CENTRAL EUROPEAN REVIEW
OF ECONOMICS AND MANAGEMENT

ISSN 2543-9472; eISSN 2544-0365

Vol. 8, No.1, March 2024, 131-154

www.cerem-review.eu
www.ojs.wsb.wroclaw.pl

Resource efficiency awareness of companies

Sıdıka BAŞÇI
Ankara Yıldırım Beyazıt University, Türkiye

Houcine SENOUSSI
Quartz Laboratory, CY Tech, France

Received: 17.08.2023, Revised: 06.09.2023, Revision: 28.09.2023, Revision: 22.02.2024, Accepted: 24.02.2024
doi: http://10.29015/cerem.981

Aim: This paper aims to identify the characteristic variables that influence firms' resource efficiency awareness and subsequently group countries based on the similarities of these influencing factors.

Design / Research methods: We utilize data from the GESIS Data Archive and Flash Eurobarometer, which conducted a survey in 2017 across 36 European countries and the United States of America. Machine learning tools are applicable to the analysis. Specifically, the Chi-squared independence test is applied to determine the impact of characteristic variables on resource efficiency awareness. Following this, unsupervised learning (clustering) algorithms are used to identify countries that exhibit similar patterns.

Conclusions/findings: The findings reveal that turnover performance over the past two years and last year's turnover significantly influence firms' resource efficiency awareness, while factors such as employee number of the company, one-person company or not (sole proprietorship), and the establishment year of the company do not seem to have a notable effect. The impact of the customer profile of the firm on resource efficiency awareness remains uncertain. Based on these dependency results, the study identifies ten potential clusters of countries with similar characteristics in terms of resource efficiency awareness and related factors.

Originality/value of the article: Machine learning methods are relatively novel approaches that have gained prominence with the rise of extensive datasets. As a result, the paper exhibits originality in terms of research methodology. Furthermore, when considering resource efficiency as a significant topic, the article holds considerable importance.

Implications of the research (if applicable): Utilizing the findings of the paper, it becomes possible to develop an application suitable for an Erasmus project. Given that resource efficiency is one of Erasmus' critical focal points in 2023, the likelihood of approval is considerably elevated.

Keywords: Energy, Resource Efficiency, Machine Learning.
JEL: F64, C55

Correspondence address: Sıdıka Başçı, Ankara Yıldırım Beyazıt University, Türkiye, E-mail: sidika.basci@aybu.edu.tr Houcine Senoussi, Quartz Laboratory, CY Tech, France, E-mail: hse@cy-tech.fr

1. Introduction

If eco-efficiency can be achieved, more value can be created with less environmental impact. The three items concerning eco-efficiency are increasing resource efficiency investment, producing more environmentally compatible green products or services, and the consumption of energy from renewable resources. Among these actions, firms can choose to increase their resource-efficient investments only if they are aware of the importance of resource efficiency. Therefore, in this paper, we aim to determine the characteristic variables of the firms that are effective in the resource efficiency awareness of the firms. Moreover, we group the countries with similar kinds characteristics. After these determinations, we suggest education policies to firms for achieving awareness which can serve to eco-efficiency.

In the literature, there are papers that use conventional econometric methods for this aim. However, within the last twenty years, the availability of big data accelerated tremendously, and this opened a debate about whether to use conventional methods or Machine Learning (ML) methods for the analysis of big data. Conventional methods start the analysis with a given theory of economics and then test the validity of this theory with the data available. However, ML methods do just the converse. They produce a theory depending on the big data available. Judge (2016), who is a pioneer in econometrics, comments in his 2016 paper about the current state of econometrics as follows:

"Looking ahead, a non-traditional econometric approach is outlined. This method recognizes that our knowledge regarding the underlying behavioral system and observed data process is complex, partial, and incomplete. It then suggests a self-organized, agent based, algorithmic-representation system that involves networks, machine learning, and an information theoretic basis for estimation, inference, model evaluation, and prediction."

We used the survey outcomes of GESIS Data Archive, Flash Eurobarometer which was conducted in 2017 to 36 European countries and the United States of America. We focused on the set of firm-specific variables and the firm's perceptions of resource efficiency for sectors of manufacturing, commerce, services, and industry.

Since this was a huge data, ML methods were applicable. Therefore, we can say that this paper contributes to this new debate with an application.

We first used the Chi-square independence test for the two sets of variables for each country. The results showed that turnover performance for the past two years and last year's turnover were effective on the resource efficiency awareness of firms, one-person company or not, and establishment year of the company are not effective on the resource efficiency awareness of the companies, and there is ambiguity for customer profile of the country.

Depending on the results of dependency, we tried to group the countries which are similar. The results of k-means, which assumes non-overlapping clusters, showed that there is no natural partition. However, after relaxing the assumption and applying Principle Component Analysis, we found ten groups.

Since turnover is an effective characteristic variable, an education program can be designed to firms having a high turnover for increasing resource awareness. Moreover, this education can be given to countries belonging to the same group at the same time.

The paper proceeds as follows. In section 2, we have the literature review. Section 3 describes the data. The methodology is introduced in section 4. Results are presented in section 5, and finally, section 6 concludes.

2. Literature review

We divide this section into two parts. In the first part, we aim to review the literature where conventional econometric methods are used to measure resource efficiency. In the second part, we give some examples of applications in economics where ML methods are used.

2.1 Conventional econometric methods

The empirical literature mostly focuses on energy efficiency investments. The results of these articles are summarized in Table 1. There are three common themes among these results. (1) the lack of energy efficiency among priorities, (2) problems of access to capital, length of return of investment, (3) lack of information.

Table 1. Literature of conventional econometric methods to measure resource efficiency

Articles	Countries Studied	Sectors Studied	Basic Findings
Velthuijsen (1995)	Netherlands, Slovakia, Czech Republic	Manufacturing Industry	Obstacles to energy efficiency investments are problems in access to capital, high risk, very long turnaround time, and poor market conditions.
de Groot et al. (2001)	Netherlands	Manufacturing Industry	Firms give relatively low priority to projects related to energy efficiency. This has a negative impact on energy efficiency investments.
Diederen et al. (2003)	Netherlands	Greenhouse	The energy market is a volatile market with high uncertainty. Therefore energy prices can be considered as a negative factor.
Anderson and Newell (2004)	Netherlands, USA	Greenhouse, Manufacturing Industry	There is a negative relation between return time, costs, lack of personnel and liquidity constraints with energy efficiency investments.
Schleich, Gruber (2008)	Germany	Service Industry and Small Industries	There are two important obstacles to energy-efficient investment. These are a lack of information on energy consumption and the inconsistency of incentives.
Schleich (2009)	Germany	Service industry and Small Industries	Obstacles to energy investments are a lack of information on energy consumption, lack of information on energy efficiency measures, time constraints, and different priorities.

Table 1. Cont…

Muthulingam et al. (2011)	Germany, USA	Service industry, Small Industries, Manufacturing Industry	Institutional hierarchy within the company shapes energy efficiency investments.
Kostka et al. (2011)	China	Manufacturing Industry	Lack of information is an obstacle to energy efficiency investments.
Trianni, Cagno (2012)	China, Italy	Manufacturing Industry	Firms have problems accessing capital for energy efficiency investments.
Delmas, Pekovic (2015)	France	Small Industries	Very few firms adopt resource efficiency strategies in perceived economic downturns compared to perceived steady or growing market conditions.
Di Maio, Rem, Balde, Polder (2017)	The Netherlands	Mining, Manufacturing Industry, Agriculture, Service Industry	The sectors that rank less resource-efficient are those where the prices of the inputs used are high.
Bodas-Freitas, Corrocher (2019)	Several countries	Several Industries	External technical and business advice plays an important role in the adoption of resource efficiency measures.

Source: Fleiter et al. (2012); Ozbugday et al. (2020).

There are four papers that used the same data set as this paper. The first one is Horbach (2016), and in the paper, there is an estimation of a Probit model for the relationship between resource-efficiency investments and sales growth performance, and the relation was found to be positive. The second one, Jov e-Llopis and Segarra-Blasco (2018) reported that high investment in eco-strategies improves sales performance. The third one, Ozbugday et al. (2019), is a descriptive examination of the attitudes of Turkish Small and Medium-sized Enterprises (SMEs) to on-site energy generation from renewable resources, resource efficiency investments, and supply of green products or services. Finally, Ozbugday et al. (2020) studied resource efficiency in terms of investments and firm performance of European SMEs.

2.2 Machine learning

Athey (2018) wrote about the growing interest in applying machine learning tools to economics. Mullainathan and Spiess (2017) stated that the main machine learning approach used in the applications is the so-called supervised learning because it revolves around the problem of prediction.

However, unsupervised learning approaches like clustering, which we used in this paper, have also been used in several works.

- Andrejovská et al. (2016) used it to categorize European Union countries according to factors affecting their agricultural production.
- Ramachandran et al. (2018) applied it to identify African enterprise groups with respect to their experiences with outages, losses, generator use, and job growth.
- Göbel and Araújo (2020) used it to investigate the ability of several macroeconomic variables to distinguish crisis economies from non-crisis ones.

The review study of Ghoddusi et al. (2019) reports 130 articles related to energy published between 2005 and 2018 which use ML methods. These are all applications in areas such as predicting energy prices (e.g. crude oil, natural gas, and power), demand forecasting, risk management, trading strategies, data processing, and analyzing macro/energy trends. As can be seen, there does not exist the theme of resource efficiency. Depending on this careful review, we can say that this paper fills a gap in the literature.

3. Data

The last wave of the Flash Eurobarometer, Small and Medium-Sized Enterprises, Resource Efficiency, and Green Markets Survey (2017) is the data used in this paper. There are 37 countries where, 36 of which are European countries, and the last one is the United States of America (USA). The survey includes a set of firm-specific variables and the firm’s perceptions of resource efficiency for sectors of manufacturing, commerce, services, and industry.

Six firm-specific variables are:

A1: Employee number of the company

A2: One-person company or not (Owner runs the business alone or not)

A3: Establishment year of the company

A4: Turnover[1] performance for the past 2 years

A5: Last year's turnover level

A6: Customer profile (consumers, companies, public administration) of the company where multiple answers are possible

Table 2. An example of the firm-specific variable from data (Turkey)

TOTAL	299
1 to 9 employees	175 59%
10 to 49 employees	83 28%
50 to 249 employees	28 9%
250 employees or more	8 2%

For example, Table 2 is for Turkey, where the sample size is 299. The company's employee numbers are reported in the table. Tables similar to the below exist for items A1 – A6 for 37 countries.

Three variables that measure the perceptions of resource efficiency are:

Q1: Actions taken to be more resource efficient (saving water, saving energy, using renewable energy, saving materials, minimizing wastes, selling the scrap materials to other companies, recycling within the company, and designing products that are easier to maintain, repair or reuse) where multiple answers are possible

Q2: Planned resource-efficient actions for the next two years (saving water, saving energy, using renewable energy, saving materials, minimizing wastes, selling the scrap materials to other companies, recycling within the company, and designing products that are easier to maintain, repair or reuse) where multiple answers are possible

[1] Turnover is an accounting concept that calculates how quickly a business conducts its operations. Most often, turnover is used to understand how quickly a company collects cash from accounts receivable or how fast the company sells its inventory.

Q3: Past two years' investment rate per year to be more resource efficient (nothing, less than 1% of annual turnover, 1–5% of annual turnover, 6–10% of annual turnover, 11–30% of annual turnover and more than 30% of annual turnover

Table 3. An example of a variable that measures the perception of resource efficiency from data

TOTAL	500
Saving water	160 32%
Saving energy	323 65%
Using predominantly renewable energy (e.g. including own production through solar panels, etc.)	134 27%
Saving materials	307 61%
Minimising waste	326 65%
Selling your scrap material to another company	129 26%
Recycling, by reusing material or waste within the company	184 37%
Designing products that are easier to maintain, repair or reuse	103 21%

4. Methodology

We first studied the dependency between each awareness variable and each characteristic variable for each country. Then, we defined groups (clusters) of countries with respect to these dependencies: a cluster must contain countries having similar characteristics in terms of dependence/independence between the firm-specific variables and awareness variables. In the following sections we describe these two steps.

4.1. Chi-squared independence test

Let us consider two categorical variables, A and B, having respectively r and s possible values, and the r×s contingency table in which the i^{th} row, j^{th} column cell contains the number n_{ij} of observations for which we have the i^{th} value of A and the j^{th} value of B.

Pearson's Chi-squared independence test (see for example Milton, Arnold 2003, chapter 15) is used to decide between the following two hypotheses:

$$\begin{cases} \text{H0: A and B are independent.} \\ \text{H1: A and B are not independent.} \end{cases}$$

For that it measures the divergence of the observed values of the frequencies (n_{ij}) from those expected under the null hypothesis (t_{ij}). The test statistic is the following:

$$D = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(nij - tij)^2}{tij}$$

D follows an approximate chi-squared distribution whose number of degrees of freedom is (r-1)*(s-1). For a given value of α such that $0 <= \alpha <= 1$, we will consider that A and B are independent if D<=C, with C defined by

$$\alpha = Prob(D > C | H_0)$$

In this work, we applied a chi-square independence test with alpha=0.05, for each country, and for each couple awareness variable/characteristic variable.

Let us consider for example the variables A3 and Q1 for the United States of America. In this case the contingency table is the following:

Table 4. Contingency table for the pair of variables A3 – Q1 – USA

	Before 2010	Between 2010 and 2016	2017 and after
Many actions	146	50	3
Some actions	62	20	0
Few actions	48	16	0
No action	20	12	2

The expected values for this pair of variables are the following:

Table 5. Expected values for the pair of variables A3 – Q1 – USA

	Before 2010	Between 2010 and 2016	2017 and after
Many actions	145	51	3
Some actions	60	21	1
Few actions	47	16	1
No action	25	9	0

Before applying the independence test, we notice that the expected values of the four cells of the last column are less than five. In this case, the test results are not reliable. Therefore, we first merge the last two columns. With the new contingency table, we have r=2 and s=4. It follows that the number of degrees of freedom of the Chi-squared distribution is 3. D is equal to 2.53. The value of C given by the Chi-squared distribution table is 7.81. It follows that the two variables are independent.

At the end of this step, each country is defined by 15 variables (x1, …, x15), with xi=1 if the corresponding variables are dependent and xi=0 if they are independent.

4.2 Clustering

Clustering is the task of dividing a collection of objects, usually represented as points in a multidimensional space, into groups (clusters) based on similarity:

objects within a cluster are more similar to each other than they are to an object belonging to another cluster (high within-clusters homogeneity, high between-clusters heterogeneity). In our work a cluster is a set of countries having similar characteristics in terms of dependence/independence between the firm-specific variables and awareness variables.

There are several techniques for clustering. These techniques differ in the way they group objects (e.g., similarity measure) and in their outputs: clusters can be disjoint or not, hard or fuzzy, and partitional or hierarchical (Jain et al. 1999).

There are also several methods for measuring the within-cluster homogeneity and the between-cluster heterogeneity, and therefore, for determining the "best number" of clusters and finding the "optimal" set of clusters (Milligan, Cooper 1985; Boone 2011).

In this work, we first applied the K-means algorithm (Forgy 1965; MacQueen 1967). This algorithm creates disjoint clusters C1, …, Ck where k is determined in advance and assigns each point to the cluster whose centroid is the closest to it. We evaluated the clustering obtained with K-means using the following two metrics:

- The coefficient of determination R2, which measures the part of the variance of our dataset explained by the clusters. A high between-cluster heterogeneity corresponds to a value of R2 close to 1.
- The homogeneity index H_k which measures the degree to which the variance within a given cluster is lower than that of the entire dataset. A high within-cluster homogeneity corresponds to values of H_k close to 0.

Table 6 shows these values for different numbers of clusters. We notice that R2 is closer to 0 than to 1, and Hk index is not close to 0. We conclude that neither the within-cluster homogeneity nor the between-cluster heterogeneity is high. In other words, regardless of the number of clusters we choose, the clustering is poor.

Table 6. Homogeneity and heterogeneity measures for different numbers of clusters obtained with K-means

Number of clusters	R2	Hk index (min–max)
2	0.16	0.81–0.88
3	0.24	0.69–0.82
4	0.31	0.48–0.78
5	0.38	0.52–0.77
6	0.43	0.45–0.66

After this first result, we applied a PCA (Principal Component Analysis) to reduce the number of variables, then we applied K-means again, but the resulting clustering was as poor as the first one.

The conclusion of these two steps is the following:

- The results of K-means, with and without dimension reduction, show that there is no natural partition of the countries into a small number of disjoint groups based on the dependence/independence between the firm-specific variables and awareness variables.

Given this conclusion, we decided to apply another approach to find a clustering of our set of countries: the one that Bail et al. (2002) introduced and applied to text clustering. This approach can be summarized as follows:

- We are given a set of items (e.g., words) I and a set of objects O (e.g., documents) such that each object is defined as a set of items.
- Given a set of items S (a subset of I, we will say an itemset), let the cover of S, denoted by cov(S), be the set of all objects containing S. It follows that each itemset S defines a cluster candidate C=cov(S) of O.
- A clustering description CD is a set {S1,…,Sk} of itemsets, such that Ucov(Si)=O, i.e. such that each object x belongs to cov(Si) for at least one itemset Si. In other words, {C1=cov(S1),…,Ck=cov(Sk)} is a clustering of O. These clusters may or may not overlap.
- To find such a clustering, we first apply the well-known algorithm apriori (Agrawal, Ramakrishnan 1994) to discover the frequent itemsets: itemsets that occur at least in a minimum number of objects. Then we build a clustering description by selecting frequent itemsets with a cover having the minimum overlap with other cluster candidates.

In our work, objects are countries, and items are the dependence/independence relations between the firm-specific variables and awareness variables, which we denote by Qj-Ai=dep and Qj-Ai=indep.

The main advantage of this approach is that it produces an understandable description of the clusters. Frequent itemsets represent essential regularities in the set of objects and each cluster is defined as the cover of a frequent itemset. In our problem, a cluster is a set of countries sharing a set of dependencies and independencies that are frequently found together.

5. Results

In this section we present our results. In subsection 5.1, we reported the dependency between each awareness variable Q_j, j=1, 2, 3 and each characteristic variable A_i, i=1, 2, ..., 6 for each country. Then in subsection 5.2, we reported groups (clusters) of countries with respect to the dependencies we observed: a cluster must contain countries having similar characteristics in terms of dependence/independence between the firm-specific variables and awareness variables.

5.1. Dependence between characteristics and awareness variables

Table A1, presented in the Appendix, is the results of the Chi-Squared Independence Test for each Qj – Aj and for each country. [2] The value 1 in the table means dependence, rejection of the null hypothesis, and the value 0 means independence, not rejecting the null hypothesis. Dependence means that the characteristic variable has an effect on the awareness of the firm.

The dependence results in `employee number of the company' (A1), and the awareness of the resource efficiency variables of the firms can be summarized as follows:

- A1 – Q1: Actions taken to be more resource efficient: There is dependence for 12 countries. There is independence for 21 countries.
- A1 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 13 countries. There is independence for 20 countries.
- A1 – Q3: Past two years' investment rate per year to be more resource efficient: There is dependence for 16 countries. There is independence for 17 countries.

[2] Countries for which it was not possible to study dependence for lack of data have been removed. The final list contains 33 countries.

The dependence results about `one-person company or not (owner runs the business)' (A2) and the awareness of the resource efficiency variables of the companies can be summarized as follows:

- A2 – Q1: Actions taken to be more resource efficient: There is dependence for 9 countries. There is independence for 19 countries. There is no decision for 5 countries
- A2 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 12 countries. There is independence for 16 countries. There is no decision for 5 countries
- A2 – Q3: Past two years' investment rate per year to be more resource efficient: There is dependence for 10 countries. There is independence for 19 countries. There is no decision for 4 countries

This variable is the only one for which we have "can't conclude" values since the data is not enough to make an independence test. Therefore, it has been removed from the analysis.

The dependence results about the `establishment year of the company' (A3) and the awareness of the resource efficiency variables of the companies can be summarized as follows:

- A3 – Q1: Actions taken to be more resource efficient: There is dependence for 11 countries. There is independence for 22 countries.
- A3 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 12 countries. There is independence for 21 countries.
- A3 – Q4: Past two years' investment rate per year to be more resource efficient: There is dependence for 14 countries. There is independence for 19 countries.

The dependence results about `turnover performance for the past two years' (A4) and the awareness of the resource efficiency variables of the companies can be summarized as follows:

- A4 – Q1: Actions taken to be more resource efficient: There is dependence for 19 countries. There is independence for 14 countries.
- A4 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 20 countries. There is independence for 13 countries.

- A4 – Q3: Past two years' investment rate per year to be more resource efficient: There is dependence for 22 countries. There is independence for 11 countries.

The dependence results about "Last year's turnover level" (A5) and the awareness of the resource efficiency variables of the companies can be summarized as follows:

- A5 – Q1: Actions taken to be more resource efficient: There is dependence for 22 countries. There is independence for 11 countries.
- A5 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 20 countries. There is independence for 13 countries.
- A5 – Q3: Past two years' investment rate per year to be more resource efficient: There is dependence for 27 countries. There is independence for 6 countries.

The dependence results in the "customer profile of the company" (A6), and the awareness of the resource efficiency variables of the companies can be summarized as follows:

- A6 – Q1: Actions taken to be more resource efficient: There is dependence for 15 countries. There is independence for 18 countries.
- A6 – Q2: Planned resource-efficient actions for the next two years: There is dependence for 17 countries. There is independence for 16 countries.
- A6 – Q3: Past two years investment rate per year to be more resource efficient: There is dependence for 18 countries. There is independence for 15 countries.

These results can be summarized in the two following ways:

- If we consider for each couple of variables Ai – Qj, the number of countries (out of 33) for which we have dependence, we have the following statistics:

Table 7. Some descriptive statistic

Minimum	11
Maximum	27
Mean	17.2
Standard deviation	4.54
Median	17

If we consider separately the three variables Q1, Q2 and Q3, we notice that the dependence is higher for Q3 (mean=19.4) then for the two other variables (mean=15.8 and 16.4 respectively).

If we consider for each country the number of couples Ai – Qj (out of 15) for which we have dependence, we have the following statistics:

Table 8. Some descriptive statistics

Minimum	2
Maximum	13
Mean	7.82
Standard deviation	2.83
Median	8

5.2 Partition of the countries into clusters

Starting from the 33 countries represented as sets of items Qi-Aj=dep and Qi-Aj=indep, we applied apriori algorithm with a minimum frequency equal to 20%. We obtained 74 frequent itemsets. Examples of frequent items are shown in Table A.2 of Appendix. Two countries do not belong to the cover of any frequent itemset: Latvia and Luxemburg.

The clustering description we obtain for the remaining 31 countries is the following:

CD = {F1, F2, F3, F4, F19, F31, F39, F40, F48, F53}

With

F1={Q1-A1=indep, Q2-A1=indep, Q3-A1=indep, Q1-A3=indep, Q1-A5=indep}

And C1=Cov(F1) = {Finland, France, Island, Moldova, Serbia, Slovak Republic, UK}

F2={Q1-A1=dep, Q2-A1=dep, Q1-A5=dep, Q4-A5=dep, Q4-A6=dep}

And C2=Cov(F2) = {Austria, Croatia, Germany, The Netherlands, Norway, Poland, Romania}

F3={Q2-A1=dep, Q3-A1=dep, Q3-A4=dep, Q3-A5=dep, Q2-A6=indep}

And C3=Cov(F3) = {Austria, Check Republic, Croatia, Hungary, Ireland, Lithuania, US}

F4={Q2-A1=dep, Q3-A1=dep, Q1-A5=dep, Q3-A5=dep, Q3-A6=dep}

And C4=Cov(F4) = {Austria, Croatia, Irland, Norway, Poland, Romania, US}

F19={Q1-A1=indep, Q2-A1=indep, Q3-A1=indep, Q2-A3=indep, Q1-A4=indep}

And C19=Cov(F19) = {Albania, Bulgaria, Denmark, Finland, France, Portugal, Serbia}

F31={Q1-A1=indep, Q2-A1=indep, Q3-A1=indep, Q2-A3=indep, Q3-A3=indep}

And C31=Cov(F31) = {Belgium, Denmark, Finland, France, Island, Portugal, Serbia}

F39={Q1-A1=indep, Q2-A1=indep, Q2-A5=dep, Q3-A5=dep, Q1-A6=indep}

And C39=Cov(F39) = {Bulgaria, Denmark, Malta, Slovenia, Spain, Sweden, Türkiye}

F40={Q2_A1=indep, Q1-A5=dep, Q2-A5=dep, Q3-A5=dep, Q1-A6=indep}

And C40=Cov(F40) = {Bulgaria, Denmark, Malta, Macedonia, Slovenia, Sweden, Türkiye}

F48={Q4-A3=indep, Q1-A4=dep, Q3-A4=dep, Q1-A5=dep, Q4-A5=dep}

And C48=Cov(F48) = {Belgium, Check Republic, Croatia, Netherlands, Ireland, Livonia, Macedonia}

F53={Q1-A1=indep, Q2-A1=indep, Q2-A3=indep, Q2-A5=dep, Q4-A5=dep}

And C53=Cov(F53) = {Bulgaria, Denmark, France, Italy, Malta, Slovenia, Türkiye}

As can be noticed, these clusters are overlapping. For example, France appears in four clusters of F1, F19, F31 and F53.

6. Conclusion

Our results related to the dependence of characteristics and awareness variables lead us to the following three conclusions:

- Turnover performance for the past two years and last year's turnover are effective on the resource efficiency awareness of the firms since the number of dependent cases is more than the number of independent cases.
- Employee number of the company, one-person company or not, and establishment year of the company are not effective on the resource efficiency awareness of the firms since the number of dependent cases is less than the number of independent cases.
- There is ambiguity for customer profile of the country since number of dependent cases is sometimes more and sometimes less than the number of independent cases.

After obtaining these dependency results and then making a clustering analysis, it showed that there are 10 possible clusters.

Depending on these results, we can suggest some education policies for increasing resource awareness of firms. Since turnover is an effective characteristic variable, an education program can be designed for firms having a high turnover. Moreover, this education can be given to countries belonging to the same cluster at the same time.

Appendix

A.1. Results of the independency tests

Table A.1. Dependency/Independency between awareness variables Qi and characteristic variables Aj

	Q1_A1	Q2_A1	Q3_A1	Q1_A3	Q2_A3	Q3_A3	Q1_A4	Q2_A4	Q3_A4
Albania	0	0	0	1	0	1	0	0	0
Austria	1	1	1	0	1	1	0	1	1
Belgium	0	0	0	0	0	0	1	1	1
Bulgaria	0	0	0	0	0	1	0	1	1
Czech Republic	1	1	1	0	0	0	1	1	1
Croatia	1	1	1	0	1	0	1	0	1
Denmark	0	0	0	1	0	0	0	1	1
Finland	0	0	0	0	0	0	0	0	1
France	0	0	0	0	0	0	0	1	1
Germany	1	1	0	1	1	1	0	1	0
The Netherlands	1	1	0	1	1	0	1	1	1
Hungary	1	1	1	1	0	1	1	0	1
Ireland	0	1	1	1	0	0	1	1	1
Island	0	0	0	0	0	0	1	1	0
Italy	0	0	1	0	0	1	1	0	0
Latvia	1	0	0	1	1	0	1	0	0
Lithuania	0	1	1	0	1	0	1	1	1
Luxemburg	1	1	0	0	0	0	0	1	0
Malta	0	0	1	1	0	0	0	0	0
Macedonia	1	0	1	1	0	0	1	0	1
Moldova	0	0	0	0	1	0	1	0	1
Norway	1	1	1	0	0	0	0	0	1
Poland	1	1	1	0	1	1	1	1	0
Portugal	0	0	0	1	0	0	0	0	0
Romania	1	1	1	0	0	1	1	1	1
Serbia	0	0	0	0	0	0	0	1	0
Slovak Republic	0	0	0	0	1	1	1	1	1
Slovenia	0	0	1	0	0	1	0	0	1
Spain	0	0	1	1	1	0	0	1	1
Sweden	0	0	1	0	1	1	1	1	0
Türkiye	0	0	0	0	0	1	1	0	1
UK	0	0	0	0	1	1	1	1	1
US	0	1	1	0	0	1	1	1	1

Table A.1 (continued)

	Q1_A5	Q2_A5	Q3_A5	Q1_A6	Q2_A6	Q3_A6
Albania	1	1	0	1	0	0
Austria	1	1	1	1	0	1
Belgium	1	0	1	1	1	0
Bulgaria	1	1	1	0	1	0
Czech Republic	1	0	1	0	0	0
Croatia	1	1	1	0	0	1
Denmark	1	1	1	0	1	0
Finland	0	0	1	1	0	1
France	0	1	1	1	0	0
Germany	1	1	1	1	1	1
The Netherlands	1	0	1	1	1	1
Hungary	0	0	1	0	0	0
Ireland	1	1	1	0	0	1
Island	0	0	1	0	1	1
Italy	0	1	1	1	0	0
Latvia	1	0	1	1	1	0
Lithuania	1	0	1	1	0	0
Luxemburg	0	0	0	1	1	0
Malta	1	1	1	0	0	0
Macedonia	1	1	1	0	1	1
Moldova	0	0	0	0	0	0
Norway	1	1	1	0	1	1
Poland	1	1	1	1	1	1
Portugal	1	0	0	1	1	0
Romania	1	0	1	1	1	1
Serbia	0	0	0	0	0	1
Slovak Republic	0	1	1	1	1	0
Slovenia	1	1	1	0	0	1
Spain	0	1	1	0	0	1
Sweden	1	1	1	0	1	1
Türkiye	1	1	1	0	1	1
UK	0	1	0	0	1	1
US	1	1	1	0	0	1

A.2. Examples of frequent itemsets

F1={Q1-A1=indep, Q2-A1=indep, Q3-A1=indep, Q1-A3=indep, Q1-A5=indep}

F2={Q1-A1=dep, Q2-A1=dep, Q1-A5=dep, Q3_A5=dep, Q3-A6=dep}

….

F16={Q1-a-A1=indep, Q2-A1=indep, Q1-A3=indep, Q3-A3=dep, Q2-A5=dep}

F17={Q1-A1=indep, Q1-A3=indep, Q3-A3=dep, Q2-A5=dep, Q4-A5=dep}

…

F73={Q1-A1=indep, Q2-A3=indep, Q1-A5=dep, Q2-A5=dep, Q3-A5=dep, Q1-A6=indep}

F74={Q2-A3=indep, Q3-A4=dep, Q1-A5=dep, Q2-A5=dep, Q3-A5=dep, Q1-A6=indep}

References

Agrawal R., Ramakrishnan S. (1994), Fast algorithms for mining association rules in large databases, in: Proceedings of VDLB 94, pp. 487–499, https://www.vldb.org/conf/1994/P487.PDF [17.03.2024].

Anderson S.T., Newell R.G. (2004), Information programs for technology adoption: the case of energy-efficiency audits, "Resource and Energy Economics", vol. no. 1, pp. 27–50.

Andrejovska A. Buleca J., Hudakova M. (2016), Categorization of the EU countries in the context of agricultural production, "Agris On-line Papers in Economics and Informatics", vol. 8 no. 4, pp. 3–14.

Athey S. (2018), The impact of machine learning on economics, in: The Economics of Artificial Intelligence: an Agenda, University of Chicago Press, Chicago, pp 507–547.

Bail F., Ester M., Xu X. (2002), Frequent term-based text clustering, in: Proceeding of the 8th International Conference on Knowledge Discovery and Data Mining, ACM, Alberta, pp. 436–442.

Bodas-Freitas I., Corrocher N. (2019), The use of external support and the benefits of the adoption of resource efficiency practices: An empirical analysis of European SMEs, "Energy Policy", no. 132, pp. 75–82.

Boone D.S. (2011), Determination of the number of clusters in a data set. A stopping rule × clustering algorithm comparison, "International Journal of Strategic Decision Sciences", no. 2, pp. 1–13.

Calinski T., Harabasz J. (1974), A dendrite method for cluster analysis, "Communications in Statistics", no. 3, pp. 1–27.

De Groot H.L., Verhoef E.T., Nijkamp P. (2001), Energy saving by firms: decision-making, barriers and policies, "Energy Economics", vol. 23 no. 6, pp. 717–740.

Di Maio F., Rem P.C., Balde K., Polder M. (2017), Measuring resource efficiency and circular economy. A market value approach, "Resources, Conservation and Recycling", no. 122, pp. 163–171.

Delmas M.A., Pekovic S. (2015), Resource efficiency strategies and market conditions, "Long Range Planning", vol. 48 no. 2, pp. 80–94.

Diederen P., Van Tongeren F., Van Der Veen H. (2003), Returns on investments in energy-saving technologies under energy price uncertainty in Dutch greenhouse horticulture, "Environmental and Resource Economics", vol. 24 no. 4, pp. 379–394.

Fleiter T., Schleich J., Ravivanpong P. (2012), Adoption of energy-efficiency measures in SMEs. An empirical analysis based on energy audit data from Germany, "Energy Policy", vol. 51, pp. 863–875.

Forgy E. (1965), Cluster analysis of multivariate data: efficiency versus interpretability of classifications, "Biometrics", no. 21, pp. 768–769.

Ghoddusi H, Creamer G.G., Rafizadeh N. (2019), Machine learning in energy economics and finance. A review, "Energy Economics", no. 81, pp. 709–727.

Gobel M., Araújo T. (2020), Indicators of economic crises. A data-driven clustering approach, "Applied Network Science", vol. 5, art. 44.

Horbach J. (2016), The impact of resource efficiency measures on performance in small and medium-sized enterprises, "Ruhr Economic Papers", no. 643. https://www.econstor.eu/bitstream/10419/146752/1/869244205.pdf [03.03.2024].

Jain A.K., Murty M.N., Flynn P.J. (1999), Data clustering. A review, "ACM Computing Surveys", vol. 31 no. 3, pp. 264–323.

Jové-Llopis E., Segarra-Blasco A. (2018), Eco-efficiency actions and firm growth in European SMEs, "Sustainability", vol. 10 no. 1, pp. 1–26.

Judge G.G. (2016), Some comments on the current state of econometrics, "Annual Reviews of Resource Economics", vol. 8, pp. 1–6.

Kostka G., Moslener U., Andreas J.G. (2011), Barriers to energy efficiency improvement. Empirical evidence from small-and-medium sized enterprises in China, Frankfurt School Working Paper Series, no. 178, https://www.econstor.eu/bitstream/10419/52681/1/678802297.pdf [03.03.2024].

MacQueen J.B. (1967), Some methods for classification and analysis of multivariate observations, in: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297, https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Some-methods-for-classification-and-analysis-of-multivariate-observations/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992 [03.03.2024].

Milligan G.W., Cooper M.C. (1985), An examination of procedures for determining the number of clusters in a data set, "Psychometrika", no. 50, pp. 159–179.

Milton J.S., Arnold J.C. (2003), Introduction to probability and statistics, 4th ed., McGraw-Hill, New York.

Muthulingam S., Corbett C.J., Benartzi S., Oppenheim B. (2011), Investment in energy efficiency by small and medium-sized firms. An empirical analysis of the adoption of process improvement recommendations, Working Paper: Anderson Graduate School of Management – Decisions, Operations, and Technology Management UC Los Angeles, https://escholarship.org/uc/item/6545t5bf [03.03. 2024].

Mullainathan S., Spiess J. (2017), Machine learning. An applied econometric approach, "Journal of Economic Perspectives", vol. 31 no. 2, pp. 87–106.

Ozbugday F.C., Findik D., Basci S., Metin Ozcan K. (2020), Resource efficiency investments and firm performance. Evidence from European SMEs, "Journal of Cleaner Production", no. 252, pp. 1–10.

Ozbugday F.C., Findik D., Basci S., Metin Ozcan K. (2020), Attitudes of SMEs towards energy efficiency and distributed generation. The Turkish case, "Regulations in the Energy Industry", pp. 147–168.

Ramachandran V., Shah M.K., Moss T.J. (2018), How do African firms respond to unreliable power? Exploring firm heterogeneity using K-Means clustering, Center for Global Development Working Paper no. 493, https://ssrn.com/abstract=3310490 [03.03.2024].

Schleich J. (2009), Barriers to energy efficiency. A comparison across the German commercial and services sector, "Ecological Economics", vol. 68 no. 7, pp. 2150–2159.

Schleich J., Gruber E. (2008), Beyond case studies. Barriers to energy efficiency in commerce and the services sector, "Energy Economics", vol. 30 no. 2, pp. 449–464.

Trianni A., Cagno E. (2012), Dealing with barriers to energy efficiency and SMEs: some empirical evidences, "Energy", vol. 37 no. 1, pp. 494–504.

Velthuijsen J.W. (1995), Determinants of investment in energy conservation, Foundation for Economic Research of the University of Amsterdam (SOE), Amsterdam.