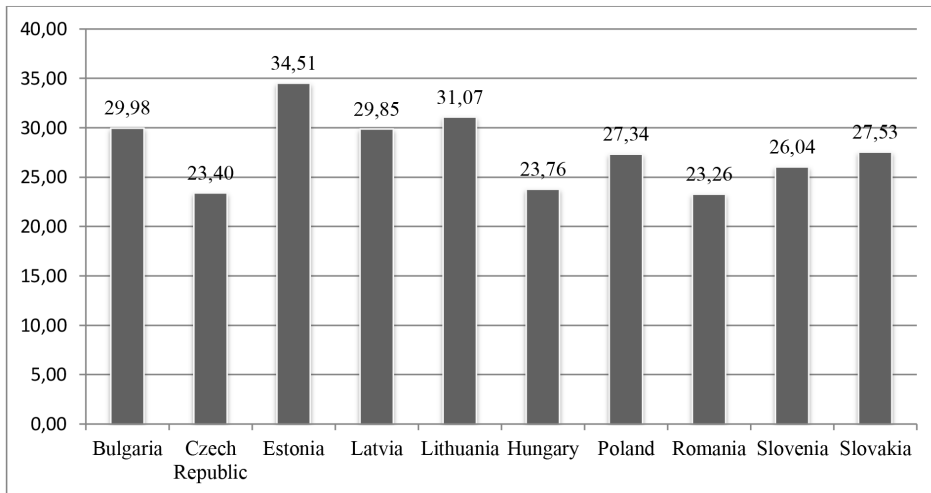


Figure 3. Total expenditure on tertiary education per student, as a percentage of GDP per capita 2013-2014 (%)

Source: Authors' calculations based on Eurostat data.

Results show that Bulgaria is one of the most inefficient in all three models (in the first and second models, it is in the penultimate place before Estonia, and in the third less efficient than Estonia and Lithuania), while, at the same time, it is second in terms of the input indicator used. This indicates that, comparatively, the country presenting a high value of total expenditure on tertiary education per student, as a percentage of GDP per capita, which, however, is not spent in the most efficient way. It should be noted that the poor performance of Bulgaria in terms of efficiency of tertiary education expenditure, is confirmed, by other studies (Jafarov and Gunnarsson, 2008; Yotova and Stefanova, 2017; Jelic and Kedzo, 2018; Stefanova, 2019). The first model has the greatest number of efficient countries (four). For this reason, the average efficiency coefficient is also the highest (0.9235) as compared to the other two models. This indicates that more countries are achieving performance that reflects the quantitative results of the expenditure incurred. The average coefficient obtained shows that, for the same amount of expenditure, one CEE country on average provides 7.65 percent less output than if it had been efficient.

According to the second and third models, reflecting the quality of tertiary education, two countries are classified as efficient, with an average efficiency coefficient of 0.8790 and 0.8545, respectively. The efficiency coefficient of the second (third) model (s) means that for the same amount of input, a country provides 12.1 percent (14.55 percent) less output than if it had been efficient.

Conclusion

The study conducted shows that, despite the use of different output indicators in the three models, the ranking of the counties (except Lithuania) is relatively close. Romania is classified as efficient in all three models applied. Lithuania and the Czech Republic have an efficiency coefficient equal to one, according to two of the models, while Slovenia has an efficiency coefficient equal to one according to the first model. According to the three models, the most inefficient country is Estonia, followed by Bulgaria. The study results are also consistent with the results of other researchers in this field.

All EU Member States from CEE included in the study show the highest efficiency coefficients in the first model. At the same time, the average efficiency coefficient, according to this model, is the highest and the number of efficient countries is the greatest. This indicates that the quantitative aspects of the results of the tertiary education expenditure incurred are higher than those related to the quality of the service provided and the labour market realisation and welfare. The existence of a supranational objective, in terms of the quantitative aspects of the results of investment in tertiary education, is one of the reasons for the results observed. However, given the ultimate goal of investing in human capital, it is necessary to strengthen the pursuit of quality in tertiary education in the Member States of Central and Eastern Europe. Increasing the positive results in these areas will favour the countries' economic development in the long run.

References

- Afonso, A. and M. St. Aubyn, 2005. Non-parametric Approaches to Education and Health Efficiency in OECD Countries. *Journal of Applied Economics*, 8 (2), 227-246.
- Afonso, A., L. Schuknecht, V. Tanzi, 2006, Public sector efficiency Evidence for new member states and emerging markets, *Working paper series* No 581/ January 2006, European Central Bank, 1-50.
- Ahec Sonje, A., M. Deskar-Skrbic and V. Sonje, 2018. Efficiency of public expenditure on education: comparing Croatia with other NMS, *MPRA Paper* No. 85152, 1-14.
- Aristovnik, Al., 2013. Relative Efficiency of Education Expenditures in Eastern Europe: A Non-parametric Approach, *Journal of Knowledge Management, Economics and Information Technology*, Vol. III, Issue 3, 1-13.
- Barro, R., 2013. Education and Economic Growth. *Annals of Economics and Finance*, 14-2(A), 277-304.

- Cooper, W., L. Seiford, J. Zhu, Data envelopment analysis, History, Models, and Interpretations. In Handbook on Data Envelopment Analysis; Springer: Boston, MA, USA, 2011, pp. 1–39.
- Dutu, R. and P. Sicari, 2016. Public Spending Efficiency in the OECD: Benchmarking Health Care, Education and General Administration, *OECD Economics Department Working Papers* No. 1278, 1-25.
- Fonchamnyo, D. C. and M. C. Sama, 2016. Determinants of public spending efficiency in education and health: evidence from selected CEMAC countries, *Journal of Economics and Finance*, 199-210.
- Herrera, S. and G. Pang, 2005. Efficiency of Public Spending in Developing Countries: An Efficiency Frontier Approach, *World Bank Policy Research*, Working Paper 3645, 1-67.
- Jafarov, Et. And V. Gunnarsson, 2008. Government Spending on Health Care and Education in Croatia: Efficiency and Reform Options, *IMF Working Paper*, WP/08/13, 1-34.
- Jelic, O. and M. Kedzo, 2018. Efficiency vs effectiveness: an analysis of tertiary education across Europe, *Public sector economics*, 42 (4), 381-414.
- Mihaiu, D. 2010. Study regarding the efficiency of public spending on education at EU level through benchmarking, *Studies in Business and Economics*, 5(3), pp. 201-214.
- Romer, P., 1986, Increasing Returns and Long-Run Growth. *The Journal of Political Economy*, 94 (5), 1002-1037.
- St. Aubyn, M., A. Pina, F. Garcia and J. Pais, 2009. Study on the efficiency and effectiveness of public spending on tertiary education, *Economic Papers* 390, 1-148.
- Stefanova, Kr., 2019, New Approach to Evaluation of the Efficiency of Higher Education Expenditure in Regard to Quality Achieved (Evidence from Selected Countries from Central and Eastern Europe), *Bulgarian Journal of Business Research*, Issue 2, 25-34.
- Toth, R., 2009. Using DEA to Evaluate Efficiency of Higher Education, *Applied Studies in Agribusiness and Commerce APSTRACT*, 79-82.
- Yotova, L. and Kr. Stefanova, 2017. Efficiency of Tertiary Education Expenditure in CEE Countries: Data Envelopment Analysis, *Economic Alternatives*, Issue 3, 352-364.
<http://ec.europa.eu/eurostat>.

*Smart, Sustainable and Inclusive Growth
Political Entrepreneurship for a Prosperous Europe*

edited by

Charlie Karlsson, Daniel Silander, Brigitte Pircher

Published by Elgar Publishing, 2019

ISBN: 978 1 78897 408 0 (cased), ISBN: 978 1 78897 409 7 (eBook), Extent: 240 pp

reviewed by Theodore P. Lianos*

In 2010, the European Commission officially launched the *Europe 2020* Project, which is described as a strategy for smart, sustainable and inclusive growth, intended to promote growth, employment and social integration in Europe. The aim of this volume is to analyze this strategy and the role of the so called European political entrepreneurship in implementing it.

The book has four parts: Europe 2020 and Study Framework, Core Actors of Europe 2020, Policies for Smart, Sustainable and Inclusive Growth, and Concluding Remarks. It contains twelve essays authored by ten experts most of whom are from or work in Sweden. All essays are well-written, concise and informative.

With respect to the EU policies described in these essays, this reader feels uneasy about two issues, namely, migration and economic growth. Regarding economic growth and the emphasis EU policy makers place on it, the authors appear to be either uninformed or to prefer to ignore the well-established fact that the limits for further growth have been exhausted. The current challenge is not GDP growth or of per capita income, but, rather, income inequality and citizens' welfare. Furthermore, the authors seem to believe that growth can be achieved without negative environmental impact. They ignore research findings showing that such 'decoupling' in a relative or absolute sense, is not a real possibility. The role of technology in increasing productivity of resources is not unlimited.

* Theodore P. Lianos, Professor Emeritus of Economics, Athens University of Economics and Business, Athens, Greece

Regarding migration policies, it appears that EU policy makers believe that Europe is underpopulated and, therefore, migration is necessary to meet labor needs. Thus, they see migrants as a cheap labor source, necessary for growth, as the case used to be in many wealthy countries in the past. At the same time, the authors ignore that migration has caused a disruption of the social web in many countries and in many areas within each country. The truth, of course, is that Europe (and the world) is overpopulated, if examined from the point of view of ecological balance. The inflows of migrants to Europe will simply exacerbate the situation. It is also clear that social discontent in many European countries is rising.

Of course, the review of a good book, such as this one, is not the proper opportunity for criticizing EU policies. However, it is hard to 'shake off' the feeling that the social, environmental and economic priorities of the European Union need to be reconsidered.

guide for authors

The articles should be written as follows:

(1) Papers must be in English.

(2) Papers for publication (two copies) should be sent to:

Mrs. Melina Petromelidou

Editorial Secretary

South-Eastern Europe Journal of Economics

ASECU, University of Macedonia, 156, Egnatia Str., 546 36 Thessaloniki, Greece

The *Journal's* phone number is (+30) 2310891793, fax: (+30) 2310891748

e-mail: asecu@uom.edu.gr

Submission of a paper will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere. The Editor does not accept responsibility for damage or loss of papers submitted. Upon acceptance of an article, author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.

(3) Papers will be considered in any form, but authors of papers accepted for publication will be expected to provide a final copy conforming to the general style of the Journal as outlined in notes 4 through 13 below.

(4) Manuscripts should be 1,5 spaced, with wide margins, and printed on one side of the paper only. All pages should be numbered in sequence. Titles and subtitles should be short. References, tables, and captions for the figures should be printed on separate pages.

(5) The first page of the manuscript should contain the following information: (i) the title; (ii) the name(s) and institutional affiliation(s) of the author(s); (iii) an abstract of not more than 100 words. A footnote on the same sheet should give the name, address, and telephone and fax numbers of the corresponding author [as well as an e-mail address].

(6) The first page of the manuscript should also contain at least one classification code according to the Classification System for Journal Articles as used by the Journal of Economic Literature; in addition, up to five key words should be supplied.

The classification system used in JEL can be found at:

http://www.aeaweb.org/journals/jel_class_system.html.

(7) Acknowledgements and information on grants received can be given in a first footnote, which should not be included in the consecutive numbering of footnotes.

(8) Footnotes should be kept to a minimum and numbered consecutively throughout the text with superscript Arabic numerals.

(9) Displayed formulae should be numbered consecutively throughout the manuscript as (1), (2), etc. against the right-hand margin of the page. In cases where the derivation of formulae has been abbreviated, it is of great help to the referees if the full derivation can be presented on a separate sheet (not to be published).

(10) References to publications should be as follows: ‘Smith (1992) reported that...’ or ‘This problem has been studied previously (e.g., Smith et al., 1969)’. The author should make sure that there is a strict one-to-one correspondence between the names and years in the text and those on the list. The list of references should appear at the end of the main text (after any appendices, but before tables and captions for figures). It should be double spaced and listed in alphabetical order by author’s name. References should appear as follows:

For monographs

Sen, A., 1970, *Collective Choice and Social Welfare*, San Francisco: Holden Day.

For contributions to collective works

Kornai, J., 1991, Stabilization and Economic Transition in Hungary: The Next Two Years, in J. de Melo and A. Sapir (eds.), *Trade Theory and Economic Reform: North, South and East*, Oxford: Basil Blackwell, 307-326.

For periodicals

Magdalinos, M., 1990, “The Classical Principles of Testing Using Instrumental Variables Estimates”, *Journal of Econometrics*, 44, 241-279.

Note that journal titles should not be abbreviated.

(11) Illustrations will be reproduced photographically from originals supplied by the author; they will not be redrawn by the publisher. Please provide all illustrations in quadruplicate (one high-contrast original and three photocopies). Care should be taken that lettering and symbols are of a comparable size. The illustrations should not be inserted in the text, and should be marked on the back with figure number, title of paper, and author’s name. All graphs and diagrams should be referred to as figures, and should be numbered consecutively in the text in Arabic numerals. Illustration for papers submitted as electronic manuscripts should be in traditional form.

(12) Tables should be numbered consecutively in the text in Arabic numerals and printed on separate sheets.

(13) Accepted papers should be submitted in electronic form in a storage media (i.e. CD, DVD, USB, etc.) with accompanying manuscript. Electronic manuscripts have the advantage that there is no need for re-setting of text, thereby avoiding the possibility of introducing errors and resulting in reliable and fast delivery of proofs. The preferred format is either .doc or .docx. Make absolutely sure that the file on the disk and the printout are identical.

(14) Page proofs will be sent to the corresponding author. Proofs should be corrected carefully; the responsibility for detecting errors lies with the author. Corrections should be restricted to instances in which the proof is at variance with the manuscript. There are neither submission fees nor page charges.