Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 12, 25–46 DOI: 10.5604/01.3001.0014.5729
Statystyka w praktyce / Statistics in practice
Otrzymano/received: 3.03.2020, zaakceptowano/accepted: 12.11.2020

Investment in research and development in the ICT sector by top European Union companies

Agnieszka Kleszcz[a]

Abstract. Multiple studies have shown that Information and Communication Technology (ICT) has boosted the growth of the global economy and improved the quality of life. At present, a significant proportion of innovations and new patents are held by companies operating in the ICT sector.

This study has two primary aims. The first of them is an analytical comparison of industrial investments in research and development (R&D) in ICT-related companies in various countries from the European Union. The second is to examine the relationship between these investments and each country's digital performance illustrated by the Digital Economy and Society Index (DESI). The main contribution of this paper is a proposal of a means to determining the quantitative relationship between R&D and DESI as well as the identification of the most important DESI component. The author focuses on the largest R&D investors from a list of 1,000 companies, created on the basis of data published by the Economics of Industrial Research and Innovation (IRI), covering the years 2013–2019. Both variables (DESI and R&D) have a clear joint group structure found by Ward's hierarchical clustering. Furthermore, an exponential law was identified predicting an increase of 18% in R&D expenditure when a 1 percentage point growth in DESI is observed. Among all the components of DESI, the applied random forest model proves human capital is the most important factor attracting R&D investments.

Moreover, a separate analysis of R&D relating to the analysed companies showed a growing trend in R&D investments with its predicted value reaching over 46 billion euro in 2021.

Keywords: ICT sector, industrial R&D investment, Digital Economy and Society Index, innovation

JEL: O30, C10

Inwestycje w badania i rozwój czołowych firm sektora ICT z krajów Unii Europejskiej

Streszczenie. Badania wskazują, że technologie informacyjno-komunikacyjne (ICT) istotnie wpływają na rozwój światowych gospodarek i poprawiają jakość życia społeczeństwa. Obecnie znaczna część innowacji oraz zgłoszonych patentów dotyczy sektora ICT. Pierwszym celem artykułu jest analiza inwestycji przemysłowych w badania i rozwój (B+R) firm należących do sektora ICT, które mają siedzibę w Unii Europejskiej. Drugi cel polega na zbadaniu związku tych inwestycji z poziomem cyfryzacji przy wykorzystaniu złożonego Indeksu Gospodarki Cyfrowej i Społeczeństwa Cyfrowego (DESI). Główny wkład badania omawianego w artykule to propozycja ustalenia ilościowej zależności pomiędzy B+R oraz DESI, a także identyfikacja najistotniejszego składnika DESI. Analizy oparto na rankingu tysiąca największych inwestorów B+R, opracowanym na podstawie danych z Economics of Industrial Research and Innovation (IRI) za lata 2013–2019. Zastosowano hierarchiczne grupowanie Warda.

[a] Uniwersytet Jana Kochanowskiego w Kielcach, Wydział Nauk Ścisłych i Przyrodniczych, Instytut Geografii i Nauk o Środowisku / Jan Kochanowski University of Kielce, Faculty of Natural Sciences, Institute of Geography and Environmental Sciences. ORCID: https://orcid.org/0000-0002-0450-5247.

W wyniku analiz stwierdzono, że obie badane zmienne mają wyraźną strukturę grupową. Ponadto zaobserwowano istnienie zależności wykładniczej przewidującej wzrost wydatków na B+R o 18%, kiedy DESI wzrośnie o 1 p.proc. Model lasów losowych posłużył do wskazania kapitału ludzkiego jako najważniejszego czynnika wpływającego na wydatki na B+R poszczególnych firm.

Przeprowadzone badanie pozwala zauważyć rosnący trend w inwestycje B+R w sektorze ICT o przewidywalnej wartości ponad 46 mld euro w roku 2021.

Słowa kluczowe: sektor ICT, inwestycje w badania i rozwój w przemyśle, Indeks Gospodarki Cyfrowej i Społeczeństwa Cyfrowego, DESI, innowacje

1. Introduction

Information and Communication Technologies (ICT) are a family of technologies that process, collect and send information in an electronic form (EPID, 2015; Investin, 2017). There is a significant relationship between ICT and the economic growth in developed countries (Jin & Cho, 2015; Sepehrdoust, 2018). Cross-sectional studies show that ICT development has played a key role in boosting the growth of productivity in developed countries (Doong & Ho, 2012; GUS, 2017; Hong, 2017; Jorgenson & Vu, 2016; Pradhan et al., 2018; Sepehrdoust, 2018; Sylwestrzak, 2018; Vu, 2013; Żelazny, 2015). ICT has also brought new forms of employment: the number of ICT specialists[1] in the EU grew by 40.0% in the years 2011 to 2019. In 2019, approximately 7.8 million persons worked as ICT specialists across the EU. The highest number (1.7 million) were employed in Germany, which accounted for 21.5% of the total ICT staff in the EU. The second largest ICT employer (14.4% of the EU total) was France, with 1.1 million ICT specialists (Eurostat, 2019).

There is a widespread agreement on the importance of innovation which is considered key drive for Europe's competitiveness, growth and jobs (Pesole, 2015). According to OECD & Eurostat (2005) 'An innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations'. Innovation has become increasingly important for company survival, growth, and profitability in rapidly changing business environments (Hilmersson & Hilmersson, 2020). At present, the ICT sector is responsible for a significant part of all innovatins in the economy. Digital innovations are revolutionising the way businesses and industries operate (EPO, 2020). Patent statistics published by the European Patent Office (EPO) show that digital technologies have taken the lead in terms of filed patent applications, and reflect their key role in the digital transformation. Patent applications in digital communications grew by 19.6%

[1] 'ICT specialists are defined as persons who have the ability to develop, operate and maintain ICT systems and for whom ICTs constitute the main part of their job' (OECD, 2004).

in 2019 (EPO, 2020). The digital megatrends include Artificial Intelligence, Internet of Things, Virtual and Augmented Reality, Big Data and Cloud technology (Liu, 2017; Schwab, 2017).

The available information is not sufficient to create a full and coherent picture of Research and Development (R&D) investment in the ICT sector, neither does it provide the particulars relating to the top R&D investors. Moreover, it fails to demonstrate clearly the factors determining such investment. Consequently, many studies present technical reports providing only basic statistical data, the evolution of the main companies' performance indicators or related background information (e.g. Hernández et al., 2018, 2019). However, some studies do analyse the internationalisation of the R&D activity (e.g. Daiko et al., 2017; Nepelski & De Prato, 2012).

This study describes investment in R&D by top EU investors from the ICT sector, with a focus on 16 countries where the company headquarters is registered. The analysis goes further by presenting the relationship between each country's volume of R&D activity and their corresponding digital performance, illustrated by the Digital Economy and Society Index (DESI).

This study has two primary aims. The first of them is an analytical comparison of industrial investments in R&D in ICT-related companies in various countries from the EU. The second is to examine the relationship between these investments and each country's digital performance illustrated by DESI.

2. Methodology

It is expected that companies based in countries with a highly-developed digital economy will invest more in R&D. This claim was studied through a statistical analysis of DESI (European Commission, 2020a) and a dataset of top R&D investors (European Commission, 2020c).

2.1. Dataset

The research examines changes in R&D investment since 2013 among the largest corporate R&D investors in the EU. For this purpose, the EU Industrial R&D Investment Scoreboard (further referred to as 'Scoreboard') from the European Commission Economics of Industrial Research and Innovation (IRI) database was analysed. The Scoreboard provides economic and financial data as well as analyses regarding the largest R&D investors in the EU. It is based on data drawn directly from each company's Annual Report. The Scoreboard is published annually and constitutes a reliable, timely benchmarking tool for making comparisons across sectors, geographical areas and companies; it also provides data which enables the monitoring and analysis of emerging investment trends and patterns. The Scoreboard refers to all of the R&D financed by a particular company through its own funds, although not necessarily carried out in the country in which the company is registered (Hernández et al., 2016, 2018, 2019).

A database of selected ICT industries from 2013 to 2019 was created. The headquarters of the EU top R&D investors from the ICT sector are based in 16 countries, however their subsidiaries could be geographically spread. Assigning companies to industrial sectors according to the existing classification systems could be problematic, because sector descriptions are often inconsistent with the firms' actual business activities, which results from the fact that many of them operate within more than one industrial area. However, companies usually indicate their main sector of activity in their annual reports.

As a part of the analysis of the ICT sector, the author extracted the companies of interest according to their sector classification ICB4 digits (Hernández et al., 2018) and the International Classification Benchmark (FTSE, 2012) from among a selection of firms from different fields. As a result, the ICT-related companies were grouped into the following industries:

- Software and Computer Services;
- Electronic and Electrical Equipment;
- Fixed Line Telecommunications;
- Mobile Telecommunication;
- Technology Hardware and Equipment.

In the period 2013–2019, out of 7,000 companies from all industries, as many as 1,684 were classified as ICT-related (24%). These companies were located in 16 countries (except 2019, when there was no firm from Portugal in the classification, leaving 15 countries that year).

2.2. Digital Economy and Society Index

The Digital Single Market strategy implemented by the European Commission (EC) aims to create digital opportunities for people and businesses and to strengthen Europe's position as the world leader in the digital economy. Further development of the digital economy is considered crucial for increasing the competitiveness of EU's economy (Eurostat, 2018). Currently, there are many indicators that parameterise the volume of digitisation at both the national and global level. Optimising the process of selecting indicators leads to an effective assessment of the dynamics of development and allows its prediction.

DESI is a composite index published every year by the European Commission EC since 2014, measuring the progress of EU countries towards a digital economy and society. The DESI measurement uses a 0–1 scale (alternatively expressed in percents). The DESI 2018 is composed of five principal policy areas (DESI dimensions) which are divided into 14 sub-dimensions, and those are finally organised into 34 indicators (Table 1). However, in order to improve the methodology and to measure the actual level of digital development in member states taking into account the latest technological developments, a number of changes were introduced to the DESI 2019. As a result, DESI 2019 consists of 44 indicators, including new ones which consider 5G readiness, Big Data, at least basic software skills, female ICT specialists, ICT graduates and Cloud or online courses, all of which are distributed throughout the five DESI dimensions.

In order to aggregate indicators expressed in different units into the dimensions and sub-dimensions of DESI, these indicators were subject to normalisation by means of the min-max method, which involves a linear projection of each indicator onto a scale between 0 and 1 (European Commission, 2018a, 2018b, 2020b; Kisielnicki, 2016).

Some dimensions, sub-dimensions and individual indicators are more relevant than others, and for this reason they have been given a greater weight in the computation of the final DESI (European Commission, 2020b, p. 14). The chart presents the overall weights attributed to the main DESI dimensions, which reflect the EU's digital policy priorities, and their short descriptions.

Chart. DESI dimensions

Dimension (with weight in %)	Sub-dimension (with number of indicators within sub-dimension)	Dimension description
1 – connectivity (25)	fixed broadband (2), mobile broadband (2), fast broadband (2), ultrafast broadband (2) and broadband prices (1)	measures the deployment of broadband infrastructure and its quality; access to fast and ultrafast broadband-based services is necessary to increase competitiveness
2 – human capital (25)	basic skills and internet use (2), advanced skills and development (2)	measures the skills needed to take advantage of the possibilities offered by digital technology
3 – use of internet service (15)	citizens' use of content (3), communication (2) and online transactions (2)	accounts for a variety of online activities, such as the consumption of online content, video calls and online shopping and banking
4 – integration of digital technology (20)	business digitalisation (5) and e-commerce (3)	measures the digitalisation of businesses and e-commerce; by adopting digital technologies, businesses can enhance their efficiency, reduce costs and engage customers and business partners to a greater extent; moreover, the internet as a sales outlet offers access to wider markets, thus providing considerable potential for growth

Chart. DESI dimensions (cont.)

Dimension (with weight in %)	Sub-dimension (with number of indicators within sub-dimension)	Dimension description
5 – digital public services (15)	e-Government (5) and e-Health (1)	measures the digitalisation of public services, focusing on e-Government and e-Health; modernisation and digitisation of public services can lead to their higher level of efficiency, benefiting the public administration, the citizens and businesses

Source: The author's work based on European Commission (2018a).

2.3. Local linear trend model

In order to analyse trends in quantitative terms and make predictions for the future, a local linear trend model (LLT) is fitted to the data. The LLT models time-series by introducing a hidden variables level (l) and slope (s), which are interpreted as the local value and local rate of changes, respectively. The observed time-series (o) is assumed to be a noisy observation of the level. Formally, the LLT is defined as below (Durbin & Koopman, 2012):

$$s_t = s_{t-1} + \epsilon_1, \qquad \epsilon_1 \sim Norm(0, \sigma_s), \tag{1}$$

$$l_t = l_{t-1} + s_t + \epsilon_2, \qquad \epsilon_2 \sim Norm(0, \sigma_l), \tag{2}$$

$$o_t = l_t + \epsilon_3, \qquad \epsilon_3 \sim Norm(0, \sigma_o), \tag{3}$$

where, for every time step t, s_t, l_t, o_t ϵ_1, ϵ_2 and ϵ_3 are random variables.

While s_t, l_t, and o_t depend on values at the previous time step $t-1$, the distribution of noises ϵ_1, ϵ_2 and ϵ_3 do not depend on time. The parameters of the model (σ_s, σ_l and σ_o) were fitted using Bayesian inference, implemented in the TensorFlow Probability library (Dillon et al., 2017). The confidence of the predictions is reported to reach a 95% confidence interval of the predictive posterior distribution.

2.4. Ward's hierarchical agglomerative clustering method

Clustering is a technique which groups similar data points so that the points in the same group are more similar to each other than to the points in other groups. The group of similar data points is called a cluster. There are many statistical techniques that utilise similarities or differences among data and variables. One of the most widespread hierarchical clustering method is the Ward method. The hierarchical clustering technique can be visualised by a tree-based representation of the observations, called a dendrogram. A hierarchical algorithm yields a dendrogram representing the nested grouping of patterns and similarity levels at which the groupings change. The clustering process is performed by merging the most similar patterns in the cluster set to form a bigger one (Bouguettaya et al., 2015; Gareth et al., 2017).

2.5. Linear regression model

The relationship between DESI and its impact on R&D investment was analysed by fitting a linear regression model. A linear regression model is a powerful statistical model of dependence between two variables. The regression equation for this model is as follows:

$$Y = \beta_0 + \beta_1 X, \quad (4)$$

where Y is the explained variable (R&D investment), while X is the predictor (value of DESI). The model parameters are β_0 and β_1.

In regression with a single independent variable, the coefficient signals how much the dependent variable is expected to increase (if the coefficient is positive) or decrease (if the coefficient is negative) when that independent variable increases by one unit.

Linear regression can be prone to providing inaccurate estimates in the presence of outliers or heteroscedastic errors. Robust Regression proves effective in overcoming the problem with outliers, but the coefficient's standard errors can still be biased due to heteroscedasticity. Considering the above, it is essential to check whether any of the above-mentioned problems exist prior to drawing any conclusions from the model. A formal test for heteroscedasticity could be performed by means of, e.g., the studentised Breusch-Pagan test.

By design, linear regression assumes that an individual observation is independent. However, this assumption can be easily violated in a number of ways, including the use of observations of the same quantity from multiple years in a single model. In this case, either the mixed-effects models could be used to increase the statistical strength of the model, or only the independent subset of the data can be used for model estimation.

Sometimes standard linear regression faces significant limitations in terms of its predictive power. An improvement can be made by reducing the complexity of the linear model, and, thus, the variance through the use of transform functions (e.g. a logarithm) (Gareth et al., 2017).

2.6. Random forest regression with feature importance

Random forest is an ensemble method based on regression trees. Multiple decision trees (DT) are trained to predict target variable values from a random subset of predictors. Each tree depends on the values of a random vector sampled independently, with the same distribution for all trees in the forest (Breiman, 2001).

Since random forest is a bagging model and every tree is trained on bootstrap samples, there are samples which are not used for training this particular tree. In the context of bagging models, they are called out-of-bag (OOB) samples and are used to estimate the generalisation error without wasting precious samples validating the model (Gareth et al., 2017).

Random forests could also be used for the assessment of feature importance. During the training process, each tree in the forest selects a predictor to split on; the order of predictors depends on the quality of the split data, measured by an impurity function (e.g. the Gini Index). The impurity function measures how well two sets of label objects are separated. The lower the impurity, the better the separation. The feature importance is measured as impurity decrease weighted by the probability of reaching a given node in the tree. The higher the impurity decrease, the more important the feature (Palczewska et al., 2013).

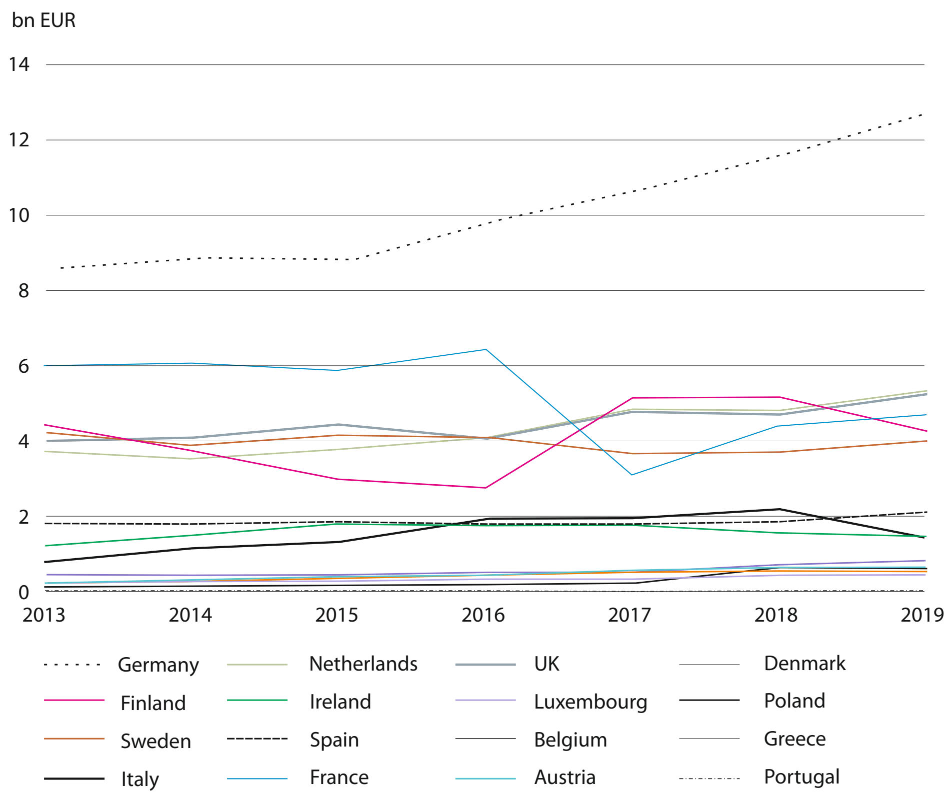
3. Statistical analysis of DESI and R&D – research results

The study analysed the volume of investment in research and development made by companies from the ICT sector of several EU member states and the relationship of that investment with each country's digital performance expressed by DESI.

3.1. Investment in R&D in the ICT sector

For a long time, the importance of investment in ICT and R&D to technical change and economic growth has been emphsised (Edquist & Henrekson, 2017). The world's most developed countries allocate the largest funds in R&D. At the same time, R&D activity has become more international in recent years in the majority of countries (De Prato & Nepelski, 2013; European Commission, 2012; Nepelski & De Prato, 2012). Investment in information and technology is essential to a country's integration with the global social system and increasing its competitive power (Bozkurt, 2015). Figure 1 illustrates the changes in R&D investments in the ICT sector over time. German companies invested the highest amounts, i.e. over 12,000 million euro in 2019, closely followed by the United Kingdom, the Netherlands and France. In 2019 the countries with the lowest investment in R&D included Greece and Poland. The remaining EU countries are not mentioned in the ranking as their investment in R&D did not achieve the required minimum; furthermore, the majority of these countries are characterised by a low DESI (Figure 4 p. 36). Among the analysed countries, the highest investments were made by the German company SIEMENS (operating in the Electronic and Electrical Equipment industry) with an R&D investment of 5,909 million euro. The second greatest investor was Nokia from Finland (from the Technology Hardware and Equipment industry) with spending on R&D amounting to 4,044 million euro. SAP from Germany (operating in the Software and Computer Services industry) invested 3,612 million euro, whereas the Swedish Ericsson (the company from the Technology Hardware and Equipment industry) earmerked 3.484 million euro for this purpose, thus placing itself fourth among the highest-investing companies.

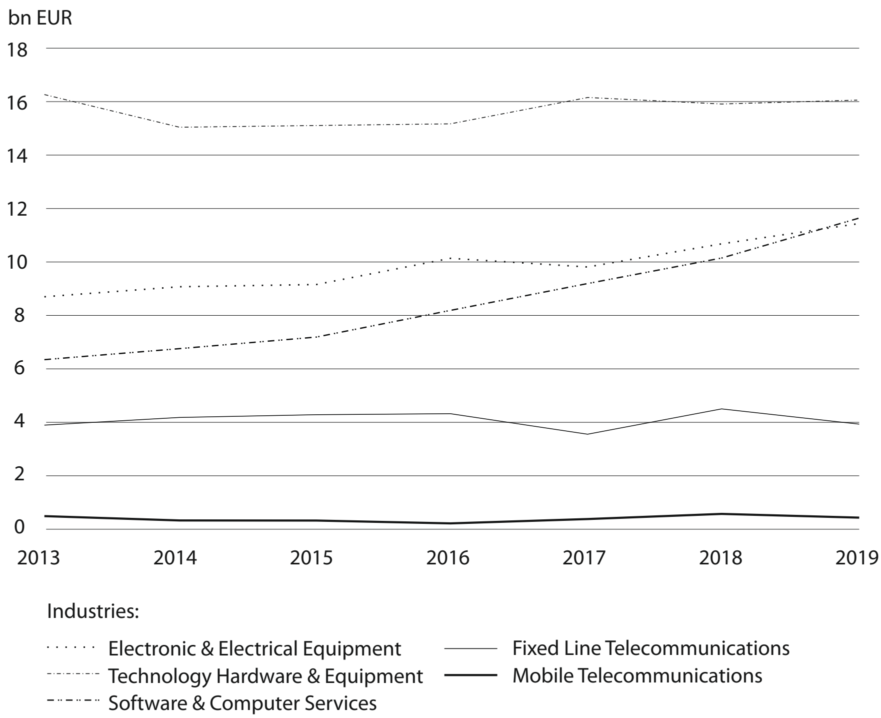
Figure 1. R&D investment in the ICT sector in selected EU countries



Source: the author's work based on IRI.

The industries of the ICT sector which invest in R&D to the greatest extent were Technology Hardware and Equipment, whose total annual investment amounted to almost 16 billion euro (Figure 2). On the other hand, the Mobile Telecommunications industry made the lowest investment in R&D.

Figure 2. R&D investment in the ICT sector by industries

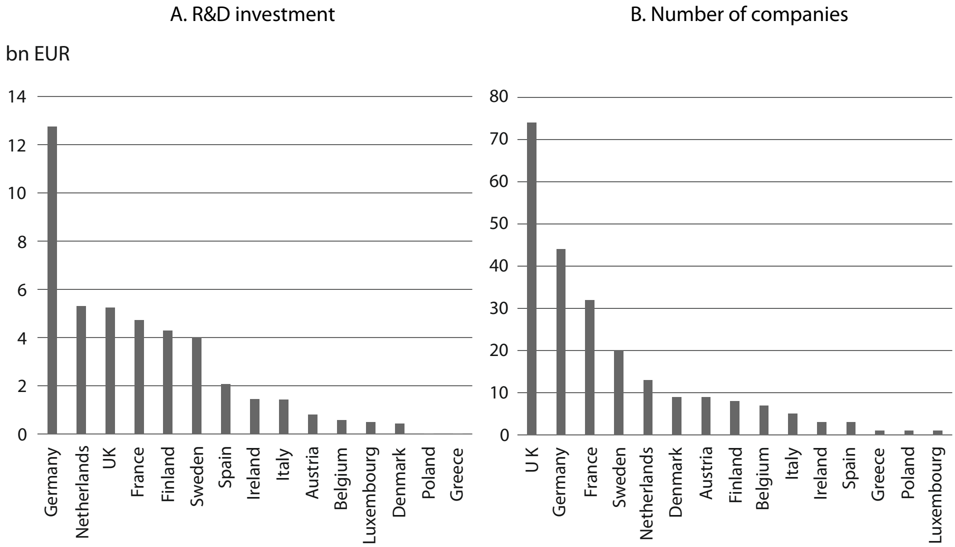


Source: the author's work based on IRI.

The greatest volume of investment in R&D made by the ICT sector in 2019 came from 15 countries: Germany, Finland, Sweden, the Netherlands, Italy, France, Spain, Ireland, the United Kingdom, Luxembourg, Austria, Belgium, Denmark, Poland and Greece. The remaining EU countries did not appear in the ranking.

As Figure 3a shows, the amount of R&D investment in the analysed companies varies across countries. Among EU member states investing the most in R&D in the ICT sector in 2019 were: Germany, the Netherlands, the United Kingdom and France. Figure 3b presents the quantitative distribution of companies from the ICT sector investing most in R&D among EU member states in 2019, which emphasises the differences between the amount of R&D investment and the quantitative distribution (e.g. in Figure 3a Germany ranks first and second in Figure 3b). Nevertheless, it is worth stressing that more than half of these companies' affiliates were located in four countries: the United Kingdom, Germany, France and Sweden.

Figure 3. R&D investment in the ICT sector and distribution of the sample of top corporate R&D performers in 2019



Source: the author's work based on calculations made with the use of EU R&D Scoreboard data.

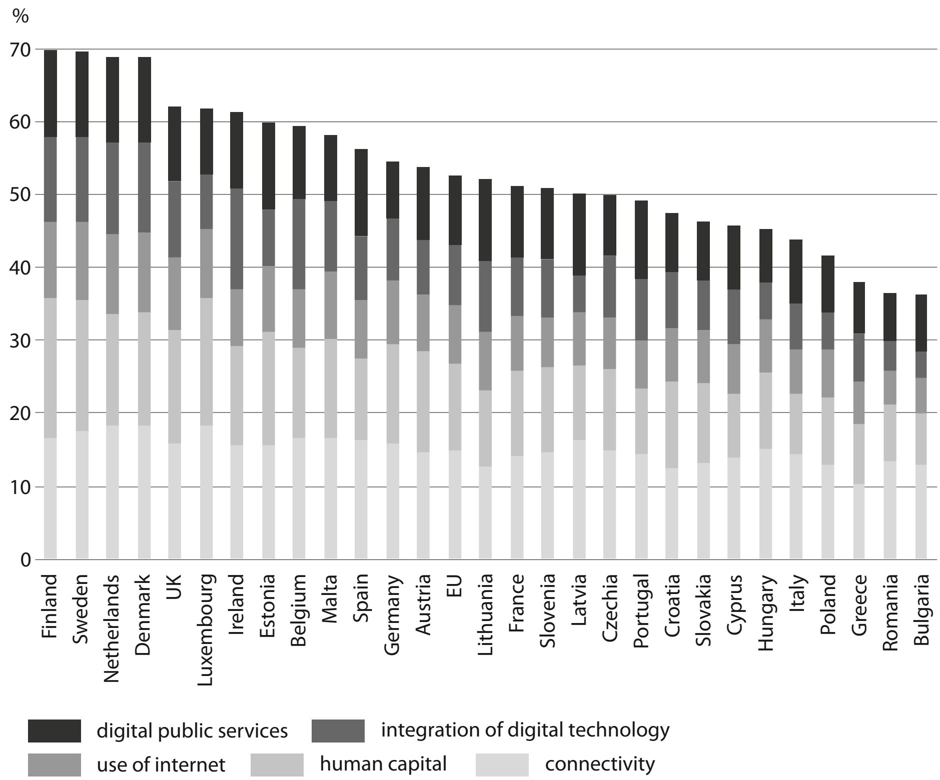
3.2. Digital Economy and Society Index

Currently, most countries which joined the EU after 2004 have a DESI below the EU average. Figure 4 (p. 36) presents DESI regarding connectivity, human capital, integra-tion of digital technology and digital public services. The countries with the highest value of DESI in 2019 include: Finland, Sweden, the Netherlands and Denmark, while the lowest values of this indicator were observed in Greece, Romania and Bulgaria.

3.3. Local linear trend model

The change in the total R&D expenses in the ICT sector in the years 2013–2019 are presented in Figure 5 (p. 37). A clear trend can be observed in the time-series. The local linear trend model shows that in the following years, investment in R&D in the ICT sector will continue to grow. The predicted investment for 2021 ranges from 40 billion euro (marking a small decrease in investment) to 53.1 billion euro. The ex-pected investment volume in 2021 is to exceed 46.7 billion euro.

Figure 4. DESI for EU countries in 2019

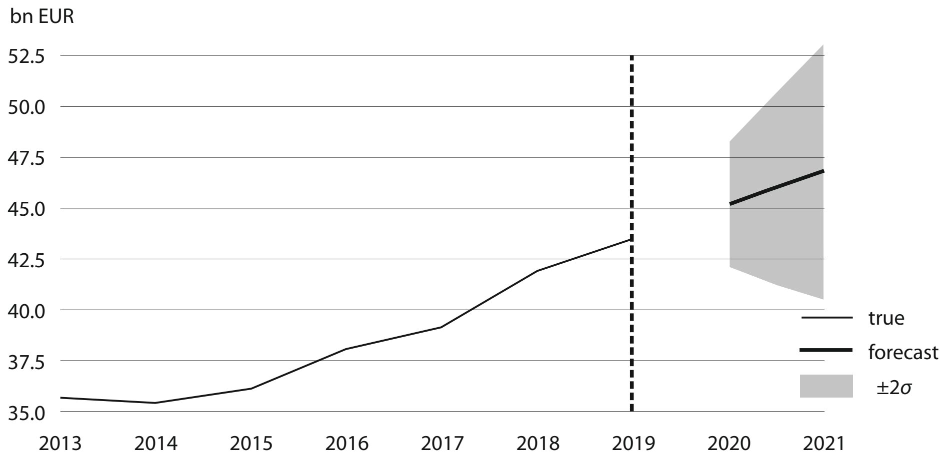


Source: the author's work based on European Commission (2019).

3.4. Ward's hierarchical agglomerative clustering method

The EU Industrial Research and Developments Scoreboard for 2019 includes all EU companies spending at least 8.6 million euro on R&D. This lowest amount was invested by Telia (a Swedish company from the Fixed Line Telecommunications industry). Out of all the 1,000 companies, which invested the largest amounts in R&D in 2019, the author classified 230 companies as belonging to the ICT sector (23.0%), using the same approach as in the case of the sub-section dataset. As Figure 3a indicates, R&D investment in the 230 analysed companies varies among individual countries. A large country like Germany (over 80 million inhabitants) is compared with countries whose population is below 1 million, e.g. Luxembourg. Thus, data normalisation on expenditure towards R&D was performed by dividing the countries by population (normalisation by population).

Figure 5. Prediction of expenses in R&D investment in the ICT sector



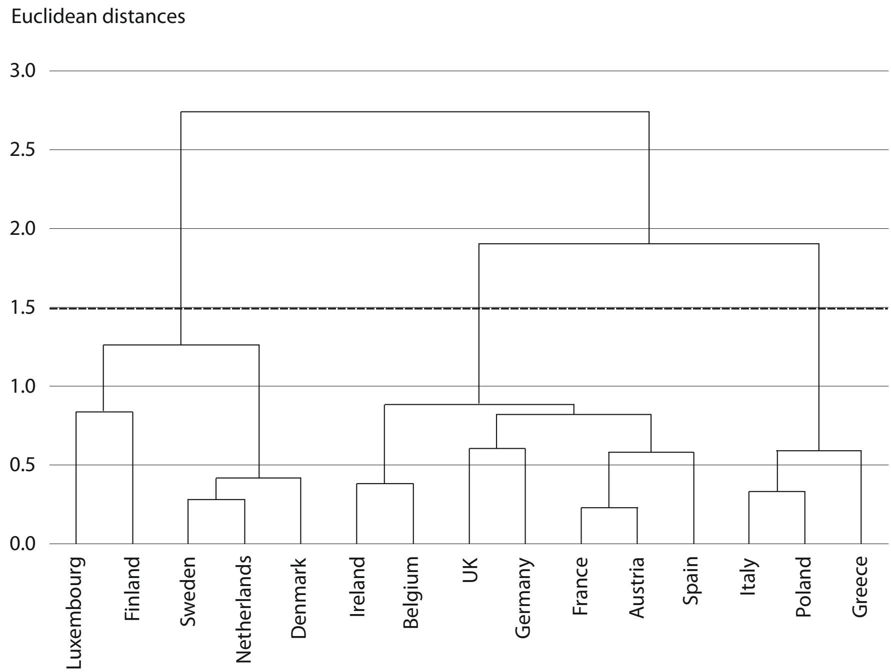
Source: the author's work.

Ward's hierarchical clustering was applied as a grouping method to identify the countries of a highly similar competitive position in term of ICT-related companies' DESI and R&D expenditure in 2019. The length of the line on the *y*-axis indicates distances between clusters – the lower the height, the more similar the clusters. The Euclidean distance was the metric used to compute the linkage. The Euclidean distance is the 'ordinary', straight-line distance between two points in Euclidean space. The calculations used standardised data on R&D expenditure by companies from the ICT sector. To compare countries, the data were organised into a two-dimensional structure (15 observations and 6 features).

Similar European countries were classified into three groups based on their average investment in R&D in the ICT sector, and five DESI sub-dimensions. The division of the countries into three clusters was based on the Ward method (an assumed binding distance of 1.5 has been chosen to obtain three groups of countries classified as: high, medium and low investment in R&D). The class labels for each observation are shown in different colours. Each leaf of the dendrogram represents one of the 15 observations. Moving higher up the tree, branches fuse either with leaves or other branches. The earlier (lower in the tree) fusions occur, the more similar the groups of observations to each other.

In Figure 6, the following groups are presented: cluster 0 – Denmark, Finland, Luxembourg, the Netherlands and Sweden; cluster 1 – Austria, Belgium, France, Germany, Ireland, Spain and the United Kingdom; cluster 2 – Greece, Italy and Poland. When examining the clusters in detail, it can be observed that they introduce ordering in multidimensional (DESI, R&D) space. Cluster 0 contains countries with the highest DESI and R&D, while cluster 2 is formed by countries with the lowest values of DESI and R&D. The above is a clear indication of the occurrence of a monotonic dependence between DESI and R&D.

Figure 6. Ward clusterisation with Euclidean distances



Source: the author's work.

3.5. Linear regression models

It can be expected that companies' investments in R&D can, at least to some extent, be explained by DESI. The Pearson's correlation coefficient confirms the occurrence of positive correlations between R&D and DESI (ρ=0.585). On this basis, a conclusion may be drawn that the independent variable can constitute a potential predictor of the studied phenomena. In order to validate this assumption, a linear model was fitted to infer the investments in R&D from the DESI index. The model is the Ordinary Least Squares (OLS) linear regression.

However, there are a few problems with this model: (i) there are two outliers (Finland and Luxembourg), (ii) for the DESI value exceeding 60 the error is much higher compared to smaller DESI values, which implies the occurrence of heteroscedasticity. Overcoming these problems involves modelling the logarithm of R&D instead of the original variable. This new model takes the following form:

$$\log \frac{\text{R\&D}}{1mln\ EUR} = \beta_0 + \beta_1 \cdot DESI + \ \varepsilon, \tag{5}$$

where ε represents model error (residuals).

The logarithm transformation reduces the variance of the dependent variable; consequently, the model can be considered homoscedastic with normally distributed residuals. This transformation has the additional benefit of enforcing a positive prediction value. Furthermore, variance reduction in logarithmic space allows the construction of a more accurate model (in the linear model with the original variables, the R-squared took the value of only 0.34 and 0.49 when outliers were removed), while the new model have the value $R^2 = 0.69$.

The p-value for the whole model obtained from the F-statistic is 0.00012, which is much below 0.05, indicating a strong significance of the result. Both coefficients β_0, β_1 are statistically significant (p-value was lower than 0.05 and equal 0.01 and 0.00012, respectively), and the residuals passed the Kolmogorov-Smirnov test for normality.

Furthermore, the hypothesis assuming that top investors registered in countries with a higher DESI spend more on R&D than other companies can be tested through the application of this model. This test is equivalent to the test checking if β_1 is positive (increasing function of DESI). The p-value for such a test equals 0.0006 (half of the p-value for the coefficient), so a conclusion can be drawn that indeed the R&D expenses are higher in the countries with a higher DESI. It should be noted that drawing conclusions regarding DESI from models fitted in a logarithm domain is admissible, as a logarithm is a monotonic function.

In the logarithm domain, the prediction equation with 95% confidence intervals for coefficients takes the following form:

$$\log \frac{\text{R\&D}}{1mln\ EUR} = \ -2.386 \pm 1.715 \ + \ 0.073 \pm 0.03 \cdot \ (DESI). \tag{6}$$

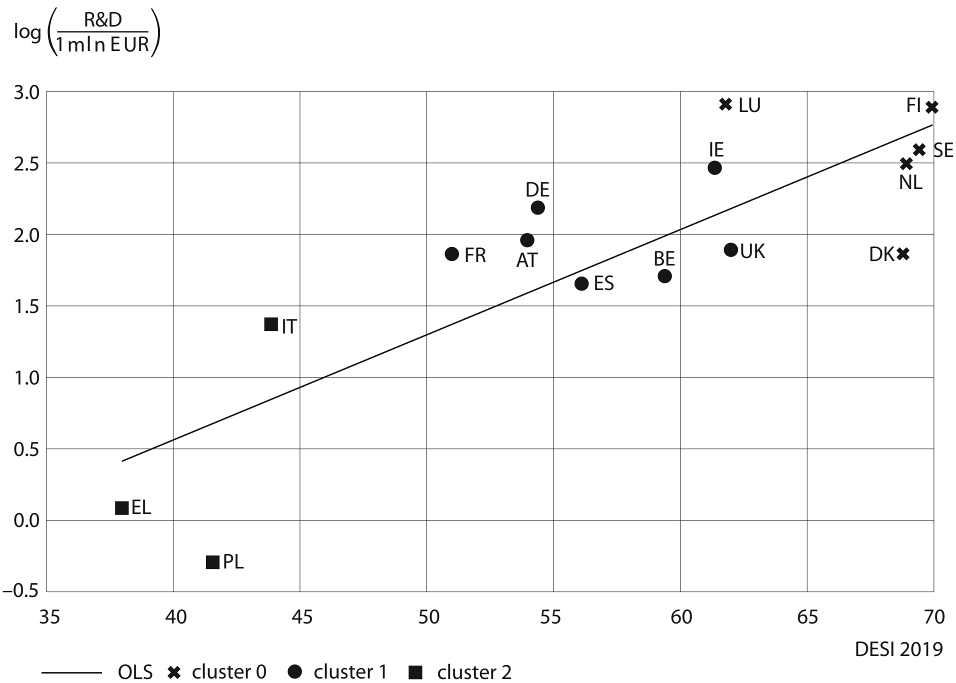
The relation, additionally supplemented with information relating to the cluster, is visualised in Figure 7. The R-square, a popular metric used in regression to show the amount of variance of Y explained by X, has the value of 0.69.

When transforming equation (6) to the original space, the following exponential law for R&D investments emerges:

$$\text{R\&D} \approx 1.183^{DESI} \times 4{,}111\ EUR. \tag{7}$$

According to this law, R&D is predicted to increase by 18.3% on average, when the DESI variable grows by one percentage point.

Figure 7. Scatter plot with regression lines and clusters



Note. AT – Austria, BE – Belgium, DK – Denmark, FI – Finland, FR – France, DE – Germany, EL – Greece, IE – Ireland, IT – Italy, LU – Luxembourg, NL – the Netherlands, PL – Poland, ES – Spain, SE – Sweden, UK – the United Kingdom.
Source: the author's work.

3.6. Analysis of significant DESI dimensions for investment in R&D

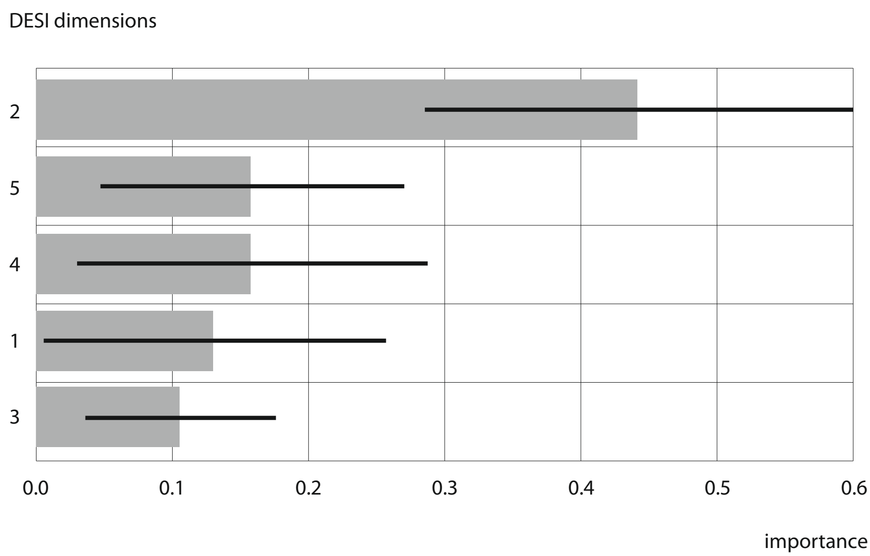
Results obtained from the regression model show that R&D investment grows along with the increase of DESI.

The random forest method is used, which provides a unique opportunity to estimate the importance of the feature (five DESI dimensions) which attracts ICT investors the most. The obtained feature importances are visualised in Figure 8. The bars represent each feature importance, namely the average value of 100 bootstrap samples. The horizontal lines represent the standard error of the estimated importance.

The most important feature is dimension 2 – human capital. This dimension consists of a number of indicators, including: ICT specialists, female ICT specialists and

ICT graduates. According to the results, access to qualified ICT specialists is the most important criterion for ICT companies to in choosing a particular country for their R&D investment. The other dimensions are similar to each other and are of a lower importance.

Figure 8. Feature importances plot for random forests



Note. 2 – human capital, 5 – digital public services, 4 – integration of digital technology, 1 – connectivity, 3 – use of internet.
Source: the author's work.

4. Summary

There is a widespread agreement that innovation is of key importance to Europe's competitiveness and growth. The rise and development of ICT has transformed both economies and societies. In recent years, the crucial role of ICT in the development of innovation has been demonstrated by the fact that a significant proportion of innovations and new patents are held by the ICT sector. Policy makers must have a range of appropriate tools at their disposal to define coherent policies, hence the need to monitor ICT's impact on economy and trends of its further development. An estimated trend modelling forecast shows that in the future years investment in R&D in the ICT sector will continue to grow. The predicted volume of investment in 2021 ranges from 42 billion euro (a small decrease) to 51.5 billion euro. The funds dedicated to investments in R&D in 2021 are expected to exceed 46 billion euro.

The analytical examination of investments in R&D, which was the first aim of the study described in this paper, focused on the largest R&D investors and was based on a sample of the top 1,000 companies which invested the largest amounts of money in R&D in the years 2013–2019, according to the EU Industrial Research and Developments Scoreboard. Thus, it can be assumed that a sample from the tail of ICT's R&D investment distribution was analysed. The degree to which top R&D investors' affiliates are diversified varies substantially among countries. The headquarters of the EU's top R&D investors from the ICT sector are situated in 16 countries (15 in 2019), with the highest concentration in three EU countries (Germany, the United Kingdom, France), however their subsidiaries were geographically spread. Disparities between countries in the basic comparison may be observed, as well the tendency to form groups, which the Ward cluster presents.

The paper constitutes a valuable contribution to the analysis of R&D investment made by ICT-related companies in the EU and its relationship with each country's digital performance expressed by the Digital Economy and Society Index (DESI). DESI is a composite index that summarises five relevant indicators relating to Europe's digital performance. Additionally, it can be used to track EU member states' evolution in the field of digital competitiveness.

The random forest model proved that the human capital DESI dimension is the most important factor attracting R&D investments.

The new approach utilising DESI to predict R&D investment is of high statistical significance, and a simple linear model applied to DESI explained 0.69 of the variance of the logarithm of R&D. A linear relation in the logarithmic scale suggests the occurrence of an exponential law between DESI and R&D, predicting an increase of 18% in R&D investment when DESI increases by 1 percentage point.

This analysis, constituting the second aim of the paper, shows that better digital performance (expressed by DESI) has a major influence on companies' investments. The majority of EU countries with the lowest DESI index have no R&D-related investment. These included: Slovenia, Latvia, Slovakia, Hungary, Cyprus, Croatia, Bulgaria and Romania (Italy, Poland and Greece were exceptions – but the last two were residual).

The further analysis of countries' similarities in terms of the DESI and R&D space was supplemented with information regarding clusters, obtained by means of the Ward method. Non-uniform variations in different clusters suggested the application of a logarithm as a form of transformation aiming to achieve variance stabilisation.

The top analysed R&D investors from the ICT sector have affiliates in 16 economies. Considering that ICT firms are global enterprises with R&D centres located all around the world, this study is limited by the lack of information on the geographical distribution of the companies' R&D centres, in the situation where they are sometimes located outside their country of origin. A dataset containing this information could extend and confirm the general applicability of the presented results. Even with the current dataset, the applicability of the presented results can be limited due to the global crisis caused by the COVID-19 pandemic. It has impacted a large number of industries, including the ICT sector, so it could have affected the forecast and results presented in article as well. On the other hand, the pandemic situation can also be an opportunity for the whole ICT industry to strenghten its position, since it would not be possible for people to work or study from home without it.

The general applicability of the model is limited to the top R&D investors (companies spending no less than 8.6 million EUR on R&D) – i.e. to the tail of R&D investment distribution; therefore, not all companies from the ICT sector are included, as the dataset does not account for R&D investments made by smaller businesses. Therefore, it is worth mentioning that if a country is not included in the IRI R&D Scoreboard database, it does not necessarily mean that there are no ICT firms investing into R&D in that country. It must be noted that assuming the given distribution of R&D investments, an estimation of the parameters for the whole population from the tail observation in the Scoreboard may be attempted. The accuracy of such estimates can be substantially increased by adding information from different sources, including Eurostat.

The EU strategy aims to boost the development in the sphere of innovation to make the EU a global economic leader. The results deriving from the presented model are useful tools for policymakers to measure the efficiency of R&D investment in ICT-related firms in Europe and to help define the right policies to attract companies to make R&D investments. An illustration of such policies being implemented can be the increase in the country's digital performance, in particular in the improvement of people's digital skills. It is especially important to increase the education of future ICT specialists, as a large number of enterprises are reporting shortages in this respect.

References

Bouguettaya, A., Yu, Q., Liu, X., Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, *42*(5), 2785–2797. http://dx.doi.org/10.1016/j.eswa.2014.09.054.

Bozkurt, C. (2015). R&D Expenditures and Economic Growth Relationship in Turkey. *International Journal of Economics and Financial Issues*, *5*(1), 188–198. https://econjournals.com/index.php/ijefi/article/view/1038/pdf.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. http://doi.org/10.1023/A:1010933404324.

Daiko, T., Dernis, H., Dosso, M., Gkotsis, P., Squicciarini, M., Vezzani, A. (2017). *World Corporate Top R&D Investors: Industrial Property Strategies in the Digital Economy*. Luxembourg: Publications Office of the European Union. https://www.oecd.org/sti/world-top-rd-investors.pdf.

De Prato, G., Nepelski, D. (2013). *The global R&D network: A network analysis of international R&D centres*. (JRC Technical Reports). http://dx.doi.org/10.2791/15651.

Dillon, J., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., Saurous, R. (2017). *Tensor Flow Distributions*. https://arxiv.org/abs/1711.10604.

Doong, S. H., Ho, S.-C. (2012). The impact of ICT development on the global digital divide. *Electronic Commerce Research and Applications*, *11*(5), 518–533. http://doi.org/10.1016/j.elerap.2012.02.002.

Durbin, J., Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Edquist, H., Henrekson, M. (2017). Swedish lessons: How important are ICT and R&D to economic growth? *Structural Change and Economic Dynamics*, *42*, 1–12. http://doi.org/10.1016/j.strueco.2017.05.004.

EPID. (2015). *Knowledge based economy*. http://europejskiportal.eu/knowledge-based-economy/.

EPO. (2020). *Digital technologies take top spot in European patent applications*. https://www.epo.org/news-events/news/2020/20200312.html.

EU. (2018). *EU R&D Scoreboard. The 2018 EU Industrial R&D Investment Scoreboard*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/131813.

EU. (2019). *EU R&D Scoreboard. The 2019 EU Industrial R&D Investment Scoreboard*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/04570.

European Commission. (2012). *Internationalisation of business investments in R&D*. http://dx.doi.org/10.2777/60978.

European Commission. (2018a). *DESI 2018 Digital Economy and Society Index. Methodological note*. http://ec.europa.eu/information_society/newsroom/image/document/2018-20/desi-2018-methodology_E886EDCA-B32A-AEFB-07F5911DE975477B_52297.pdf.

European Commission. (2018b). *Indeks gospodarki cyfrowej i społeczeństwa cyfrowego (DESI) 2018. Sprawozdanie krajowe dotyczące Polski*. https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi.

European Commission. (2019). *The Digital Economy and Society Index (DESI)*. https://digital-agenda-data.eu/.

European Commission. (2020a). *Digital Economy and Society Index*. https://digital-agenda-data.eu/datasets/desi.

European Commission. (2020b). *Digital Economy and Society Index (DESI). Methodological note*. https://ec.europa.eu/digital-single-market/en/desi.

European Commission. (2020c). *EU Industrial R&D Investment Scoreboard 2019 (EU 1000)*. https://iri.jrc.ec.europa.eu/data.

Eurostat. (2018). *Digital economy and society statistics – enterprises*. https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_enterprises.

Eurostat. (2019). *ICT specialists in employment*. https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment#Number_of_ICT_specialists.

FTSE. (2012). *International Classification Benchmark*. http://www.ftse.com/products/downloads/ICBStructure-Eng.pdf.

Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2017). *An Introduction to Statistical Learning with Applications in R*. Springer. http://doi.org/10.1007/978-1-4614-7138-7.

GUS. (2017). *Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2013–2017. Information society in Poland. Results of statistical surveys in the years 2013–2017*. https://stat.gov.pl/files/gfx/portalinformacyjny/en/defaultaktualnosci/3417/1/4/1/information_society_in_poland._results_of_statistical_surveys_in_the_years_2013-2017.pdf.

Hernández, H., Grassano, N., Tübke, A., Potters, L., Gkotsis, P., Vezzani, A. (2016). *EU R&D Scoreboard. The 2015 EU Industrial R&D Investment Scoreboard*. Luxembourg: Publications Office of the European Union. http://doi.org/10.2791/15792.

Hernández, H., Grassano, N., Tübke, A., Potters, L., Gkotsis, P., Vezzani, A. (2018). *The 2018 EU Industrial R&D Investment Scoreboard*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/131813.

Hernández, H., Grassano, N., Tübke, A., Amoroso, S., Csefalvay, Z., Gkotsis, P. (2019). *The 2019 EU Industrial R&D Investment Scoreboard*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2760/04570.

Hilmersson, F. P., Hilmersson, M. (2020). Networking to accelerate the pace of SME innovations. *Journal of Innovation & Knowledge*. http://dx.doi.org/10.1016/j.jik.2020.10.001.

Hong, J. (2017). Causal relationship between ICT R&D investment and economic growth in Korea. *Technological Forecasting and Social Change*, *116*, 70–75. http://dx.doi.org/10.1016/j.techfore.2016.11.005.

Investin. (2017). *Growth perspectives for Polish ICT sector by 2025*. https://en.parp.gov.pl/storage/publications/pdf/2017_ict_sector_by_2025_en.pdf.

Jin, S., Cho, C. M. (2015). Is ICT a new essential for national economic growth in an information society? *Government Information Quarterly*, *32*(3), 253–260. https://doi.org/10.1016/j.giq.2015.04.007.

Jorgenson, D. W., Vu, K. M. (2016). The ICT revolution, world economic growth, and policy issues. *Telecommunications Policy*, *40*(5), 383–397. https://doi.org/10.1016/J.TELPOL.2016.01.002.

Kisielnicki, J. (2016). Innowacyjność gospodarki polskiej na tle wybranych krajów Unii Europejskiej i świata. *Studia Ekonomiczne*, (281), 67–79. https://www.ue.katowice.pl/fileadmin/user_upload/wydawnictwo/SE_Artyku%C5%82y_271_290/SE_281/06.pdf.

Liu, C. (2017). International Competitiveness and the Fourth Industrial Revolution. *Entrepreneurial Business and Economics Review*, *5*(4), 111–133. http://doi.org/10.15678/EBER.2017.050405.

Nepelski, D., De Prato, G. (2012). Internationalisation of ICT R&D: a comparative analysis of Asia, the European Union, Japan, United States and the rest of the world. *Asian Journal of Technology Innovation*, *20*(2), 219–238. http://doi.org/10.1080/19761597.2012.741392.

OECD. (2004). *OECD Information Technology Outlook*. http://www.oecd.org/sti/ieconomy/37620123.pdf.

OECD., Eurostat. (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. OECD Publishing. https://ec.europa.eu/eurostat/cache/metadata/Annexes/inn_cis8_esms_an3.pdf.

Palczewska, A., Palczewski, J., Robinson, R. M., Neagu, D. (14–16.08.2013). *Interpreting random forest classification models using a feature contribution method*. IEEE 14th International Conference on Information Reuse & Integration, San Francisco.

Pesole, A. (2015). *How much does ICT contribute to innovation output? An analysis of the ICT component in the innovation output indicator*. Luxembourg: Publications Office of the European Union. http://dx.doi.org/10.2791/5451991.

Pradhan, R. P., Mallik, G., Bagchi, T. P. (2018). Information communication technology (ICT) infrastructure and economic growth: A causality evinced by cross-country panel data. *IIMB Management Review*, *30*(1), 91–103. https://doi.org/10.1016/J.IIMB.2018.01.001.

Schwab, K. (Ed.). (2017). *The Global Competitiveness Report 2017–2018*. World Economic Forum. http://www3.weforum.org/docs/GCR2017-2018/05FullReport/TheGlobalCompetitivenessReport2017–2018.pdf.

Sepehrdoust, H. (2018). Impact of information and communication technology and financial development on economic growth of OPEC developing economies. *Kasetsart Journal of Social Sciences*, *40*(1). https://doi.org/10.1016/J.KJSS.2018.01.008.

Sylwestrzak, M. (2018). Wpływ ICT na wzrost gospodarczy w krajach Unii Europejskiej w latach 2006–2016 (Impact of ICT on economic growth in the European Union in the years 2006–2016), *Ekonomiczne Problemy Usług*, *1*(131), 361–369. http://dx.doi.org/10.18276/epu.2018.131/1-35.

Vu, K. M. (2013). Information and Communication Technology (ICT) and Singapore's economic growth. *Information Economics and Policy*, *25*(4), 284–300. https://doi.org/10.1016/j.infoecopol.2013.08.002.

Żelazny, R. (2015). Determinanty rozwoju społeczeństwa informacyjnego – implikacje dla rozwoju gospodarczego. *Studia Ekonomiczne*, (213), 45–58. https://www.sbc.org.pl/dlibra/publication/edition/171967?id=171967 & from=publication.