TALLINN UNIVERSITY OF TECHNOLOGY

School of Business and Governance

[redacted]

THE EFFECT OF TAX BURDEN ON ECONOMIC GROWTH

Multivariate statistics project

Study group: TAAM11

Tallinn 2022

TABLE OF CONTENTS

RESEARCH PROBLEM

High tax burden may have many negative effect on country's economic growth. For example, it could weaken the investment environment by driving away foreign and/or institutional investors, cause people to travel to neighbouring countries to work or for shopping, increase shadow economy or informal employment. Therefore, the impact of tax burden on economic growth has raised questions whether the changes in taxation could be more growth-oriented and less growth inhibiting to contribute to development (Sinevičienė, 2016, pp. 283-284).

Sinevičienė researched the link between tax burden and economic growth by conducting an empirical analysis of the European Union countries except Croatia. The study covered annual data of 2003-2012 of implicit tax rates used as a measure of tax burden. Reasoning for using implicit tax rates instead of tax revenues as a percentage of gross domestic product (hereinafter GDP) is that according to European Commission the implicit tax rates are more accurate providing insight whether the tax rates are high or the tax base large. Additional indicators used in the study were government debt and government revenue as a percentage of GDP, GDP per capita and growth rate of real GDP. (*Ibid.*, pp. 284, 287-288)

As countries strive for a steady economic growth, it is important to consider all aspects of it. However, due to the focus of this project, only tax burden is taken into account. For that reason, the analysis is not providing a complete perspective of the economic growth of the countries analysed. Research methods used in the Sinevičienė (2016, p. 287) study were descriptive statistics analysis, cluster analysis and correlation analysis. First, the cluster analysis was performed to classify European Union countries into clusters using GDP per capita as the grouping variable. Next, the descriptive analysis of the clusters was presented. Lastly, correlation analysis was performed, and results analysed.

Respectively to the previously conducted study, following research questions are raised:

1. What is the optimal number of clusters of European Union Member States concerning tax burden and economic growth?
2. Is there a link between tax burden and economic growth?

The first research problem is exploratory, and the second research question is inferential. Consequently, two methods are used for analysis: clustering and principal component analysis.

DATA

The data is obtained from Eurostat database, whereas implicit tax rate indicators calculated for the annual report of taxation trends in the European Union were downloaded from Eurostat webpage (Eurostat, 2022a). The project is based on 2018 data to avoid the impact of recent economic shocks (COVID and energy crisis, Russian-Ukrainian war, recession). However, more recent data is also available and could be used in future projects.

Variables used in the analysis were mostly the same as in the Sinevičienė (2016) study:

1. implicit tax rate on consumption as % of total tax revenue of the category[1];
2. implicit tax rate on labour as % of total tax revenue of the category;
3. implicit tax rate on capital as % of total tax revenue of the category;
4. GDP per capita as chain linked volumes (2015) in million euro;
5. growth rate of real GDP chain linked volumes as % change on previous period;
6. government consolidated deficit[2] as % of GDP;
7. government consolidated debt as % of GDP.

All variables were extracted from Eurostat database and webpage, no variables were calculated.

Descriptive statistics of the data in 2018 is presented in Table 1. The table shows that there are quite large differences between countries thus proving the need for clustering as in the Sinevičienė (2016) study also found.

Table 1: Descriptive statistics of European Union Member States, Iceland, Norway in 2018.

Variable	N	Mean	Standard deviation	Median	Min	Max
Implicit tax rate on consumption	29	19,2	2,7	18,7	13,9	24,7
Implicit tax rate on capital	26	22,7	10,7	23,7	6,1	52,4
Implicit tax rate on labour	28	34,6	5,4	34,3	23,0	42,7
Growth rate of real GDP	28	3,2	1,9	2,8	0,8	8,5

[1] Implicit tax rate of the category (consumption, labour and capital) is calculated as a ratio of total tax revenue of the corresponding category to a proxy of the potential tax base defined using the production and income accounts of the national accounts (Directorate-General for Taxation and Customs Union, European Commission, 2022, p. 299).
[2] Sinevičienė (2016) study uses government revenue as % of GDP. However, in this project government deficit as % of GDP is used instead because it is one of Maastricht criteria's that European Union Member States must comply, whereas government revenue as % of GDP is not (Eurostat, 2022b).

GDP per capita	28	480 015,1	756 208,0	192 157,3	12 165,8	3 207 750,8
Government consolidated deficit, % of GDP	27	-0,2	1,6	-0,1	-3,6	3,0
Government consolidated debt, % of GDP	27	65,6	39,8	61,3	8,2	186,4

Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

The dataset selected enables to cluster different countries and describe these clusters by a number of variables. The high dimensionality of the data also provides good basis to perform principal component analysis and correlation analysis. As a result, the dataset can be used to implement selected methods to answer the research questions.

METHODOLOGY

In accordance with the literature and research questions clustering with hierarchical clustering and k-means and principal component analysis were chosen as statistical methods for the project. These methods were accompanied by correlation analysis and descriptive statistics. Clustering method provides an answer to the first research question and a basis for the second research question. Principal component analysis, correlation analysis and descriptive statistics of clusters provide an answer to the second research question.

Analysis with different statistical methods was performed using R[3]. First, hierarchical clustering was performed to find the optimal number of clusters. Before clustering, the data was prepared by converting it into a matrix and scaled. It should be noted that some of the variables were not available for some countries, hence these observations were dropped[4]. After that, different linkage methods (complete, single, average, ward) were analysed to find the method that gives the best clusters by calculating the agglomerative coefficient of each method. Finally, hierarchical clustering with Euclidean distance matrix was performed using the selected method and optimal number of clusters. Secondly, k-means clustering was performed using the optimal number of clusters identified with hierarchical clustering to verify the results. Thereafter, descriptive statistics of clusters was produced to analyse the groups rather than individual countries.

Next, principal component analysis was performed to produce a low-dimensional representation of high-dimensional data. The same data matrix was used as an input for principal component analysis as in hierarchical clustering. An indicator of total variance explained by each principal component was also calculated and plotted to visualize how much variance each principal component explains. In addition, a biplot was made to visualize the results of principal component analysis.

Lastly, correlation matrix was produced to analyse the correlation between different variables and verify findings of previous methods.

[3] The R script used for analysis is provided as a separate file.
[4] Iceland and Norway data was dropped because the first lacked implicit tax rate, GDP and government deficit and debt data and the latter lacked government deficit and debt data for 2018.

RESULTS AND CONCLUSIONS

The first research question was about finding the optimal number of clusters of European Union Member States that could describe tax burden and economic growth the best possible way given the input data. Firstly, the best linkage method for hierarchical clustering was Ward's method with agglomerative coefficient of 0.769. The optimal number of clusters resulted in 2 (gap = 0.152)[5] (see Appendix 1). Hierarchical clustering using the optimal number of clusters and linkage method resulted in a dendogram. However, k-means clustering produced somewhat different result (see Figure 1).

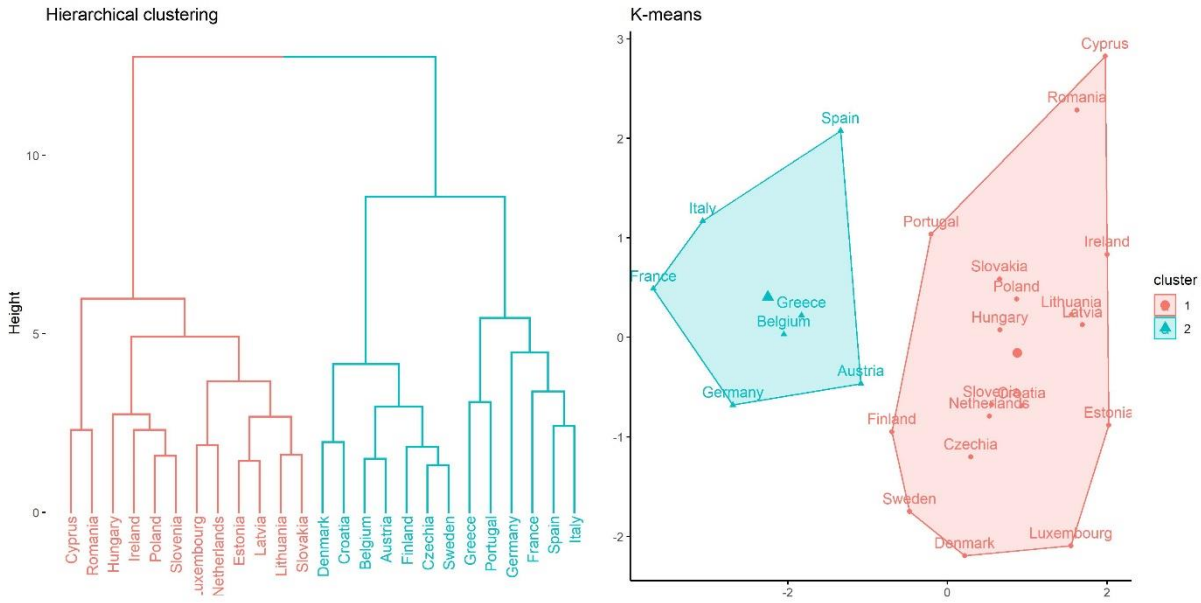


Figure 1. Dendogram (left) and K-means (right) with optimal number of clusters. Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

Majority of the countries were clustered differently with k-means than with hierarchical clustering and only Czechia, Denmark, Croatia, Portugal, Finland and Sweden remained in the same cluster with both methods.

[5] Cluster 1 gap = 0.148, cluster 3 gap = 0.145.

The second research question was about the link between tax burden and economic growth. To provide an answer, descriptive statistics of the clusters was compiled using the results of hierarchical clustering. Tables 2 and 3 show the results of descriptive statistics by these clusters.

Table 2: Descriptive statistics of the first cluster.

Cluster	First cluster					
Country	Belgium, Czechia, Denmark, Germany, Greece, Spain, France, Croatia, Italy, Austria, Portugal, Finland, Sweden					
Variable	**N**	**Mean**	**Standard deviation**	**Median**	**Min**	**Max**
Implicit tax rate on consumption	13	18,9	3,1	18,1	13,9	24,7
Implicit tax rate on capital	13	29,9	8,4	28,3	19,6	52,4
Implicit tax rate on labour	13	37,8	4,5	39,7	29,2	42,7
Growth rate of real GDP	13	2	0,7	2	0,9	3,2
GDP per capita	13	833 931,5	989 566,2	367 756,8	49 825,6	3 207 750,8
Government consolidated deficit, % of GDP	13	-0,3	1,4	-0,1	-2,6	1,9
Government consolidated debt, % of GDP	13	86,1	44,1	74,1	32,1	186,4

Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

Table 3: Descriptive statistics of the second cluster.

Cluster	Second cluster					
Country	Estonia, Ireland, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Netherlands, Poland, Romania, Slovenia, Slovakia					
Variable	**N**	**Mean**	**Standard deviation**	**Median**	**Min**	**Max**
Implicit tax rate on consumption	12	19,2	2,5	19,3	14,8	23,1
Implicit tax rate on capital	12	14,2	6	14,3	6,1	24,2
Implicit tax rate on labour	12	32,7	3,4	32,4	26	38,9
Growth rate of real GDP	12	4,6	1,9	4,3	1,2	8,5
GDP per capita	12	180 981,8	227 760,1	72 880,5	21 362,2	742 788,8
Government consolidated deficit, % of GDP	12	-0,4	1,8	-0,4	-3,6	3
Government consolidated debt, % of GDP	12	48,8	24,3	49,1	8,2	98,1

Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

The first cluster can be described as high economic development cluster, where GDP per capita is high and growth rate of real GDP is low. The second cluster can be described as high growth rate

cluster, where the average GDP per capita is 4,6 times lower than the first cluster average, while growth rate of real GDP is much higher. The tax burden on consumption is relatively the same between two clusters. However, there is a distinct difference between the clusters regarding the tax burden on capital. Countries in the first cluster tax capital 15,7 percentage points more on average than countries in the second cluster. The second cluster also tax labour 5,1 percentage points less on average than countries in the first cluster. Concerning government deficit, the mean values do not differ significantly between two clusters. However, government debt is much higher in the first cluster.

To visualize the results of principal component analysis, a biplot was plotted (see Figure 2).

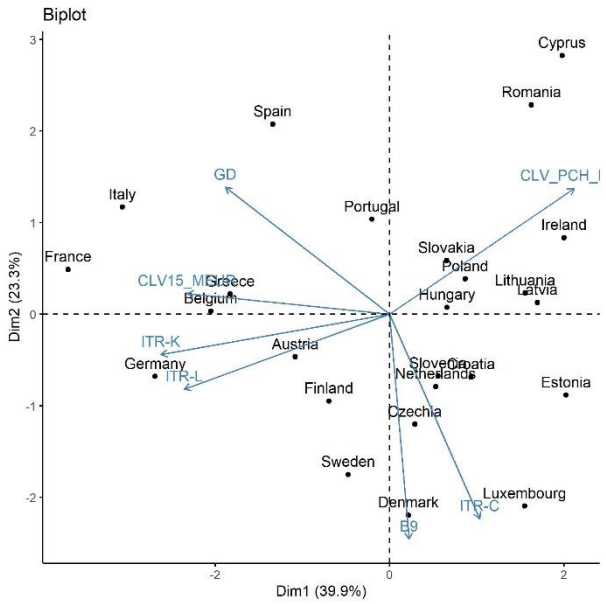


Figure 2. Principal component analysis biplot.

Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

The biplot in Figure 2 shows that for example the highest real GDP growth rate can be associated with Cyprus, Romania and Ireland, the highest implicit tax rate on capital is observed in countries such as France, Belgium and Germany. These assumptions are also in line with the data. Additionally, total variance explained by each principal component was plotted on a scree plot (see Appendix 2). The scree plot indicates that the first principal component explains 39,9% of total variance in the dataset, the second principal component explains 23,3%, the third 12,8% and the rest fall below 1% of total variance explained. Therefore, the biplot in Figure 2 explains nearly 2/3 of the total variance in the data and should be considered when looking for patterns in the data.

Lastly, correlation matrix was produced to look for any correlation between the variables (see Appendix 3). It can be assumed that there is a positive correlation between tax burden on capital and tax burden on labour (0.655). However, there is also a negative correlation between tax burden on capital and on labour and real GDP growth rate (-0,495 and -0,469 respectively), yet a positive correlation between GDP per capita (0,56). These conclusions are also supported by the cluster analysis.

The findings of this project correspond with the general results of Sinevičienė (2016) study. Thus, it can be concluded that there could be a link between tax burden and economic growth, as countries with high economic development have higher overall tax burden (first cluster) and countries with lower economic development have lower overall tax burden (second cluster). However, the economic growth of the already highly developed countries is much smaller, and the government debt is much higher than the countries with lower economic development.

There were also some limitations of this analysis. Firstly, the dataset was restricted with only one year. To provide a more in-depth analysis a longer timeseries should be considered and outliers in the data should be analysed further. Secondly, the clustering methods provided somewhat different results and should be analysed further to arrive at the best number of clusters. Furthermore, the principal components used in the analysis explained about 2/3 of the total variance. Depending on the research, this ratio might be improved. Despite these limitations, the results of this analysis were generally in line with the previous study, hence the overall conclusions can be trusted to some extent.

REFERENCES

Directorate-General for Taxation and Customs Union, European Commission. (2022). *Taxation Trends in the European Union.* Luxembourg: Publications Office of the European Union. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/f85da28f-f5be-11ec-b976-01aa75ed71a1

Eurostat. (2022a). *Data on Taxation*. Retrieved December 22, 2022, from European Commission: https://taxation-customs.ec.europa.eu/taxation-1/economic-analysis-taxation/data-taxation_en

Eurostat. (2022b). *Excessive Deficit Procedure*. Retrieved December 22, 2022, from Eurostat: https://ec.europa.eu/eurostat/web/government-finance-statistics/excessive-deficit-procedure

Eurostat. (2022c). gov_10dd_edpt1: Government deficit/surplus, debt and associated data. Retrieved December 22, 2022, from https://ec.europa.eu/eurostat/web/main/data/database

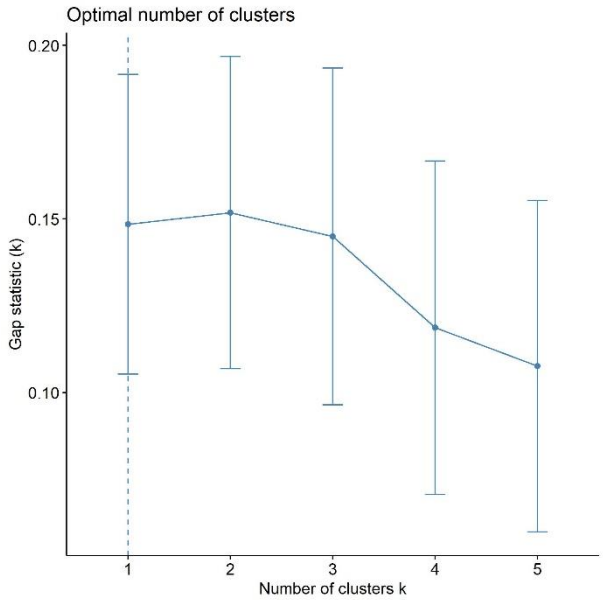
Eurostat. (2022d). nama_10_gdp: GDP and main components (output, expenditure and income). Retrieved December 22, 2022, from https://ec.europa.eu/eurostat/web/main/data/database

Eurostat. (2022e). tec00115: Real GDP growth rate - volume. Retrieved December 22, 2022, from https://ec.europa.eu/eurostat/web/main/data/database

Sinevičienė, L. (2016). Tax Burden and Economic Development: The Case of the European Union Countries. (M. Bilgin, & H. Danis, Eds.) *Entrepreneurship, Business and Economics, 2*, 283-298. doi:https://doi.org/10.1007/978-3-319-27573-4_19

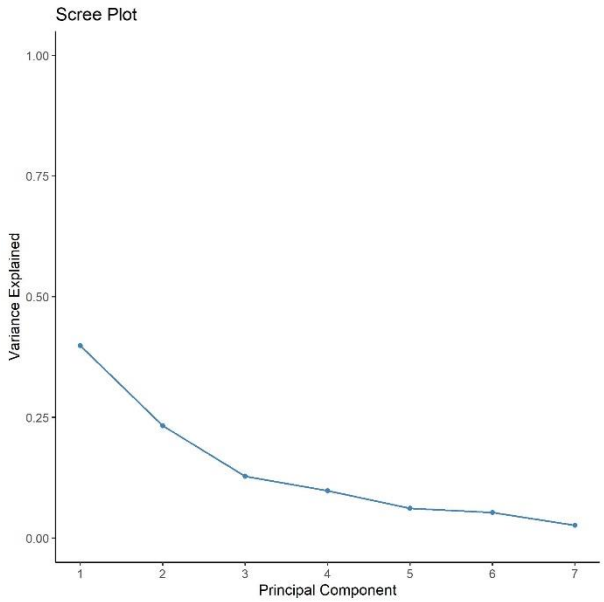
APPENDICES

Appendix 1. Optimal number of clusters.



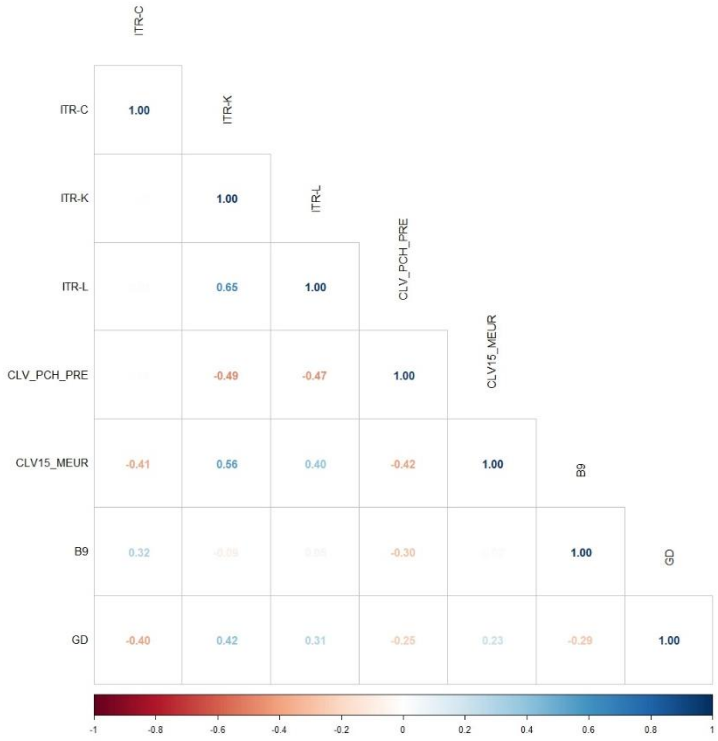
Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

Appendix 2. Scree plot of total variance explained by each principal component.



Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.

Appendix 3. Correlation between the variables.



Source: Eurostat, 2022a; Eurostat, tables tec00115, nama_10_gdp, gov_10dd_edpt1; author's calculations.