



International Journal of Data Science and Big Data Analytics

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Analysis of innovation with data science: The case of Greece

Thanasis Zoumpikas¹, Manolis Vavalis^{2,3*}, Elias Houstis³

¹Department of Mathematics and Computer Science, University of Barcelona, Barcelona, Spain. E-mail: thanasis.zoumpikas@ub.edu

²Department of Mathematics, University of California, San Diego, USA. E-mail: evavalis@ucsd.edu

³Department of Electrical and Computer Engineering, University of Thessaly, Volos, Greece. E-mail: enh@e-ce.uth.gr

Article Info

Volume 1, Issue 1, February 2021

Received : 17 November 2020

Accepted : 25 December 2020

Published : 05 February 2021

doi: [10.33472/AFJBS.3.1.2021.20-42](https://doi.org/10.33472/AFJBS.3.1.2021.20-42)

Abstract

The purpose of this research study is two-fold. First to evaluate and compare the innovativeness of Greece relative to the European Union using indicators from the European innovation scoreboard and second to propose practices and techniques concerning the utilization of machine learning for modeling and analyzing innovation in general. Systematic analysis is conducted regarding the over-performance and the under-performance of Greece and the trends of these indicators over the years through statistical techniques and methods. Machine learning and advanced statistical methods are incorporated to ascertain the most important features that drive the variation of the summary innovation score of the European Union and Greece. Clusters and groups of correlated indicators are also specified. Our study provides preliminary explanations and evidence to help the country value its advantages and deal with its disadvantages. It also paves the way towards an analysis of innovation approach that has the potential to help us elucidate certain issues and deeply understand observations concerning Greece and beyond.

Keywords: Innovation, Data science, Data analysis, Machine learning, Statistics

© 2021 International Journal of Data Science and Big Data Analytics. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

Innovation is one of the two fundamental functions of an organization (Drucker, 1954). It is the procedure of translating an idea or invention into a good or service that creates value or for which customers will pay. The idea must be able to be replicated at an economical cost and must satisfy a specific need. Innovation implicates the intentional application of information, creativity, and lead in deriving greater or different values from resources (Henderson and Lentz, 1995). It involves all processes by which new ideas are generated and converted into useful products. It also includes the developing of new sources of supply with raw materials (Schumpeter and Opie, 1934). In business, innovation is the outcome of applied ideas by the company to further satisfy the requirements and expectations of the customers.

Innovation may also be characterized as a process that provides company, suppliers and consumers with added value and a degree of novelty, creating new processes, products, technologies, services and new forms of marketing. Also, it is the adoption of new or significantly improved elements to create added value to the organization directly or indirectly for its customers (Carnegie, Roderick and Business Council of Australia, 1993). In a social context, innovation

* Corresponding author: Manolis Vavalis, Department of Mathematics, University of California, San Diego, USA. E-mail: evavalis@ucsd.edu

aids in the development of new methods for alliance creation, joint venturing, flexible work hours, and the creation of buyers' purchasing power. It is synonymous with risk-taking. Organizations and companies that develop new revolutionary products and services take on the great risk because they create new markets.

Innovation's role in market development and coordination is an intrinsic one. The value of innovative technologies, including product development, management practices, ways of doing work and beyond, is critical in all human fields (Tohidi and Jabbari, 2012).

On the other hand, the industry is crucial for competitiveness and innovation is a key factor in this regard (Industrial policy | Internal Market, Industry, Entrepreneurship and SMEs). The industry commonly accounts for around 80% of a country's exports. Some 65% of research and development (R&D) investment in the private sector comes from manufacturing. Industrial modernization must therefore be broad in every country and include: the successful marketing of product and service innovations, the industrial exploitation of innovative manufacturing technologies and innovative business models.

Organizations that prioritize innovation are likewise the individuals with the most astonishing increase in turnover. About 79% of companies that have introduced at least one innovation since 2011 experienced an increase of more than 25% in their turnover by 2014 (Innobarometer | Internal Market, Industry, Entrepreneurship and SMEs).

Specifically, innovation policy targets to a high extent the Small and Medium-sized Enterprises (SMEs). Small companies face constraints to innovation or the commercialization of its innovations. About 63% of firms with fewer than 9 employees reported having introduced at least one innovation since 2011, compared to 85% of firms with 500 or more employees. Some 71% of companies with fewer than nine employees found it difficult to commercialize their innovations due to a lack of financial resources, compared to 48% of companies with 500 or more employees (Innovation | Internal Market, Industry, Entrepreneurship and SMEs).

This study concentrates on the European Union (EU) and Greece. The European Commission offers numerous tools designed to map, monitor and evaluate the performance of the EU in various areas of innovation. The information provided helps EU, national and regional policy makers and practitioners to evaluate their performance and policies and learn about new patterns and potential market opportunities that can guide policy making based on evidence (Monitoring innovation Internal Market, Industry, Entrepreneurship and SMEs).

In this paper, the country level innovation performance is examined by analyzing the data from the European Innovation Scoreboard (EIS), specifically the version of 2018. Indicators form the basis of the annual EIS which provides a benchmarking of the research and innovation operation of the EU Member States. It specifically offers comparisons on relative strengths and limitations regarding the research and innovation systems in the country level. It effectively offers assistance to Member States in identifying areas on which they need to concentrate their efforts to improve their success on innovation.

The main indicator is the Summary Innovation Index that summarizes the range of different indicators of innovation and measures the total innovation performance. For detailed information on the definitions, the explanations and the methods of calculation for above mentioned indicators the reader is referred to (European Commission, 2018).

The specific purpose of this paper is to compare innovativeness of Greece versus the EU average using the indicators provided by the EIS for the period 2010-2017. A brief analysis on the systematic over-performance and under-performance of Greece and the trends of these indicators over the years is provided. Furthermore, machine learning and statistical techniques are used to identify the key features that influence the fluctuation of the summary innovation score at the EU and Greece level.

The rest of this paper is organized as follows. Next, selected studies that are related to our efforts are presented. Section 3 contains our methodologies concerning the data collection, the pre-processing of the data, its statistical analysis, and the machine learning process. Section 4 provides information about our indicators and 5 the predictive trend and machine learning analysis of our data. Section 6 summarizes our results concerning the performance of Greece against the EU average and presents our concluding remarks. Links to detailed data and further supporting information are provided in Appendix.

2. Related work

As innovation analysis is considered fundamental and one of the most important keys for the economic growth of each country, many researchers from different fields focus their efforts on studying various aspects of innovation. During the past few decades, a significant amount of research work has been conducted using advanced tools and indicators examining innovation.

SMEs consist a principal part of the industry and economy in all modern countries. In 1990, despite the fact that only little was known about SME innovation activities, Hyvärinen Lisa reviewed the definitions of innovation technology and factors on the background of innovation activities of SMEs. In the aforementioned research work, various concepts approaching innovativeness of SMEs and their contribution to total innovation are explained (Hyvärinen, 1990).

There exists a rather limited amount of scientific work, to be presented below, regarding the Greek innovativeness and economic performance. Researchers have used empirical analysis to show the unfriendliness of the Greek private sector to invest in R&D and the low productivity of innovation (Beneki et al., 2012). Others investigate the impact of the indicator R&D activity on operational performance of SMEs extending the objective on the operational performance of SMEs in the small, open Greek economy (Eleftherios et al., 2016).

Additional studies focus on the significance and awareness of a set of established strategic influences of technological innovation in the context of European newly-industrialized countries. Research studies such as (Souitaris, 2001), provided evidence from interviews conducted on Greek manufacturing firms (mainly SMEs) measuring their innovation rate as well as key performance indicators. Using statistical analysis tools, they summarize and highlight the most important indicators that have major importance influence of innovation. This study also indicates that the Greek institutional context had insufficient important influences of innovation and the highly innovative companies were the ones to overcome barriers such as the low supply of technology and other innovation obstacles.

Moreover, such studies developed also at regional level and provide evaluation of the numerous policy instruments used by regional governments in Europe to promote innovation activity in SMEs (Asheim et al., 2003). Scientists try to find patterns of innovation in regional innovation structures which are becoming increasingly diverse, complex and nonlinear. To address these issues, they use multi-output models (Hajek and Henriques, 2017).

There is a variation in methods utilized by researchers in order to forecast or analyze in depth innovation or specifically indicators of innovation. A wide variety of machine learning and deep learning algorithms is commonly used. Advanced machine learning methods, such as ensemble decision trees, are utilized in (Hajek and Stejskal, 2015). Where they demonstrate the use of ensembles of decision trees to model the intrinsic nonlinear characteristics of the innovation process and apply their method for predicting innovation activity to chemical companies. In addition, other studies use non-linear methods based on Artificial Intelligence, namely neural networks (Paz-Marín et al., 2012; Hajek and Henriques, 2017; Chien et al., 2010; Wang and Chien, 2006; Saberi and Yusuff, 2012). In the aforementioned study (Wang and Chien, 2006), they forecast innovation performance using a neural network model with fuzzy rules and provide evidence from Taiwanese manufacturing industry. They also implement an adaptive neuro-fuzzy inference system to measure the innovation performance through technical information resources and innovation objectives. In (Saberi and Yusuff, 2012), they develop an Artificial Neural Network classification method and prediction model that can assist companies especially SMEs in evaluating Advanced Manufacturing Technology implementation contributing to innovation.

In addition, the fact that decision makers often group the object of their analysis into homogeneous classes lead them to use clustering algorithms, see for example (Paz-Marín et al., 2012; Klimova et al., 2016; Roszko-Wójtowicz and BiaBek, 2018; Saberi and Yusuff, 2012). Others prefer more traditional techniques based on statistical analysis and equation modeling (Kalapouti et al., 2017; Jan van den Ende and Timo van Balen, 2017).

Please note that there exist a variety of data utilized that may come from a single source or by combining multiple data sources. They might rely on traditional methods like interviews (Souitaris, 2001) and well-formed databases obtained from Eurostat's official website (Kalapouti et al., 2017; Roszko-Wójtowicz and BiaBek, 2018; Roszko-Wójtowicz and BiaBek, 2017), World Bank Database, SCImago Journal (Paz-Marín et al., 2012). or data obtained from companies providing business services, such as ICAP (Beneki et al., 2012).

We should highlight the research from Rotterdam School of Management regarding the innovativeness of the Netherlands compared to EU countries (Jan van den Ende and Timo van Balen, 2017). They statistically compared the Netherlands versus EU in indicators of innovation using data from EIS database. Briefly, the methods they utilized are generalized least squares regression for trend estimation and statistics.

This study concerns a review of several of those methods described above, the implementation of the associated statistical techniques and machine learning models and their application on indicators time series data. It assesses systematic over-performance and under-performance of Greece relative to EU countries and compares the trends of Greece and EU regarding the composite and simple indicators of innovation using data from EIS database. Moreover, it presents the application of machine learning models to identify, in a model based sense, the most crucial features-indicators influencing the volatility of summary innovation index.

It is worth to point out that (Jan van den Ende and Timo van Balen, 2017) is the study most closely related to ours. They tackle the same problem of analyzing innovativeness in the country-level through a very interesting, innovative and effective methodology. This paper is inspired from their analysis of the Netherlands relative to the EU regarding innovation performance. Besides, we consider a similar approach to analyze the case of Greece. Furthermore, it adds scientific value by involving machine learning to estimate the most important indicators, combining more information on this study. Besides, further analysis of indicators utilizing principal components analysis and advanced data visualization methods emphasizes the importance of this study.

3. Analytics work-flow and overall methodology

The basic elements of our overall methodology for our study are presented. The related architectural structure of our analytics workflow, which consists of the following three main stages, is graphically depicted in Figure 1.

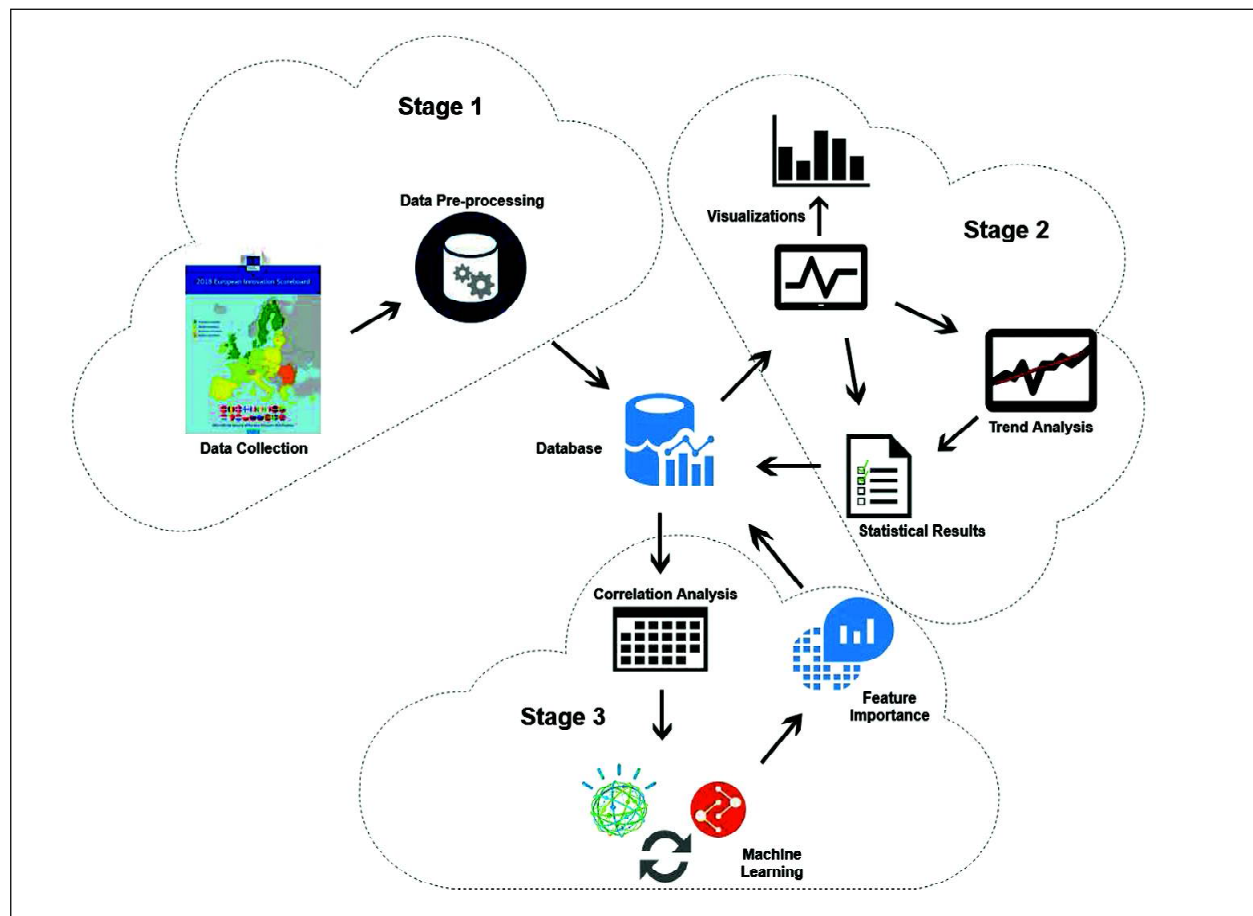


Figure 1: Analytics work-flow and methodology

3.1. First stage – pre-processing

Data is acquired from the 2018 database edition of the EIS, containing Greece and the EU average for the period 2010-2017. It is collected from EIS website and it is of high quality with minor missing observations for Greece. The EIS 2018 database comprises of many dimensions. This study utilizes composite indicators and each individual indicator for the years 2010-2017.

EIS 2018 database is processed to choose only indicators and composite indicators regarding the EU and Greece. The data is preprocessed and the two indicators (“Foreign doctorate students as a % of all doctorate students” and “Employment in fast-growing enterprises (% of total employment)”) are dropped from our analysis, because of missing values for Greece. A time series data from 2010 to 2017 is built using the standardized scores given by the database, for each indicator and composite indicator. In total, time series data comprising of 25 indicators and 11 composite indicators (including the summary innovation index) are collected. A short description of these indicators, together with their long acronyms used in the presentation of our study is given in 6 below, while their details are available at (European Commission, 2018; Jan van den Ende and Timo van Balen, 2017).

3.2. Second stage – statistical analysis

The second stage consists of statistical analysis of the aforementioned time series data. Each indicator and composite indicator of Greece versus the EU average over the time frame (2010-2017) is visualized. The visualization process includes two graphs per indicator (or composite indicator), the actual values of time series and the percentage change each year. Then we comment on the graphs and provide statistical test to measure the systematic out-performance or under-performance of Greece compared to the EU. These comparisons are concluded from a two samples t_{test} assuming unequal variances between the (non-missing) data observations over the studied period. This method is commonly known as Welch’s t_{test} (Welch, 1947). A hypothesis H_0 is stated to test whether Greece outperforms on average the EU. For the aforementioned test, we accept or reject the H_0 according to the t_{stat} , the test statistic of the test, and p_{value} , or probability value, is the likelihood of finding the observed, or more extreme results, when the study question’s null hypothesis (H_0) is true. The t_{stat} , p_{value} and our decision based on an alpha level $\alpha = 0.05$ are reported, i.e., 95% statistical significance. Since H_0 is a one-tailed test, it is rejected when $p_{value}/2 \leq 0.05$.

In addition, the trend-lines (assumed to be linear) statistically analyzed for each indicator (or composite indicator) of Greece and of the EU using generalized least squares regression with a correction for autocorrelation on the years in Greece and EU level. The generalized least-squares (GLS) regression method with auto-regressive errors (AR(1)), namely GLSAR (McKinney et al., 2011), is used for linear regression lines. Also, the slope coefficient of the trend-line (b), its standard error (se) and the level of significance of the trend (p_{value}) are reported respectively. Then, the trend-lines of Greece and the EU in a statistical manner using a z_{test} are compared, following (Paternoster et al., 1998) approach.

The motivation and the details of our approach for comparing two trend-lines (linear regression lines), are given in (Paternoster et al., 1998). It is again a hypothesis testing, where $H_0: b_1 = b_2$, i.e., $b_1 - b_2 = 0$ and the alternative hypothesis $H_1: b_1 \neq b_2$, i.e., $b_1 - b_2 \neq 0$. An alpha level of statistical significance equal to $\alpha = 0.05$ is used, i.e., 95% statistical significance. Thus, insignificance is concluded for $p_{value} > 0.05$ (two-tailed test). Again, the slope coefficient of the trend-line (b), its standard error (se), the z_{score} and the level of significance of the difference in trend (p_{value}) are reported respectively.

3.3. Third stage – machine learning

The third stage includes the machine learning part of our study. We use the following approach in order to estimate the contribution of each indicator to the final performance of Greece and EU innovation, i.e., to what degree it affects Greece or EU overview innovation indexes.

The analysis of importance is conducted only in simple indicators, because we want to study the basis of the problem rather than using the composite indicators, which contain groups of simple indicators. Firstly, a correlation analysis of the indicators is carried out to avoid including the strongly correlated features in our further analysis, providing the heat-maps for interpretation purposes. Highly correlated features are dropped from our analysis, i.e., correlation coefficient above 0.90 ($r > 0.90$). This is a common tactic to avoid multi-collinearity issues and get better results from model based feature importance and generally in machine learning algorithms.

In addition, the relationship of the indicator correlation matrices of Greece, the EU and the Netherlands is examined utilizing an advanced statistical test, the Mantel test (Mantel, 1967) and the Frobenius norm of the matrices. Mantel test is a non-parametric statistical test and computes the statistical significance of the correlation through permutations of the rows and columns of one of the input matrices. The test statistic is the Pearson correlation coefficient r . Specifically, a two-sided Mantel test is utilized with 10000 permutations to identify the correlation between the two matrices. Please recall, that the Frobenius norm, commonly known as the Euclidean norm, is a matrix norm of an $m \times n$ matrix A defined by

$$\sqrt{\sum_{i=1}^m \sum_{j=1}^n |\alpha_{ij}|^2}$$

Then, a vector (v) is constructed that models as follows the fluctuation of the summary innovation index over the time period (2010-2017). A sliding window that covers a single year over the time frame starting from the beginning is implemented. If the value of the summary innovation index in currently considered year is higher than the value for the previous year we set $v_i = 1$, otherwise we set $v_i = 0$. Thus, it is a binary classification problem with the indicators as features and the vector v as target-label. A 3-fold cross validation is used on our data to train four machine learning models for classification. The models are Logistic Regression, Random Forest, Extra-Trees and Support Vector Machines. For each model, the estimates of feature importance across all three folds of cross validation are averaged to get a better estimate of model based feature importance. In order to get a final summary of the most important features for Greece and EU, each of the aforementioned model based feature importance is expressed in percentage feature importance for each

model and then we average on percentages across on four models. For homogeneity and averaging purposes, our choice is to express feature importance in percentage values, because the procedure and values for calculating the most important features are different in each model. Finally, we discuss the most important indicators which drive and affect the most the fluctuation of summary innovation index at Greece and EU level.

4. Data analysis

The visualizations of time series data and percentage change regarding the years 2010-2017 are provided for each composite indicator and indicator used in this study were presented in Appendix.

4.1. Composite indicators

Table 1 displays inferential statistics on the composite indicators time series data. For presentation purposes, all values are rounded to certain significant digits.

The summary innovation index of EU is higher than Greece over the whole period of our study. EU outperforms Greece in total innovation. Observing the percentage change graph of this composite indicator in Figure 2, there is a 12% decrease of innovation in Greece level on year 2014 compared to previous year. However both EU and Greece show upward trend in summary innovation index from 2014 till now.

It is worth to highlight the composite indicator “Innovators” in Figure 2, where Greece outperforms the EU average in all years from 2010 to 2017. We point out that this composite indicator is rather important as it is comprised of three simple indicators. In a period of Greece’s economic crisis, there is a systematic over-performance on average of Greece versus the EU in the share of companies that have produced developments on the market or within their organizations, including innovative processes and products, marketing and organizational innovators, and innovative SMEs in-house.

Following, there is a statistical hypothesis testing of the systematic over- or under-performance of Greece compared to the EU average. A hypothesis is stated for testing. The mean of the EU and Greece values is denoted μ_{eu} and μ_{gr} respectively. We state the null hypothesis $H_0: \mu_{eu} \geq \mu_{gr}$ and the alternative hypothesis $H_1: \mu_{eu} < \mu_{gr}$. With a two samples statistical t_{test} assuming unequal variances, we test whether the EU indicator has greater value in average than Greece indicator. In this way, a test is performed for over-performance or under-performance of Greece compared to the EU average. Table 1 contains the composite indicators, the t_{stat} of the t_{test} , the p_{value} and our decision on H_0 based on 95% statistical significance ($\alpha = 0.05$).

As we can clearly see, we have strong evidence to reject the null hypothesis H_0 ($p_{value}/2 \leq 0.05$), that the EU average value is higher than Greece value. This means, that we statistically confirm the systematic over-performance of Greece versus EU in the composite indicator “innovators”.

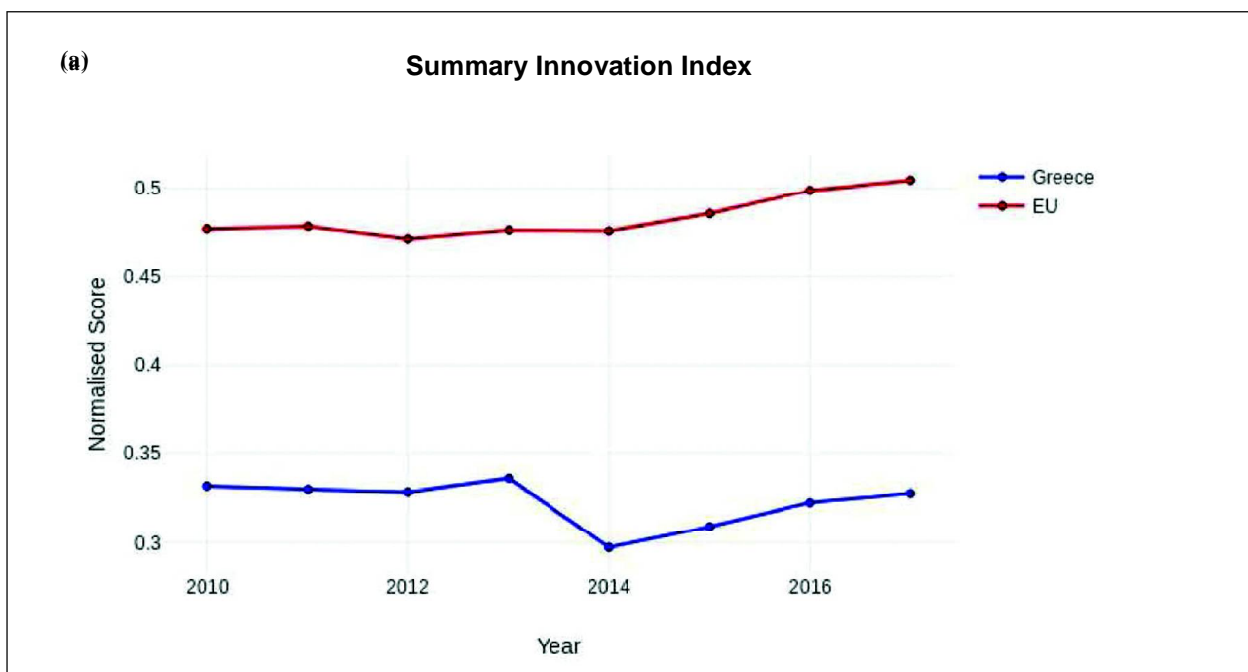
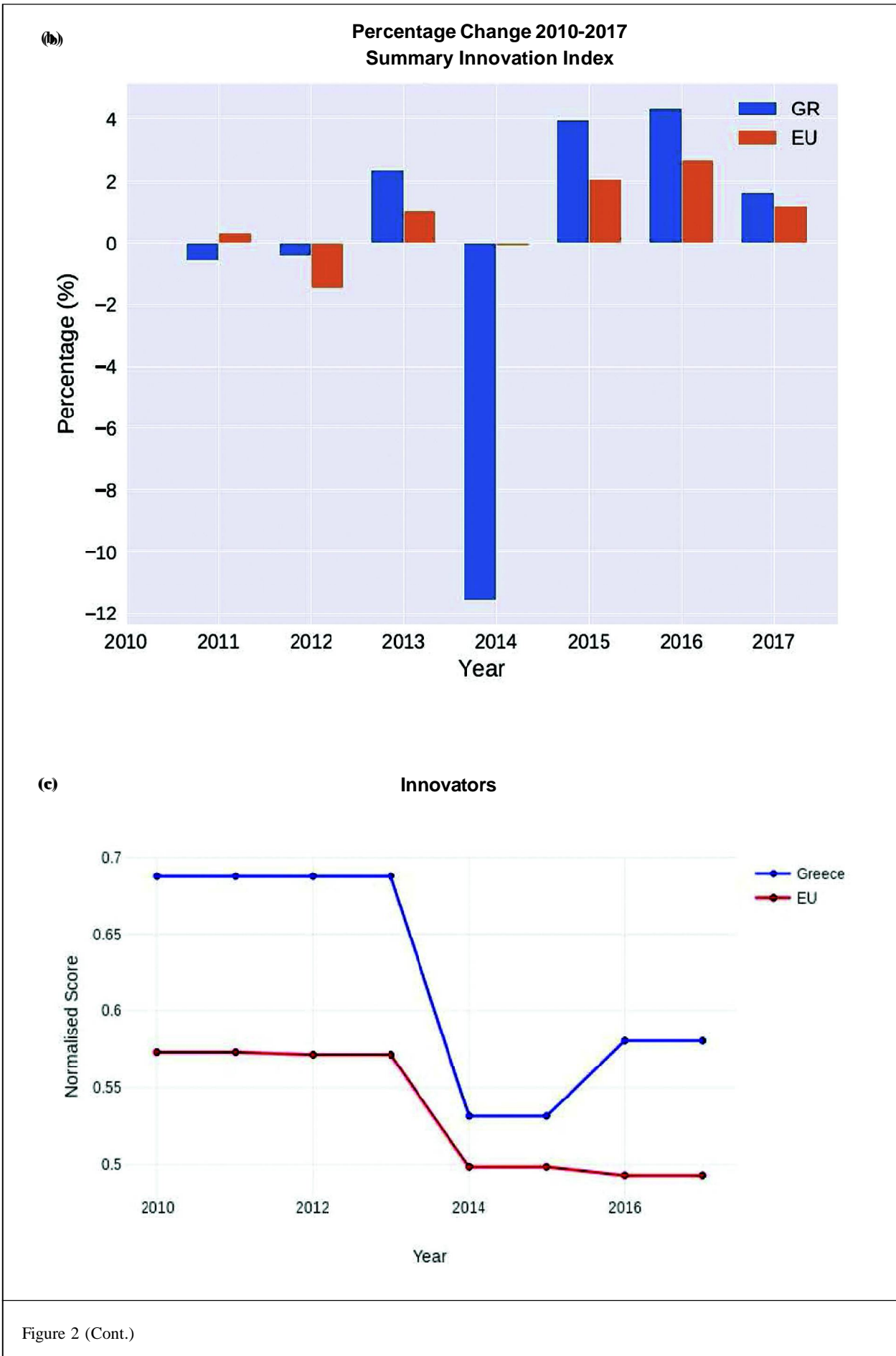


Figure 2: On the top: Summary innovation index (a) and its percentage change (b). On the bottom: Innovators (c) and its percentage change (d).



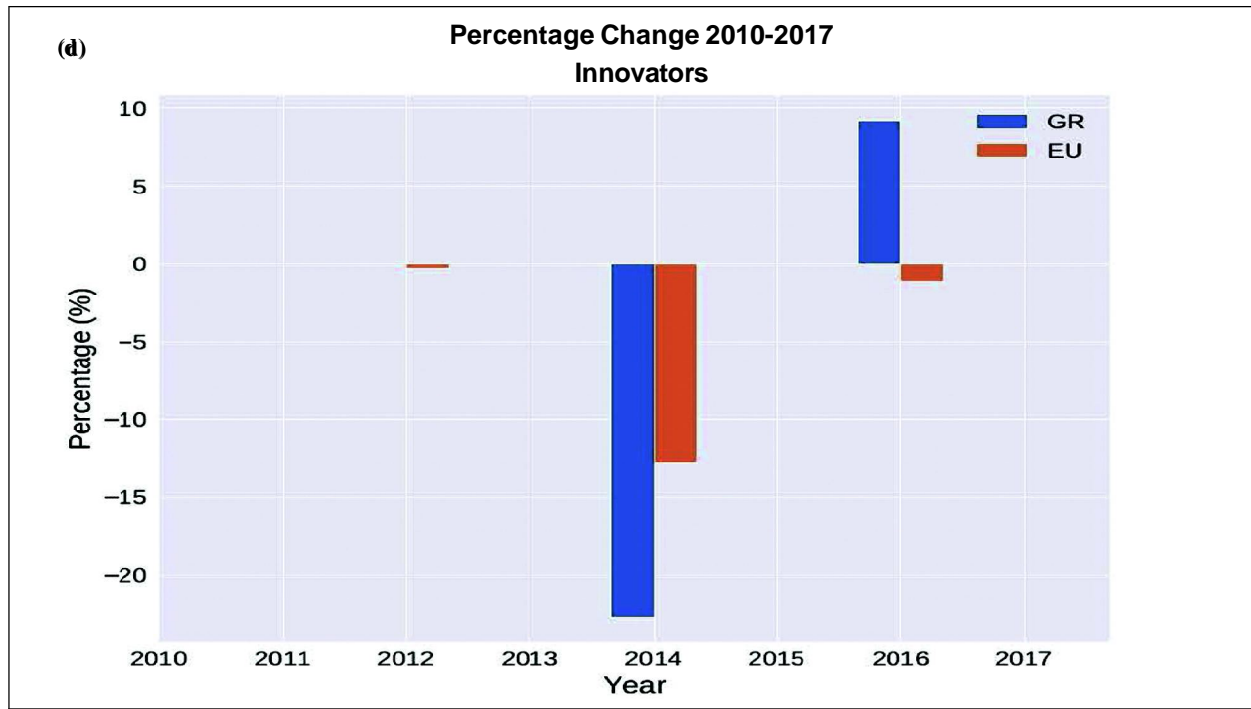


Figure 2 (Cont.)

4.2. Indicators

The Table 1 displays inferential statistics on the indicators time series data. All values in this table are rounded to 3 decimal values.

Following, there is a focus on indicators where Greece outperforms the EU with strong statistical evidence, according to Table 1. As clearly seen, Greece exceed the EU in average in six indicators, namely “Innovative SMEs collaborating with others (% of SMEs)”, “Non-R&D innovation expenditures (% of turnover)”, “Sales of new-to-market and new-to-firm innovations as % of turnover”, “International scientific co-publications per million population”, “SMEs introducing marketing or organizational innovations as % of SMEs” and “Percentage population aged 25-34 having completed tertiary education”. The later three of them are visualized in Figure 3.

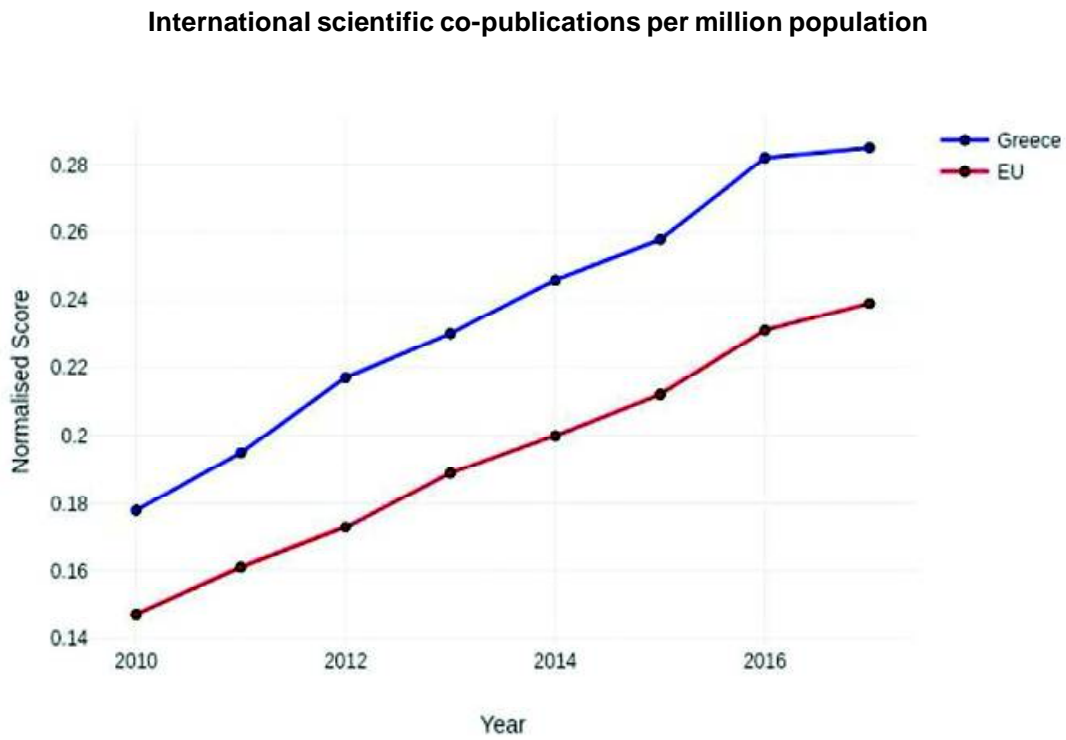
The statistical hypothesis testing is displayed in Table 1, which shows the indicators, the t_{stat} of the t_{test} , the p_{value} and our decision on H_0 based on 95% statistical significance ($\alpha = 0.05$). For the effectiveness of the presentation of our data analysis the names of the indicators are truncated into long acronyms and are presented in Appendix.

Table 1: Index of Table: Composite Indicators (the first eleven rows), Indicators (the rest). First three columns: t -test on H_0 . Fourth column: Performance of Greece (GR) vs European Union (EU) (\blacktriangledown : Lower, \blacktriangleup : Higher). Fifth and sixth column: Trend of GR and of the EU (+ : Positive, - : Negative). Last two columns: Correlation analysis decision on indicators of GR and EU (\times : Drop, \checkmark : Keep)

| | | t-test on H_0 | | Performance | Trend | | Correlate | |
|-------------------------|-------|-----------------|--------------|----------------------|-------|----|-----------|----|
| | | p-Value | H_0 | | GR | EU | GR | EU |
| Summary_Innovation_I | 25.49 | .000 | \checkmark | \blacktriangledown | | - | | |
| Human_Resources | 10.14 | .000 | \checkmark | \blacktriangledown | + | | | |
| Research_Systems | 4.34 | .001 | \checkmark | \blacktriangledown | + | + | | |
| Innovation-friendly_env | 14.44 | .000 | \checkmark | \blacktriangledown | | | | |
| Finance_and_support | 15.45 | .000 | \checkmark | \blacktriangledown | + | | | |
| Firm_investments | 14.94 | .000 | \checkmark | \blacktriangledown | | - | | |

| Table 1 (Cont.) | | | | | | | | |
|----------------------------|-------|-------------------------|-------|-------------|-------|----|-----------|----|
| | | <i>t</i> -test on H_0 | | Performance | Trend | | Correlate | |
| | | <i>p</i> -Value | H_0 | | GR | EU | GR | EU |
| Innovators | -2.98 | .013 | ✗ | ↗ | | | | |
| Linkages | 7.20 | .000 | ✓ | ↘ | | | | |
| Intellectual_assets | 23.95 | .000 | ✓ | ↘ | + | + | | |
| Employment_impacts | 14.61 | .000 | ✓ | ↘ | | | | |
| Sales_impacts | 5.34 | .001 | ✓ | ↘ | - | - | | |
| Broadband_penetration | 8.25 | .000 | ✓ | ↘ | | | ✓ | ✓ |
| Venture_capital | 12.58 | .000 | ✓ | ↘ | - | - | ✓ | ✓ |
| Design_applications | 29.72 | .000 | ✓ | ↘ | + | + | ✓ | ✓ |
| Trademark_apps | 7.91 | .000 | ✓ | ↘ | + | + | ✗ | ✓ |
| Employment_activities | 8.43 | .000 | ✓ | ↘ | | | ✗ | ✗ |
| Enterprises_training | 8.05 | .000 | ✓ | ↘ | | | ✓ | ✓ |
| Innovative_Smes | -6.44 | .000 | ✗ | ↗ | | | ✓ | ✓ |
| International_publications | -2.36 | .034 | ✗ | ↗ | + | + | ✗ | ✗ |
| Knowledge_exports | 5.83 | .001 | ✓ | ↘ | - | - | ✓ | ✗ |
| New_doctorate_grads | 8.00 | .000 | ✓ | ↘ | | | ✓ | ✗ |
| Non_rd | -2.85 | .013 | ✗ | ↗ | | | ✓ | ✗ |
| Opportunity_enterpre | 27.90 | .000 | ✓ | ↘ | | | ✓ | ✓ |
| Pct_patent | 65.42 | .000 | ✓ | ↘ | | + | ✓ | ✓ |
| Percentage_tertiary_edu | -3.58 | .005 | ✗ | ↗ | + | + | ✗ | ✗ |
| Percentage_lifelong_lea | 37.95 | .000 | ✓ | ↘ | + | + | ✗ | ✓ |
| Private_co_funding | 8.68 | .000 | ✓ | ↘ | | | ✓ | ✓ |
| Public_private_pubs | 33.78 | .000 | ✓ | ↘ | | | ✓ | ✓ |
| Rd_business | 32.70 | .000 | ✓ | ↘ | + | | ✗ | ✗ |
| Rd_public | 6.43 | .000 | ✓ | ↘ | + | + | ✗ | ✓ |
| Sales | -2.12 | .071 | ✗ | ↗ | - | - | ✗ | ✓ |
| Scientific_pubs | 14.34 | .000 | ✓ | ↘ | | | ✗ | ✗ |
| Smes_in_house | -0.97 | .355 | ✓ | ↘ | | | ✗ | ✓ |
| Smes_marketing | -4.72 | .001 | ✗ | ↗ | - | - | ✗ | ✗ |
| Smes_product | -1.57 | .143 | ✓ | ↘ | | | ✗ | ✗ |
| Exports_technology | 41.01 | .000 | ✓ | ↘ | | | ✓ | ✓ |

(a)



(b)

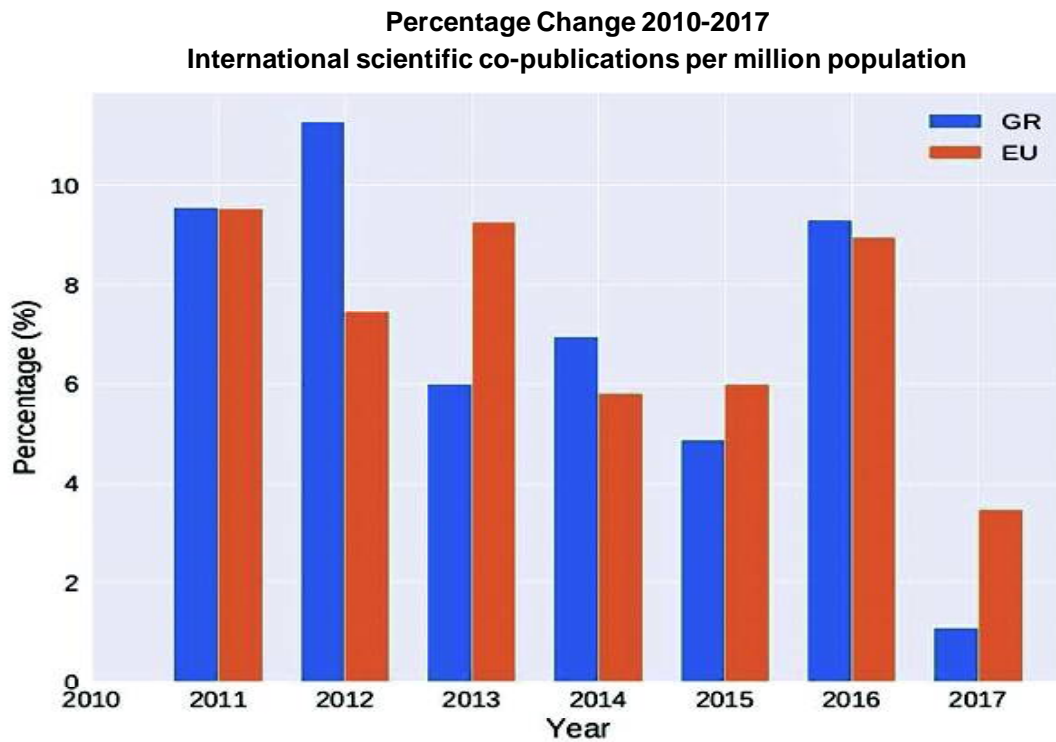
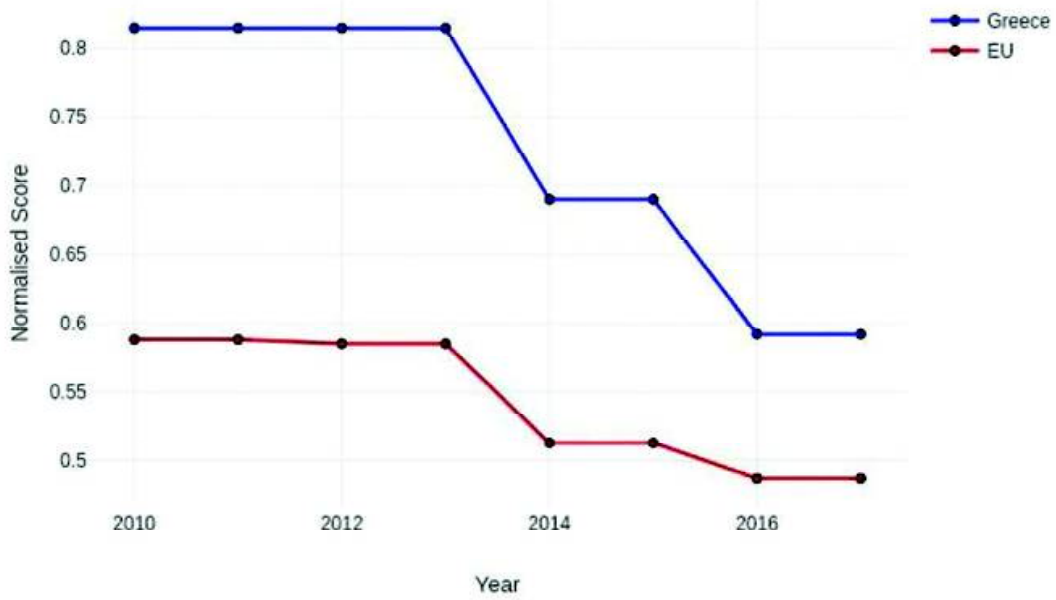


Figure 3: On the top: International scientific copublications per million population (a) and its percentage change (b) On the middle: SMEs introducing marketing or organizational innovations as% of SMEs (c) and its percentage change (d). On the bottom: Percentage population aged 25-34 having completed tertiary education (e) and its percentage change (f)

(c)

SMEs introducing marketing or organisational innovations as % of SMEs



(d)

**Percentage Change 2010-2017
SMEs introducing marketing or organisational innovations as % of SMEs**

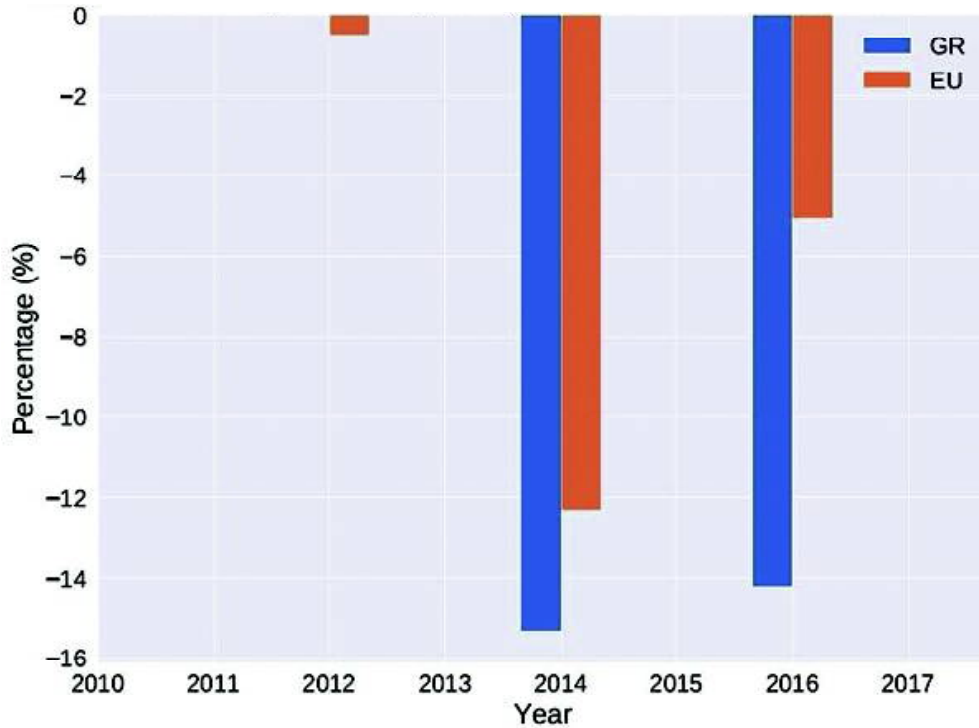
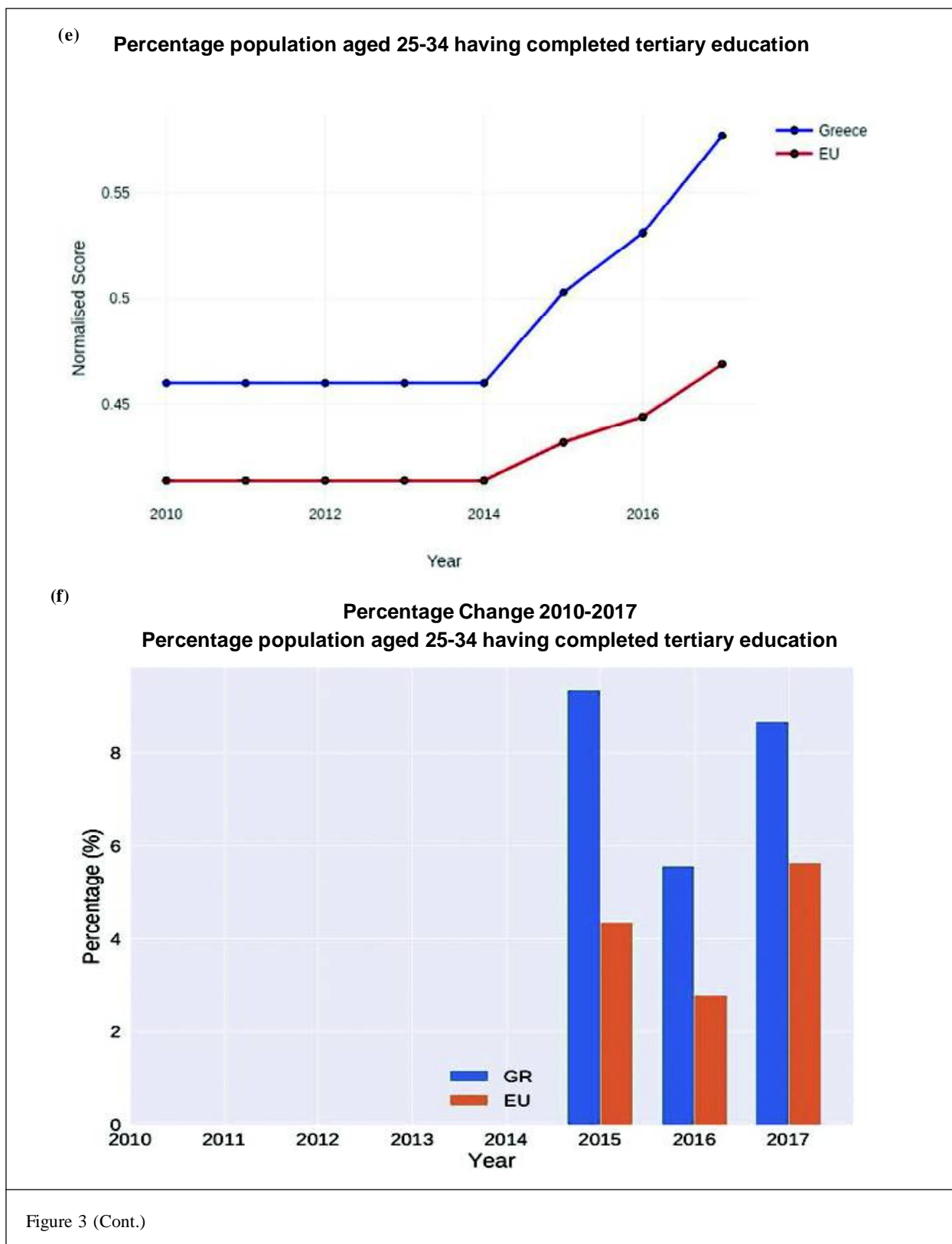


Figure 3 (Cont.)



5. Predictive analytics

5.1. Trend analysis of indicators

This section provides trend analysis for indicators of Greece and the EU average. In addition, the two trend-lines are statistically compared. In Table 2 trend-line statistics of composite indicators for Greece and the EU are observed, indicating the upward or downward trends. All values in this table are rounded to 3 (2 in the case of z_{score}) significant decimal digits.

Table 2: Trend-line statistics - Greece and European Union and difference between them. With gray color, we denote statistical significance at 95% level ($\alpha = 0:05$)

| Trend-line | Greece | | | European Union | | | Difference | | | |
|-------------------------|----------|-----------|---------------|----------------|-----------|---------------|--------------|-----------|----------------|----------------|
| | <i>b</i> | <i>se</i> | <i>pvalue</i> | <i>b</i> | <i>se</i> | <i>pvalue</i> | <i>bdiff</i> | <i>se</i> | <i>z-score</i> | <i>p-value</i> |
| Summary_Innovation_I | -0.002 | .003 | .575 | .005 | .001 | .011 | -0.007 | .003 | -2.23 | .013 |
| Human_Resources | .011 | .002 | .008 | .014 | .002 | .000 | -0.003 | .003 | -1.09 | .139 |
| Research_Systems | .011 | .001 | .001 | .007 | .001 | .000 | .004 | .002 | 2.25 | .012 |
| Innovation-friendly_env | .005 | .005 | .393 | .017 | .006 | .032 | -.013 | .008 | -1.62 | .053 |
| Finance_and_support | .019 | .004 | .004 | .016 | .006 | .054 | .004 | .007 | .520 | .302 |
| Firm_investments | .002 | .003 | .446 | .013 | .003 | .010 | -.010 | .004 | -2.49 | .006 |
| Innovators | -.025 | .010 | .062 | -.017 | .004 | .008 | -.008 | .011 | -7.17 | .237 |
| Linkages | .001 | .004 | .785 | .003 | .004 | .419 | -.002 | .005 | -3.86 | .350 |
| Intellectual_assets | .016 | .002 | .000 | -.002 | .001 | .121 | .018 | .002 | 9.43 | .000 |
| Employment_impacts | .009 | .005 | .149 | -.001 | .002 | .727 | .010 | .006 | 1.73 | .042 |
| Sales_impacts | -.045 | .010 | .005 | .006 | .002 | .060 | -.051 | .010 | -5.18 | .000 |
| Broadband_penetration | .010 | .006 | .144 | .035 | .009 | .010 | -.025 | .010 | -2.39 | .008 |
| Venture_capital | -.007 | .002 | .011 | .035 | .013 | .048 | -.042 | .013 | -3.14 | .001 |
| Design_applications | .015 | .002 | .000 | -.003 | .001 | .068 | .019 | .002 | 8.43 | .000 |
| Trademark_apps | .030 | .003 | .000 | .005 | .001 | .006 | .025 | .003 | 7.09 | .000 |
| Employment_activities | .009 | .005 | .151 | .006 | .001 | .000 | .003 | .005 | .506 | .307 |
| Enterprises_training | -.010 | .010 | .372 | .016 | .005 | .021 | -.026 | .011 | -2.35 | .009 |
| Innovative_Smes | .010 | .008 | .244 | .010 | .009 | .298 | .000 | .012 | .030 | .488 |
| International_public | .015 | .001 | .000 | .013 | .000 | .000 | .002 | .001 | 2.04 | .021 |
| Knowledge_exports | -.042 | .004 | .000 | .006 | .001 | .003 | -.048 | .004 | -11.88 | .000 |
| New_doctorate_grads | .004 | .003 | .235 | .031 | .006 | .004 | -.027 | .007 | -4.01 | .000 |
| Non_rd | .005 | .007 | .498 | .014 | .007 | .114 | -.008 | .010 | -8.07 | .210 |
| Opportunity_enterpre | -.001 | .005 | .896 | -.000 | .003 | .937 | -.000 | .006 | -.062 | .475 |
| Pct_patent | .003 | .001 | .069 | -.006 | .001 | .007 | .009 | .002 | 4.75 | .000 |
| Percentage_tertiary_edu | .019 | .004 | .007 | .009 | .002 | .009 | .010 | .005 | 2.15 | .016 |
| Percentage_lifelong_lea | .009 | .002 | .008 | .001 | .000 | .039 | .008 | .002 | 3.70 | .000 |
| Private_co_funding | -.003 | .008 | .683 | -.001 | .001 | .240 | -.002 | .008 | -2.47 | .403 |
| Public_private_pubs | -.004 | .003 | .314 | .001 | .002 | .698 | -.004 | .004 | -1.18 | .118 |
| Rd_business | .011 | .003 | .011 | .008 | .001 | .001 | .003 | .003 | 1.06 | .144 |
| Rd_public | .046 | .008 | .003 | -.004 | .001 | .042 | .050 | .008 | 5.90 | .000 |

Table 2 (Cont.)

| Trend-line | Greece | | | European Union | | | Difference | | | |
|--------------------|----------|-----------|---------------|----------------|-----------|---------------|--------------|-----------|----------------|----------------|
| | <i>b</i> | <i>se</i> | <i>pvalue</i> | <i>b</i> | <i>se</i> | <i>pvalue</i> | <i>bdiff</i> | <i>se</i> | <i>z-score</i> | <i>p-value</i> |
| Sales | -.100 | .030 | .021 | -0.02 | .006 | .809 | -.098 | .031 | -3.20 | .001 |
| Scientific_pubs | .006 | .003 | .085 | .003 | .001 | .009 | .003 | .003 | 1.09 | .139 |
| Smes_in_house | -.012 | .014 | .445 | -.012 | .005 | .063 | .001 | .015 | .038 | .485 |
| Smes_marketing | -.044 | .007 | .001 | -.020 | .004 | .002 | -.024 | .008 | -3.10 | .001 |
| Smes_product | -.018 | .015 | .279 | -.018 | .005 | .012 | -.000 | .015 | -.016 | .494 |
| Exports_technology | .006 | .005 | .312 | .014 | .003 | .005 | -.008 | .006 | -1.37 | .086 |

We start with the composite indicators and we highlight the summary of innovation score. The observable trends in Summary innovation index in aforementioned tables are positive for the EU, but negative for Greece. The score of Greece decreases by factor of 0.2% per year, while the score of the EU increases 0.5% per year. However, the slope of Greece’s trend-line is not significant at 95% statistical significance ($p_{value} = 0.575$), while EU is statistical significant ($p_{value} = 0.011$). Please note that the difference between the trend-lines of Greece and EU is calculated as described in three using a statistical z_{test} . The difference of Greece and EU trend-lines in summary innovation index is statistically significant.

Let us focus next on statistically significant trends ($p_{value} < 0.05$). For Greece, there is upward trend in human resources, research systems, finance and support, intellectual assets, and sale impacts. The first two exhibit an increment by a factor of 1.1% every year, while finance and support and intellectual assets show 1.9% and 1.6% respectively. Sale impacts of Greece decrease by a factor of 4.5% every year.

EU shows upward trend in summary innovation index, human resources, research systems, innovation-friendly environment, firm investments and innovators. EU human resources and firm investments show an increment of 1.4% and 1.3% every year respectively. Research systems and innovation-friendly environment display increase in score of 0.7% and 1.7% every year respectively. However, innovators in EU average present a decrease by 1.7%.

The trends of Greece’s composite indicators versus EU are compared in a statistical manner in Table 2, according to the methodology described above. Significant difference is observed in trend-lines between EU and Greece in summary innovation index, research systems, firm investments, intellectual assets, employment impacts and sales impacts. We report the slope difference b_{diff} , the standard error se , the z_{score} of the statistical test and the p_{value} , in Table 2.

Next we move to simple indicators. The upward or downward trends are indicated in both trend-lines of Greece and EU and then the two trend-lines are statistically compared. We concentrate again on indicators with $p_{value} < 0.05$, indicated by the gray shaded values in tables.

Greece’s R&D expenditure in the public sector and Trademark applications are growing in a fast pace. However, Sales of new-to-market and new-to-firm innovations of Greece are rapidly decreasing.

The rate of change in indicators of EU is characterized more stable than of Greece. This is expected since EU indicator scores are average values of EU countries. However we would like to highlight the fast pace of increment in broadband penetration, venture capitals and new doctorate graduates. EU SMEs introducing product or process innovations is decreasing by a factor of 2.0% year by year.

5.2. Indicator importance

This section presents and analyzes the results of our methodology obtained through machine learning techniques concerning the importance of the indicators.

For visualization and interpretation purposes, heat-maps are utilized for pairwise correlation analysis. Figures 4 and 5 show correlation heat-maps for the EU and Greece indicators respectively.

Furthermore, a Mantel test is used to statistically investigate the correlation between our two correlation matrices, following the methodology presented in section 3 above. The correlation coefficient of these two matrices is $r = 0.08$ while $p_{value} = 0.13$. This specific value of the coefficient r shows that there is no correlation between these two matrices, i.e., between the correlations of indicators of Greece and the EU. However, the results are not statistical significant since

the $p_{value} > 0.05$ in a two-sided test. Thus, we are unable to come to a statistical significant conclusion from the comparison of these two tables which seem not to be correlated. This observation might require further elucidation and commenting.

Also, the indicators correlation matrices of the Netherlands and the EU are compared using the Mantel test. In this case $r = 0.17$ and $p_{value} = 0:04$. Therefore, there exist statically significant results at 95% level of significance, that the EU and the Netherlands may have a weak correlation. However, $r = 0.17$ is a rather small coefficient and commonly displays no correlation. For completeness and comparison purposes, the correlation of the matrices related to Greece and the Netherlands are examined. Here we get $r = 0.07$, but we cannot conclude no-correlation between those two, due to the low statistical evidence ($p_{value} = 0.07$).

Moreover, another metric is utilized to measure the degree of differences in these correlation matrices, the Frobenius norm of each matrix, as presented in section 3. We point out that the degree of difference between these correlation matrices as measured is the same. Specifically, we have $norm_{EU-GR} = 19.75$, $norm_{EU-NL} = 18.42$ and $norm_{GR-NL} = 18.94$.

The features with 0.90 or more correlation are removed from our study to prevent multi-collinearity problems. The indicators that have been dropped are marked with the symbol 7 in the last two columns in table 1 for the Greece and EU cases respectively. After dropping these features we calculate indicator importance as described in section 3. The importance of the top indicators of Greece versus the EU may be graphically compared in Figure 6. If we focus only on the top-five important features affecting the most the fluctuation of summary innovation value for EU and Greece we identify only two that are in common; the venture capital and the Design applications.

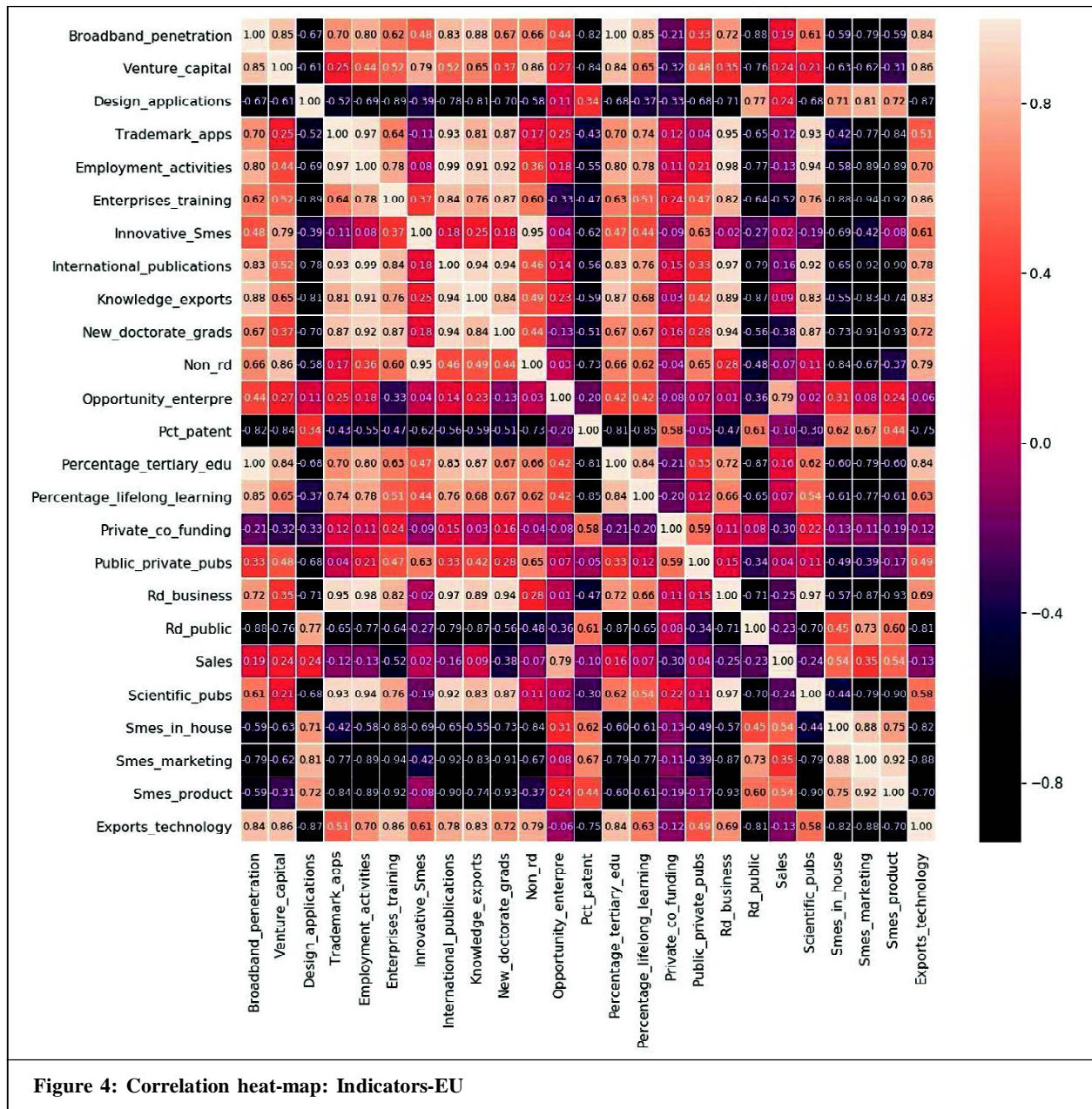


Figure 4: Correlation heat-map: Indicators-EU

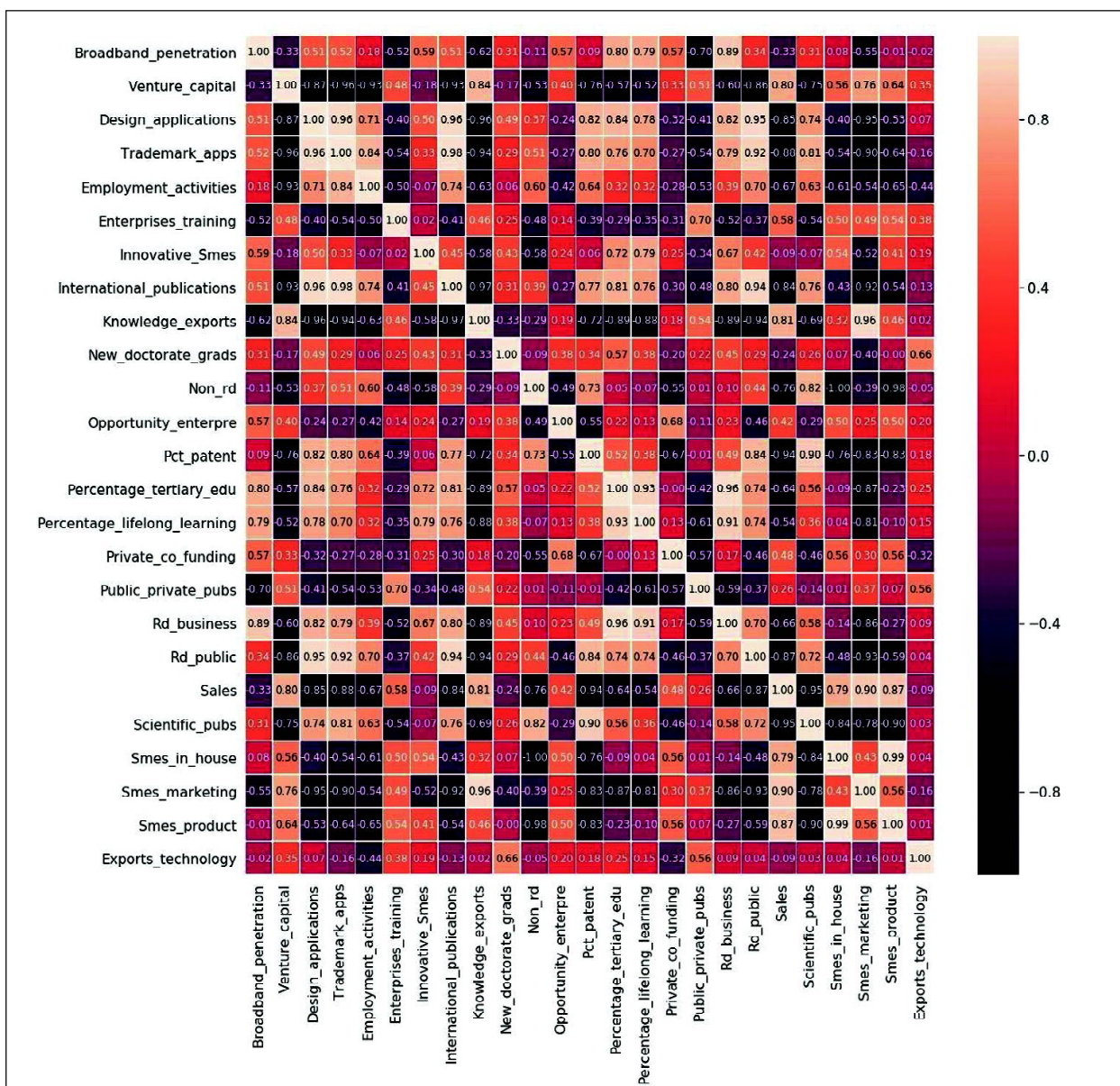


Figure 5: Correlation heat-map: Indicators-Greece

Venture capital plays a significant role in both the Greek and EU innovation performance. In fact, it is the catalyst for start-ups and entrepreneurship. It facilitates innovations and enables them to be evolved into marketable products. It allows the financing of business ideas that would otherwise have no chance of gaining access to the capital required. In addition, designing innovative applications and products plays surely a fundamental role in country’s innovation output.

According to our results, with the exceptions of venture capital and designing applications and products, EU average gives a high credit in high level features to increase innovation such as exports of medium and high technology products, broadband penetration and public-private copublications. It is meaningful that exports of technology products play an important role in innovation output. In addition, research connections and productive collaboration projects between researchers from the private sector and researchers from the public sector, resulting in academic publications, are significant in the increment of innovation. Actually research is mostly what drives innovation. Also, facilities and especially high speed internet and networking consolidate the e-potential of EU. To realize the full e-potential of Europe depends on establishing the conditions for the flourishing of electronic commerce and the Internet. Broadband penetration plays an interesting role in innovation output.

Greece seems to rely also significantly on well-educated people and innovative SMEs in order to increase innovation output. New doctorate graduates and people 25-34 having completed tertiary education play certainly an important role on innovation output. In fact, innovative ideas come mostly from educated people. SMEs in Greece represent 99.9% of

the total private sector of the country. Specifically, micro enterprises (1-5 employees and below 1mil. revenues) represent about 96,6% of the private sector and about 56% of the total employment of the Greek economy. In those terms, SMEs are the most significant part of the Greek and European economies, affecting directly both the financial and the social aspects of economic life. Innovative SMEs collaborating with other enterprises or institutions are an important indicator of Greece's innovation. The flow of information and knowledge between public research institutions and companies, and between companies and other firms, tends to be important for the development of innovation.

6. Synopsis and concluding remarks

The presented analysis focus on providing comparisons of Greece and EU average concerning innovation. It assesses systematic over-performance or under-performance of Greece compared to EU. It report the trends of indicators, upwards (positive trend) or downwards (negative trend), and the importance of each indicator to total innovation output regarding Greece and EU.

Columns "Performance" and "Trend" in Table 1, summarize the innovativeness of Greece compared to EU. Please note that every empty cell denotes not enough statistical evidence to decide (statistical insignificant). The missing indicators for Greece are the Foreign doctorate students as a % of all doctorate students and Employment in fast-growing enterprises (% of total employment).

Trend-line analysis is commonly used as a forecasting tool. In this direction, Tables 1, 2, provide us with statistical evidence to make predictions on innovativeness of Greece about the years to come. In Table 3, we highlight the five most important indicators, according to our methodology, affecting the fluctuation of the summary innovation index.

Our data suggest that Greece should take actions to increase its innovation output by focusing not only on the top-five important indicators of Greece level, indicated above. It should also try to follow the model of EU towards the increment of indicators highlighted in Table 3.

The efforts in this paper is based on the utilization of the data from EIS version 2018. Specifically, the normalized scores of composite indicators and simple indicators from 2010-2017 of Greece and EU average are selected. Data charts of indicators and percentage change each year are presented in the data analysis. Greece with EU average in country-level innovativeness is compared by utilizing statistics and hypothesis testing. Over-performance of Greece versus EU is found in the composite indicator of Innovators and in simple indicators, namely innovative SMEs collaborating with others, international scientific copublications, non-R&D innovation expenditures, percentage population aged 25-34 having completed tertiary education, sales of new-to-market and new-to-firm innovations and SMEs introducing marketing or organizational innovations.

In addition, the comparative analysis of the linear trend-lines of Greece and EU indicators in a statistical manner, utilizing generalized least-squares regression method with auto-regressive errors is provided. Greece shows positive significant trend relative to the EU in composite indicators, namely Research systems, Intellectual assets and Employment impacts, and in indicators, namely Design applications, Trademark applications, International scientific co-publications, PCT patent applications, Percentage population aged 25-34 having completed tertiary education, Percentage population aged 25-64 involved in lifelong learning, R&D expenditure in the public sector. For more information on systematic over-performance or under-performance of Greece versus EU and trend-line analysis, please view tables 1 and 2.

The impact of indicators on the volatility of the summary innovation index is calculated by using four well-known classifier models to build and employ a model based feature significance analysis for Greece and the EU. Specifically the models are the Logistic Regression, Random Forest, Extra-Trees and Support Vector Machines. Indicator correlation analysis provides us with evidence to exclude some indicators from our modeling. Calculated by using four well-known classifier models to build and employ a model based feature significance analysis for Greece and the EU.

The analysis shows that there is no significant association between the correlation matrices of Greece and the EU. However, the result is statistical insignificant. Thus, we do not have enough statistical evidence to conclude no association between these two. For completion and comparison purposes, we examine the correlation between Greece and the Netherlands, and the EU and the Netherlands. The degree of difference between those in terms of indicators correlation is to a great extends the same.

The fact that only relatively limited data instances are available leads us to cross-validate training of models, specifically 3-fold cross-validation. The estimate of feature importance of each fold is kept and then average for each model. Finally, after transforming values of importance in percentage values for each model the percentage values are summarize. For further information on indicator importance of EU and Greece, please view Figure 6, or the more compact Table 3.

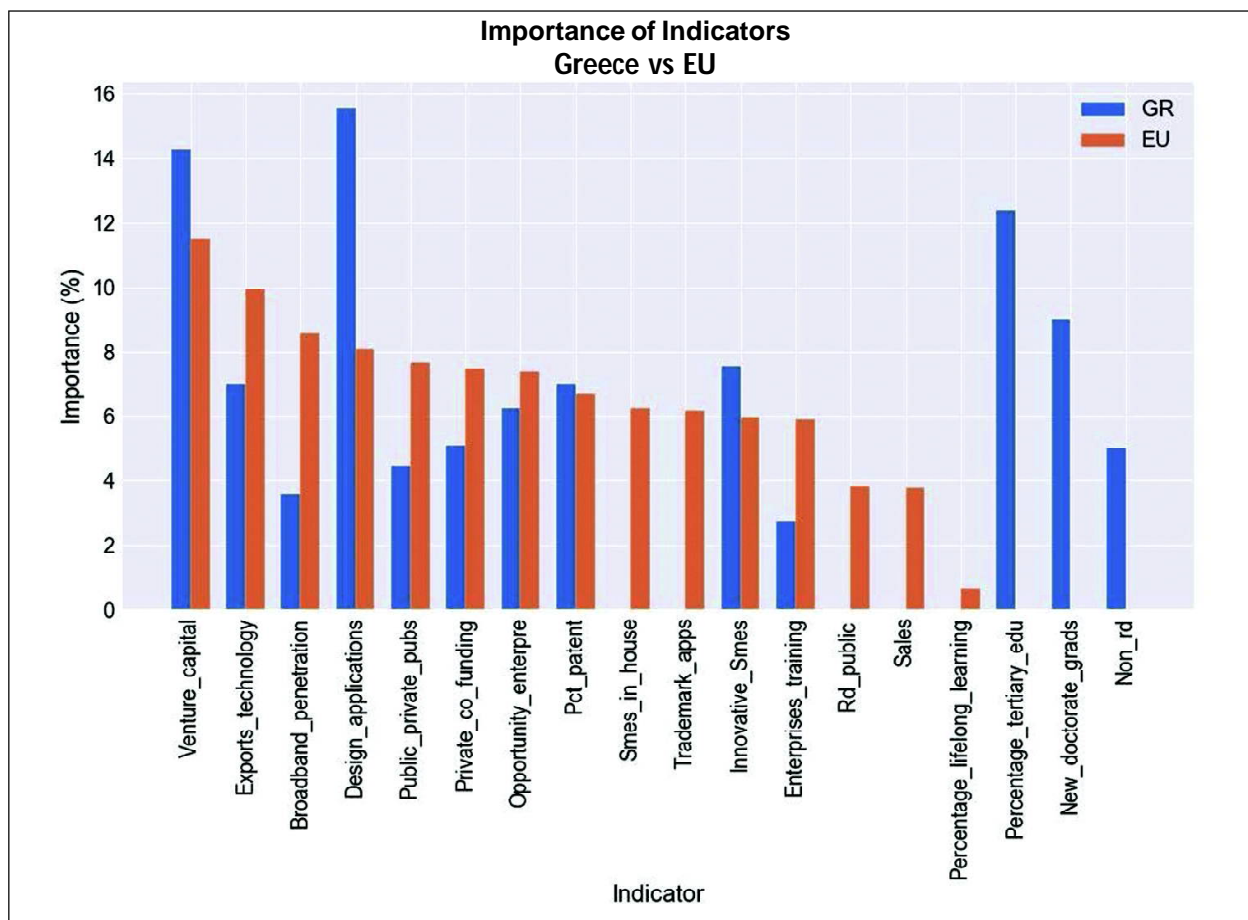


Figure 6: Importance of indicators-Greece vs European Union (EU)

| Greece | | EU | |
|-------------------------|------------|-----------------------|------------|
| Indicator | Importance | Indicator | Importance |
| Design_applications | 16% | Venture_capital | 12% |
| Venture_capital | 14% | Exports_technology | 10% |
| Percentage_tertiary_edu | 12% | Broadband_penetration | 9% |
| New_doctorate_grads | 9% | Design_applications | 8% |
| Innovative_Smes | 8% | Public_private_pubs | 8% |

In an effort to further investigate our results, a Principal Component Analysis (PCA) has been carried out for the EU and Greek correlation matrices, as proposed in (Jolliffe and Cadima, 2016). It turns out that in both cases the eigenvalues are, as expected, non-negative real numbers. Most of them are zero. Then non-zero eigenvalues are plotted in Figure 7. Furthermore, all the eigenvectors are real with all of their element in the interval $(-1, 1)$.

The first two principal components cumulatively characterize about 75% of the variance of the data in each case. Figure 8 shows the two biplots of the EU and Greek case. A biplot is a display that attempts to represent both the

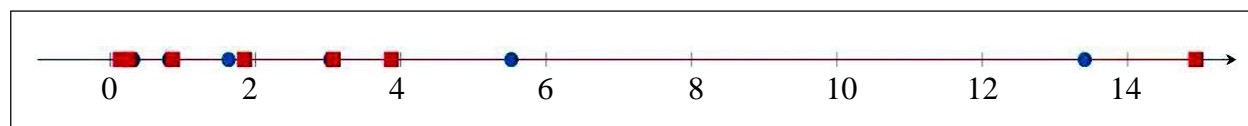


Figure 7: Plot of all non zero eigen values for the EU (red) and the Greek (blue) indicators correlation matrices

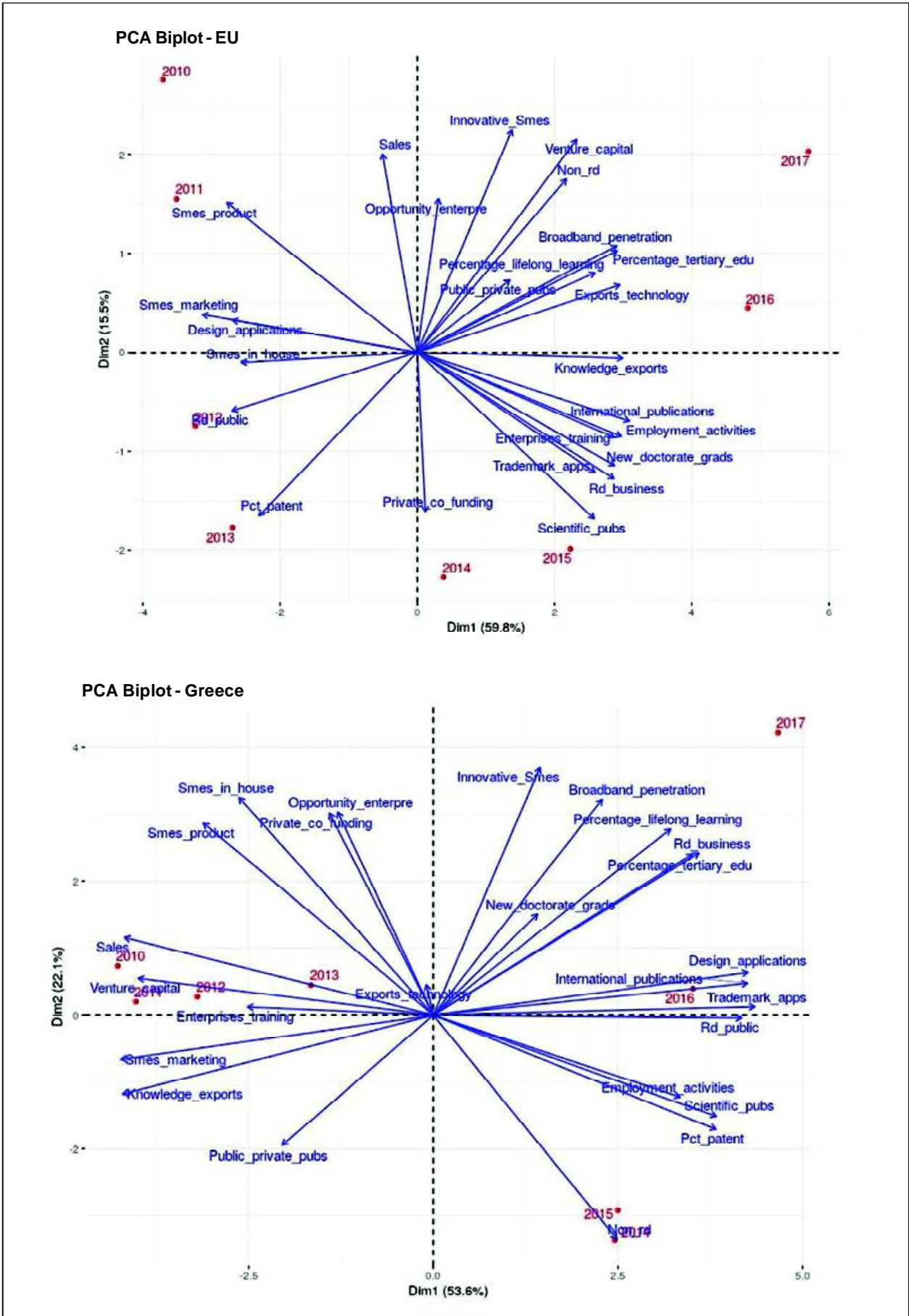


Figure 8: PCA - Biplots of Indicators of the EU and Greece.

observations and variables of multivariate data in the same plot and their contribution to the principal components (Jolliffe and Cadima, 2016). The axes of the above-mentioned figures are the principal components 1 and 2. Also, positively correlated variables point to the same side of the plot, while negatively correlated variables point to opposite sides of the graph. On both graphs, we are able to recognize groups of positive and negative correlated indicators.

Figure 8 clearly exhibits the fundamentally different behavior of indicator data of Greece versus the EU through the years. Clusters can be easily identified in the case of Greece. For example, we can identify four clusters-groups. The first one containing data from the years 2010 to 2013, the second one data from 2014 and 2015, and the third and fourth 2016 and 2017 respectively. On the other side, in the case of the EU, if we want to discriminate four clusters on data, we can say that cluster one contains data from 2010 and 2011, cluster two 2012 and 2013, cluster three and four 2014, 2015 and 2016, 2017 respectively.

Regardless the above observations, the complete Eigen analysis of our correlation matrices seems to be a well justified interesting subject. It is though beyond the scope of this paper. It is foreseen that our research study has the potential to provide explanations and evidences to assist the country assess its strengths and weaknesses regarding its innovation performance and as an extension its economic growth. By comparisons with EU countries (average), we display the position of Greece relative to the EU.

The center of gravity of the scientific merit of our work is closer to the area of data science rather than to the economic analysis. Therefore, a rigorous and systematic analysis of our results at the economic and development level is surely needed. Nevertheless, such a study is beyond the scope of this paper.

Further future work concerns deeper analysis of particular, existing or new, mechanisms of innovation. The utilization of data from different and multiple data sources as mentioned in the studies presented in section 2 will further contribute to our study. Data sources such as interviews of executive managers or chief executives of enterprises or academics from numerous institutions will allow us to examine in depth particular indicators of innovation. Micro data from Eurostat is expected to also enhance and clear our vision by adding significant new insight.

Please note that a similar study at the regional level for Greece, analyzing and comparing the innovation performance of regions using the indicators provided by the Regional Innovation Scoreboard database is underway and it will be presented elsewhere.

Acknowledgments

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code:T1EDK-02161). Also, this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860843".



Marie Skłodowska-Curie
Actions

References

- Asheim, Bjørn et al. (2003). *Regional Innovation Policy for Small-Medium Enterprises*. Edward Elgar Publishing. DOI: 10.4337/9781781009659.
- Beneki, Christina, Dimitrios Giannias, and George Moustakas (2012). Innovation and economic performance: the case of Greek SMEs. In: *Regional and Sectoral Economic Studies*. 12(1), 43-54.
- Carnegie, Roderick and Business Council of Australia. (1993). *Managing the innovating enterprise: Australian companies competing with the world's best*. Business Library. 427. ISBN: 1863501517. URL: <https://catalogue.nla.gov.au/Record/1573090>
- Chien, Shih-Chien, Tai-Yue Wang, and Su-Li Lin (2010). Application of neuro-fuzzy networks to forecast innovation performance – The example of Taiwanese manufacturing industry. *Expert Systems with Applications*. 37(2), 1086-1095. DOI: 10.1016/J.ESWA.2009.06.107. <https://www.sciencedirect.com/science/article/>
- Drucker, Peter F. (Peter Ferdinand) (1954). *The practice of management*. 1st ed. New York: Harper & Row, p. 404.
- Eleftherios, Aggelopoulos., Nikolaos, Eriotis., Antonios, Georgopoulos., and Anastasios, Tsamis (2016). RD activity and operating performance of small and medium-sized enterprises (SMEs): The case of a small open economy. *Journal of Accounting and Taxation*. 8(4), 40-50. DOI: 10.5897/JAT2016.0233.
- European Commission (2018). *DocsRoom - European Commission EIS 2018 Methodology Report*. <https://ec.europa.eu/docsroom/documents/30081>
- Hajek, Petr., and Roberto, Henriques (2017). Modelling innovation performance of European regions using multi-output neural networks. *PLOS ONE* 12(10), Ed. by María Carmen Díaz Roldán, e0185755. DOI: 10.1371/journal.pone.0185755. <https://dx.plos.org/10.1371/journal.pone.0185755>
- Hajek, Petr., and Jan, Stejskal (2015). Predicting the innovation activity of chemical firms using an ensemble of decision trees. In: *2015 11th International Conference on Innovations in Information Technology (IIT)*. IEEE. 35-39. ISBN: 978-1-4673-8509-1. DOI: 10.1109/INNOVATIONS.2015.7381511.
- Henderson, John C., and Christine, M.A. Lentz (1995). Learning, working, and innovation: a case study in the insurance industry. *Journal of Management Information Systems*. 12(3), 43-64. DOI: 10.1080/07421222.1995.11518090.
- Hyvärinen, Liisa (1990). Innovativeness and its indicators in small- and medium-sized industrial enterprises. *International Small Business Journal: Researching Entrepreneurship*. 9(1), 64-79. DOI: 10.1177/026624269000900106. <http://journals.sagepub.com/doi/10.1177/026624269000900106>
- Industrial policy | Internal Market, Industry, Entrepreneurship and SMEs. https://ec.europa.eu/growth/industry/policy_en
- Innobarometer | Internal Market, Industry, Entrepreneurship and SMEs. https://ec.europa.eu/growth/industry/innovation_en
- Jan van den Ende and Timo van Balen (2017). *Innovativeness of the Netherlands relative to EU countries*. Tech. rep. Rotterdam, the Netherlands: Rotterdam School of Management Erasmus University.
- Jolliffe, Ian T., and Jorge, Cadima (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*. 374(2065), 1-16. DOI: 10.1098/rsta.2015.0202. URL: <http://www.ncbi.nlm.nih.gov/pubmed/26953178> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4792409>
- Kalapouti, Kleoniki et al. (2017). Measuring efficiency of innovation using combined data envelopment analysis and structural equation modeling: Empirical study in EU regions. *Annals of Operations Research*. 1-24. DOI: 10.1007/s10479-017-2728-4.
- Klimova, Nataliya, Oleg Kozyrev., and Eduard, Babkin (2016). *Innovation in Clusters*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-21109-1. <http://link.springer.com/10.1007/>
- Mantel, N. (Feb. 1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27(2), 209-220. URL: <http://www.ncbi.nlm.nih.gov/pubmed/6018555>
- McKinney, Wes., Josef Perktold., and Skipper, Seabold (2011). *Time Series Analysis in Python with statsmodels*. Proceedings of the 10th Python in Science Conference. Ed. by Stéfan van der Walt and Jarrod Millman. 96–102.
- Monitoring innovation | Internal Market, Industry, Entrepreneurship and SMEs. URL: https://ec.europa.eu/growth/industry/innovation/facts-figures_en
- Paternoster, Raymond et al. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*. 36(4), 859–866. DOI: 10.1111/j.1745-9125.1998.tb01268.x.

- Paz-Marín, Mónica de la, Pilar, Campoy-Muñoz, and César, Hervás-Martínez (2012). Non-linear multiclassifier model based on Artificial Intelligence to predict research and development performance in European countries. *Technological Forecasting and Social Change*. 79(9), 1731-1745. DOI: 10.1016/J.TECHFORE.2012.06.001. <https://www.sciencedirect.com/science/article/pii/S0040162512001485?via%3Fihp>
- Roszek-Wójtowicz, Elzbieta., and Jacek, BiaBek (2017). Evaluation Of The Eu Countries' Innovative Potential - Multivariate Approach. *Statistics in Transition*. New Series 18.1, 167-180. ISSN: 1234-7655. DOI: 10.21307/stattrans-2016-064. http://www.exeley.com/statistics_in_transition/doi/10.21307/stattrans-2016-064— (2018). “Diverse approaches to the multidimensional assessment of innovation in the European union”. In: *Acta Oeconomica* 68(4), 521-547. DOI: 10.1556/032.2018.68.4.3.
- Saberi, Sara., and Rosnah, Mohd. Yusuff (2012). Neural network application in predicting advanced manufacturing technology implementation performance. *Neural Computing and Applications*. 21(6), 1191-1204. DOI: 10.1007/s00521-010-0507-0.
- Schumpeter, Joseph, A. and Redvers, Opie. (1934). *The theory of economic development; an inquiry into profits, capital, credit, interest, and the business cycle*, Harvard University Press, 255. URL: <http://www.hup.harvard.edu/catalog.php?isbn=9780674879904>
- Souitaris, Vangelis (2001). Strategic influences of technological innovation in Greece. In: *British Journal of Management*. 12(2), 131-147. DOI: 10.1111/1467-8551.00190. <http://doi.wiley.com/10.1111/1467-8551.00190>
- Tohidi, Hamid., and Mohammad, Mehdi, Jabbari (2012). The important of innovation and its crucial role in growth, survival and success of organizations. *Procedia Technology*. 1, 535-538. DOI: 10.1016/J.PROTCY.2012.02.116. <https://www.sciencedirect.com/science/article/pii/S221201731200117X>
- Wang, Tai-Yue., and Shih-Chien, Chien (2006). Forecasting innovation performance via neural networks—a case of Taiwanese manufacturing industry. *Technovation*. 26(5-6), 635-643. DOI: 10.1016/J.TECHNOVATION.2004.11.001. <https://www.sciencedirect.com/science/article/pii/S0166497204002135>
- Welch, B. L. (1947). The generalization of ‘student’s’ problem when several different population variances are involved. In: *Biometrika*. 34(1-2), 28-35. DOI: 10.1093/biomet/34.1-2.28.

Appendix

In this section, we provide the tables of our analysis and further information. Following, we make available all the visualizations and graphs utilized in our analysis in figshare. (<https://doi.org/10.6084/m9.figshare.9844907.v1>)

Long acronyms of names of indicators:

Summary_Innovation_I Summary innovation index.

Innovation-friendly_env Innovation-friendly environments.

Broadband_penetration Broadband penetration.

Venture_capital Venture capital (% of GDP).

Design_applications Design applications per billion GDP (in PPS).

Trademark_apps Trademark applications per billion GDP (in PPS).

Employment_activities Employment in knowledge-intensive activities (% of total employment).

Enterprises_training Enterprises providing training to develop or upgrade ICT skills of their personnel.

Innovative_Smes Innovative SMEs collaborating with others (% of SMEs).

International_publications International scientific co-publications per million population.

Knowledge_exports Knowledge-intensive services exports as % of total services exports.

New_doctorate_grads New doctorate graduates per 1000 population aged 25-34.

Non_rd Non-R&D innovation expenditures (% of turnover).

Opportunity_enterpre Opportunity-driven entrepreneurship (Motivational index).

Pct_patent PCT patent applications per billion GDP (in PPS).

Percentage_tertiary_edu % population aged 25-34 having completed tertiary education.

Percentage_lifelong_lea % population aged 25-64 involved in lifelong learning.

Private_co_funding Private co-funding of public R&D expenditures (% of GDP).

Public_private_pubs Public-private co-publications per million population.

Rd_business R&D expenditure in the business sector (% of GDP).

Rd_public R&D expenditure in the public sector (% of GDP).

Sales Sales of new-to-market and new-to-firm innovations as % of turnover.

Scientific_pubs Scientific publications among the top 10% most cited publications worldwide as % of total scientific publications of the country.

Smes_in_house SMEs innovating in-house as % of SMEs.

Smes_marketing SMEs introducing marketing or organizational innovations as % of SMEs.

Smes_product SMEs introducing product or process innovations as % of SMEs.

Exports_technology Exports of medium and high technology products as a share of total product exports.

Cite this article as: Thanasis Zoumpikas, Manolis Vavalis, Elias Houstis (2021). [Analysis of innovation with data science: The case of Greece. *International Journal of Data Science and Big Data Analytics*. 1\(1\), 20-42. doi: 10.51483/IJDSBDA.1.1.2021.20-42.](#)