The Impact of New Doctorate Graduates on Innovation Systems in Europe

Leogrande, Angelo and Costantiello, Alberto and Laureti, Lucio

LUM University Giuseppe Degennaro, LUM University Giuseppe Degennaro, LUM University Giuseppe Degennaro

6 September 2022

Angelo Leogrande[1]*, *Alberto Costantiello*[2]*, *Lucio Laureti*[3]*,
*LUM University Giuseppe Degennaro, Casamassima, Bari, Puglia, Italy, EU

The Impact of New Doctorate Graduates on Innovation Systems in Europe

Abstract

In this article we investigate the determinants of "*New Doctorate Graduates*" in Europe. We use data from the EIS-European Innovation Scoreboard of the European Commission for 36 countries in the period 2010-2019 with Pooled OLS, Dynamic Panel, WLS, Panel Data with Fixed Effects and Panel Data with Random Effects. We found that "*New Doctorate Graduates*" is positively associated, among others, with "*Human Resources*" and "*Government Procurement of Advanced Technology Products*" and negatively, associated among others, with "*Total Entrepreneurial Activity*" and "*Innovation Index*". We apply a clusterization with k-Means algorithm either with the Silhouette Coefficient either with the Elbow Method and we found that in both cases the optimal number of clusters is three. Furthermore, we use the Network Analysis with the Distance of Manhattan, and we find the presence of seven network structures. Finally, we propose a confrontation among ten machine learning algorithms to predict the value of "*New Doctorate Graduates*" either with Original Data-OD either with Augmented Data-AD. Results show that SGD-Stochastic Gradient Descendent is the best predictor for OD while Linear Regression performs better for AD.

Keywords: Innovation, and Invention: Processes and Incentives; Management of Technological Innovation and R&D; Diffusion Processes; Open Innovation.

JEL Classification: O30; O31, O32; O33; O36.

1. Introduction-Research Question

The following article analyzes the variables that have an impact in determining the number of new PhDs in Europe in the context of technological innovation. The analysis is above all quantitative, metric and has a machine learning approach. PhDs are needed to make countries competitive from the point of view of technological innovation and research and development. Research doctorates are necessary to be competitive at national level in the knowledge and information economy.
In this regard, it is necessary to underline that in the context of the technological competition that juxtaposes the US and China, Europe is far behind. In fact, there are no big players in Europe operating in the IT sectors. This lack is also due to sociological and institutional factors. From a sociological point of view, in fact, in Europe there is an abundance of small and medium-sized enterprises and a sort of prejudice towards large companies especially on those listed on financial markets.
From the point of view of the institutions it is necessary to consider that the European Commission to defend European free and competitive markets often is hostile in respect to fusions, mergers, and

[1] Assistant professor of Economics at LUM University Giuseppe Degennaro and Researcher at LUM Enterprise s.r.l. Email: leogrande.cultore@lum.it, Strada Statale 100 km 18, Casamassima, Bari, Puglia, Italia.
[2] Professor of Economics at LUM University Giuseppe Degennaro. Email: costantiello@lum.it. Strada Statale 100 km 18, Casamassima, Bari, Puglia, Italia.
[3] Professor of Economics at LUM University Giuseppe Degennaro. Email: laureti@lum.it. Strada Statale 100 km 18, Casamassima, Bari, Puglia, Italia.

acquisition of large companies. However, these sociological and institutional limitations impede the creation of big corporations that might best be able to give job prospects to PhDs.
Indeed, one of the problems of PhDs is employment. Only part of the PhDs is employed in universities. The remainder is instead employed in small and medium-sized enterprises and in public institutions. However, notoriously, small and medium-sized enterprises do not have the necessary organizational structures to do research and development, and public institutions are far behind in technological innovation. It follows that much of the skills and technical-scientific knowledge of research doctorates is lost due to the lack of productive organizations able to use their abilities for profit and market purposes.
On the other hand, however, there are also problems within the university system for the efficient organization of research doctorates. In fact, to create knowledge that is in line with the market and with the general trends of scientific research systems, universities tend to continuously innovate programs and create new scientific disciplines. However, such efforts are often in vain as the doctorate remains excessively tied to a type of academic training that prevents doctorates from being useful in businesses and institutions. To solve this problem, universities, alongside the traditional doctorate, also offer professional doctorates which aim to be closer to the needs of enterprises. There are many expectations regarding the ability of professional doctorates to innovate the university system and the business system together. However, the ability of research doctorates to be useful also depends on the presence of an entrepreneurial and industrial system that is pro-actively interested in technical-scientific knowledge.
Furthermore, it is also necessary to consider that even where there are good opportunities for cooperation between universities and businesses, the system of public institutions may not necessarily be able to encourage such collaborations. There are therefore three elements necessary to ensure that the research doctorate can be useful to an economic-territorial context, namely: a good academic orientation, an entrepreneurial system attentive to scientific research, and institutions ready to promote forms of public collaboration.
It is also necessary to consider that there are disciplines, namely STEM subjects, in which there is a chronic lack of research doctorates compared to both job opportunities and the needs of universities, institutions and the market. It is therefore necessary that policy makers continue to invest in the formation of doctoral programs to ensure that universities have the necessary resources to offer high-level training. However, there is also a need for a broader work of a cultural and value nature that can open companies to the participation of doctorates in production processes. Finally, institutions should also be more careful to make the best use of the human capital produced by research doctorates. The knowledge economy will increasingly require the production of research doctorates in new subjects, at the frontier of science, as happens for example for the various contaminations of artificial intelligence with the various traditional disciplinary sectors. However, to ensure that these market demands are met, it is necessary that there are policy makers able to recognize the strategic role of research doctorates in the production of added value at the country level in universities, companies and in public institutions.
The article continues as follows: the second paragraph presents a brief analysis of the scientific literature, the third paragraph discusses the econometric results, the fourth paragraph contains clustering with the k-Means algorithm, the fifth paragraph presents the network analysis, the sixth paragraph compares ten machine learning algorithms for prediction, the seventh paragraph contains prediction with Augmented Data and machine learning algorithms, the eighth paragraph concludes.

2. Literature Review

The following paragraph presents a brief analysis of some bibliographic references necessary to introduce the topic of the role of research doctorates in the university system and in production systems with attention to technological innovation and research and development.

[1] refer to the statistics relating to the relationship between the condition of work and the number of research doctorates in the USA in the period between 1997 and 2016. The authors verify that approximately 3,124 doctorates were obtained in the period considered. However, the growth in the number of doctoral degrees was mainly due to international doctoral students who grew by 156% compared to US doctoral students. From an ethnic point of view, 85% of doctorates are white or Asian, while Hispanics and blacks are in the minority. 76% of PhDs received job offers. The salaries of doctorates who have decided to remain in the academy equal to 59,484.00 dollars, are lower than both the doctorates who entered the industry sector with a salary of 84,918.00 dollars, in the government with a salary of 69,970 dollars or in educational institutions with an income of $ 81,428. [2] consider the reasons that push students to choose a professional doctorate instead of a traditional academic doctorate or PhD path. The authors also consider the implications that this choice has from a professional and career promotion point of view. The results show that doctoral students choose professional doctorates for reasons that are related to the flexibility of training programs. Furthermore, professional doctorates present both the possibility of dealing with theoretical issues and the possibility of tackling practical issues. Finally, professional doctorates are also able to give greater importance and prospects to the professional career of the participants. [3] analyze the case of four women who attended the first edition of the Professional Doctorate program of the University in the Southwest of England. The authors analyze their own story in the doctoral program evaluating their experience also with other doctoral colleagues. In the STEM disciplines, students are also increasingly acquiring PhDs [4]. However, the number of research doctorates remains below the number required by the market, or which would be necessary to allow adequate technology transfer and growth in the production of services and products based on research and development. Furthermore, the continuous investment in science and technology increases even more the need to produce and share knowledge in the STEM field. The response of universities to this climate consists in orienting doctoral programs more and more towards the transfer of knowledge and the acquisition of new skills with very significant changes in terms of the offer of research doctorates compared to traditional mainstream proposals. [5] address the issue of emigration of Italian doctorates abroad. The authors focus on PhDs in the disciplines of Economic Science, Finance and Business Management in the period 2008-2010. Doctoral students who tend to go abroad are those who are both the best and those below the average for research performance. On the other hand, PhDs that have average research performance tend to stay in Italy. The best-performing PhDs in terms of research tend to go abroad to top universities to improve their careers. On the other hand, below-average Italian PhDs who emigrate do not necessarily seek to work in an academic environment and often choose to work in non-academic institutions simply expressing the need to find work. The authors conclude the article by contrasting the widespread idea that only high-performing doctorates choose the path of emigration abroad. [6] addresses the issue of employment of workers with a PhD outside the university-academic world. The authors consider the case of Finland. In fact, universities are no longer able to absorb research doctorates. Many doctorates therefore must find employment outside the university world. However, the transition from the academic world to the world of industry and business is made complex by the inefficiencies of the labor market. The analysis is based on the career path of 1183 academics. The results show that there are not many opportunities for PhDs outside the university world except in corporate roles that are closely related to research and development. Furthermore, employers who hire doctorates expect them to be able to bring specific scientific and technological knowledge and to be able to forge relevant relationships with the academic world. [7] analyze the role of the learning environment with a qualitative approach of doctoral students and postgraduates in the disciplines of dental studies. The authors analyzed a total of 20 participants of which 16 in the last year of the doctorate and 4 postgraduates in dentistry. The authors asked the

interviewees to identify the elements of efficiency and inefficiency in the training path. The results show that the students identified as effective for learning: intervention planning, patient treatment methodologies, complexity of clinical cases presented, orientation to clinical research, clinical training. The students themselves identified the following elements as ineffective in the context of doctoral and professional learning, namely: absence of an adequate number of clinical cases, excessive emphasis on research, lack of adequate facilities in dental schools. The authors suggest paying attention to aspects of the learning environment within doctoral and professional studies to enhance students' ability to learn effectively. [8] refer to the professional doctorates introduced to meet the challenges of the twenty-first century and built trying to meet the needs of the university and of the profession. Both the university world and the world of professional activities are changing rapidly. The professional doctorate seeks to offer the skills necessary to be competitive despite the changes. The authors consider the professional doctorate as the most relevant university innovation among high-level qualifications and as the most significant output of the liberal approach to the education sector. The analysis proposed through the dialogue between a research doctor and a doctoral student addresses the issue of the new challenges of university education and how the professional doctorate can ensure the achievement of relevant educational objectives. [9] consider the case of modern education in medical disciplines in China using official data. The authors compare education in medical disciplines with education in other scientific disciplines. The data used were collected by the Chinese Ministry of Education in the period between 1998 and 2012. The results show an increase in students. In particular, the number of PhD students increased 5 times in the period considered. The authors consider the importance of investing in the development of medical professions with a modern approach to the university education system.

[10] considers the performance of Poland in the field of technological innovation in connection with countries that are defined as "*Moderate Innovators*" based on the use of the definitions of the European Innovation Scoreboard-EIS of the European Commission. The results of the analysis show that Poland's weak performance compared to its peers is also due to the lack of an adequate number of PhDs. [11] considers the role of Work Integrated Learning-WILs for PhDs in Australia. Work-Integrated Learning methods are very popular with graduates even if they are rarely used by PhDs. However, only half of Australian PhDs find space in the high-level education sector. Research doctorates who fail to enter the university can play a very important role in supporting the transfer of knowledge to companies and institutions. The author therefore presents a qualitative comparison of the views of PhDs who attended Griffith University. The results show that PhDs recognize the growing role of methods related to Work-Integrated Learning WIL as a tool for increasing the employability of PhDs in the industrial sector. [12] analyze the case of the role of doctoral studies for management training. The training of PhDs has received little attention from universities. The analysis seeks to identify methodologies for assessing the quality of the research doctorate. The authors verify that the value attributed by students to the doctorate differs significantly from the value attributed by teachers to the doctorate. In fact, while on the one hand the students in the evaluation of the research doctorate give value to personal, professional, and organizational issues, on the other hand the teachers in the evaluation of the research doctorate consider above all the methodological aspects and the epistemological characteristics. Furthermore, the authors emphasize that the organization of research doctorates tends to be oriented more to the aims of university institutions than to professional, market and industry needs. The authors suggest expanding the value of the doctorate by considering not only the needs of teachers but also the needs of students and the demands of the production and industrial system.

3. The Econometric Model

We have estimated the following model for 36[4] countries using data from the European Innovation Scoreboard-EIS of the European Commission:

$$\begin{aligned}\boldsymbol{NewDoctorateGraduatesInEurope}_{it} &= \boldsymbol{a_1} + \boldsymbol{b_1}(\boldsymbol{AttractiveResearchSystems})_{it} \\ &+ \boldsymbol{b_2}(\boldsymbol{BasicSchoolEntrepreneurialEducationAndTraining})_{it} \\ &+ \boldsymbol{b_3}(\boldsymbol{EmploymentMHTManufacturingKISServices})_{it} \\ &+ \boldsymbol{b_4}(\boldsymbol{EnteprisesProvidingICTTraining})_{it} + \boldsymbol{b_5}(\boldsymbol{FirmInvestments})_{it} \\ &+ \boldsymbol{b_6}(\boldsymbol{ForeignDoctorateStudents})_{it} \\ &+ \boldsymbol{b_7}(\boldsymbol{GovernmentProcurementOfAdvancedTechnologyProducts})_{it} \\ &+ \boldsymbol{b_8}(\boldsymbol{HumanResources})_{it} + \boldsymbol{b_9}(\boldsymbol{InnovationIndex})_{it} \\ &+ \boldsymbol{b_{10}}(\boldsymbol{InnovativeSalesShare})_{it} + \boldsymbol{b_{11}}(\boldsymbol{LifelongLearning})_{it} \\ &+ \boldsymbol{b_{12}}(\boldsymbol{MediumAndHighTechProductExports})_{it} \\ &+ \boldsymbol{b_{13}}(\boldsymbol{MostCitedPublications})_{it} \\ &+ \boldsymbol{b_{14}}(\boldsymbol{NonR\&DInnovationExpenditures})_{it} \\ &+ \boldsymbol{b_{15}}(\boldsymbol{ProductOrProcessInnovators})_{it} + \boldsymbol{b_{16}}(\boldsymbol{TertiaryEducation})_{it} \\ &+ \boldsymbol{b_{17}}(\boldsymbol{TotalEntrepreneurialActivity})_{it} + \boldsymbol{b_{18}}(\boldsymbol{VentureCapital})_{it}\end{aligned}$$

Where $\boldsymbol{i = 36}$ and $\boldsymbol{t = [2010; 2019]}$.

The European Innovation Scoreboard-EIS defines New Doctorate Graduates as a variable that considers the number of doctorates in the population between 25 and 34 years old. This variable analyzes the offer of doctorates in all fields of training. We found that "*New Doctorate Graduates*" is positively associated with:

- *Human Resources:* is a measure that considers three elements: "New Doctorate Graduates", "Population Aged 25-34 with Tertiary Education", and "Lifelong Learning" [13]. There is therefore a positive relationship between the "Human Resources" variable and the "New Doctorate Graduates" variable. This relationship can be better understood considering that the "New Doctorate Graduates" variable is a constituent component of the "Human Resources" variable. It obviously follows that in countries where significant investments are made in human resources there is also a significant increase in the value of the number of people who acquire doctoral degrees. Investment in human capital is an essential element to ensure that an economy can be oriented towards technological innovation and research and development. And in this process a very important role is played by research doctorates who can create increasingly technologically advanced products and services [14].
- *Government Procurement of Advanced Technology Products:* is a variable that considers the capacity of public procurement to promote technological innovation through the demand for goods and services. This variable has a range between 1 and 7. The value is 1 if the purchase choice of the public administration depends exclusively on the price. The value is maximum or 7 if the purchases of the public administration depend exclusively on the degree of technological innovation of the products and services. There is a positive relationship between the value of the state's ability to express a demand for advanced technological services and products and the presence of PhDs. This relationship means that if the state uses public

[4] Countries are: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, UK.

spending to purchase technologically advanced products and services then it also offers a very significant incentive for the diffusion of research doctorates.

- *Medium and High-Tech Product Exports:* refers to the export of media and other technology products as a percentage of total exports [15] . It is therefore a measure that shows the ability of the European Union to market its products and services with a high level of technological innovation and research and development at an international level. The ability of countries to export medium and high technology products are also the most competitive countries from a production point of view. Furthermore, exports of medium and high-tech products also have the potential to generate higher levels of well-paid employment. There is therefore a positive relationship between the ability to export medium and high technology products and the spread of research doctorates. This relationship derives from the fact that the possibility of producing medium and high technology goods and services also depends on the presence of research doctorates who also operate within applied industrial research.
- *Product or Process Innovators:* is a variable that refers to the number of small and medium-sized enterprises that have introduced product and process innovations for the company or for the market [16]. In this regard, a distinction is made between product innovation and process innovation. Product innovation is the introduction on the market of a new or improved good based on the capacity of use, the components and the subsystems used. Process innovation means the reorganization of production activities using technologies or organizational interventions. This value is reported to the total amount of Small and Medium Enterprises. The ability of small and medium-sized enterprises to innovate in technology also indicates the ability of enterprises to be more productive, competitive, and able to generate good employment. There is a positive relationship between the value of process and product innovations and the number of PhDs at country level. This relationship indicates that small and medium-sized enterprises that are engaged in technological innovation need and benefit from the presence of PhDs.
- *Most-cited Publications:* is a variable that takes into account 10% of the most cited publications worldwide as a percentage of the country's total scientific publications [17]. The indicator is a measure of the efficiency of the research system at national level. In fact, where publications are cited more than it is believed that there is a higher quality of research. There is therefore a positive relationship between the number of the most cited publications at international level and the number of research doctorates at national level. Research doctorates are in fact trained for scientific research and tend to produce international publications. It therefore follows that the greater the research doctorates, the greater the scientific publications at an international level and the possibility of placing publications in the top 10% at an international level also increases. Obviously, if there are more PhDs then it is also possible to deduce that there are more investments in scientific research.
- *Public-Private co-Publications:* is an indicator that considers public-private co-publications for millions of inhabitants [18]. However, publications from the medical-health sector are excluded from the analysis. This indicator takes into consideration the relationships existing between the public and the private sector with the aim of generating scientific publications as output. The collaboration between researchers operating in the private sector, generally operating in applied or industrial research, and researchers operating in the public sector, generally oriented towards basic research, can increase the quality of research and generate an important impact in terms of technological innovation at an industrial level. There is a positive relationship between public-private cooperation in scientific research and the presence of research doctorates. This positive relationship can best be understood by generally

considering that both researchers working in the private sector and researchers working in the public sector generally hold doctoral degrees.

- *Enterprises Providing ICT Training:* is an indicator that considers the number of companies that invest in the IT training of their staff [19] . The indicator is built as a ratio. The numerator is the number of firms that have provided training to develop the IT skills of their employees and the denominator is the total number of firms. IT skills are very relevant for technological innovation in the digital field within companies. Therefore, if companies invest in the technological and IT training of their employees, this investment can be understood as a signal of the orientation towards innovation of companies at the country level. There is a positive relationship between the number of companies offering investments in terms of IT training for employees [20] and the presence of research doctorates at national level. This relationship can be understood considering that the two variables are part of the same phenomenon or the overall investment that countries make in human resources to be able to generate technological innovation both in companies, with ICT training [21] and in universities, with the training of research doctorates.
- *Non-R&D Innovation Expenditure:* is a variable that takes into consideration the innovation that is achieved without investments in research and development as a percentage of turnover. It is a variable considered as a ratio between the value spent on innovation not deriving from research and development as a percentage of the total value of the turnover of all companies. There are investments in innovation that do not require research and development such as investments in equipment and machinery, in the acquisition of patents, and in the dissemination of new ideas and methodologies relating to production. There is a positive relationship between investment in technological innovation that does not involve investment in research and development and the number of research doctorates at national level. This relationship indicates that even technological innovations that do not have implications in terms of research and development still require a qualified human capital that recognizes the implications of innovations in terms of added value in the broader context of the information economy.
- *Innovative Sales Share:* is an indicator that takes into consideration the impact of technological innovations in terms of sales as a percentage of turnover [22]. The indicator is calculated based on a ratio. In the numerator is the sum of the total turnover from new or improved products. The denominator is the total turnover for all firms. This variable therefore considers the impact of technological innovations in terms of turnover. The indicator takes into consideration those that are absolute innovations, or innovations that are such for the market in a broad sense, which also considers relative innovations, or innovations that are such only for the company. There is a positive relationship between the impact of innovation on sales and the number of PhDs. This relationship can be better understood considering that innovation in companies is often achieved by qualified human capital with research doctorates.
- *Foreign Doctorate Students:* is the number of foreign doctorates as a percentage of total doctorates [23]. This is a relationship established with the number of PhD students coming from abroad and the total number of PhD students as the denominator. The indicator is calculated at the country level. The indicator considers the presence of foreign doctorates as an indicator of knowledge efficiency. The growth of foreign doctorates allows countries to increase the supply of researchers and increase the human capital employed in scientific research. Obviously, there is a positive relationship between the value of foreign doctorates and the number of total research doctorates at country level. It should be noted that the

countries that invest most in the training of PhDs also tend to attract many foreign PhDs with funding and scholarships.

- *Venture Capital:* is an indicator that considers the expenses in the rich capital as a percentage of the Gross Domestic Product [24]. It is an indicator built as a relationship. In the numerator there are the Venture Capital expenses or as financial investments in the risk capital of companies in terms of equity [25]. The denominator, on the other hand, is the gross domestic product. Venture capital operations are generally carried out with reference to the support of start-ups. Investment activities carried out by business angels must also be included in this category. The growth of venture capital is an indicator of the dynamism of the financial and entrepreneurial systems in the production of new companies, above all tech-oriented start-ups. In fact, for companies that introduce new products and services that are generally very risky, third-party risk capital is the only tool to finance the start-up of the business. There is a positive relationship between the value of venture capital and the value of PhDs. It follows that in countries where there are much more developed financial markets there is also the possibility of having greater investments that can support the training courses of research doctorates.

The Estimation of New Doctorate Students in Europe												
		Dynamic Panel		**Pooled OLS**		**WLS**		**Fixed Effects**		**Random Effects**		**Average**
	New Doctorate Graduates	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	
	const	-0,681727		0,545328		0,217265		0,485182		0,454612		0,20413
A1	*Attractive Research Systems*	-0,217195	***	-0,118726	***	-0,100736	***	-0,144066	***	-0,142414	***	-0,1446
A4	*Basic-School Entrepreneurial Education and Training*	-0,457651	***	-0,207173	***	-0,236607	***	-0,217731	***	-0,216973	***	-0,2672
A11	*Employment MHT Manufacturing KIS Services*	-0,247201	***	-0,308098	***	-0,267629	***	-0,268902	***	-0,27592	***	-0,2736
A15	*Enterprises Providing ICT Training*	0,161033	***	0,163878	***	0,146395	***	0,155161	***	0,156019	***	0,1565
A18	*Firm Investments*	-0,333236	*	-0,244541	***	-0,203621	***	-0,334343	***	-0,326293	***	-0,2884
A19	*Foreign Doctorate Students*	0,0780362	**	0,0578045	***	0,0755263	***	0,0723461	***	0,0713054	***	0,071
A22	*Government Procurement of Advanced Technology Products*	1,88971	***	2,7748	***	2,43812	***	2,02214	***	2,08052	***	2,24106
A23	*Human Resources*	2,38176	***	2,32023	***	2,26283	***	2,26956	***	2,26966	***	2,30081
A24	*Innovation Index*	-0,945528	***	-1,62125	***	-1,38305	***	-1,06802	***	-1,11208	***	-1,226
A26	*Innovative Sales Share*	0,0802836	**	0,141125	***	0,108812	***	0,100028	***	0,102899	***	0,10663
A32	*Lifelong Learning*	-0,430177	***	-0,469942	***	-0,476951	***	-0,479579	***	-0,476432	***	-0,4666

A35	*Medium and High-Tech Product Exports*	0,342385	***	0,447548	***	0,384598	***	0,371249	***	0,380387	***	0,38523
A36	*Most-Cited Publications*	0,212033	***	0,313291	***	0,281285	***	0,201501	***	0,213282	***	0,24428
A38	*Non-R&D Innovation Expenditure*	0,155542	**	0,165696	***	0,146432	***	0,153267	***	0,153656	***	0,15492
A44	*Product or Process Innovators*	0,2112	***	0,412022	***	0,33529	***	0,266795	***	0,280083	***	0,30108
A45	*Public-Private co-Publications*	0,133941	***	0,20767	***	0,176661	***	0,197455	***	0,197611	***	0,18267
A53	*Tertiary Education*	-0,736037	***	-0,633606	***	-0,632968	***	-0,629736	***	-0,628869	***	-0,6522
A55	*Total Entrepreneurial Activity*	-0,627072	***	-1,47415	***	-1,07341	***	-1,06672	***	-1,09802	***	-1,0679
A59	*Venture Capital*	0,0652905	***	0,105366	***	0,0693387	***	0,0536658	***	0,0568507	***	0,0701
	A37(-1)	0,0602356	***									

We also found that "*New Doctorate Students*" is negatively associated with:

- *Attractive Research Systems:* is an indicator that considers three elements, namely international scientific publications [26], the most cited scientific publications, and the presence of foreign doctorates [27]. There is a negative relationship between the attractiveness of the research system and the presence of PhDs. This relationship turns out to be counterfactual. In fact, one would expect that with the growing attractiveness of the national research system, the number of research doctorates will also increase. However, we must also consider that the value of the coefficient of this variable is equal to -0.1446274 as the average of the various estimated econometric models.
- *Basic-School Entrepreneurial Education and Training:* is a variable that considers entrepreneurial training carried out at school level. The indicator measures the level of training in business management of small and medium-sized enterprises that is present in primary and secondary school. There is a negative relationship between the value of entrepreneurial training present at school level and the presence of PhDs. This relationship may be since countries that offer business culture training at school level have a greater capacity to be employed immediately and therefore have less incentives to study, leading to the acquisition of tertiary qualifications such as that relating to Ph.D.
- *Employment MHT Manufacturing KIS Services:* is the percentage of employment [28] in medium and high-tech companies that use services with a high content of knowledge. There is a negative relationship between the value of employment in medium and high-tech manufacturing companies that use services with a high level of knowledge and the number of doctorates at national level. This relationship appears to be partly counterfactual. In fact, one might think that employment growth in medium and high-tech manufacturing industries is positively connected with the number of research doctorates. However, it is possible that the training of research doctorates is of little use to the production needs of companies operating in manufacturing sectors which are also highly knowledge intensive. The motivation can be found in the fact that PhDs generally have a type of academic training that does not necessarily

meet the productive needs of companies even when they operate in knowledge-intensive sectors.

- *Firm Investments:* is a variable that considers three indicators, namely: investments in research and development, investments in technological innovation not in research and development, investments that companies make for the training of employees in IT skills [29]. There is a negative relationship between the value of firms' investments in technological innovation and the value of research doctorates. This negative relationship may seem counterfactual. However, it must be considered that most of the research doctorates are funded substantially not by the expenditure of companies in technological innovation, but rather by public expenditure on academic-university research. It follows therefore that the fact that companies invest in technological innovation does not necessarily imply an increase in expenditure on university research that can increase the number of research doctorates.
- *Lifelong Learning:* is a variable that considers the percentage of the population aged between 25 and 64 who is committed to lifelong learning [30]. It is a variable that considers training and education that are relevant for the acquisition of new skills that can be used in the world of work. There is a negative relationship between the value of lifelong learning and the value of research doctorates at national level. That is, the fact that there are people involved in lifelong learning does not have an impact on the growth in the number of research doctorates. It follows that although lifelong learning is a desirable fact and brings benefits for innovation and the population, it does not have a positive impact on the growth of PhDs.
- *Tertiary Education:* is a variable that considers the population aged between 25 and 34 with tertiary education [31]. This variable takes into consideration both graduates in scientific disciplines and graduates in the humanities as technological innovation requires the contribution of multidisciplinary knowledge. There is a negative relationship between the value of the population with tertiary education and the number of research doctorates. It follows that the growth of students who have a tertiary education does not necessarily generate a growth in the number of research doctorates. This relationship means that specific investments are needed to ensure that more doctoral pathways are offered for graduates.
- *Total Entrepreneurial Activity:* it is an indicator that considers the percentage of the population aged between 18 and 64 who owns a business or has founded a business. There is a negative relationship between the value of the diffusion of entrepreneurship and the number of PhDs. This relationship can be better clarified considering that generally small and medium-sized enterprises and start-ups are created by people who do not necessarily have educational qualifications. Furthermore, the fact that there are companies does not necessarily guarantee that they have the capacity to generate technological innovation and research and development. In fact, generally the enterprises that carry out technological innovation and research and development are not small and medium-sized enterprises but rather large enterprises [32].
- *Innovation Index:* is the global indicator of technological innovation at the country level [33]. There is a negative relationship between the value of the innovation index and the number of research doctorates. This relationship may appear paradoxical. However, it must be considered that the growth of the innovation index is not only connected to the presence of research doctorates. In fact, to make the innovation index grow, it is also necessary that there are institutions capable of accompanying and encouraging innovation. Furthermore, innovation also requires an evolved entrepreneurial system, of medium and large companies, a value of exports of products and services at an international level. It follows that if a country wants to increase the innovation index it is not enough to invest in research doctorates, it is

also necessary to invest in institutions and create the conditions to create companies that are able to create new products and services based on research and development [34].

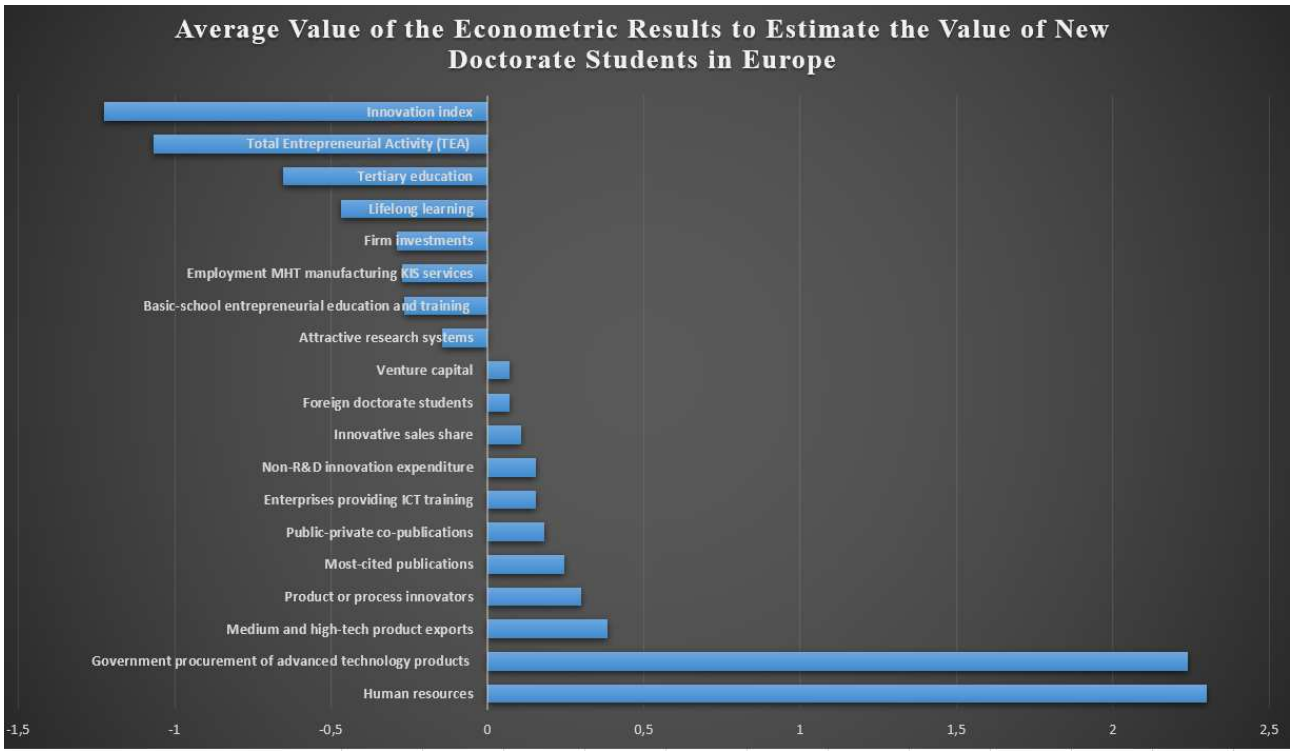


Figure 1. Average Value of the Econometric Results to Estimate the Value of New Doctorate Students in Europe.

4. Clusterization with k-Means Algorithms: a Confrontation between the Elbow Method and the Silhouette Coefficient

A clustering is carried out below using the k-Means algorithm optimized using the Silhouette coefficient. Clustering is carried out to verify if there are groupings among the various European countries in terms of new doctoral students. The following clusters are identified below, namely:

- *Cluster 1*: Iceland, Luxembourg, Spain, Ireland, Slovakia, Estonia, Norway, Austria, Italy, the Netherlands, Israel, Belgium, Portugal, Czech Republic;
- *Cluster 2*: Poland, Cyprus, North Macedonia, Malta, Hungary, Turkey, Bosnia, Montenegro, Latvia, Serbia, Bulgaria, Lithuania, Greece, Croatia, Romania;
- *Cluster 3*: Sweden, Denmark, United Kingdom, Switzerland, Germany, Slovenia, Finland, France.

If we look at the median value of the individual clusters, we can verify the following ordering: C3 = 128.72> C1 = 88.51> C2 = 25.25. The analysis from a geographical point of view shows the dominance of the Scandinavian countries except for Norway and with the addition of Switzerland, Germany, France, and the UK. In second place are the Iberian Peninsula, Italy, Ireland, and the small countries of Central Europe. The countries of Eastern Europe close the ranking and are in the last place for value of "New Doctorate Students".

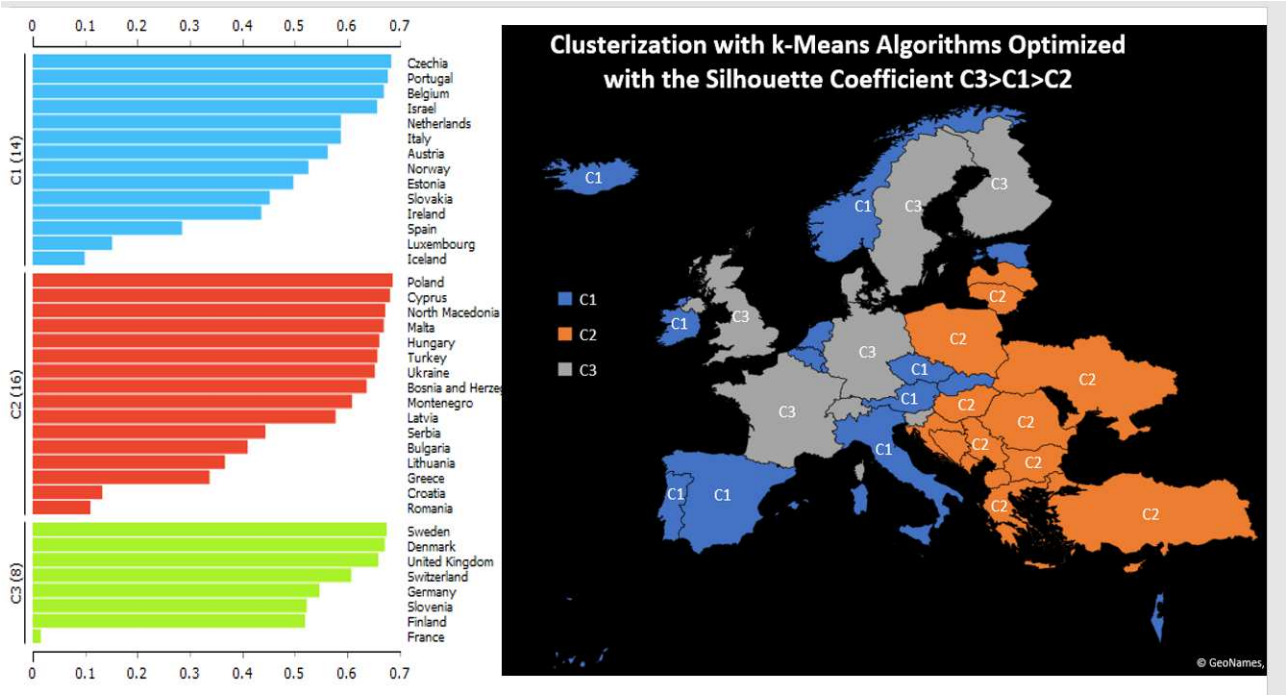


Figure 2. Clusterization with k-Means algorithms optimized with the Silhouette Coefficient.

However, to have a counter-proof of the efficiency of the chosen clustering model, a comparison was also made with the use of the Elbow method. The Elbow method confirms the presence of three clusters. There is therefore a convergence between the Elbow method and the Silhouette coefficient in determining the optimal number of clusters using the k-Means algorithm.

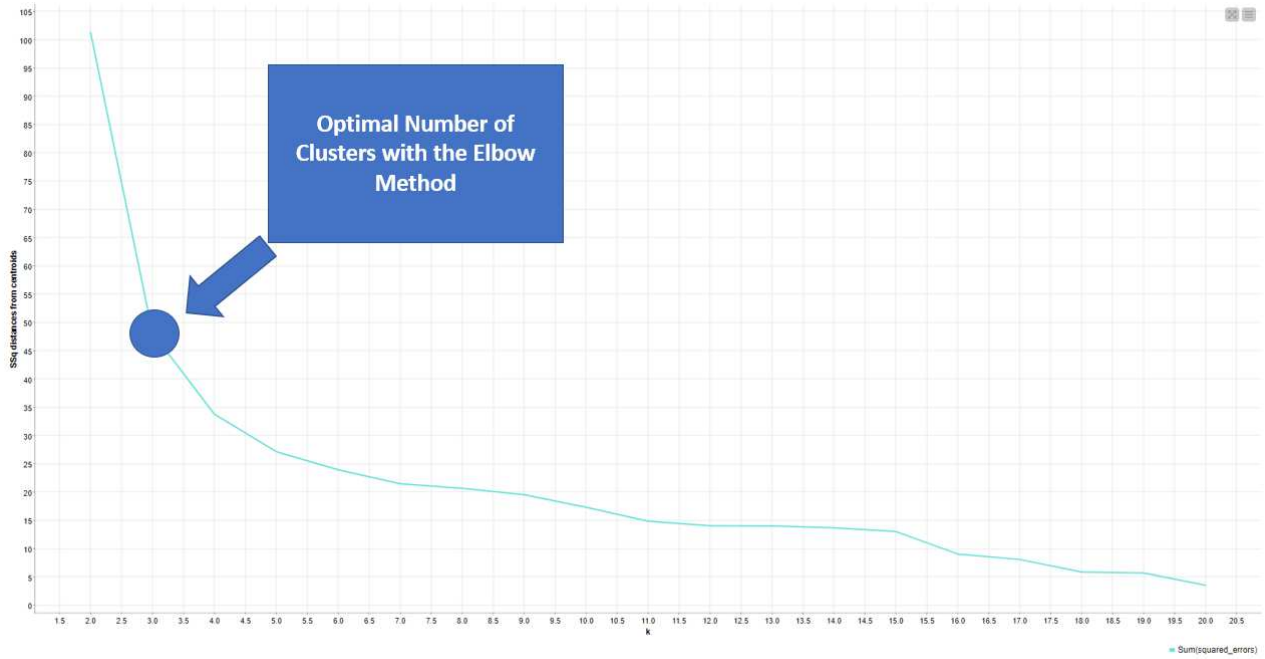


Figure 3. Clusterization with the k-Means optimized with the Elbow Method.

5. Network Analysis with the use of the Manhattan Distance

A network analysis is carried out below with the use of the Manhattan Distance. In this regard, four complex network structures and three simplified network structures are identified. There is a complex network structure between Belgium, Portugal, the Czech Republic, Israel, the Netherlands, Estonia, Italy. Particularly:

- Belgium has a connection with Portugal for a value of 0.23 units;
- Portugal has a connection with Belgium for a value of 0.23 units, with the Czech Republic for a value of 0.33, with Israel with a value of 0.29;
- The Czech Republic has a connection with Portugal for a value of 0.33 and with Israel for a value of 0.34;
- Israel has a connection with Portugal for a value of 0.29, with the Czech Republic for a value of 0.34, with the Netherlands for a value of 0.3 and with Italy for a value of equal to 0.3;
- Italy has a connection with Israel for a value of 0.3, with Estonia for a value of 0.24 and with the Netherlands for a value of 0.00;
- The Netherlands has a connection with Israel for a value of 0.3, with Italy for a value of 0.00 and with Estonia for a value of 0.24;
- Estonia has a connection with the Netherlands for a value of 0.24 and with Italy for a value of 0.24.

There is a complex network structure between North Macedonia, Poland, Cyprus, Malta, Turkey, Bosnia, and Montenegro. Particularly:

- Montenegro has a connection with Bosnia for a value of 0.21;
- Bosnia has a connection with Montenegro for a value equal to 0.21, and with Turkey equal to a value of 0.25;
- Turkey has a connection with Bosnia for a value of 0.25 and with Turkey for a value of 0.24;
- Malta has a connection with Turkey for a value of 0.24, with Cyprus for a value of 0.33, and with Poland for a value of 0.33;
- Cyprus has a connection with Malta for a value of 0.33, and with Poland for a value of 0.23;
- Poland has a connection with Malta for a value of 0.33, with Cyprus for a value of 0.23 and with North Macedonia for a value of 0.29;
- North Macedonia has a connection with Poland for a value of 0.29.

There is a complex network structure between Serbia, Greece and Bulgaria, namely:

- Serbia has a connection with Greece for a value of 0.34;
- Greece has a connection with Serbia for a value of 0.34 and with Bulgaria for a value of 0.29;
- Bulgaria has a connection with Greece for a value of 0.29.

There is a relationship between Bulgaria, Ukraine and Latvia. Particularly:

- Bulgaria has a connection with Ukraine for a value of 0.12 units;
- Ukraine has a connection with Bulgaria for a value of 0.12 units and with Latvia for a value of 0.31 units;

There are simplified network relationships between the following countries:

- Germany has a connection with Finland for a value of 0.17;
- Lithuania has a connection with Croatia for a value of 0.33;
- Austria has a connection with Ireland for a value of 0.29.

As is evident from the analysis, the country with the greatest number of connections is Israel. Furthermore, it is possible to verify that many of the countries that make up cluster 2, which, as demonstrated in the previous paragraph, are the least efficient cluster, are connected in a complex network structure. This condition highlights the fact that the countries of cluster 2 that substantially

coincide with Eastern Europe are closely connected in their evolutionary dynamics of the "New Doctorate Students" variable. The countries of the cluster of excellence, on the other hand, that is the countries of cluster 3, appear to be devoid of interconnections measured on the basis of the distance from Manhattan.

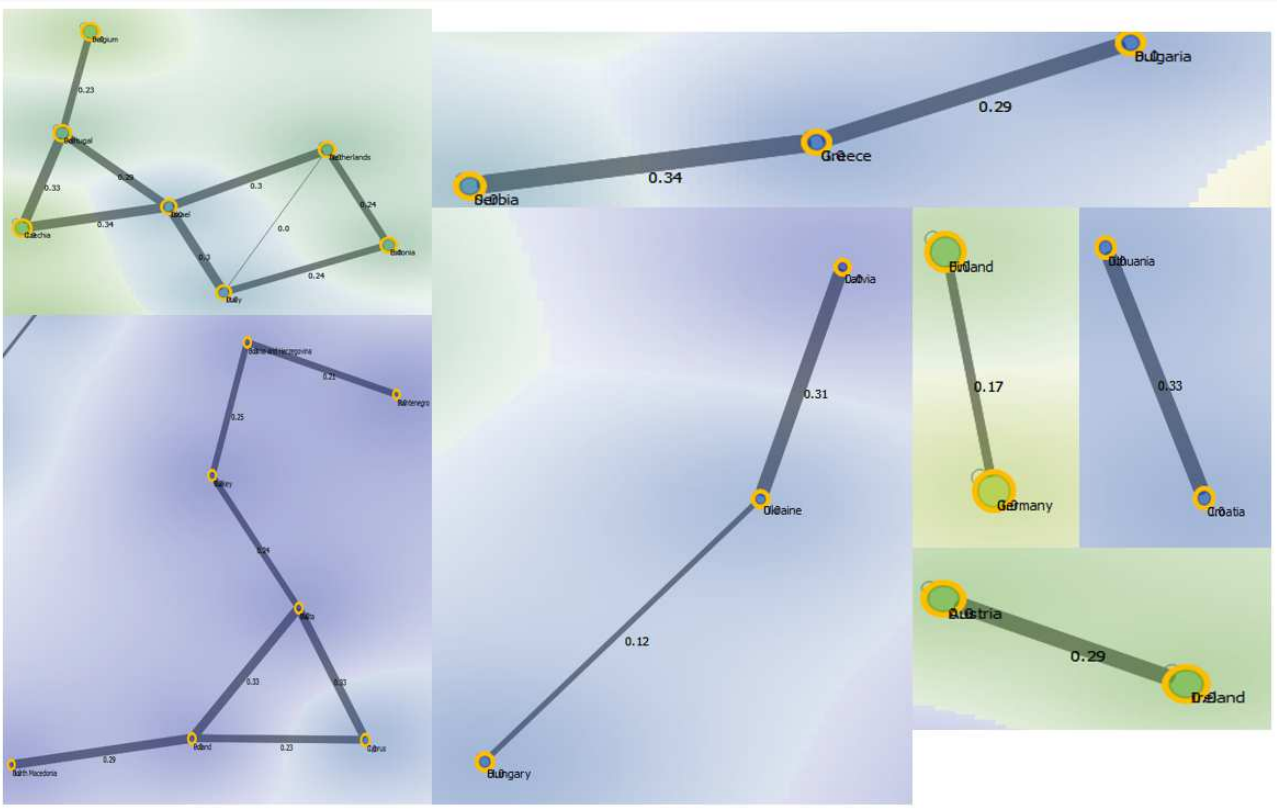


Figure 4. Network analysis structure with the use of the Distance of Manhattan.

6. Machine Learning with Original Data-OD

An analysis is carried out below using a set of algorithms used for prediction through machine learning. There are ten algorithms that are used in a comparison based on the analysis of the MAE, RMAE, MSE, and R-Squared. The algorithms have been trained with the use of 80% of the data that have been used for training while the remainder is made for prediction. The ranking of the algorithms is determined as follows:

- SGD-Stochastic Gradient Descendent in first place with a payoff value of 5
- Linear Regression with a payoff value of 7;
- Gradient Boosting with a payoff value of 13;
- Random Forest with a payoff value of 17;
- Tree and AdaBoost with a payoff value of 21
- kNN with a payoff value of 28;
- SVM-Support Vector Machine with a payoff value of 32;
- Constant with a payoff value of 36;
- Neural Network with a payoff value of 40.

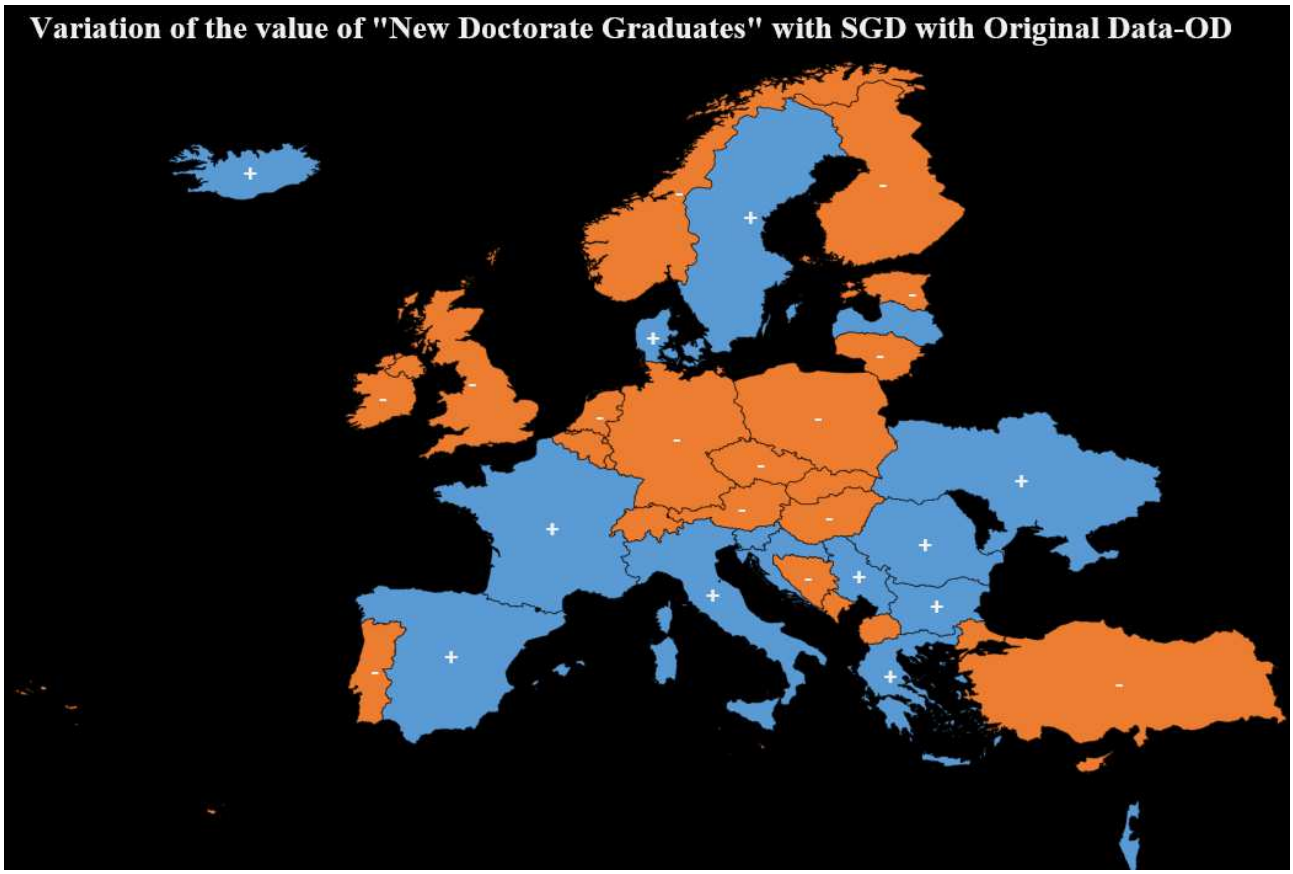


Figure 5. Variation of the value of "New Doctorate Graduates" with SGD with Original Data-OD. The sign + indicates an increase in the value of New Doctorate Graduates, while the sign – represents a negative variation of the investigated variable.

Therefore, the best performing algorithm in predictive terms is the. By applying the SGD-Stochastic Gradient Descendent it is possible to predict the growth in the number of "New Doctorate Graduates" in the following countries, namely:

- Greece with a variation from an amount of 42.56 units up to a value of 62.86 units or equal to a value of 20.31 units equal to an amount of 47.71%;
- Iceland with a variation from an amount of 42.56 units up to a value of 59.33 units or a variation equal to an amount of 16.78 units equal to a value of 39.42%;
- Slovenia with a variation from an amount of 88.51 units up to a value of 122.81 units or equal to a value of 34.3 units equal to a value of 38.75%;
- Latvia with a variation from an amount of 19.58 units up to a value of 24.17 units or equal to a value of 4.59 units equal to an amount of 23.44%;
- Croatia with a variation from an amount of 42.56 units up to a value of 51.15 units or equal to a value of 8.59 units equal to a value of 20.19%;
- Serbia with a variation from an amount of 54.05 units up to a value of 64.61 units or equal to a value of 10.56 units equal to an amount of 19.54%;
- Bulgaria with a variation from an amount of 42.56 units up to a value of 50.05 units or equal to an amount of 7.49 units equal to a value of 17.6%;
- Israel with a variation from an amount of 70.72 units up to a value of 83.07 units or equal to a value of 12.35 units equal to an amount of 17.47%;
- Sweden with a variation from an amount of 122.98 units or equal to an amount of 144.27 units equal to a variation of 21.29 units equal to a value of 17.31%;

- France with a variation from an amount of 100 units up to a value of 114.65 units or equal to a value of 14.65 units equal to a value of 14.65%;
- Italy with a variation from an amount of 65.53 units up to a value of 73.02 units or equal to a variation of 7.49 units equal to an amount of 11.43%;
- Spain with a variation from an amount of 145.95 units up to a value of 157.97 units or equal to a value of 12.01 units equal to an amount of 8.23%;
- Denmark with a variation from an amount of 134.47 units up to a value of 141.67 units or equal to a variation of 7.21 units equal to a variation of 5.36%;
- Romania with a variation from an amount of 19.58 units up to a value of 20.27 units equal to an amount of 0.69 units equivalent to a value of 3.52%;
- Ukraine with a variation from an amount of 25.25 units up to a value of 25.57 units or equal to an amount of 0.32 units equal to a value of 1.25%.

The Stochastic Gradient Descent-SGD algorithm also provides for the reduction of the future trend of the value of "*New Doctorate Graduates*" for the following countries, namely:

- Slovakia with a variation from an amount of 77.02 units up to a value of 75.85 units or equal to a value of -1.18 units equal to a value of -1.53%;
- Poland with a variation from an amount of 19.58 units up to a value of 19.03 units or equal to a variation of -0.54 units equal to a value of -2.78%;
- Finland with a variation from an amount of 122.98 units up to a value of 119.34 units or equal to a value of -3.64 units equal to a value of -2.96%;
- Austria with a variation from an amount of 100 units up to a value of 95.39 units or equal to a value of -4.61 units equal to a value of -4.61%;
- Netherlands with a variation from an amount of 77.02 units up to a value of 73.19 units or equal to a value of -3.83 units equal to a value of -4.97%;
- North Macedonia with a variation from an amount of 8.09 units up to a value of 7.67 units or equal to a value of -0.42 units equal to a value of -5.21%;
- Germany with a variation from an amount of 134.47 units up to a value of 126.83 units or equal to a value of -7.64 units equal to a value of -5.68%;
- Hungary with a variation from an amount of 31.07 units up to a value of 29.14 units or equal to a variation of -1.93 units equal to a value of -6.2%;
- Switzerland with a variation from an amount of 180.42 units up to a value of 168.35 units or equal to an amount of -12.07 units equal to a value of -6.69%;
- Estonia with a variation from an amount of 77.02 units up to a value of 70.89 units or equal to a variation equal to a value of -6.13 units equal to a value of -7.96%;
- Czech Republic with a variation from an amount of 100 up to a value of 91.62 units or equal to a value of -8.38 units equal to an amount of -8.38%;
- Turkey with a variation from an amount of 19.58 units up to a value of 17.93 units or equal to a value of -1.65 units equal to a value of -8.43%;
- Norway with a variation from an amount of 88.51 units up to a value of 80.49 units or equal to a variation of -8.02 units equal to a value of -9.07%;
- Lithuania with a variation from an amount of 42.56 units up to a value of 38.54 units equal to a variation of -4.02 units equal to a value of -9.44%;
- Malta with a variation from an amount of 19.58 units up to a value of 17.29 units equal to an amount of -2.29 units equal to a value of -11.69%;

- Ireland with a variation from an amount of 111.49 units up to a value of 97.7 units or equal to a value of -13.79 units equal to a variation of -12.37%;
- Luxembourg with a variation from an amount of 111.49 units up to a value of 95.87 units or equal to a value of -15.62 units equal to a value of -14.01%;
- Bosnia with a variation from an amount of 12.56 units or equal to a value of 10.63 units or equal to a variation of -1.93 units equal to an amount of -15.38%;
- United Kingdom with a variation from an amount of 180.42 units up to a value of 152.13 units or equal to a variation of -28.3 units equal to an amount of -15.68%;
- Portugal with a variation from an amount of 88.51 units up to a value of 74.57 units or equal to a value of -13.95 units equal to a value of -15.76%;
- Belgium with a variation from an amount of 100 units up to a value of 77.48 units or equal to a variation of -22.52 units equal to a value of -22.52%;
- Cyprus with a variation from an amount of 31.07 units up to a value of 18.89 units or equal to a variation of -12.18 units equal to a value of -39.19%;
- Montenegro with a variation from an amount of 3.69 units up to a value of -0.24 units or equal to a value of -3.92 units equal to a value of -106.48%.

From a geographical point of view, the value of New Doctorate Graduates appears to be increasing in the countries of southern Europe except for Portugal, in most Eastern European countries and in some countries of Northern Europe. However, in most Northern European countries the algorithm predicts a reduction in the value of the "*New Doctorate Graduates*".

Predictions of the Value of "*New Doctorate Graduates*" in Europe with SGD and Original Data									
Country	**2021**	**SGD**	**Abs Var**	**% Var**	***Country***	**2021**	**SGD**	**Abs Var**	**% Var**
Greece	42,56	62,86	20,31	47,71	*Netherlands*	77,02	73,19	-3,83	-4,97
Iceland	42,56	59,33	16,78	39,42	*North Macedonia*	8,09	7,67	-0,42	-5,21
Slovenia	88,51	122,81	34,3	38,75	*Germany*	134,47	126,83	-7,64	-5,68
Latvia	19,58	24,17	4,59	23,44	*Hungary*	31,07	29,14	-1,93	-6,2
Croatia	42,56	51,15	8,59	20,19	*Switzerland*	180,42	168,35	-12,07	-6,69
Serbia	54,05	64,61	10,56	19,54	*Estonia*	77,02	70,89	-6,13	-7,96
Bulgaria	42,56	50,05	7,49	17,6	*Czechia*	100	91,62	-8,38	-8,38
Israel	70,72	83,07	12,35	17,47	*Turkey*	19,58	17,93	-1,65	-8,43
Sweden	122,98	144,27	21,29	17,31	*Norway*	88,51	80,49	-8,02	-9,07
France	100	114,65	14,65	14,65	*Lithuania*	42,56	38,54	-4,02	-9,44
Italy	65,53	73,02	7,49	11,43	*Malta*	19,58	17,29	-2,29	-11,69
Spain	145,95	157,97	12,01	8,23	*Ireland*	111,49	97,7	-13,79	-12,37
Denmark	134,47	141,67	7,21	5,36	*Luxembourg*	111,49	95,87	-15,62	-14,01
Romania	19,58	20,27	0,69	3,52	*Bosnia and Herzegovina*	12,56	10,63	-1,93	-15,38
Ukraine	25,25	25,57	0,32	1,25	*United Kingdom*	180,42	152,13	-28,3	-15,68
Slovakia	77,02	75,85	-1,18	-1,53	*Portugal*	88,51	74,57	-13,95	-15,76
Poland	19,58	19,03	-0,54	-2,78	*Belgium*	100	77,48	-22,52	-22,52
Finland	122,98	119,34	-3,64	-2,96	*Cyprus*	31,07	18,89	-12,18	-39,19
Austria	100	95,39	-4,61	-4,61	*Montenegro*	3,69	-0,24	-3,92	-106,48

7. Machine Learning with Augmented Data-AD

Subsequently, the prediction data obtained using the original data was added to the time series to obtain a further prediction with the augmented data. Also in this case, 80% of the data was used for the training of the algorithm and the remaining 20% was used for the actual prediction. The performance of the algorithms was achieved through the minimization of statistical errors or MAE, RMSE, MSE and through the maximization of R-squared. The following algorithm order was therefore obtained, that is:

- Linear Regression with a payoff value of 4
- SGD-Stochastic Gradient Descendent with a payoff value of 8;
- Gradient Boosting with a payoff value of 12;
- Random Forest with a payoff value of 16;
- AdaBoost with a payoff value of 20;
- kNN with a payoff value of 24;
- Tree with a payoff value of 28;
- SVM with a payoff value of 32;
- Constant with a payoff value of 36;
- Neural Network with a payoff value of 40.

Therefore, the best performing algorithm is Linear Regression. Through the use of the Linear Regression algorithm it is possible to predict an increase in the value of the "New Doctorate Graduates" for the following countries through the application of the increased data, that is:

- Cyprus with a variation from 18.89 up to a value of 21.65 units or a variation equal to an amount of 2.76 units equal to a value of 14.61%;
- Malta with a variation from an amount of 17.29 units up to a value of 19.58 units or an amount equal to 2.29 units equivalent to a value of 13.28%;
- Poland with a variation from an amount of 19.03 units up to a value of 20.42 units or equal to an amount of 1.39 units equal to a value of 7.35%;
- Turkey with a variation from an amount of 17.93 units up to a value of 19.21 units or equal to a value of 1.28 units equal to an amount of 7.18%;
- Romania with a variation from an amount of 20.27 units up to a value of 21.25 units or equal to a value of 0.98 units equal to a value of 4.83%;
- Denmark with a variation from an amount of 141.67 units up to a value of 145.94 units or equal to a value of 4.27 units equal to a value of 3.01%;
- Ukraine with a variation from an amount of 25.57 units up to a value of 26.23 units or equal to a value of 0.66 units equal to an amount of 2.59%;
- Italy with a variation from an amount of 73.02 units up to a value of 74.88 units or equal to a value of 1.86 units equal to a value of 2.55%;
- Estonia with a variation from an amount of 70.89 units up to a value of 72.32 units or equal to a value of 1.43 units equal to an amount of 2.01%;
- Greece with a variation from an amount of 62.86 units up to a value of 64.04 units equal to an amount of 1.18 units equal to a value of 1.88%;
- Belgium with a variation from an amount of 77.48 units up to a value of 78.39 units or equal to a value of 1.45 units equal to a value of 1.87%;
- Israel with a variation from an amount of 83.07 units up to a value of 84.58 units or equal to a value of 1.51 units equal to a value of 1.82%;

- Sweden with a variation from an amount of 144.27 units up to a value of 146.89 units or equal to a value of 2.62 units equal to a value of 1.81%;
- Norway with a variation from an amount of 80.49 units up to a value of 81.93 units or equal to a value of 1.44 units equal to a value of 1.79%;
- Finland with a variation from an amount of 119.34 units up to a value of 121.191 units or equal to a value of 1.851 units equal to a value of 1.55%;
- France with a variation from an amount of 114.65 units up to a value of 116.32 units or equal to a value of 1.67 units equal to a value of 1.45%;
- Switzerland with a variation from an amount of 168.35 units up to a value of 170.73 units or equal to a value of 2.38 units equal to a value of 1.41%;
- Germany with a variation from an amount of 126.83 units up to a value of 127.52 units or equal to a value of 0.69 units equal to an amount of 0.54%;
- Hungary with a variation from an amount of 29.14 units up to a value of 29.28 units or equal to a value of 0.14 units equal to an amount of 0.48%;
- Montenegro with a variation from an amount of -0.24 units up to a value of 2.6 units equal to an amount of 2.85%.

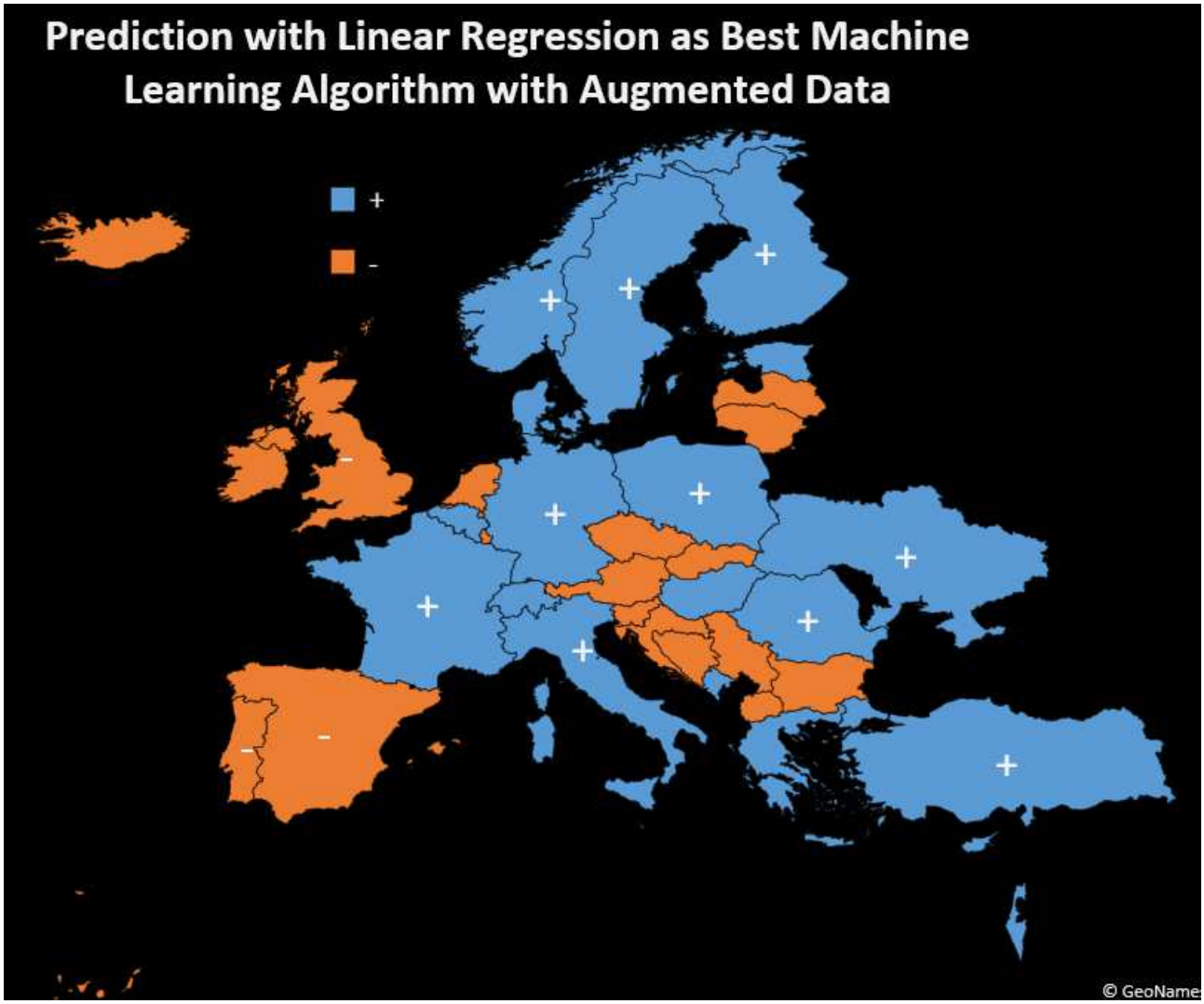


Figure 6. Prediction with Linear Regression as Best Machine Learning Algorithm with Augmented Data.

Through the application of the best performer algorithm or Linear Regression it is possible to predict a reduction in the value of "New Doctorate Graduates" in the following countries, namely:

- United Kingdom with a variation from an amount of 152.13 units up to a value of 151.88 units or equal to a value of -0.24 units equal to a value of -0.16%;
- Portugal with a variation from an amount of 74.57 units up to a value of 74.15 units or equal to a value of -0.41 units equal to a value of -0.55%;
- Luxembourg with a variation from an amount of 95.87 units up to a value of 95.10 units or equal to a value of -0.76 units equal to a value of -0.79%;
- Netherlands with a variation from an amount of 73.19 units up to a value of 72.28 units or equal to a value of 0.90 units equal to a value of -1.23%;
- Iceland with a variation from an amount of 59.33 units up to a value of 58.52 units or equal to a value of -0.8 units equal to a value of -1.35%;
- Ireland with a variation from an amount of 97.7 units up to a value of 96.22 units or equal to a variation of -1.47 units equal to a value of -1.50%;
- Austria with a variation from an amount of 95.39 units up to a value of 93.26 units equal to a value of -2.1 units equal to a value of 2.22%;
- Slovakia with a variation from an amount of 75.85 units up to a value of 74.02 units equal to an amount of 1.82 units or equal to a value of 2.40%;
- Bulgaria with a variation from an amount of 50.05 units up to a value of 48.80 units or equal to a value of -1.24 units equal to a value of 2.48%;
- North Macedonia with a variation from an amount of 7.67 units up to a value of 7.44 units or equal to a value of -0.22 units equal to a value of 2.93%;
- Czech Republic with a variation from an amount of 91.62 units up to a value of 88.73 units or equal to a variation of -2.88 units equal to an amount of -3.14;
- Serbia with a variation from an amount of 64.61 units up to a value of 62.56 units or equal to a value of -2.05 units equal to a value of -3.17%;
- Spain with a variation from an amount of 157.97 units up to a value of 151.48 units or equal to a value of -5.49 units equal to a value of -4.10%;
- Latvia with a variation from an amount of 24.17 units up to a value of 22.79 units or equal to a value of -1.37 units equal to a value of -5.67%;
- Croatia with a variation from an amount of 51.15 units up to a value of 47.95 units or equal to a variation of 3.19 units equal to a value of 6.25%;
- Lithuania with a variation from an amount of 38.54 units up to a value of 34.03 units or equal to a value of -4.5 units equal to an amount of -11.70%;
- Slovenia with a variation from an amount of 122.81 units up to a value of 107.488 units equal to a variation of -15.32 units equal to a value of -12.47%;
- Bosnia with a variation from an amount of 10.63 units up to a value of 9.28 units or equal to a value of -1.34 units equal to a value of -12.67%.

Prediction with Linear Regression as Best Machine Learning Algorithm with Augmented Data									
Country	**SGD**	**Linear Regression**	**Abs Var**	**% Var**	***Country***	**SGD**	**Linear Regression**	**Abs Var**	**% Var**
Montenegro	-0,24	2,62	2,86	1089,00	*Hungary*	29,14	29,28	0,14	0,49
Cyprus	18,89	21,65	2,76	14,61	*United Kingdom*	152,13	151,88	-0,25	-0,16
Malta	17,29	19,59	2,30	13,29	*Portugal*	74,57	74,16	-0,41	-0,55

Poland	19,03	20,43	1,40	7,35	*Luxembourg*	95,87	95,11	-0,76	-0,80
Turkey	17,93	19,22	1,29	7,18	*Netherlands*	73,19	72,28	-0,91	-1,24
Romania	20,27	21,25	0,98	4,84	*Iceland*	59,33	58,53	-0,80	-1,35
Denmark	141,67	145,94	4,27	3,01	*Ireland*	97,70	96,23	-1,47	-1,51
Ukraine	25,57	26,23	0,66	2,60	*Austria*	95,39	93,26	-2,13	-2,23
Italy	73,02	74,88	1,86	2,55	*Slovakia*	75,85	74,03	-1,82	-2,40
Estonia	70,89	72,32	1,43	2,02	*Bulgaria*	50,05	48,80	-1,25	-2,49
Greece	62,86	64,04	1,18	1,88	*North Macedonia*	7,67	7,44	-0,23	-2,94
Belgium	77,48	78,94	1,46	1,88	*Czechia*	91,62	88,74	-2,88	-3,15
Israel	83,07	84,59	1,52	1,83	*Serbia*	64,61	62,56	-2,05	-3,17
Sweden	144,27	146,89	2,62	1,82	*Spain*	157,97	151,48	-6,49	-4,11
Norway	80,49	81,94	1,45	1,80	*Latvia*	24,17	22,80	-1,37	-5,67
Finland	119,34	121,19	1,85	1,55	*Croatia*	51,15	47,95	-3,20	-6,25
France	114,65	116,32	1,67	1,46	*Lithuania*	38,54	34,03	-4,51	-11,70
Switzerland	168,35	170,73	2,38	1,41	*Slovenia*	122,81	107,49	-15,32	-12,48
Germany	126,83	127,52	0,69	0,55	*Bosnia and Herzegovina*	10,63	9,28	-1,35	-12,68

8. Conclusions

This article has investigated the innovational determinants of the variable "*New Doctorate Students*" in Europe. We have used data from the European Innovation Scoreboard-EIS of the European Commission. We performed a series of econometric models to find the statistical relationships among the variables i.e. Pooled OLS, Dynamic Panel, WLS, Panel Data with Fixed Effects, Panel Data with Random Effects. We found that "*New Doctorate Graduates*" is positively associated, among others, with "*Human Resources*" and "*Government Procurement of Advanced Technology Products*" and negatively, associated among others, with "*Total Entrepreneurial Activity*" and "*Innovation Index*". Furthermore, we have applied a clusterization with k-Means algorithm. We find that the optimal number of cluster is three. The more advanced countries in the sense of "*New Doctorate Graduates*" are the Scandinavians, UK, Germany, France, and Switzerland while the Eastern European countries are the worst. The network analysis has showed that while countries of the first and second clusters have share essentially a common framework, the countries in the top cluster i.e. the third, have no network connection at all. Finally, we propose an analysis among ten machine learning algorithms to predict the value of "*New Doctorate Graduates*" either with Original Data-OD either with Augmented Data-AD. We found that the best predictor for OD is SGD-Stochastic Gradient Descendent while the best predictor for AD is Linear Regression. Doctorate programs and Doctorate students are essential to promote the knowledge economy and to give the European Union a competitive chance in the Sino American techwar.

9. Figure Index

10. References

[1] J. Randall, J. A. Rios e H. J. Jung, «A Longitudinal analysis of doctoral graduate supply in the educational measurement field,» *Educational Measurement: Issues and Practice,* vol. 40, n. 1, pp. 59-68, 2021.

[2] L. M. Dos Santos e H. F. Lo, «The Development of Doctoral Degree Curriculum in England: Perspectives from Professional Doctoral Degree Graduates,» *International Journal of Education Policy and Leadership,* vol. 6, n. n6., p. 13, 2018.

[3] H. J. Goodall, V. A. Huggins, L. A. Webber e K. L. Wickett, «From student to graduate: Four learners' perspectives of the professional doctorate journey,» *Management in Education,* vol. 31, n. 4, pp. 180-186, 2017.

[4] L. C. Wickramasinghe e J. G. Borger, «The new age of the PhD: Transforming the PhD from a product to a process,» *Journal of Life Sciences,* vol. 2, n. 1, 2020.

[5] M. Cattaneo, P. Malighetti e S. Paleari, «The Italian brain drain: cream and milk,» *Higher Education,* vol. 77, n. 4, pp. 603-622, 2019.

[6] A. Haapakorpi, «Doctorate holders outside the academy in Finland: academic engagement and industry-specific competence,» *Journal of Education and Work,* vol. 30, n. 1, pp. 53-68, 2017.

[7] J. Subramanian e W. M. Thomson, «The learning environment in professional doctorate and postgraduate dental education: a qualitative study,» *European Journal of Dental Education,* vol. 21, n. 4, pp. 261-271, 2017.

[8] B. Wadham e N. Parkin, «Strange new world: Being a professional and the professional doctorate in the twenty-first century,» *Innovations in Education and Teaching International,* vol. 54, n. 6, pp. 615-624, 2017.

[9] J. Hou, Z. Wang, X. Liu, Y. Luo, S. Sabharwal, N. Wang e Q. Meng, «Public health education at China's higher education institutions: a time-series analysis from 1998 to 2012,» *BMC public health,* vol. 18, n. 1, pp. 1-8, 2018.

[10] A. Zygmunt, «An analysis of innovation framework conditions between Poland and the other Moderate Innovators countries,» *Vision,* pp. 1455-1464, 2020.

[11] F. Valencia-Forrester, «Internships and the PhD: Is This the Future Direction of Work-Integrated Learning in Australia?,» *International Journal of Work-Integrated Learning,* vol. 20, n. 4, pp. 389-400, 2019.

[12] V. Anderson e J. Gold, «The value of the research doctorate: A conceptual examination,» *The International Journal of Management Education,* vol. 3, n. 100305, p. 17, 2019.

[13] A. Leogrande, A. Massaro e A. M. Galiano, «The Determinants of Human Resources in European Countries During the Period 2010-2019,» *American Journal of Humanities and Social Sciences Research (AJHSSR),* vol. 4, n. 9, pp. 145-171, 2020.

[14] A. Leogrande e A. Costantiello, «Human Resources in Europe. Estimation, Clusterization, Machine Learning and Prediction,» *American Journal of Humanities and Social Sciences Research (AJHSSR),,* 2021.

[15] A. Leogrande, A. Costantiello e L. Laureti, «The Export of Medium and High-Tech Products Manufactured in Europe,» *SSRN,* 2022.

[16] A. Costantiello, A. Leogrande e L. Laureti, «The corporate innovation in Europe,» *SSRN ,* n. 3964043, 2021.

[17] A. Leogrande, A. Costantiello e L. Laureti, «k-Means Clusterization and Machine Learning Prediction of European Most Cited Scientific Publications,» *SSRN,* 2022.

[18] A. Costantiello, L. Laureti, A. Leogrande e M. Matarrese, «The Innovation Linkages in Europe,» *SSRN ,* n. 3983218, 2021.

[19] L. Laureti, A. Costantiello, M. Matarrese e A. Leogrande, «Enterprises Providing ICT Training in Europe,» *SSRN,* 2022.

[20] A. Leogrande, N. Magaletti, G. Cosoli, V. Giardinelli e A. Massaro, «ICT Specialists in Europe,» *SSRN,* 2022.

[21] A. Leogrande, N. Magaletti, G. Cosoli, V. Giardinelli e A. Massaro, «The Determinants of Internet User Skills in Europe,» *SSRN ,* n. 4113155, 2022.

[22] A. Costantiello, L. Laureti, G. De Cristoforo e A. Leogrande, «The Innovation-Sales Growth Nexus in Europe,» *SSRN ,* n. 3933407, 2021.

[23] L. Laureti, A. Costantiello, M. Matarrese e A. Leogrande, «Foreign Doctorate Students in Europe,» *SSRN ,* n. 4032975, 2022.

[24] A. Leogrande, A. Costantiello e L. Laureti, «The Impact of Venture Capital Expenditures on Innovation in Europe,» *SSRN ,* n. 3930697, 2021.

[25] L. Laureti, A. Costantiello e A. Leogrande, «The Finance-Innovation Nexus in Europe,» *IJISET - International Journal of Innovative Science, Engineering & Technology,* vol. 7, n. 12, pp. 11-55, 2020.

[26] A. Leogrande, A. Costantiello, L. Laureti e M. Matarrese, «International Scientific Co-Publications in Europe,» *SSRN ,* n. 4117970, 2022.

[27] A. Leogrande, A. Massaro e A. M. Galiano, «The attractiveness of European research systems,» *American Journal of Humanities and Social Sciences Research (AJHSSR),* vol. 4, n. 10, pp. 72-101, 2020.

[28] A. Costantiello e A. Leogrande, «The Innovation-Employment Nexus in Europe,» *,American Journal of Humanities and Social Sciences Research (AJHSSR),* vol. 4, n. 11, pp. 166-187, 2021.

[29] A. Costantiello, L. Laureti e A. Leogrande, «The Determinants of Firm Investments in Research and Development,» *International Virtual Academic Conference Education and Social Sciences Business and Economics,* 2021.

[30] A. Costantiello, L. Laureti e A. Leogrande, « The Determinants of Lifelong Learning in Europe,» *SSRN,* 2022.

[31] A. Leogrande, G. Birardi, A. Massaro e A. M. Galiano, «Italian Universities: Institutional Mandate and Communitarian Engagement,» *European Journal of Educational Management,* vol. 2, n. 2, pp. 85-110, 2019.

[32] A. Costantiello, L. Laureti e A. Leogrande, «The SMEs Innovation in Europe,» *SSRN ,* n. 3964059, 2021.

[33] A. Leogrande, L. Laureti e A. Costantiello, «The Innovation Index in Europe,» *SSRN ,* n. 4091597, 2022.

[34] A. Leogrande, A. Massaro e A. M. Galiano, «The Determinants of Innovation in European Countries in the period 2010-2019,» *American Journal of Humanities and Social Sciences Research (AJHSSR),* vol. 4, n. 8, pp. 91-126, 2020.

11. Appendix

11.1 Econometric Results

		Dynamic Panel		**Pooled OLS**		**WLS**		**Fixed Effects**		**Random Effects**	
		Coefficient	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**	**Coefficient**	**P-value**
	const	-0,681727		0,545328		0,217265		0,485182		0,454612	
A1	*Attractive research systems*	-0,217195	***	-0,118726	***	-0,100736	***	-0,144066	***	-0,142414	***
A4	*Basic-school entrepreneurial education and training*	-0,457651	***	-0,207173	***	-0,236607	***	-0,217731	***	-0,216973	***

A11	*Employment MHT manufacturing KIS services*	-0,247201	***	-0,308098	***	-0,267629	***	-0,268902	***	-0,27592	***
A15	*Enterprises providing ICT training*	0,161033	***	0,163878	***	0,146395	***	0,155161	***	0,156019	***
A18	*Firm investments*	-0,333236	*	-0,244541	***	-0,203621	***	-0,334343	***	-0,326293	***
A19	*Foreign doctorate students*	0,0780362	**	0,0578045	***	0,0755263	***	0,0723461	***	0,0713054	***
A22	*Government procurement of advanced technology products*	1,88971	***	2,7748	***	2,43812	***	2,02214	***	2,08052	***
A23	*Human resources*	2,38176	***	2,32023	***	2,26283	***	2,26956	***	2,26966	***
A24	*Innovation index*	-0,945528	***	-1,62125	***	-1,38305	***	-1,06802	***	-1,11208	***
A26	*Innovative sales share*	0,0802836	**	0,141125	***	0,108812	***	0,100028	***	0,102899	***
A32	*Lifelong learning*	-0,430177	***	-0,469942	***	-0,476951	***	-0,479579	***	-0,476432	***
A35	*Medium and high-tech product exports*	0,342385	***	0,447548	***	0,384598	***	0,371249	***	0,380387	***
A36	*Most-cited publications*	0,212033	***	0,313291	***	0,281285	***	0,201501	***	0,213282	***
A38	*Non-R&D innovation expenditure*	0,155542	**	0,165696	***	0,146432	***	0,153267	***	0,153656	***
A44	*Product or process innovators*	0,2112	***	0,412022	***	0,33529	***	0,266795	***	0,280083	***
A45	*Public-private co-publications*	0,133941	***	0,20767	***	0,176661	***	0,197455	***	0,197611	***
A53	*Tertiary education*	-0,736037	***	-0,633606	***	-0,632968	***	-0,629736	***	-0,628869	***
A55	*Total Entrepreneurial Activity (TEA)*	-0,627072	***	-1,47415	***	-1,07341	***	-1,06672	***	-1,09802	***
A59	*Venture capital*	0,0652905	***	0,105366	***	0,0693387	***	0,0536658	***	0,0568507	***
	A37(-1)	0,0602356	***								

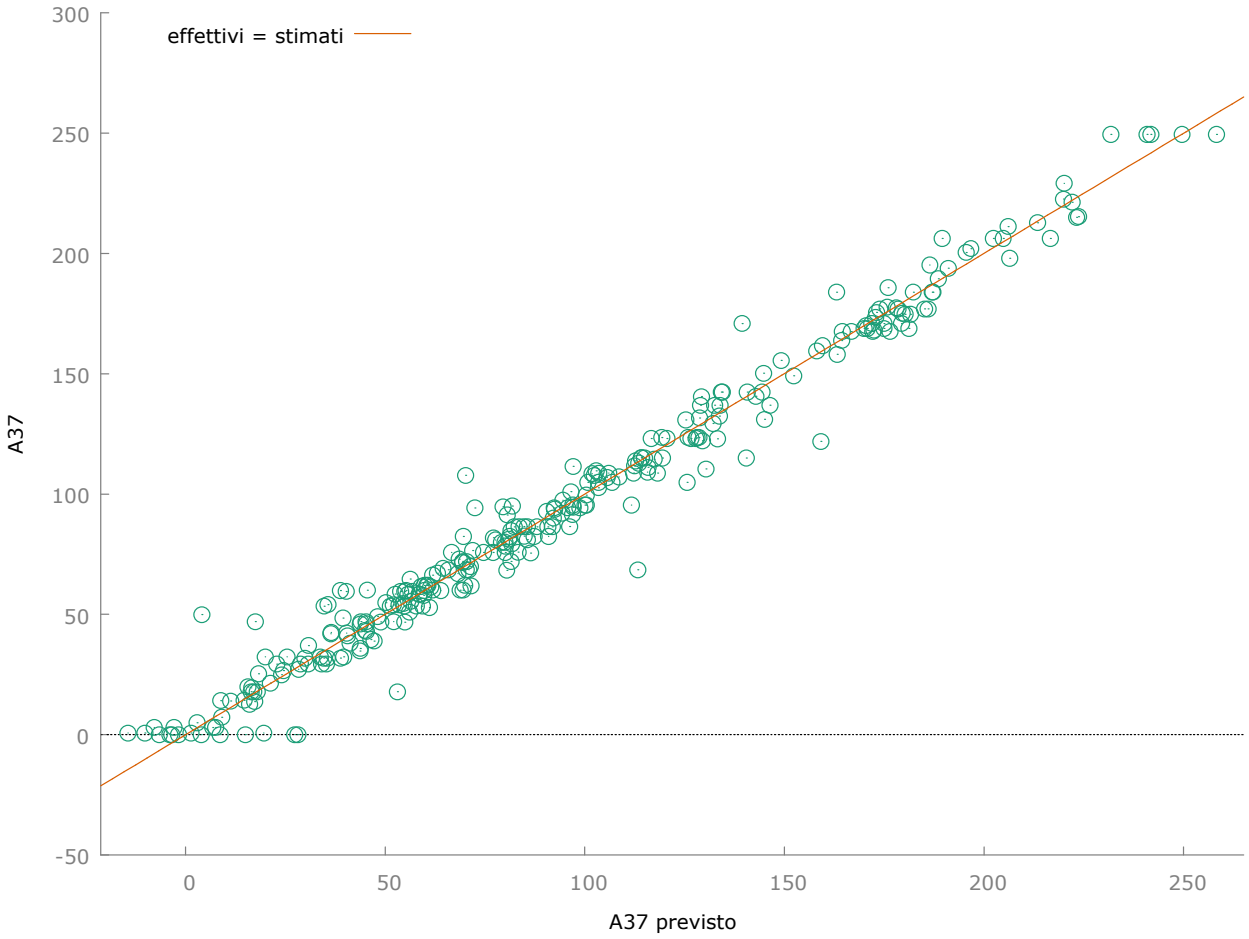
Modello 47: Panel dinamico a un passo, usando 286 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche: minimo 7, massimo 8
Matrice H conforme ad Ox/DPD
Variabile dipendente: A37

	Coefficiente	*Errore Std.*	*z*	*p-value*	
A37(-1)	0,0602356	0,0232773	2,588	0,0097	***
const	−0,681727	0,540321	−1,262	0,2071	
A1	−0,217195	0,0706545	−3,074	0,0021	***
A4	−0,457651	0,0869745	−5,262	<0,0001	***
A11	−0,247201	0,0849785	−2,909	0,0036	***
A15	0,161033	0,0492832	3,267	0,0011	***

A18	−0,333236	0,180519	−1,846	0,0649	*
A19	0,0780362	0,0372704	2,094	0,0363	**
A22	1,88971	0,502125	3,763	0,0002	***
A23	2,38176	0,190952	12,47	<0,0001	***
A24	−0,945528	0,339914	−2,782	0,0054	***
A26	0,0802836	0,0367255	2,186	0,0288	**
A32	−0,430177	0,0815905	−5,272	<0,0001	***
A35	0,342385	0,0798837	4,286	<0,0001	***
A36	0,212033	0,0762522	2,781	0,0054	***
A38	0,155542	0,0672644	2,312	0,0208	**
A44	0,211200	0,0628070	3,363	0,0008	***
A45	0,133941	0,0513295	2,609	0,0091	***
A53	−0,736037	0,0820703	−8,968	<0,0001	***
A55	−0,627072	0,196921	−3,184	0,0015	***
A59	0,0652905	0,0211419	3,088	0,0020	***

Somma quadr. residui 24849,58 E.S. della regressione 9,683593

Numero di strumenti = 34
Test per errori AR(1): z = -1,77932 [0,0752]
Test per errori AR(2): z = -0,379163 [0,7046]
Test di sovra-identificazione di Sargan: Chi-quadro(13) = 12,5507 [0,4831]
Test (congiunto) di Wald: Chi-quadro(20) = 30630,9 [0,0000]

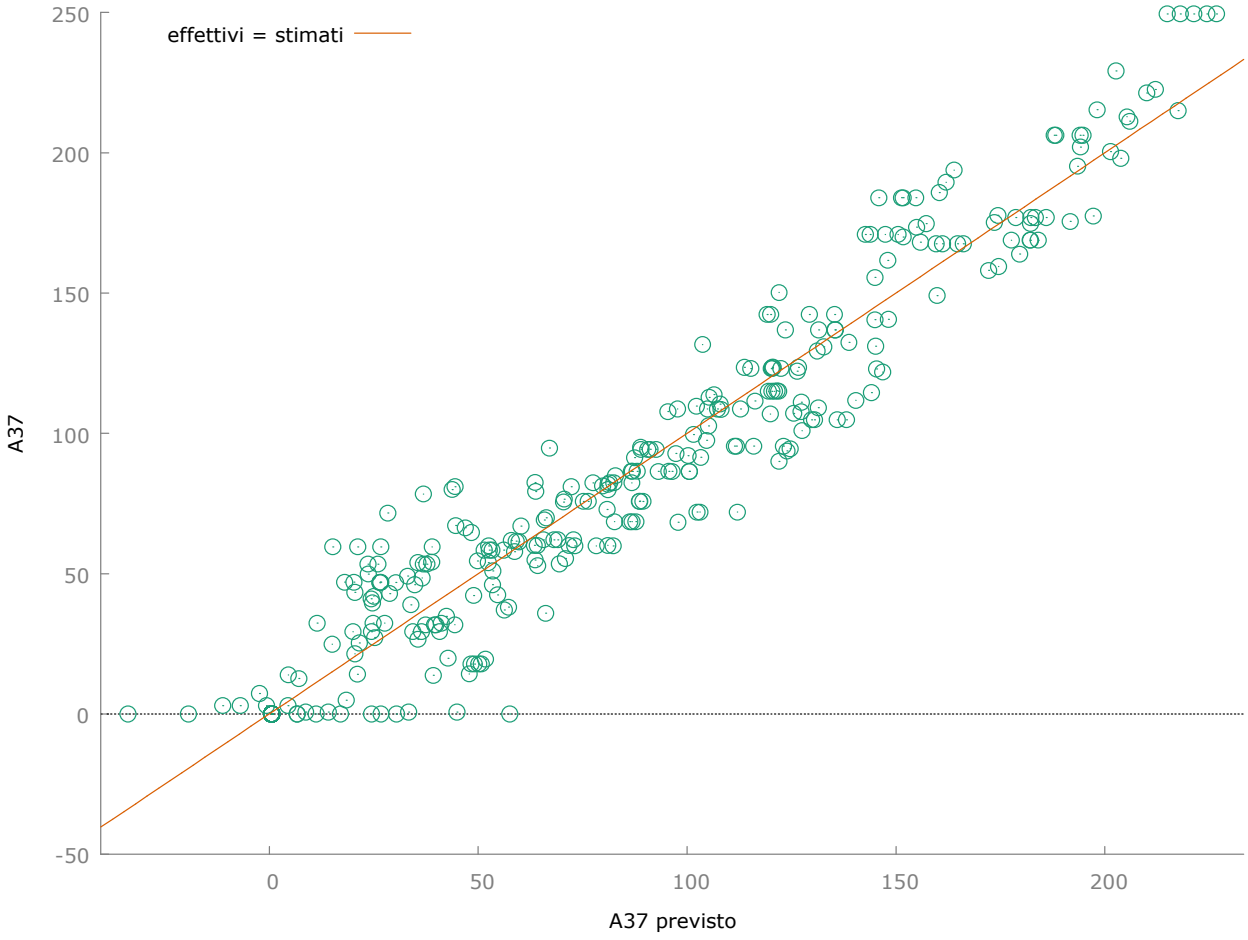


Modello 48: Pooled OLS, usando 357 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche: minimo 8, massimo 10
Variabile dipendente: A37

	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	0,545328	1,83514	0,2972	0,7665	
A1	−0,118726	0,0393996	−3,013	0,0028	***
A4	−0,207173	0,0402967	−5,141	<0,0001	***
A11	−0,308098	0,0319241	−9,651	<0,0001	***
A15	0,163878	0,0178278	9,192	<0,0001	***
A18	−0,244541	0,0438698	−5,574	<0,0001	***
A19	0,0578045	0,0213078	2,713	0,0070	***
A22	2,77480	0,230607	12,03	<0,0001	***
A23	2,32023	0,0733368	31,64	<0,0001	***
A24	−1,62125	0,165945	−9,770	<0,0001	***
A26	0,141125	0,0312290	4,519	<0,0001	***
A32	−0,469942	0,0335410	−14,01	<0,0001	***
A35	0,447548	0,0370249	12,09	<0,0001	***
A36	0,313291	0,0627752	4,991	<0,0001	***
A38	0,165696	0,0197236	8,401	<0,0001	***

A44	0,412022	0,0503281	8,187	<0,0001	***
A45	0,207670	0,0234553	8,854	<0,0001	***
A53	−0,633606	0,0280605	−22,58	<0,0001	***
A55	−1,47415	0,231128	−6,378	<0,0001	***
A59	0,105366	0,0223072	4,723	<0,0001	***

Media var. dipendente	75,12776	SQM var. dipendente	65,70561
Somma quadr. residui	91762,63	E.S. della regressione	16,50129
R-quadro	0,940295	R-quadro corretto	0,936929
F(19, 337)	279,3376	P-value(F)	3,0e-193
Log-verosimiglianza	−1497,098	Criterio di Akaike	3034,195
Criterio di Schwarz	3111,750	Hannan-Quinn	3065,042
rho	0,876570	Durbin-Watson	0,343931



Modello 49: WLS, usando 357 osservazioni
Incluse 36 unità cross section

Variabile dipendente: A37

Pesi basati sulle varianze degli errori per unità

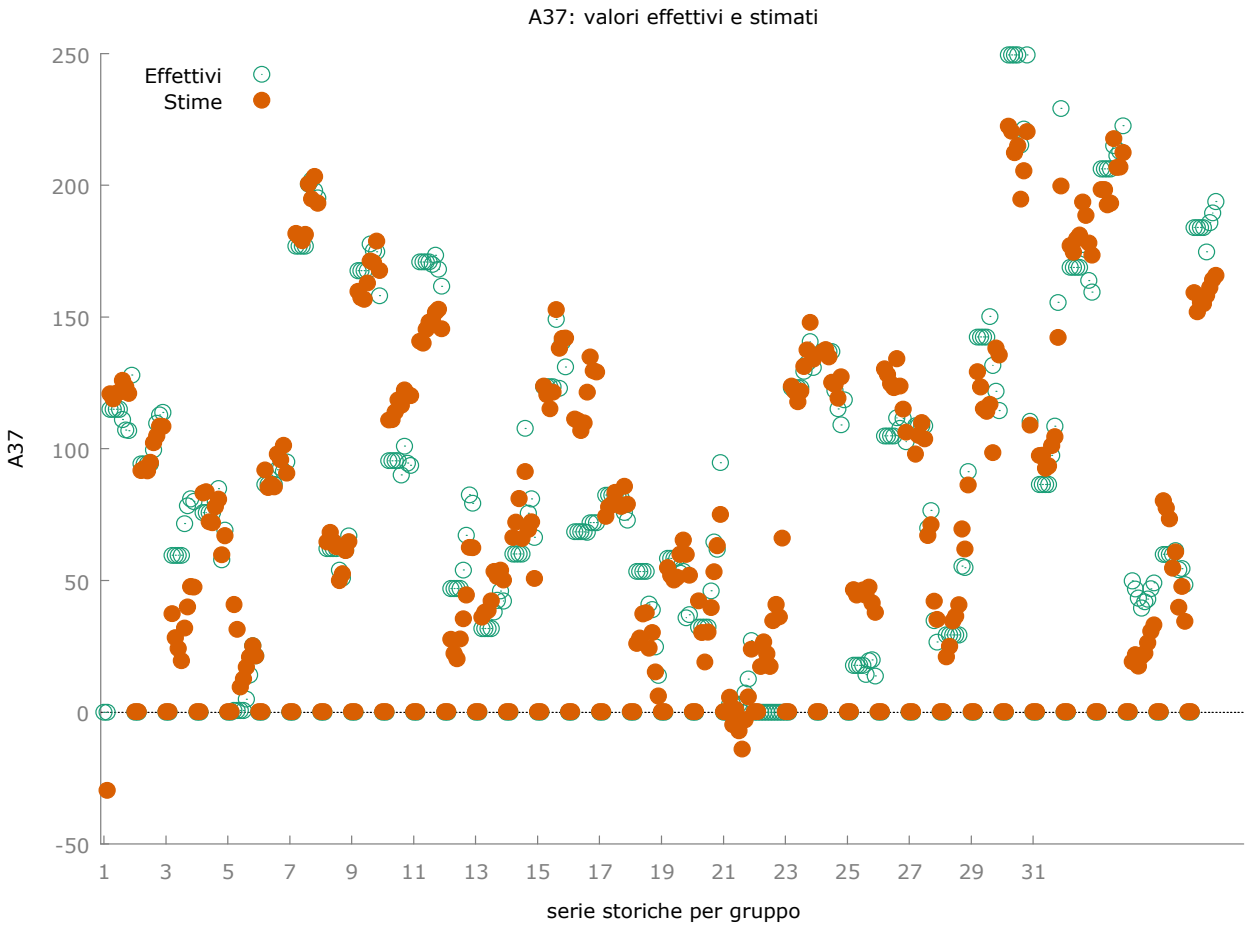
	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	0,217265	0,977003	0,2224	0,8242	
A1	−0,100736	0,0254414	−3,960	<0,0001	***
A4	−0,236607	0,0305183	−7,753	<0,0001	***
A11	−0,267629	0,0223745	−11,96	<0,0001	***
A15	0,146395	0,0124072	11,80	<0,0001	***
A18	−0,203621	0,0316127	−6,441	<0,0001	***
A19	0,0755263	0,0155333	4,862	<0,0001	***
A22	2,43812	0,167210	14,58	<0,0001	***
A23	2,26283	0,0555004	40,77	<0,0001	***
A24	−1,38305	0,123087	−11,24	<0,0001	***
A26	0,108812	0,0199584	5,452	<0,0001	***
A32	−0,476951	0,0205070	−23,26	<0,0001	***
A35	0,384598	0,0260395	14,77	<0,0001	***
A36	0,281285	0,0384143	7,322	<0,0001	***
A38	0,146432	0,0137182	10,67	<0,0001	***
A44	0,335290	0,0303953	11,03	<0,0001	***
A45	0,176661	0,0167147	10,57	<0,0001	***
A53	−0,632968	0,0219216	−28,87	<0,0001	***
A55	−1,07341	0,115167	−9,320	<0,0001	***
A59	0,0693387	0,0129610	5,350	<0,0001	***

Statistiche basate sui dati ponderati:

Somma quadr. residui	320,6421	E.S. della regressione	0,975428
R-quadro	0,981497	R-quadro corretto	0,980454
F(19, 337)	940,8703	P-value(F)	8,2e-279
Log-verosimiglianza	−487,3883	Criterio di Akaike	1014,777
Criterio di Schwarz	1092,331	Hannan-Quinn	1045,624

Statistiche basate sui dati originali:

Media var. dipendente	75,12776	SQM var. dipendente	65,70561
Somma quadr. residui	100037,7	E.S. della regressione	17,22927



Modello 50: Effetti fissi, usando 357 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche: minimo 8, massimo 10
Variabile dipendente: A37

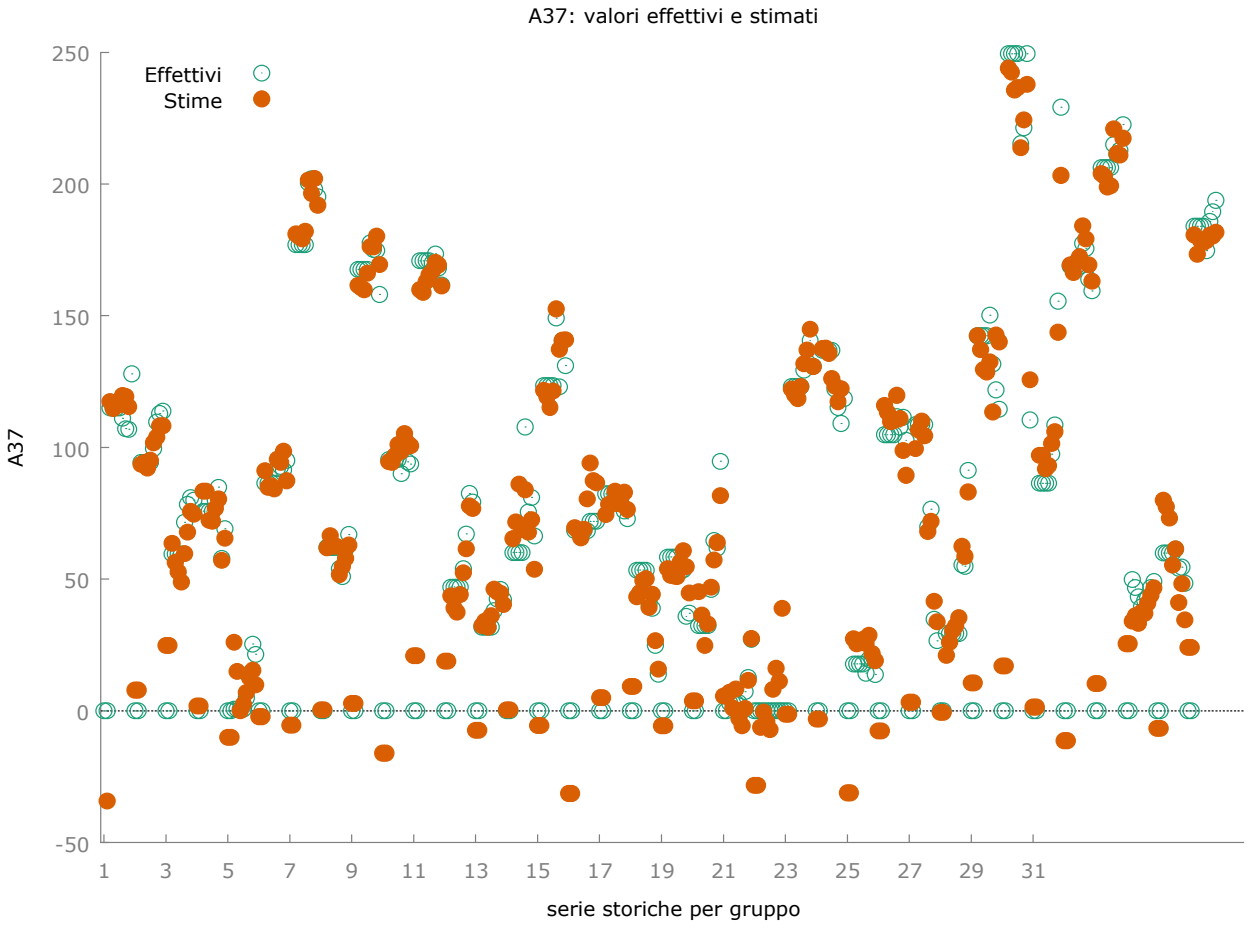
	Coefficiente	*Errore Std.*	*rapporto t*	*p-value*	
const	0,485182	1,24066	0,3911	0,6960	
A1	−0,144066	0,0516311	−2,790	0,0056	***
A4	−0,217731	0,0517276	−4,209	<0,0001	***
A11	−0,268902	0,0366204	−7,343	<0,0001	***
A15	0,155161	0,0221558	7,003	<0,0001	***
A18	−0,334343	0,0529291	−6,317	<0,0001	***
A19	0,0723461	0,0255455	2,832	0,0049	***
A22	2,02214	0,213739	9,461	<0,0001	***
A23	2,26956	0,0780304	29,09	<0,0001	***
A24	−1,06802	0,154680	−6,905	<0,0001	***
A26	0,100028	0,0274720	3,641	0,0003	***
A32	−0,479579	0,0380194	−12,61	<0,0001	***
A35	0,371249	0,0408351	9,091	<0,0001	***
A36	0,201501	0,0652482	3,088	0,0022	***
A38	0,153267	0,0221076	6,933	<0,0001	***

A44	0,266795	0,0455925	5,852	<0,0001	***
A45	0,197455	0,0274152	7,202	<0,0001	***
A53	−0,629736	0,0340795	−18,48	<0,0001	***
A55	−1,06672	0,196368	−5,432	<0,0001	***
A59	0,0536658	0,0180249	2,977	0,0031	***

Media var. dipendente	75,12776	SQM var. dipendente	65,70561
Somma quadr. residui	34496,02	E.S. della regressione	10,68762
R-quadro LSDV	0,977555	R-quadro intra-gruppi	0,954505
LSDV F(54, 302)	243,5793	P-value(F)	5,8e-219
Log-verosimiglianza	−1322,460	Criterio di Akaike	2754,920
Criterio di Schwarz	2968,196	Hannan-Quinn	2839,749
rho	0,491180	Durbin-Watson	0,776713

Test congiunto sui regressori -
Statistica test: F(19, 302) = 333,479
con p-value = P(F(19, 302) > 333,479) = 5,50002e-190

Test per la differenza delle intercette di gruppo -
Ipotesi nulla: i gruppi hanno un'intercetta comune
Statistica test: F(35, 302) = 14,3242
con p-value = P(F(35, 302) > 14,3242) = 8,16843e-046



Modello 51: Effetti casuali (GLS), usando 357 osservazioni
Incluse 36 unità cross section
Lunghezza serie storiche: minimo 8, massimo 10
Variabile dipendente: A37

	Coefficiente	*Errore Std.*	*z*	*p-value*	
const	0,454612	3,06076	0,1485	0,8819	
A1	−0,142414	0,0477839	−2,980	0,0029	***
A4	−0,216973	0,0480115	−4,519	<0,0001	***
A11	−0,275920	0,0344124	−8,018	<0,0001	***
A15	0,156019	0,0206042	7,572	<0,0001	***
A18	−0,326293	0,0493582	−6,611	<0,0001	***
A19	0,0713054	0,0238775	2,986	0,0028	***
A22	2,08052	0,204931	10,15	<0,0001	***
A23	2,26966	0,0738368	30,74	<0,0001	***
A24	−1,11208	0,148132	−7,507	<0,0001	***
A26	0,102899	0,0264453	3,891	<0,0001	***
A32	−0,476432	0,0357000	−13,35	<0,0001	***
A35	0,380387	0,0384211	9,900	<0,0001	***
A36	0,213282	0,0617914	3,452	0,0006	***
A38	0,153656	0,0208113	7,383	<0,0001	***
A44	0,280083	0,0437700	6,399	<0,0001	***
A45	0,197611	0,0255492	7,735	<0,0001	***
A53	−0,628869	0,0318464	−19,75	<0,0001	***
A55	−1,09802	0,189471	−5,795	<0,0001	***
A59	0,0568507	0,0174394	3,260	0,0011	***

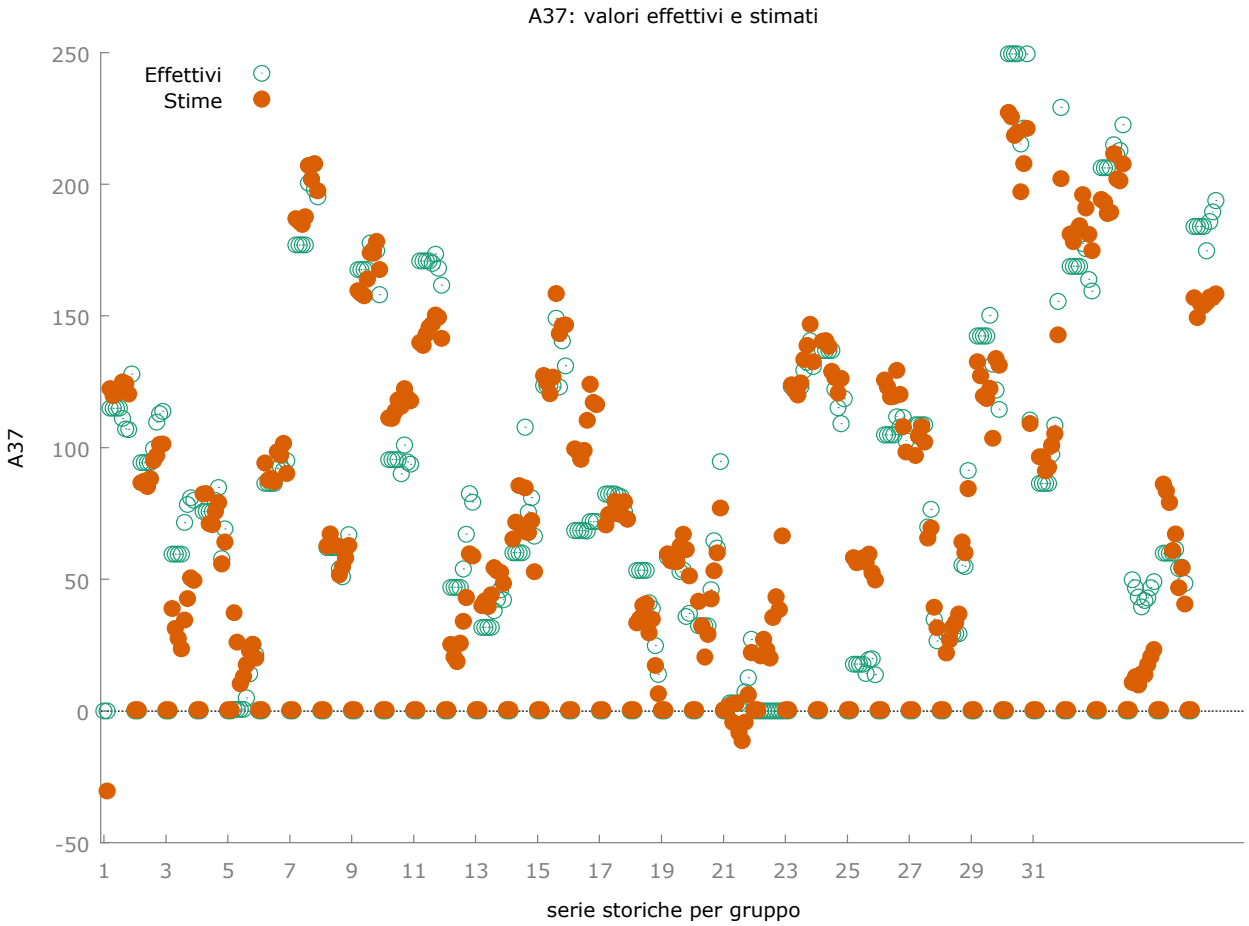
Media var. dipendente	75,12776	SQM var. dipendente	65,70561
Somma quadr. residui	101110,4	E.S. della regressione	17,29576
Log-verosimiglianza	−1514,414	Criterio di Akaike	3068,827
Criterio di Schwarz	3146,382	Hannan-Quinn	3099,674
rho	0,491180	Durbin-Watson	0,776713

Varianza 'between' = 295,461
Varianza 'within' = 114,225
theta medio = 0,806189

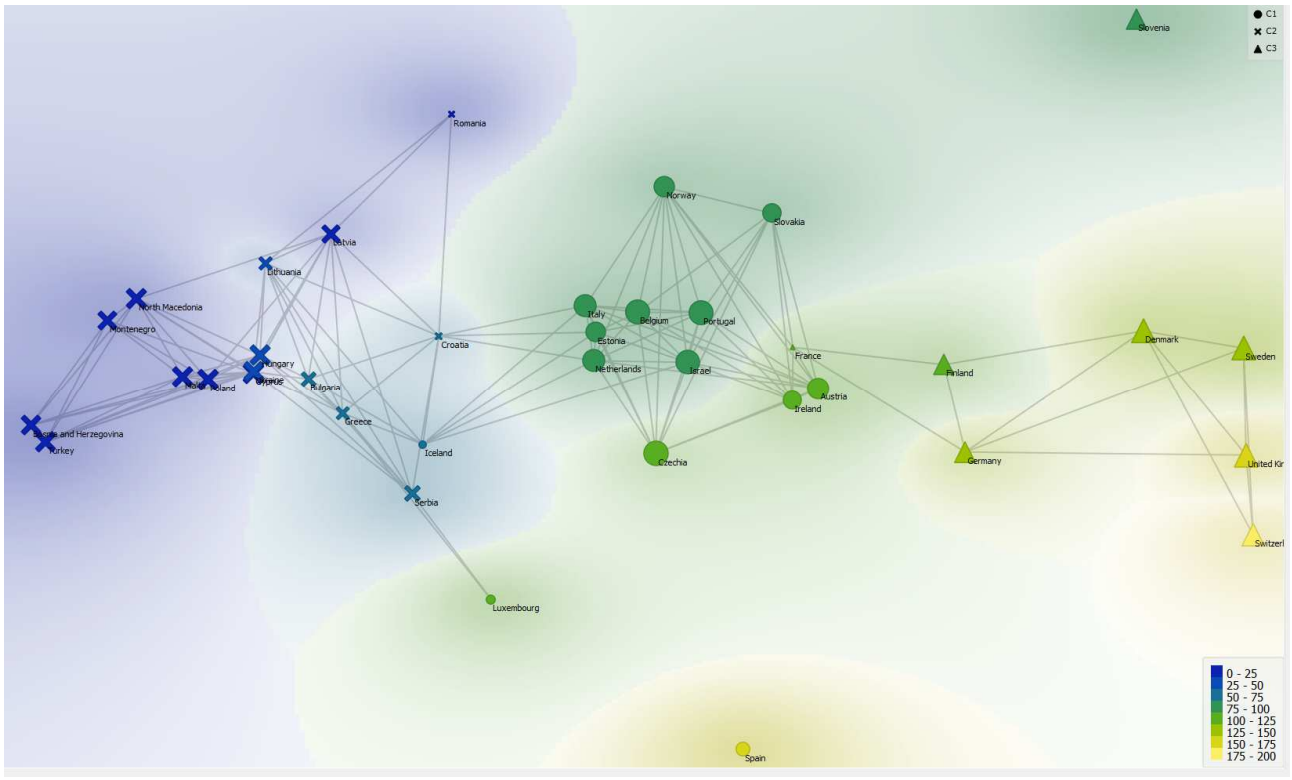
Test congiunto sui regressori -
Statistica test asintotica: Chi-quadro(19) = 6829,8
con p-value = 0

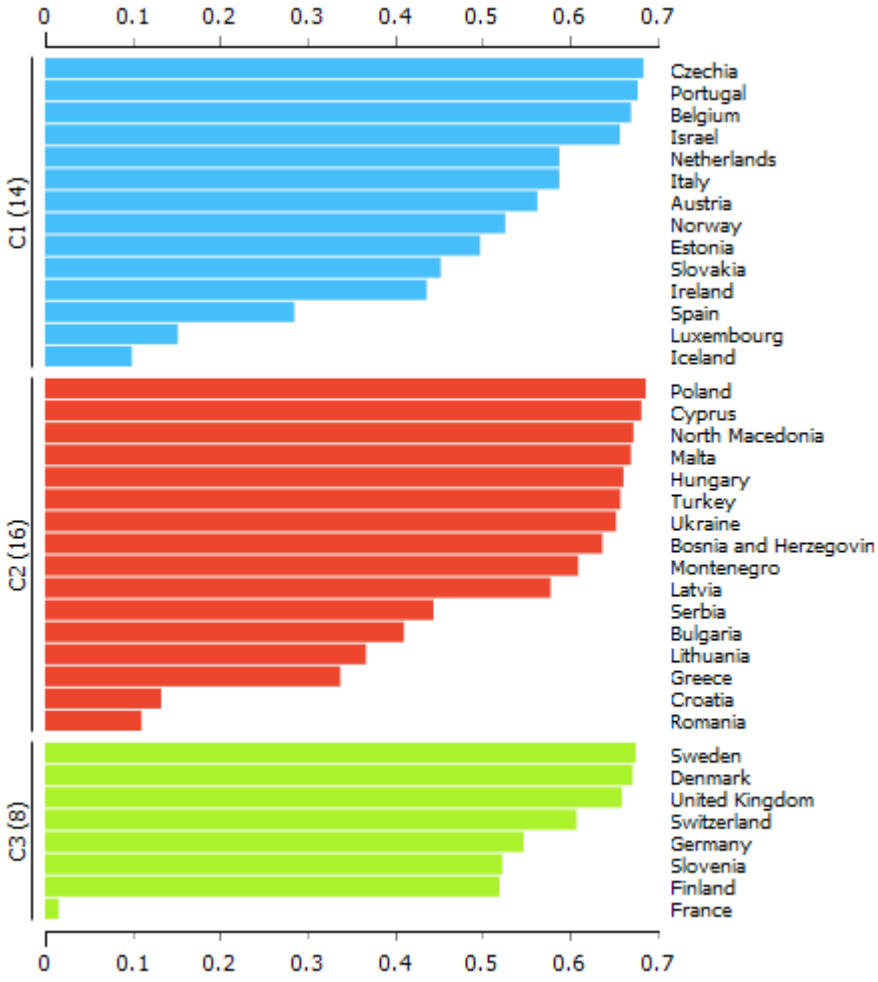
Test Breusch-Pagan -
Ipotesi nulla: varianza dell'errore specifico all'unità = 0
Statistica test asintotica: Chi-quadro(1) = 457,777
con p-value = 1,46469e-101

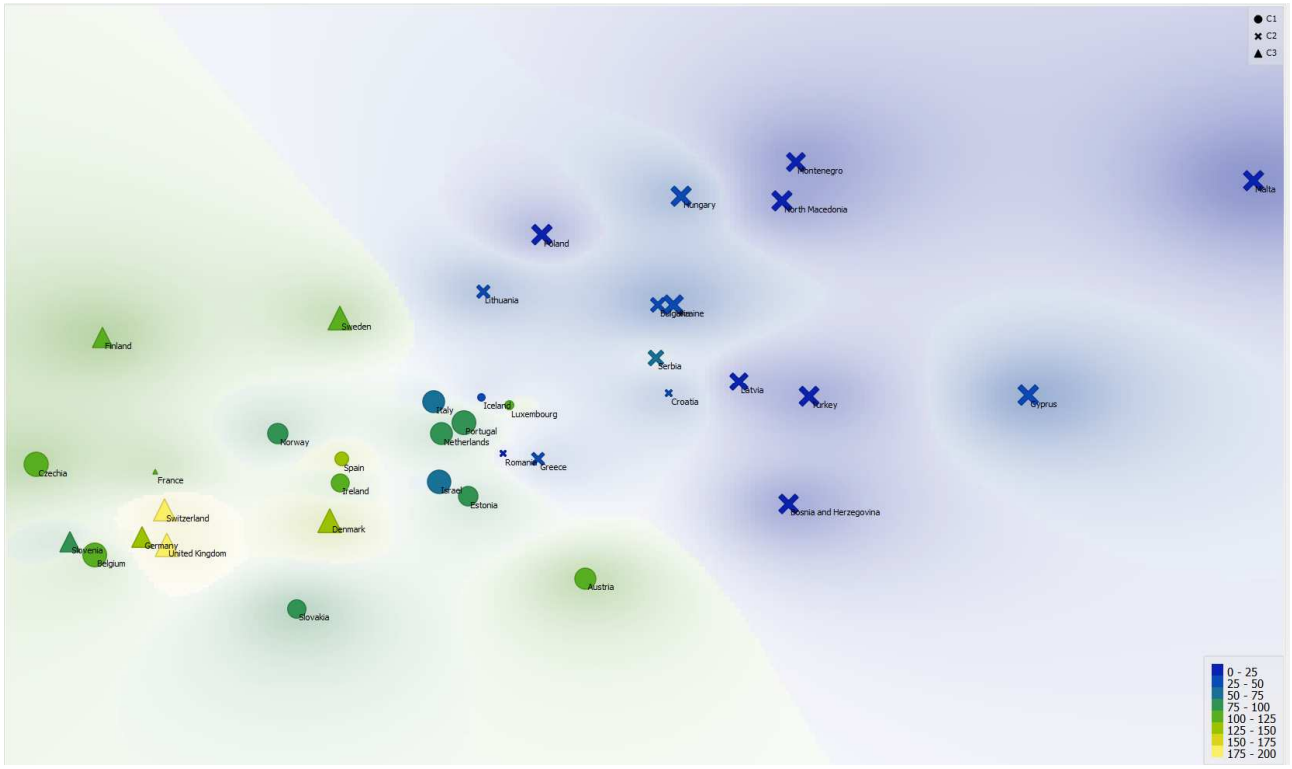
Test di Hausman -
Ipotesi nulla: le stime GLS sono consistenti
Statistica test asintotica: Chi-quadro(19) = 6,83531
con p-value = 0,995041

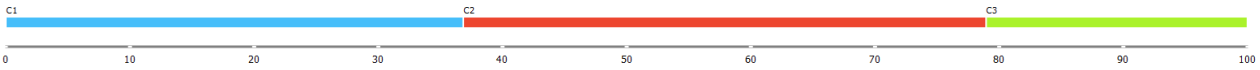


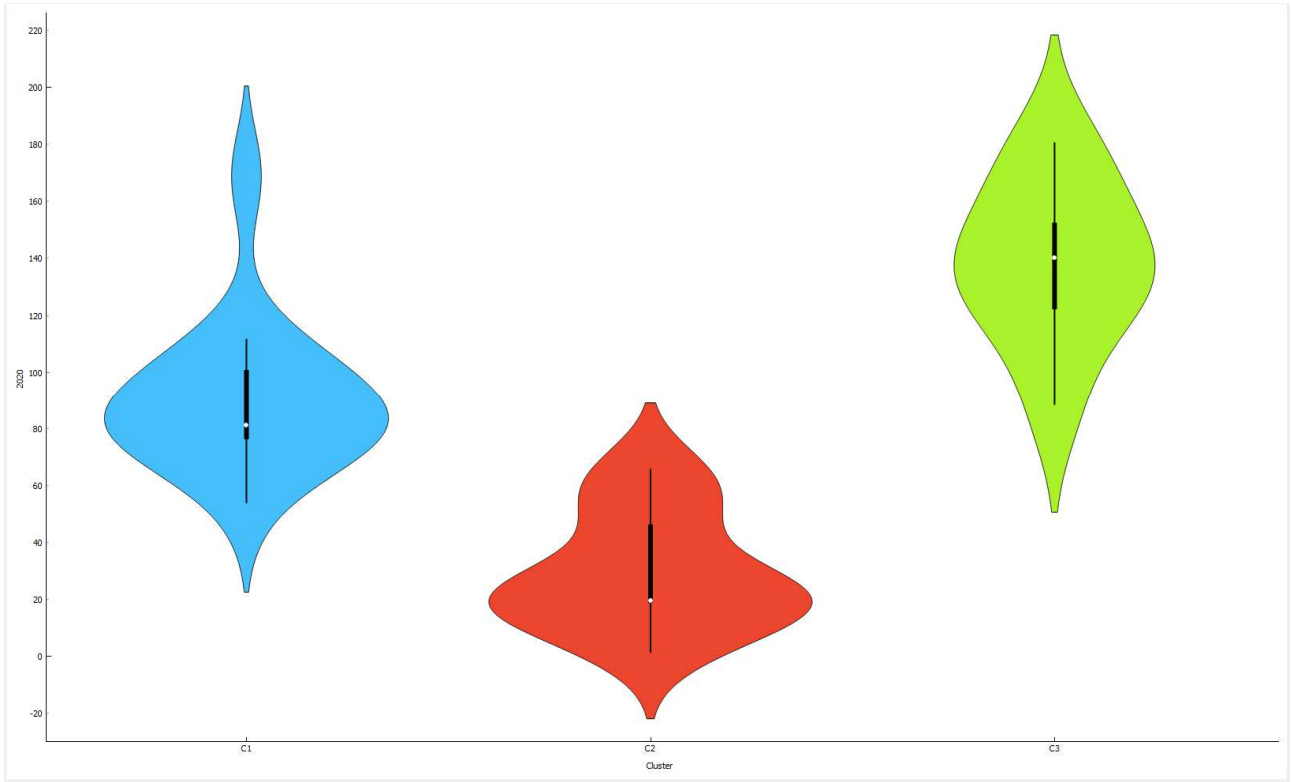
11.2 Clusterization with the k-Means Algorithm

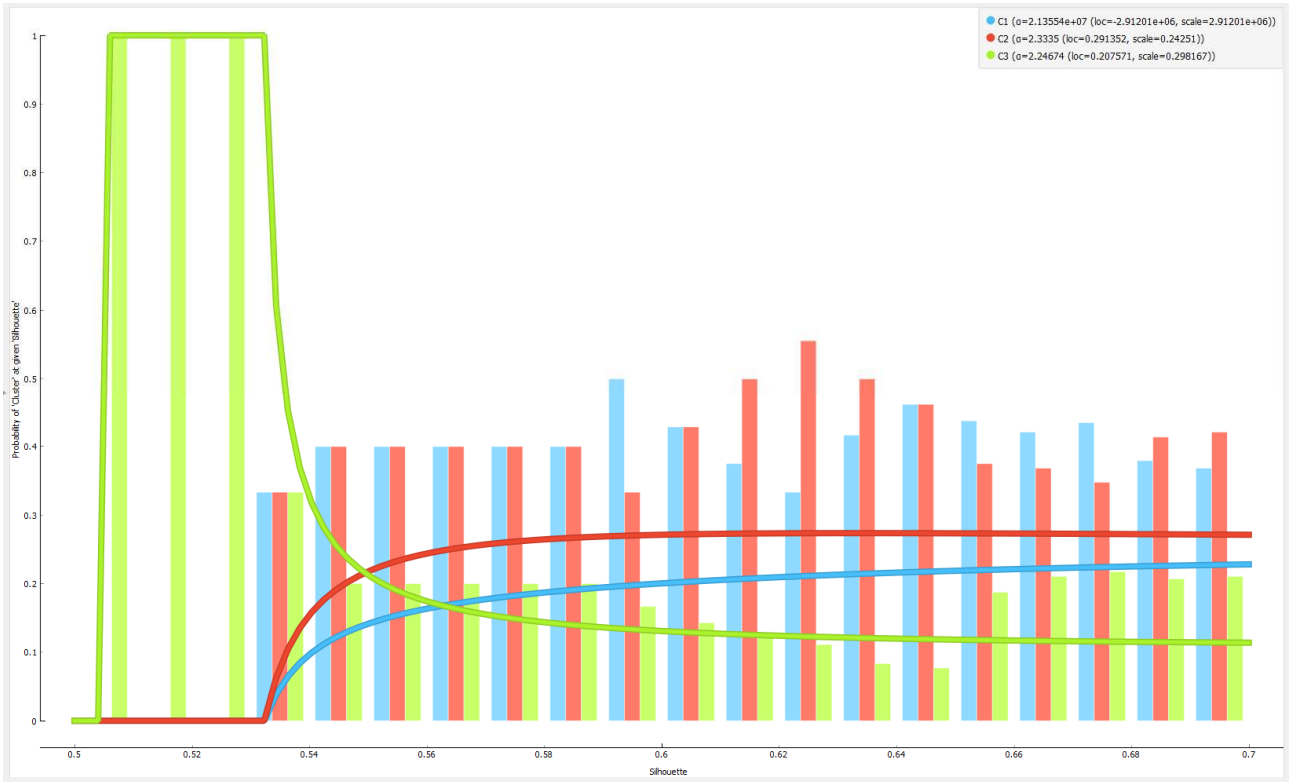

C1
C2
C3
Slovenia
Romania
Norway
Slovakia
Latvia
Lithuania
North Macedonia
Montenegro
Italy
Belgium
Portugal
Croatia
Estonia
Hungary
Netherlands
Israel
France
Finland
Denmark
Sweden
Austria
Ireland
Bulgaria
Greece
Iceland
Czechia
Germany
Serbia
Luxembourg
Switzerl
United Kir
Spain
0 - 25
25 - 50
50 - 75
75 - 100
100 - 125
125 - 150
150 - 175
175 - 200


0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
C1 (14)
Czechia
Portugal
Belgium
Israel
Netherlands
Italy
Austria
Norway
Estonia
Slovakia
Ireland
Spain
Luxembourg
Iceland
C2 (16)
Poland
Cyprus
North Macedonia
Malta
Hungary
Turkey
Ukraine
Bosnia and Herzegovin
Montenegro
Latvia
Serbia
Bulgaria
Lithuania
Greece
Croatia
Romania
C3 (8)
Sweden
Denmark
United Kingdom
Switzerland
Germany
Slovenia
Finland
France

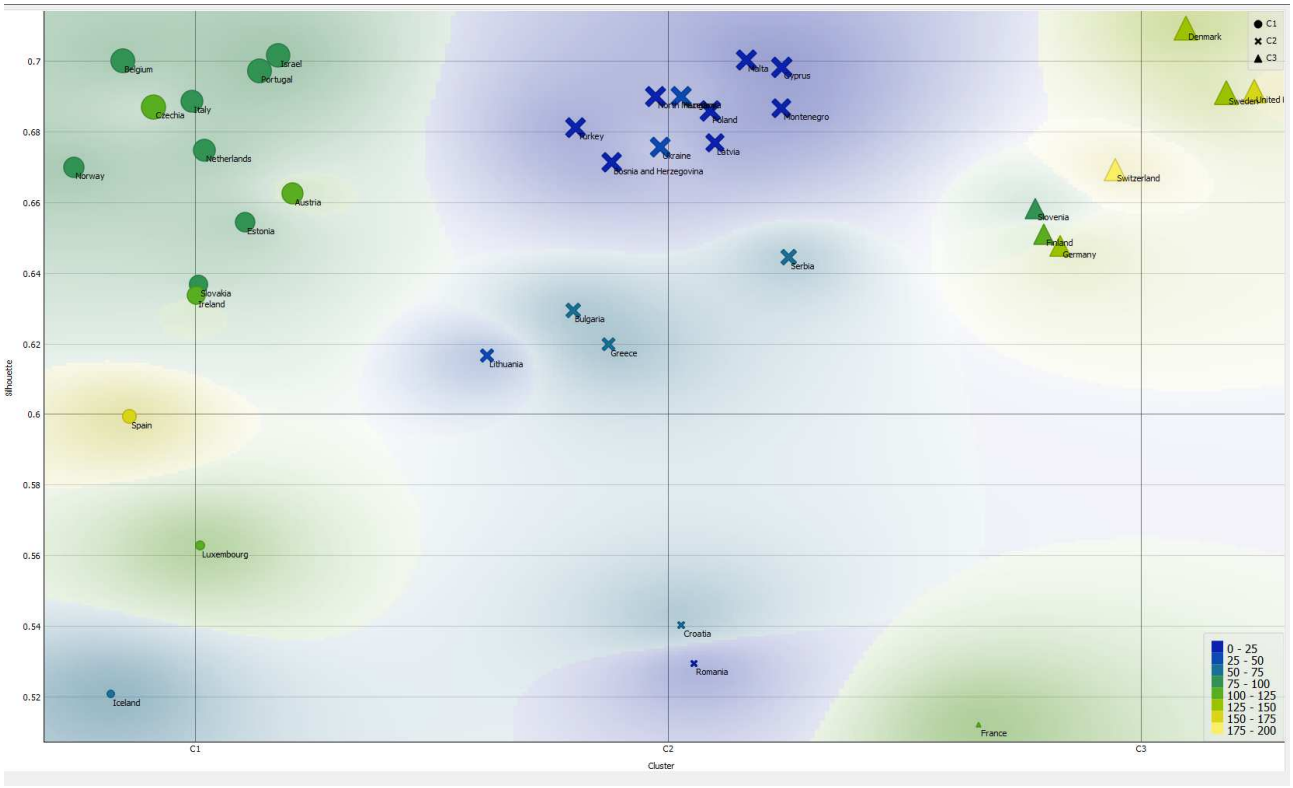

C1
C2
C3
Montenegro
Malta
Hungary
North Macedonia
Poland
Lithuania
Sweden
Finland
Serbia
Latvia
Croatia
Turkey
Cyprus
Italy
Iceland
Luxembourg
Portugal
Netherlands
Romania
Greece
Norway
Spain
Ireland
Israel
Estonia
Czechia
France
Switzerland
Denmark
Slovenia
Germany
United Kingdom
Belgium
Bosnia and Herzegovina
Austria
Slovakia
0 - 25
25 - 50
50 - 75
75 - 100
100 - 125
125 - 150
150 - 175
175 - 200

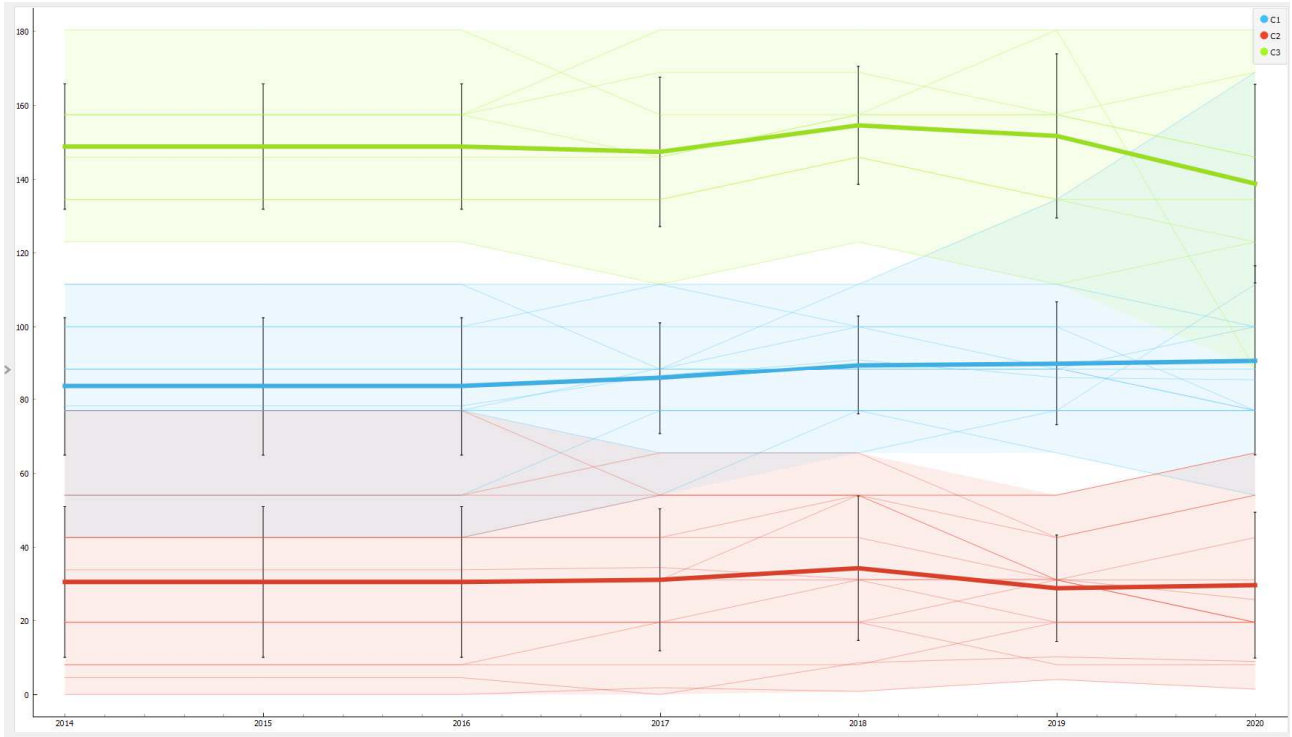

C1
C2
C3
0
10
20
30
40
50
60
70
80
90
100

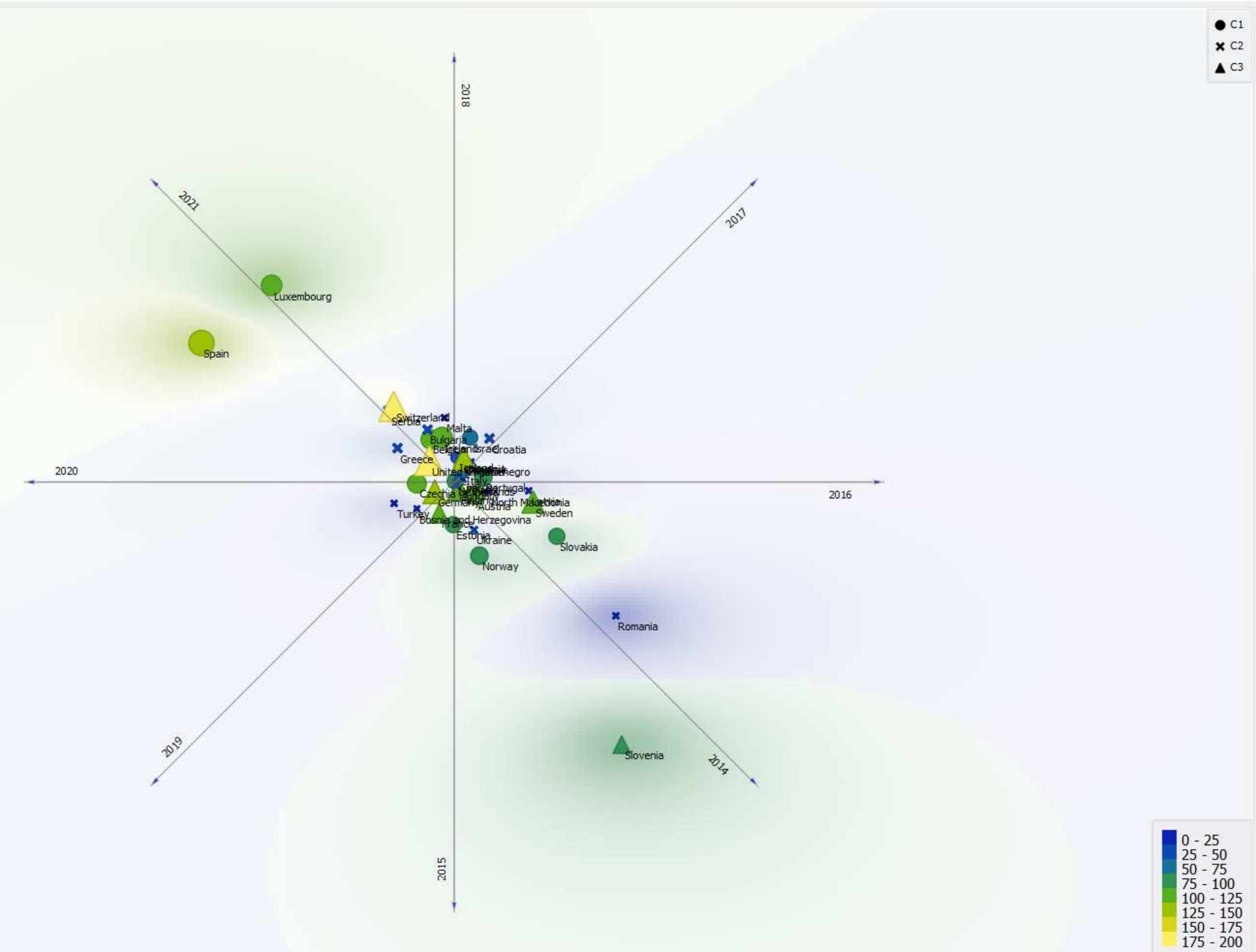

220
200
180
160
140
120
100
80
60
40
20
0
-20
2000
C1
C2
C3
Cluster

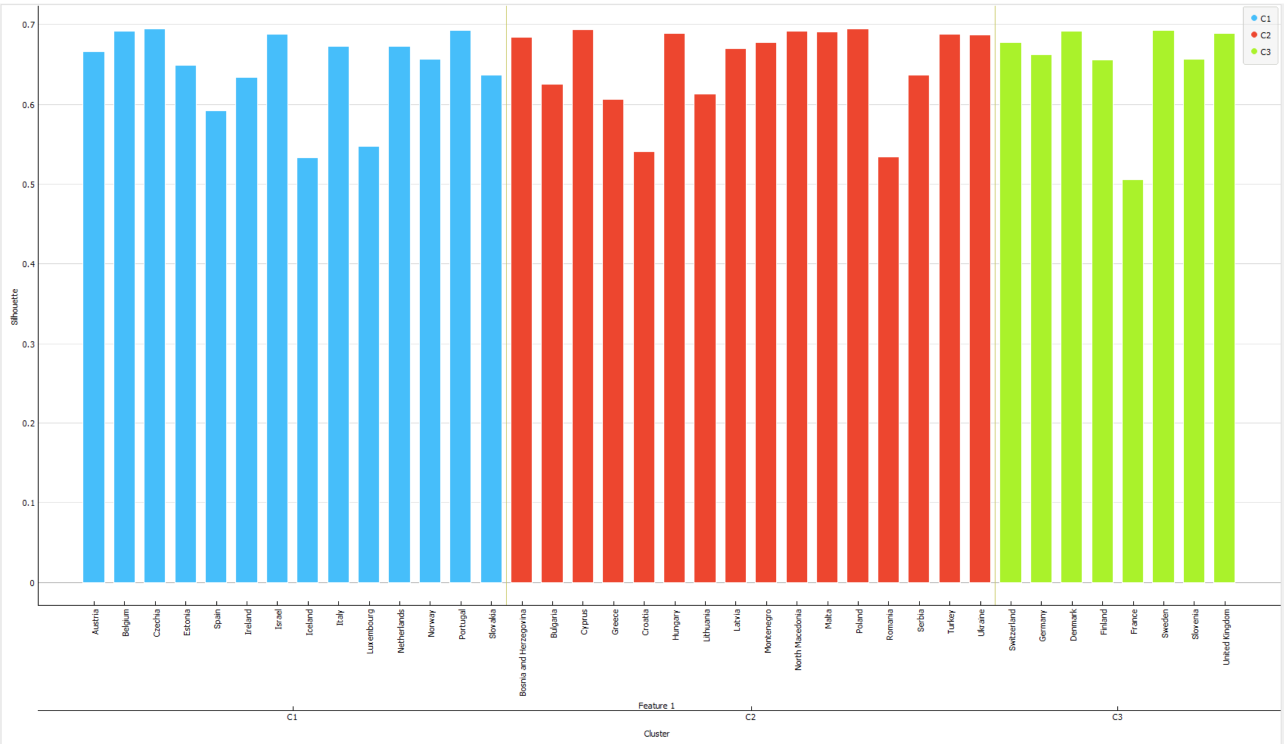


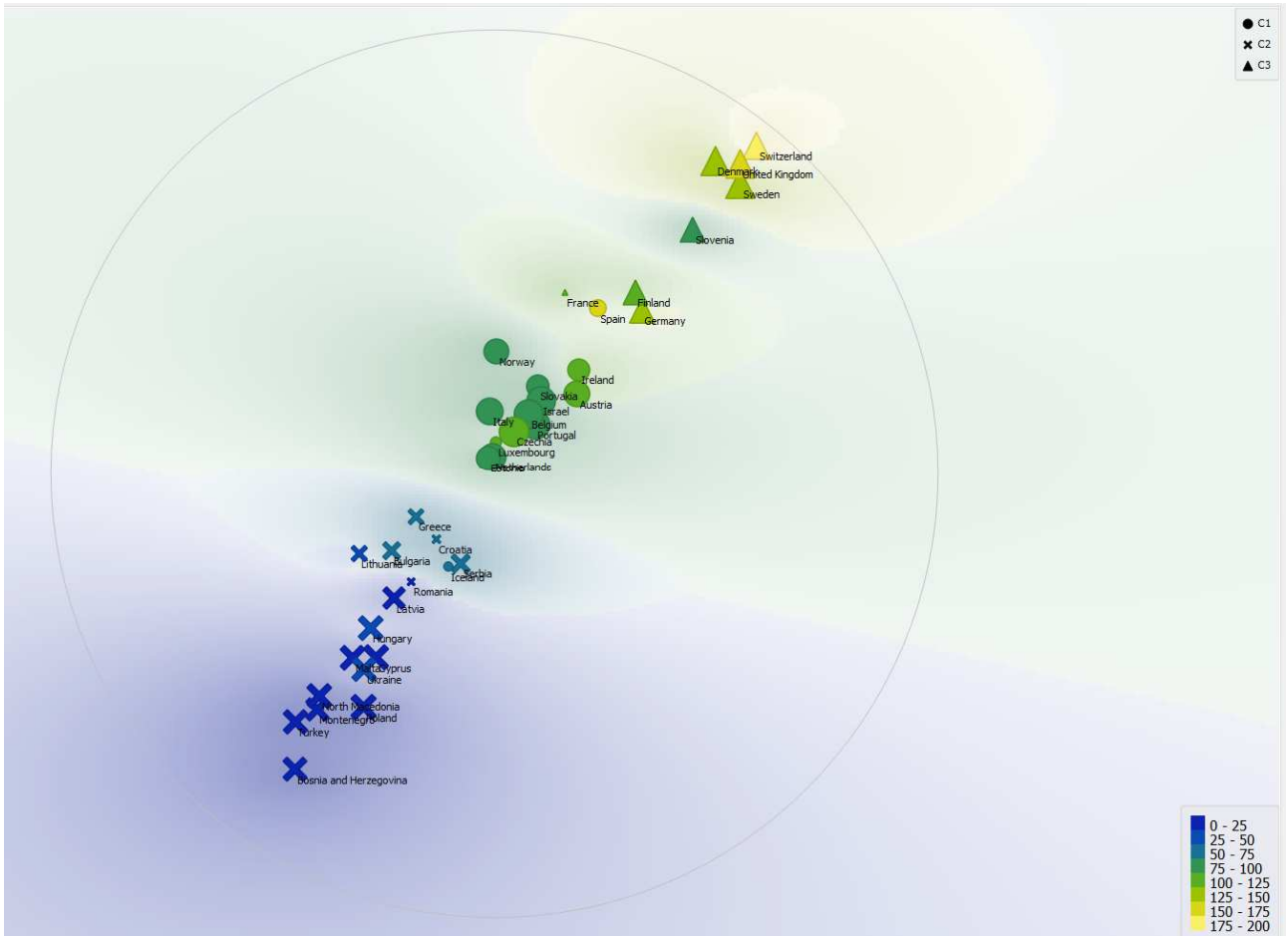
	2021	Feature 1	Cluster	Silhouette	2014	2015	2016	2017	2018	2019	2020
1	100	Austria	C1	0.666199	100	100	100	100	100	100	100
3	100	Belgium	C1	0.691387	88.5114	88.5114	88.5114	88.5114	100	100	77.0228
7	100	Czechia	C1	0.694318	88.5114	88.5114	88.5114	88.5114	88.5114	88.5114	100
10	77.0228	Estonia	C1	0.64928	77.0228	77.0228	77.0228	65.5342	65.5342	77.0228	77.0228
12	145.954	Spain	C1	0.591834	77.0228	77.0228	77.0228	88.5114	111.489	134.466	168.932
17	111.489	Ireland	C1	0.633546	100	100	100	111.489	111.489	111.489	100
18	70.7154	Israel	C1	0.688376	78.2721	78.2721	78.2721	86.4959	91.0114	85.9208	85.3764
19	42.5569	Iceland	C1	0.533027	54.0455	54.0455	54.0455	77.0228	77.0228	65.5342	54.0455
20	65.5342	Italy	C1	0.67234	77.0228	77.0228	77.0228	77.0228	77.0228	77.0228	77.0228
22	111.489	Luxembourg	C1	0.547732	42.5569	42.5569	42.5569	54.0455	77.0228	77.0228	111.489
27	77.0228	Netherlands	C1	0.67234	77.0228	77.0228	77.0228	77.0228	77.0228	77.0228	77.0228
28	88.5114	Norway	C1	0.657	111.489	111.489	111.489	88.5114	88.5114	88.5114	88.5114
30	88.5114	Portugal	C1	0.692878	88.5114	88.5114	88.5114	88.5114	88.5114	88.5114	77.0228
35	77.0228	Slovakia	C1	0.637033	111.489	111.489	111.489	111.489	100	88.5114	77.0228
2	12.5633	Bosnia and Her...	C2	0.68371	4.55936	4.55936	4.55936	0	8.59525	10.2135	8.9559
4	42.5569	Bulgaria	C2	0.625719	42.5569	42.5569	42.5569	54.0455	54.0455	42.5569	54.0455
6	31.0683	Cyprus	C2	0.693829	19.5797	19.5797	19.5797	19.5797	19.5797	31.0683	19.5797
11	42.5569	Greece	C2	0.606294	42.5569	42.5569	42.5569	42.5569	54.0455	54.0455	65.5342
15	42.5569	Croatia	C2	0.54049	54.0455	54.0455	54.0455	65.5342	65.5342	42.5569	54.0455
16	31.0683	Hungary	C2	0.689187	31.0683	31.0683	31.0683	31.0683	31.0683	31.0683	31.0683
21	42.5569	Lithuania	C2	0.612906	54.0455	54.0455	54.0455	54.0455	54.0455	31.0683	42.5569
23	19.5797	Latvia	C2	0.669575	42.5569	42.5569	42.5569	42.5569	42.5569	31.0683	19.5797
24	3.6855	Montenegro	C2	0.677359	0	0	0	1.8565	0.844975	4.07884	1.45405
25	8.09107	North Macedonia	C2	0.691828	19.5797	19.5797	19.5797	19.5797	19.5797	8.09107	8.09107
26	19.5797	Malta	C2	0.691121	8.09107	8.09107	8.09107	19.5797	19.5797	19.5797	19.5797
29	19.5797	Poland	C2	0.694817	19.5797	19.5797	19.5797	19.5797	31.0683	19.5797	19.5797
31	19.5797	Romania	C2	0.533862	77.0228	77.0228	77.0228	54.0455	54.0455	31.0683	19.5797
32	54.0455	Serbia	C2	0.636321	31.0683	31.0683	31.0683	31.0683	54.0455	54.0455	65.5342
36	19.5797	Turkey	C2	0.688378	8.09107	8.09107	8.09107	8.09107	8.09107	19.5797	19.5797
37	25.2512	Ukraine	C2	0.687252	33.8217	33.8217	33.8217	34.4672	31.2479	31.3347	25.7197
5	180.42	Switzerland	C3	0.677043	157.443	157.443	157.443	180.42	180.42	180.42	180.42
8	134.466	Germany	C3	0.662636	134.466	134.466	134.466	134.466	145.954	134.466	134.466
9	134.466	Denmark	C3	0.691655	145.954	145.954	145.954	145.954	157.443	157.443	145.954
13	122.977	Finland	C3	0.655758	134.466	134.466	134.466	134.466	145.954	134.466	122.977
14	100	France	C3	0.505738	122.977	122.977	122.977	111.489	122.977	111.489	122.977
33	122.977	Sweden	C3	0.692346	157.443	157.443	157.443	168.932	168.932	157.443	145.954
34	88.5114	Slovenia	C3	0.656375	180.42	180.42	180.42	157.443	157.443	180.42	88.5114
38	180.42	United Kingdom	C3	0.688866	157.443	157.443	157.443	145.954	157.443	157.443	168.932

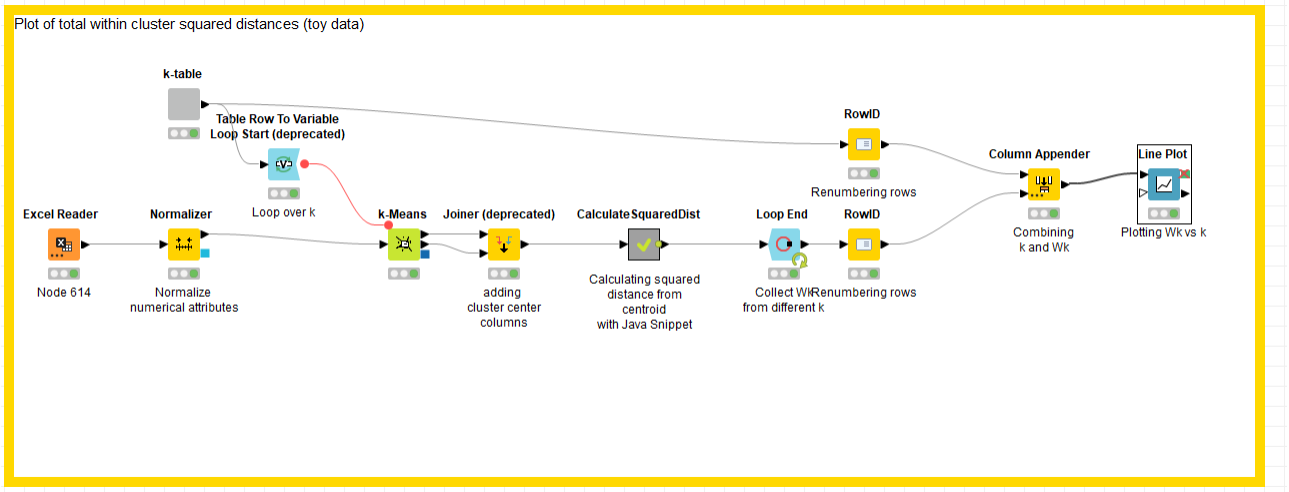

C1
C2
C3
Denmark
Belgium
Israel
Portugal
Czechia
Italy
Netherlands
Norway
Austria
Estonia
Slovakia
Ireland
Spain
Luxembourg
Iceland
Malta
Cyprus
Poland
Montenegro
Turkey
Ukraine
Latvia
Bosnia and Herzegovina
Switzerland
Slovenia
Finland
Germany
Serbia
Bulgaria
Greece
Lithuania
Croatia
Romania
France
Sweden
0 - 25
25 - 50
50 - 75
75 - 100
100 - 125
125 - 150
150 - 175
175 - 200
silhouette
Cluster


C1
C2
C3
2014
2015
2016
2017
2018
2019
2020


C1
C2
C3
2018
2021
2017
2020
2016
2019
2014
2015
Luxembourg
Spain
Slovakia
Norway
Romania
Slovenia
0 - 25
25 - 50
50 - 75
75 - 100
100 - 125
125 - 150
150 - 175
175 - 200


Silhouette
Feature 1
Cluster
C1
C2
C3


C1
C2
C3
Switzerland
Denmark
United Kingdom
Sweden
Slovenia
France
Finland
Spain
Germany
Norway
Ireland
Slovakia
Austria
Israel
Italy
Belgium
Portugal
Czechia
Luxembourg
Greece
Croatia
Lithuania
Bulgaria
Iceland
Romania
Latvia
Hungary
Ukraine
Montenegro
Turkey
Bosnia and Herzegovina
0 - 25
25 - 50
50 - 75
75 - 100
100 - 125
125 - 150
150 - 175
175 - 200


Plot of total within cluster squared distances (toy data)
k-table
Table Row To Variable Loop Start (deprecated)
Loop over k
RowID
Renumbering rows
Column Appender
Combining k and Wk
Line Plot
Plotting Wk vs k
Excel Reader
Node 614
Normalizer
Normalize numerical attributes
k-Means
Joiner (deprecated)
adding cluster center columns
CalculateSquaredDist
Calculating squared distance from centroid with Java Snippet
Loop End
Collect Wk from different k
RowID
Renumbering rows