Erasmus School of Economics

Master Thesis Economics and Business

Specialization Policy Economics

Debt Burden Effect on Economic Growth: Investigation with the help of clustering and Fixed Effects

638730

Ioanna Petroula Karlits

Supervisor: Prof. Agnieszka Markiewicz

Second assessor.

Abstract

The current thesis studies aspects of the economy of Eurozone countries, with the exception of Croatia which is not included in the analysis as it entered the Eurozone after 2018.

Through the method of agglomerative clustering the convergence of economies was studied. As a result of this process, two groups of countries were created based on the level of economic growth and a number of factors that are considered important in the theory of economic growth. The one group includes three former Soviet countries (Estonia, Lithuania, Latvia) and the other one includes Austria, Belgium, France, Netherlands, Finland, Germany, Ireland, Luxembourg, Malta, Greece, Slovak Republic, Slovenia, Cyprus, Italy, Portugal and Spain (16 countries).

As the ex-Soviet countries are atypical countries, in the rest of the analysis there was a focus only on the 16 countries. More specifically, the impact of variables on economic growth was examined (those variables are recognized by economic theory as factors of economic growth). Additionally, emphasis is given to identifying the debt threshold over which there is an additional negative effect on economic growth. Based on the results this is observed with the threshold of 80%.

Keywords: hierarchical agglomerative clustering, fixed effects, Eurozone, debt threshold

Contents

Abbreviation Table

Abbreviation	Meaning	Page
ECSC	European Coal and Steel Community	4
EEC	European Economic Community	4
GDP	Gross Domestic product	5
SGB	Stability and Growth Pact	5
ECB	European Central Bank	5
FE	Fixed effects	6
MST	minimal spanning tree	7
HT	hierarchical tree	7
GCI	global competitiveness index	8
EFI	economic freedom index	8
GII	global innovation index	8
DBSCAN	density-based spatial clustering of application with noise	8
HDI	human development index	9
CPI	corruption perceptions index	9
OLS	Ordinary least squares	13
GMM	Generalized method of moments	15
VAR	vector autoregression model	15
FORD	fields of research and development	19

Note: The table includes abbreviations utilized in the research. The meaning of each one is given in the second column. The third column refers to the page where the variable appears for the first time in the text

1. Introduction

After the end of the Second World War, in order to prevent future military conflicts between European countries, European politicians initiated the process of creating today's European Union. So, in 1952 six countries, France, Belgium, Netherlands, Luxembourg, West Germany, and Italy, created the European Coal and Steel Community (ECSC), following the Schuman Declaration. So, in spite of the fact that the aim of European Integration was a political one, the means to achieve this, were economical (*History of the EU, EU Pioneers | European Union*, n.d.). The next step to this project of European Integration was the creation of the European Economic Community (EEC) in 1957 by the Treaty of Rome, with France, Belgium, Netherlands, Luxembourg, West Germany and Italy. EEC adopted a common agricultural policy (1962) and established a custom union (1968). In the following years, EEC gained new Country-members, Denmark, Ireland, United Kingdom, Greece, Portugal, and Spain. Then, in 1993 the treaty of Maastricht created the European Union which was enlarged by new members. Today, European Union has 27 members (*History of the EU, EU Pioneers | European Union*, n.d.).

As was already mentioned, in spite of the fact that the initial aim of European Integration was a political one, this has been realized through economic procedures, aiming at the European economic integration with main objective the creation of a European Single Market. Since the main obstacles to achieve this goal are the (many) differences in national regulations, a very serious process began in 1986 with the adoption of the Single European Act in order to sort out these differences. This process led in 1993 to the creation of the European Single Market, with the Treaty of Maastricht. The single market ensured the free movement of people, goods, services, and money (*History of the EU, EU Pioneers | European Union*, n.d.).

The last step was the adoption of a single European currency, the euro, on the 1st of January 1999, by 11 countries (Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain), followed by Greece in 2001. On the 1st of January 2002, euro notes and coins became the legal currency in these 12 EU countries (*History of the EU, EU Pioneers | European Union*, n.d.).

The process of the adoption of the single currency started in April 1989, when it was suggested by Delor's Report that the economic and monetary union should be accomplished in three stages;

The first stage was initiated in July 1990, with the introduction of complete freedom of capital movements. Afterwards, the second stage in January 1994 followed, with the coordination of the central banks and monetary policies of the country members for the preparations to adopt a single currency. During this stage, the Stability and Growth Pact (SGB) was adopted. The SGB set two limits on Member States: a country's budget deficit must not exceed 3% of gross domestic product (GDP) and national debt cannot surpass 60% of GDP (but then, the SGB undergoes changes). In the third stage which was initiated in January 1999, there was modification of the exchange rates of the existing at that time Member States. Finally, during this stage, there was the realization of the common Monetary policy with the coordination of the European Central Bank (ECB). On 1 January 1999, the third and final stage of EMU commenced with the irrevocable fixing of the exchange rates of the currencies of the 11 Member States initially participating in the Monetary Union and with the conduct of a single monetary policy under the responsibility of the ECB. In the following years, the number of Member States increased till recently.

The creation of the European Single market with a single currency was expected to reinforce the economy of the country-members through economies of scale, because of the greater market area, eliminate the transaction costs, decrease the differences between the prices of the same good in the countries of the single market area and reduce the volatility of the exchange rates between these countries. On the other hand, the adoption of the common currency implies that each state cannot exercise its own economic policy for the necessary macroeconomic adjustments (Taylor & Mankiw, 2014). This was the case with the crisis of 2008, where Ireland, Portugal, Spain, Italy, and Greece faced more severe economic problems compared to other Eurozone countries.

This differentiation inspired me to study how a grouping of Eurozone countries can be formed, according to the differences and similarities of their economy. More specifically, emphasis is given to economic growth and a set of variables concerning various factors of it.

Besides, after the financial crisis of 2008, there is growing interest in the relationship of debt with economic growth. So, in this paper, there is an attempt to determine the debt threshold for countries that are more similar based on the grouping.

Additionally, I investigate how gross fixed capital formation research and development expenditure and growth rate of employees affect economic growth for this specific group.

To study the question of the grouping formation the method of the agglomerative average clustering with average linkage is applied. The result of this process is the dendrogram, based on

which countries will be organized into groups/clusters. The countries that belong to the same group are the ones that are more similar compared to the others which are not included in the same cluster. Based on the groups fixed effects (FE) is applied for a subset of variables of the ones applied in the clustering. The variables of the econometric model are gfcf, r&d, gremployees, and debt. Those are the variables of evaluation interest concerning the way they affect grgdp (economic growth).

From the clustering process, the former Soviet countries were separated from the rest of the Eurozone countries, as countries that had to make a new beginning, as independent countries, after their liberation from the Soviet Union. Consequently, by investigating the relationship between the (independent) variables of interest and economic growth (dependent), the expected positive relationship between the growth rate of employees and growth rate of GDP is verified, while the quadratic effect of debt on growth is observed beyond the threshold of 80%.

This research has the following structure. It starts with the literature review in section 2, followed by section 3 where the data of the thesis are explained. In section 4 the methodology is explained. Afterwards, in section 5 the results are elaborated. Finally, in section 6 the conclusions of the research are presented.

2. Literature Review

2.1 Cluster analysis and analysis of convergence among the European countries

One crucial condition for an optimal currency area is the homogeneity of the consisting countries (Mankiew & Taylor, 1014), concerning a lot of macroeconomic issues (Mankiew & Taylor, 1014). So, because of the differences that exist between the countries of the Eurozone, convergence has been of great importance among the Eurozone construction process.

A useful method to analyze convergence is the agglomerative hierarchical clustering. Using this method, Irac and Lopez (2015) worked on classification of the first 12 country members of the Eurozone trying to identify patterns and trends of structural convergence between these countries over time. The analysis used 27 different indicators from 1997 to 2012. This time interval was separated in two periods: 1997 to 2007 (pre-crisis period) and 2007 to 2012 (post – crisis period). The clustering method resulted in forming two clearly different structural clusters: the "South Country Group" (with 4 countries, Greece, Italy, Portugal and Spain) and the "Other Country Group" (Austria, Belgium, Finland, France, Germany, Ireland, Luxemburg and the Netherlands). But instead of convergence, the paper states that there is no evidence of between-cluster convergence for the years 1999 to 2012, a result that indicates the existence of structural problems for European integration.

In another approach Kantar, Deviren and Keskin, (2014) studied the relationships between the Eurozone countries according their debt to GDP ratio, for the period 2000 to 2011.

This investigation was carried out with the use of hierarchical structure methods (minimal spanning tree, (MST) and hierarchical tree, (HT)). The period 2000 to 2011 was divided in two sub-periods 2000 – 2004 and 2005 – 2011, because in the year 2004 European Union has been enlarged. The MST was created by transforming the correlation coefficient between any two countries (that is between the debt to GDP ratios of the two countries) into a distance between them. This transformation is realized through an appropriate function as a metric. Then, the construction of HT follows based on MST. The analysis resulted in the construction of clustered structures of trees which led to the formation of different clusters of countries based on their debt

to GDP ratio. Then, in order to examine the statistical reliability of the links of MSTs and HTs, the bootstrap technique was applied. Moreover, the clustering linkage procedure was applied for a clearer observation of the cluster structure. The above analysis showed that the countries of the Eurozone which were most affected by the economic crisis of 2009 were found in the same cluster.

Proença, Neves, Dias and Martins (2021) used a clustering methodology in order to classify countries into clusters based on their sovereign debt ratings as provided by the three main rating agencies. The study covers the cases of 32 European countries and provides explanation for their specific classification into (financial) strong or weak groups and identifies the relevant determining factors. The analysis is based on panel data estimation using an ordered Probit approach. for two sub-periods, the first from 2001 to 2008 and the second from 2009 to 2016. This split into the two sub – periods before and after the financial crisis, made it possible to examine the possible changes in the different factors which determine the inclusion of the countries into the different groups. It was found that in the second period (after the financial crisis) the main determinants for the classification of a developing country were the unemployment rate, as well as the membership in the Eurozone. Nevertheless, the rich countries' main determinant factor proved to be the inflation.

Pełka (2018) focuses on the grouping European Union states based on how innovative they are, by using the Density-based spatial clustering of application with noise (DBSCAN) combined with ensemble methods. According to the results of the research four clusters were derived. The first one includes the countries which are "innovation leaders", i.e. the United Kingdom Belgium, Austria, France, Germany, Norway, Denmark, Finland, the Netherlands, Ireland, Slovenia and Sweden. The second group includes the "mid-low" countries concerning innovation, i.e. Croatia, Hungary, Poland, Slovakia, Romania and Bulgaria,

In the third and fourth group there are countries which are considered "moderate innovators" . In the third group Italy, the Czech Republic and Spain, are included . Finally Greece and Portugal form the fourth group.

Onuferová, Čabinová and Matijová (2020) attempted to assess the development of the European Union countries concerning the economic and social sector, by using hierarchical clustering with Ward linkage to group the countries based on the global competitiveness index (GCI), economic freedom Index (EFI), global innovation index (GII), corruption perceptions

index (CPI) and human development index (HDI). The authors applied the technique for both 2011 and 2018, to understand the change of the countries in time concerning those 5 variables. From the clustering of the year 2011 four groups were created; group 1 which contains the "Economic leaders": Sweden, Netherlands, Finland, and Denmark, group 2 which includes the "Economically advanced countries": Luxembourg, Ireland, the United Kingdom, Germany, France, Belgium, and Austria, group 3 which consists of the "Economically averaged countries" *:* Malta, Slovenia, Portugal, Poland, Spain, Estonia, and Cyprus and group 4 which contains the "Economically limited countries": Lithuania, Hungary, the Czech Republic, Italy, Slovakia, Latvia, Croatia, Greece, Romania, and Bulgaria. From the clustering of the year 2018 the following groups appeared; The countries called "Economic leaders" (group 1)*:* Sweden, the Netherlands, the United Kingdom, Germany, Finland, and Denmark. The "Economically advanced countries" (group 2) : Luxembourg, Ireland, Estonia, France, Belgium, and Austria. The "Economically average countries" (group 3)*:* Poland, Lithuania, Latvia, Spain, Slovenia, Portugal, the Czech Republic, Malta, and Cyprus and the final group "*Economically limited countries*"(group 4): Greece, Croatia, Italy, Slovakia, Hungary, Romania, and Bulgaria. From the clustering of 2011 Romania and Bulagaria were the most economically similar countries, but Greece and Sweden were the most dissimilar ones. From the clustering of 2018 it was concluded that Lithuania and Latvia were the most similar ones contrary to Greece and Denmark which were the most dissimilar ones.

Concerning the development it was concluded that 22 out of the 28 countries (all countries except, Lithuania, the United Kingdom, Estonia, Czech Republic, Latvia, and Germany) remained stable over time while the rest of them had improved in their development i.e. their corruption was reduced and their competitiveness was augmented.

Mazurek (2014) investigated the grouping of European Union member states and two other countries; Iceland and Norway, based on five variables: unemployment, inflation, labor productivity and growth of GDP, for the time period 2008-2012. The main clustering technique used in this research was K-means. Based on the clustering it was found that 6 clusters were created.

The bigger one consisted of 14 countries of Central and West Europe. The states of this group had similar economic development.

One of the clusters contained together the South European countries (Portugal, Greece and Spain), as well as Ireland which is rational according to the author as those countries were affected by most of the financial crisis of 2008, as they were most severely affected concerning the economic downturn, they had the highest deficit and unemployment.

Another separate cluster includes Hungary, the Balkan and Baltic countries which were also affected by the crisis. The common characteristics of this group were the economic decline, the unemployment and inflation levels above the average, the relatively low debt compared to the average of the group, and the low productivity.

The author also mentions that Poland and Slovakia were placed in the same cluster as they did not experience economic depression. The author also mentions that those countries have one of the lowest productivity from the total sample of countries and high levels of unemployment.

Another cluster consisted of Norway and Luxembourg, which had increased growth of GDP, reduced unemployment, "medium inflation", high-level debt and productivity.

Finally, there was a cluster that contained one country; Iceland. During the period studied by the author, Iceland had an economic decline low level of unemployment, high inflation debt, and productivity.

Another kind of classification of the Eurozone countries into clusters, concerning the causes of the financial crisis was presented by Amado (2022). The purpose of this paper is to investigate how the Eurozone financial crisis occurred by focusing on the period 1996 to 2007 for the 11 European initial Euro countries plus Greece. At first, based on the literature review, carried out by the author, three fundamental causes arose: (I) competitiveness, (II) differences in flows between European North and South countries and (III) excess of public and private spending. Then analysis of each cause contributed to the identification of a series of explanatory variables i.e. real effective exchange rate and the weight of the manufacturing sector in total employment. The different values of these variables for each one of the Eurozone countries were used in order to separate these countries into clusters with hierarchical cluster analysis. The key description that emerged was the distinction between core and periphery countries. In the last part of the paper it is argued that the explanation for the emergence of the three fundamental causes of the crisis lies in the convergence of nominal long-term interest rates in the periphery cluster countries, relative to the core cluster countries.

The idea to investigate the formation of groups of countries based on specific economic characteristics has been realized not only via the agglomerating clustering method, but via complex systems analysis. So, dealing with the same topic as in the previous work (the public debt crisis in Europe), Matesanz and Ortega (2015) examined the quarterly public debt to GDP ratios for the European countries for the 2000 – 2014 period, using methods derived from the complex systems analysis. The purpose of the study was to investigate whether the various countries form networks between themselves in relation to public debt and the see how these networks are evolving. Special attention was given to the effect of the global financial crisis on the evolution of these networks. The analysis showed that the evolution of the public debts tend to move in parallel direction. This trend increases the global connectivity of the countries in the network and especially in the times of financial crisis, countries with similar level of public debt to GDP ratios tend to form more interconnections.

The convergence of the European countries in relation to the formation of the Eurozone has been studied by Ignatov (2023), who examined which convergent determinants are important in the processes of the formation of the EU from 1999 to 2021. The analysis, through the K – means clustering algorithm, revealed that specific aspects of the institutional and fiscal framework of the EU countries (like worsening demographic dynamics rising indebtedness and insufficient regulatory quality, are factors of heterogeneity which point rather to divergence processes.

The existence of heterogeneity in the European Union has been studied from another point of view by Cavenaile and Dubois (2011). This analysis focuses on the differences between the 15 Western European country members of European Union and the new country members from Central and Eastern Europe. The study investigated the process of convergence of the countries belonging to European Union from 1990 to 2007 and used the growth model of Solow, from which a convergence equation is derived. Then a panel methodology to this equation showed that these two groups of countries (first set of countries and newcomers after 2000) differed also in terms of convergence. This result gives a proof of the heterogeneity that exists among the countries of European Union.

2.2 Variables of evaluation interest concerning economic growth

In the literature there is a large number of publications on economic growth and the factors that affect it. For the needs of the thesis, focus will be given on the following factors: Debt, Gross Fixed Capital Formation, Growth rate of Employment and Research and Development.

2.2.1 Debt

Below the relationship of general government debt and economic growth will be analyzed.

Cecchetti, Mohanty and Zampolli (2011) investigate the effect of household, non-financial corporate and government debt on economic growth in the long run and short run by using the panel data analysis fixed effects. Focusing their research on a sample of 18 OECD countries and a time period of 31 years (1980-2010) they find that beyond the threshold 85% of government debt as a percentage of GDP there is negative impact of debt on economic growth.

After the threshold a 10 percentage increase of debt as a fraction of GDP per capita decreases economic growth by 10-15 basis points. The measurement of economic growth (long run) is represented by a five year overlapping average of GDP per capita growth and the debt is a percentage of GDP. Furthermore the authors mention ways through which debt affects growth. More precisely, in low levels debt contributes to capital rise and "allocative efficiency" which increase economic growth. The authors attribute the negative relationship on the ageing of population which results on higher government spending and reduction in tax revenue (increase in deficit) and consequently to an increase in public debt. In addition they ascribe this negative relationship to the increase in dependency ratios, which decrease economic growth over the threshold.

In addition, they find that high public debt affect negatively credit flows and as a result economic growth. Another result of the study is that government debt affects economic growth negatively both during periods of crisis and periods of stability. However debt has smaller negative impact on growth during the non-crisis period.

Afonso and Alves (2014)) examine as well the relationship of general gross government debt with economic growth (represented by the growth rate of real GDP) in the long term and short

term. Their study concerns 14 European countries for the years 1970-2012. The authors use the panel data techniques of OLS, GLS, 2SLS and FE. They also observe as Cecchetti et al (2011) a nonlinear relationship between debt and growth. Nevertheless, they find a different threshold is 74% compared to the previous paper. Most probably this is due to the different methodology technique used in each of the papers. Another investigation about this non-linear relationship comes from Égert (2013) (whose focus is on 20 countries and time period 1960-2009. The author finds that after the general government as a percentage of GDP threshold of 50% (approximately) debt affects negatively growth.

The threshold of debt varies across studies. According to Égert (2013) the size of debt effects and the threshold depend substantially on the country and time period. For instance the level of the threshold may depend on the way governments allocate debt to handle exogenous shocks (Checherita & Rother, 2010).

Checherita and Rother (2010) also attempted to find the relationship between government debt and economic growth by studying 12 Eurozone countries (France, Germany, Italy, Luxembourg, Austria, Ireland, Belgium, Finland, Greece, Netherlands, Spain and Portugal,) for the period 1970-2011, by using models such as FE. The authors found that before a certain threshold point economic growth is affected positively by debt while after the threshold debt affects negatively growth. More precisely they that beyond 90-100% of debt to GDP, debt affects negatively economic growth. Nevertheless, they mention that this negative effect may commence from lower levels of debt to GDP at 70-80%, which indicates that countries should be even more careful concerning their debt policies.

2.2.2 Gross Fixed Capital Formation

The gross fixed capital formation was mentioned as an important factor for economic growth by Solow (Solow, 1956), Romer (Romer, 1986), and Barro (Barro, 1990). Regarding the relationship of gross fixed capital formation and economic growth, Lymonova (2019) found via the ordinary least square (OLS) method for the member states of the Euro area during the period 2002-2017 that gross fixed capital formation affects positively growth.

Nevertheless, Mehrara and Musai (2013, September) by using panel techniques found the opposite result, i.e., they found that economic growth solely affects gross fixed capital formation. The results refer to the Middle East and North Africa region.

Ncanywa and Makhenyane (2016) investigated the effect of gross fixed capital formation on economic growth in the region of South Africa from 1960 to 2014. The authors applied the Johansen co-integration and the vector error correction model. According to their research, it was found that there is a bidirectional positive relationship between the two variables.

2.2.3 Growth rate of Employment

The Growth rate of employees refers to the growth rate of employment.

Sodipe and Ogunrinola (2011) observed a negative relationship between employment growth rate and economic growth, contrary to the relationship between employment level and growth. Those conclusions were drawn based on the OLS methodology applied for Nigeria.

Maestas, Mullen and Powell (2023, April 1) detected also this positive relationship by capturing the negative relationship between population aging and economic growth. More precisely, they found that for every 10% increase in the population over the age of 60 years old there is a 5.5% decrease in the GDP per capita. According to the authors, one-third of this decrease is caused by the decrease in employment growth. The research was focused on the USA for the time period 1980-2010.

By studying European Union countries from 1996 until 2019 (with the multifactorial regression methodology), Soava, Mehedintu, Sterpu and Raduteanu (2020) concluded that the relationship between growth and employment rate is positive.

2.2.4 Research and Development

By using a “multivariate regression” Sylwester (2001) attempted to specify the relationship between economic growth and research and development in 20 countries that belong to the OECD. Based on their results, there is no “strong association” between growth and research and development. This result is compared with the positive relationship between those two variables in the group of G-7. One possible explanation according to the author is that research and

development plays a more deterministic role in the growth of countries that have high research and development.

Guloglu and Tekin (2012) by examining the relationship between economic growth and research and development expenditure with the contribution of three different econometric techniques; vector autoregression model (VAR), generalized method of moments (GMM) and FE Guloglu and Tekin (2012), concluded that research and developent expenditure affects positively economic growth. Those results were derived based on a sample of 13 developed countries of the OECD group for the time period commencing in 1991 and ending in the year 2007. The authors also mentioned that their results are compliant with both "technology-push" and "demand-pull" theories. The former theory refers to the fact that research and development is the driving force through which new ideas and products are created and as a result economic growth augments. The latter theory expresses the point of view that a higher market expansion is correlated with higher innovation as research and development is "demand-led".

Furthermore, Abdulkadir Afriana and Azis (2020) examined this relationship by using as a research sample 33 OECD countries. The conclusions derived by the authors are that research and development impacts negatively economic growth.

3. Data

For this research, data on an annual basis were initially derived from the OECD, World Bank, and IMF, concerning the period 1996-2018. Filling the missing data was realized by applying the method of linear interpolation. This method helped in estimating the missing values of a variable by calculating the linear equation between the value before the missing value and the value after the gap.

The variables used in the clustering are:

- growth rate of GDP (grgdp)
- natural log of GDP per capita (lngdp),
- growth rate of Employees (gremployees)
- gross Fixed Capital Formation (gfcf)
- research and development (r&d)
- government Effectiveness Indicator or government effectiveness (goveff)
- debt (debt)
- tax revenue (taxrev)
- inflation (inflation)

The variables used in the econometric model are:

- growth rate of GDP (grgdp)
- gross Fixed Capital Formation (gfcf)
- research and development (r&d)
- growth rate of employees (gremployees)
- Two dummy variables concerning debt (dummydebt, dummydebtsq). I included two dummy variables through a trial and error process:

Initially I included only debt, as a variable it did not work, the debt t-ratio was statistically insignificant. Then I tried to differentiate between higher and lower debt-burdened countries invoking a debt to GDP threshold, thus including a dummydebt variable; it remains statistically insignificant. Then I used two dummy variables one for the countries above the threshold (dummydebtsq) to catch any quadratic debt effect and one dummy variable for the countries below the debt threshold (dummydebt), to catch the "flat" effect of average debt on growth over the whole period. This produces statistically significant estimators for both debt dummies.

In other words (more specifically):

-When I included the debt variable, I did not get statistically significant estimator for the debt.

-When I included a dummy debt variable, only for the debt-burdened countries (above the threshold of 80%), I also did not get a statistically significant estimator for the debt.

-When I diversified between countries below and countries above the debt threshold using the quadratic debt variable for the countries above the threshold and a proportional (not flat) dummy variable for the countries below the threshold, I got a significant estimator for the dummysquare variable above the threshold, but insignificant dummy debt estimator for the countries below the threshold.

- But when I used a dummy variable proportional to the squared debt for the countries above the threshold and an intercept-like (flat) dummy variable for the rest of the countries (below the threshold), then I got statistically significant effects for both debt dummy variables.

-I tried different debt thresholds and the best result I got was using the 80%.

(For the specification of the dummy variables, see below in the data section and for the explanation see in the results.)

Below, the definitions of the variables are provided:

Growth rate of GDP

Economic growth in the current research is measured "by the change in the volume of the economy's output", where volume of GDP is "the sum of value added, measured at constant prices, by households, government, and industries in the economy" (World bank *Glossary | DataBank*, n.d.)

GDP per Capita

GDP per capita is gross domestic product divided by midyear population.

Natural log of GDP per capita

The use of ln(GDP per capita) helps in transforming GDP per capita data into a mathematical form which is more helpful for statistical analysis.

Growth rate of Employees

The Growth rate of employees refers to the growth rate of employment. The term employment refers to the number of people who are currently occupied. The variable also takes into account seasonal unemployment, due to the fact that labor force, which is used for the calculation of the variable has a tendency to change annually due to the type of unemployment.

Gross Fixed Capital Formation

Gross fixed capital formation or elsewhere Gross Domestic Fixed Investment includes many terms such as land improvements, machinery, and equipment. It also includes the building of infrastructure such as schools, hospitals, industrial buildings, and private residents. The term also includes the net acquisition of valuables. (World Bank *Glossary | DataBank*, n.d.).

Research and development

The variable Research and Development Expenditure refers to the amount allocated for Research and Development as a percentage of GDP. It should also be noted that this variable includes capital and current expenditures concerning the sectors of Business enterprise, Government, Higher education, and Private non-profit (World Bank *Glossary | DataBank*, n.d.).

Additionally, it is important to define the meaning of Research and Development. According to the OECD (2015) definition:

“Research and experimental development … comprise creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge….. R&D covers basic research, applied research, and experimental development”.

Basic research refers to the experimental or theoretical analysis realized to gain novel knowledge. In this case, researchers do not aim for a specific application or utilization of their work. Nonetheless, Applied Research is empirical work, realized to discover insights, for the purpose of applying them in practice. Experimental development refers to the systematic work, based on knowledge derived from previous studies, which contributes to the creation of new products or processes. This type of work can also enhance present products or processes. As indicated by the fields of research and development (FORD) classification Research and development is classified according to the following sectors:

- Natural sciences
- Engineering and technology
- Medical and health sciences
- Agricultural and veterinary sciences
- Social sciences
- Humanities and the arts.

Government Effectiveness Indicator or government effectiveness

Government effectiveness Indicator according to the World Bank (WGI-Documents, n.d.) list represents the quality of the state's bureaucracy and institutional effectiveness as well as how excessive is red tape. Furthermore, government effectiveness refers to the quality of education, health services, drinking water, sanitation, transportation, maintenance and waste disposal, and road infrastructure. It also includes three significant terms:

- Infrastructure disruption which reflects the likelihood of disruption to and/or inadequacy of infrastructure for transport, including due to terrorism/insurgency, strikes, politically motivated shutdowns, natural disasters; infrastructure includes (as relevant) roads, railways, airports, ports, and customs checkpoints.
- State failure: the risk the state is unable to exclusively ensure law and order, and the supply of basic goods such as food, water, infrastructure, and energy, or is unable to respond to or manage current or likely future emergencies, including natural disasters and financial or economic crises.
- Policy instability: The risk the government's broad policy framework shifts over the next year, making the business environment more challenging. This might include more onerous employment or environmental regulation; local content requirements; import/export barriers, tariffs, or quotas; other protectionist measures; price controls or caps; more "political" control of monetary policy, or simply more direct intervention into the operations and decisions of private companies, etc

Debt

Based on the Eurostat (*Glossary:government debt*)) General Government Debt is defined as presented below:

"General government gross debt, also known as Maastricht debt or public debt, is the nominal (face) value of total gross debt outstanding at the end of the year and consolidated between and within the government subsectors.

General government gross debt comprises outstanding stocks of liabilities in the financial instruments currency and deposits, debt securities and loans at the end of the reference period".

Tax revenue

Following the definition of the World Bank (*Glossary | DataBank*, n.d.): "Tax revenue refers to compulsory transfers to the central government for public purposes. Certain compulsory transfers such as fines, penalties, and most social security contributions are excluded. Refunds and corrections of erroneously collected tax revenue are treated as negative revenue."

Inflation

Following the World Bank (*Glossary | DataBank*, n.d.) definition, inflation "indicates the rate of price change in the economy as a whole".

In the current research, some of the variables were derived intact from the official databases mentioned above, while others were modified or created based on these official data sources. More precisely, the variable of the growth rate of GDP (grgdp), the natural log of GDP per capita (lngdp), dummydebt, dummydebtsq and growth rate of employees (gremployees) were created. The variables gross fixed capital formation (gfcf), research and development (r&d), the government effectiveness indicator or government effectiveness (goveff), debt, tax revenue (taxrev), and inflation were derived from the official databases mentioned above.

The variable of growth rate of GDP (constant) is based on the formula:

$$\text{grGDP} = \left(\frac{GDP_t}{GDP_{t-1}} - 1\right) \times 100 \quad (1)$$

The base year is 2015 and the GDP per capita is in US dollars. However, the variable is not in US dollar units, but in percentage. The variable lngdp is based on the formula:

$$lngdp = \ln(\text{GDP per capita constant}_t) \quad (2)$$

The base year is 2015 and the GDP per capita is in US dollars. The variable growth rate of Employees is based on the formula:

$$\text{gremployees} = \left(\frac{\text{employment}_t - \text{employment}_{t-1}}{\text{employment}_{t-1}}\right) \times 100 \quad (3)$$

and is in percentage units.

Where employment: $\text{employment} = \text{labor Force} - \text{number of unemployed people}$ (4)

where labor force: a variable from the World Bank and includes the people who are currently employed and people who are unemployed but seeking work. Labor force also includes first-time job-seekers. The other term of equation (4) the number of unemployed people:

$$\text{number of unemployed people} = (\text{unemployment \% of total labor force}) \times \frac{\text{labor force}}{100} \quad (5)$$

The variable government effectiveness is derived from the WGI database of the World Bank. The data are accumulated from survey institutes, think tanks, non-governmental and international organizations, as well as private firms. It should be noted that the government effectiveness Indicator has a range of values from -2.5 to 2.5 (Kaufmann, Kraay & Mastruzzi, 2011 September). So, in order to be converted in the range of 0 to 100 as all the other variables the equation y=20x+50 was used where x represents the indicator from the database. The reason for this calculation is for the clustering and not so much for the econometric equation followed.

Gross fixed capital formation, research and development expenditure and tax revenue are represented as a percentage of GDP. Furthermore, debt represents the general government gross debt as a percentage of GDP.

Moreover, the variable inflation is measured by the annual growth rate of the GDP implicit deflator. The GDP implicit deflator is the fraction of GDP in current country currency to GDP in constant country currency (World Bank, 2023).

The variable dummydebt takes values 0 or 1 under certain conditions presented below:

$$dummydebt = \begin{cases} 1 \; if \; debt \leq 80 \\ 0 \; if debt > 80 \end{cases} \quad (6)$$

Where: 80 is the value of the threshold (debt 80% of GDP)

The variable dummydebtsq is calculated with the contribution of a helpful dummy variable (dummyhelp) and the variable debt squared (debtsq).

$$dummydebtsq = debtsq \times dummyhelp \quad (7)$$

Where:

$$dummyhelp = \begin{cases} 1 \; if \; debt > 80 \\ 0 \; if \; debt \leq 80 \end{cases} \quad (8)$$

Where: 80 is the value of the threshold (debt 80% of GDP)

and

$$\text{debtsq} = \frac{debt_t}{GDP_t} \times \frac{debt_t}{GDP_t} \times 100 \quad (9)$$

The variable dummydebtsq takes values as follows:

$$dummydebtsq = \begin{cases} 0 & if \; dummyhelp = 0 \\ \neq 0 & if \; dummyhelp = 1 \end{cases} \quad (10)$$

It should be mentioned that for the application of the clustering technique, data were scaled by dividing each variable with its standard deviation. By scaling the data, all features are made to be equally important in the clustering process, which can improve the effectiveness of the clustering algorithm. As a result, higher variance variables do not distort the procedure of the clustering. Contrary to the clustering case data are unscaled in the regression.

Concerning the quality of the data, it is possible that measurement error exists, as it is possible that data are underreported, in all the variables of interest. For instance, due to the black economy the variable GDP may not be calculated accurately. This can result to an imprecise recording, due

to human error or unwillingness of the citizens to reveal their true point of view for personal reasons.

4. Methodology

4.1 Clustering

In this research, the agglomerative hierarchical clustering technique is used. According to this methodology, primarily each country is considered one cluster on its own. At each step of the process, the clusters that are more similar to each other are grouped with other clusters. This process happens consecutively until one group is created (stopping criteria are not used, as there should not be intervention in the number of clusters). The result of this procedure is pictured by a graph called dendrogram.

The mathematical procedure of how this result is achieved is analyzed in the following paragraphs. Nonetheless, before elaborating it should be mentioned that the distance measurement is the Euclidean distance and that the dissimilarity distance of clusters is approached with the average linkage. With this type of dissimilarity distance, the distance between two clusters is determined as the average distance of the countries that belong in one cluster with the countries that belong to other clusters. I chose the average linkage, because this method takes into account the cluster structure meaning that every country has the same importance in the clustering process. The cluster structure is not taken into consideration in other types of clustering linkage techniques, such as those of single and complete linkage. Unlike the average linkage methodology, the single linkage defines the distance between clusters as the minimum distance between countries and the complete linkage as the highest distance between the countries (Carvalho, Munita & Lapolli, 2019).

Suppose that:

- l : indicator of the change in variables: {1,2,..., m}
- t : indicator of the years: $\{t_1,\ldots,t_n\}$

- j , k : countries
- C_j, C_k : clusters
- $|C_j|, |C_k|$: number of countries included in the clusters C_j, C_k

The formula of the dissimilarity distance between clusters is:

$$D(C_j, C_k) = \frac{1}{|C_j| \times |C_k|} \sum_{\substack{j \in C_j \\ k \in C_k}} D(j,k) \quad (11)$$

Where $D(j,k)$ is the Euclidian distance:

$$D(j,k) = \sqrt{\sum_{t=t_1}^{t_n} \sum_{l=1}^{m} (x_{jlt} - x_{klt})^2} \quad (12)$$

Suppose that:

- x_{jlt}: the value of variable l in year t of the county j
- x_{klt}: the value of variable l in year t of the county k

In the case when a country is a cluster on its own, then the Euclidean distance is equal to the dissimilarity distance.

In order to explain analytically the specific clustering procedure, it is good to use a straightforward example which illustrates its logic. The example is representative of how the chosen clustering technique works (more complicated and less complicated).

In the example there are 3 variables (l ∈{1,2,3 }) 2 years (t ∈ {1,2} and 4 countries (A, B, C, D). At the beginning of the procedure each country is a cluster on its own. The methodology starts by the calculation of Euclidean distances $D(j,k)$ for all combination of countries: $D(A,B), D(A,C), D(A,D), D(B,C), D(B,D), D(C,D)$. The countries with the smaller Euclidean distance, A and B in our case, are grouped together. So, there are AB, C, D. Following this step the dissimilarity distance $D(C_j, C_k)$ is calculated to merge the clusters. So $D(AB,C), D(AB,D), D(C,D)$ are calculated. It should be noted here that the dissimilarity distance $D(C,D)$ is the same as the Euclidean distance. Once all dissimilarity distances are calculated, the

clusters which have the smaller dissimilarity distance are grouped together. Supposing that the clusters AB and C have the smaller average distance (according to formula 6). As a result, AB and C are put together. Finally D merges with ABC. The dendrogram below indicates analytically the gradual merging of clusters, firstly A and B and secondly C with AB. Finally, D is merged with ABC.

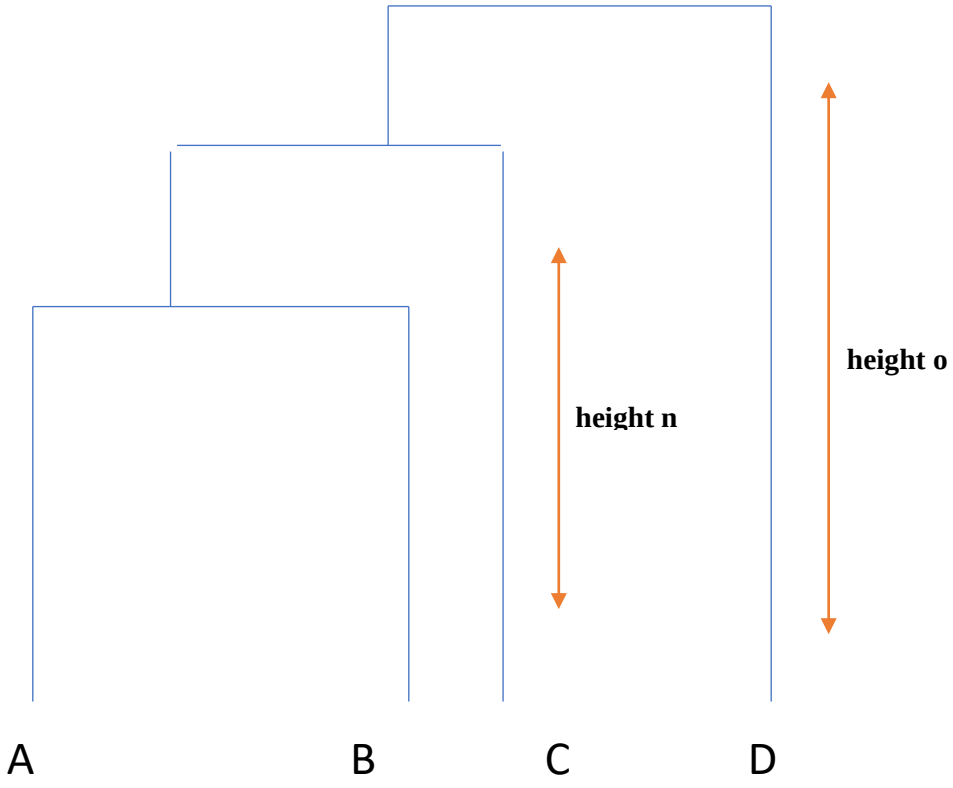

Figure 4.1 Dendrogram example

The vertical axis (height) of the dendrogram informs about the level of difference between clusters. As the height of the branches increases (vertical lines), the bigger is the difference among the clusters. For instance, C is closer to AB than D is, as height o is bigger than heigh n.

For evaluating the cluster analysis the cophenetic coefficient is used to test the clustering. (Palacio-Niño & Berzal, 2019)

4.1.1 Cophenetic Coefficient

The cophenetic coefficient indicates the correlation between the actual distances (in our case the Euclidean distance) and the cophenetic distances. Cophenetic distance of two countries (the height of the dendrogram) is considered as the minimum dissimilarity distance of a country with another one when they are merged in a cluster.

The formula of the cophenetic coefficient is presented below:

$$c = \frac{\sum_{j<k}[D(j,k) - \bar{D}][e(j,k) - \bar{e}]}{\sqrt{\sum_{j<k}(D(j,k) - \bar{D})^2 \sum_{j<k}(e(j,k) - \bar{e})^2}} \quad (13)$$

- j,k: countries
- $D(j,k)$: Euclidean distance
- $\bar{D}$: mean of Euclidean distances
- e(j,k) : cophenetic distance
- $\bar{e}$: mean of cophenetic distances

The cophenetic coefficients are used for evaluating the quality of the cluster. According to Saraçli, Doğan, N., and Doğan, S. (2013) the cophenetic coefficient measures the validity of the dendrogram, as it measures the degree of maintaining the pairwise actual distances. The cophenetic coefficient takes values belonging to the range [-1,1]. The higher is its value and closer to 1, the better is the quality of the clustering. Higher values indicate better efficiency in preserving the actual distances.

4.2 Regression Methodology

Fixed effects are used to investigate how gross fixed capital formation, research and development, growth rate of employees, and debt (independent variables) affect economic growth. The model is represented by the equation below:

$$grGDP_{it} = \beta_0 + \beta_1 gfcf_{it} + \beta_2 r\&d_{it} + \beta_3 gremployees_{it} + \beta_4 dummydebt_{it} + \beta_5 dummydebtsq_{it} + A_i + \delta_t + \varepsilon_{it} \quad (14)$$

I use the explanatory variable the gfcf and r&d to account for the effects of capital accumulation and innovation on economic growth respectively. The third variable (gremployees) is included in order to account for the relationship between the growth rate of employment and economic growth. The variables dummydebt and dummydebtsq are used to detect whether there is a threshold of 80% of debt as a percentage of GDP for the specific group studied in the current research. In other words, they are used to find if there is a non-linear relationship between economic growth and debt around the level of 80%. The intuition of whether this threshold exists will be based on both the coefficient and the statistical significance of the variables.

All variables vary over time t (on an annual basis) and over countries (i), except from A_i and δ_t. A_i is the country fixed effects. This term represents unobserved country characteristics that remain unchanged in time, such as geographical position. The time-fixed effect δ_t represents a common characteristic of all countries, which alters over the years, like the European Central Bank interest rate instrument, from 1999 onwards, the related with the interest rate instrument euro-dollar exchange rate over the years and the world oil prices, etc.

A subset of the variables included in the clustering were used in the econometric model and not all of them. More specifically, tax revenue was not included in the econometric equation as it has a lot of fluctuation and its effect is incorporated, in the long run in the debt accumulation. Inflation did not enter the econometric equation as although it was necessary in the clustering process to differentiate among countries, once I excluded the three a-typical countries and wanted to run a regression on the rest 16 countries it had no role to play, since I wanted to explain real (not nominal) effects on real GDP growth (for instance, the ECB central interest rate is indeed related to real economic growth and is also a tool to fight inflation. Therefore, I could have included it in the regression. Actually, I tried to include the ECB central interest rates, but it proved non-statistically significant, while not improving R square). Finally, the variables lngdp and goveff were deducted from the model as they were insignificant and their incorporation (separately) in the model resulted in the reduction of R^2.

5 Results

5.1 Clustering Results

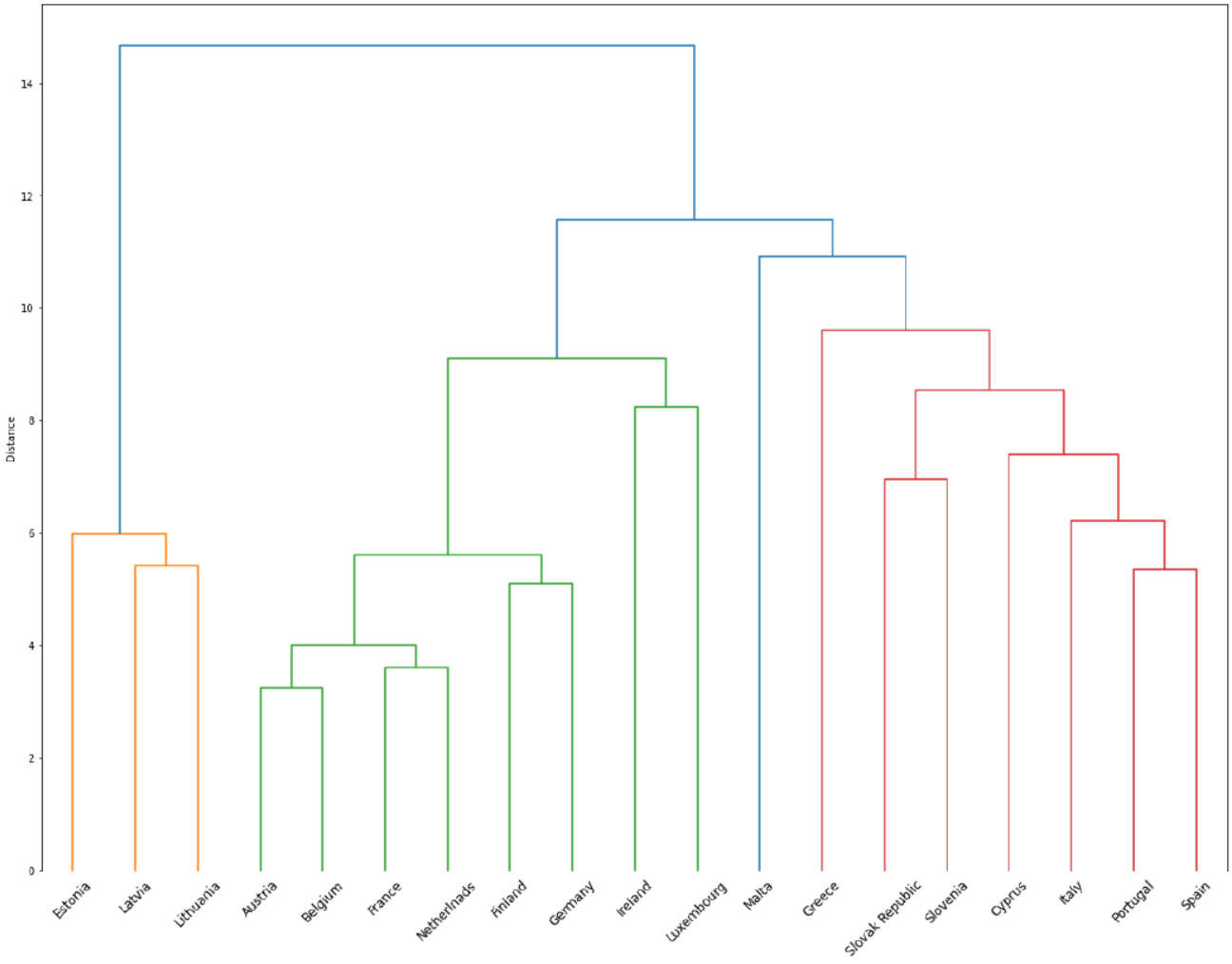

Figure 5.1.1: Dendrogram of 19 Eurozone countries

From the dendrogram above it is noticeable that there are two main groups of countries with the greatest distance between them. The one group (group A) includes the countries: Estonia, Latvia, Lithuania and the second one (group B) includes Austria, Belgium, France, Netherlands, Finland, Germany, Ireland, Luxembourg, Malta, Greece, Slovak Republic, Slovenia, Cyprus, Italy, Portugal and Spain. So further (econometric) focus should be given to those two groups indicated by the graph. Nonetheless, Latvia, Estonia and Lithuania, all of them former parts of the Soviet

Union, will not be studied, as they are atypical country cases, even though they have had significant economic progress in the last 30 years. As can be seen in the graphs below, at the initial stages of the period in question, Latvia, Lithuania, and Estonia lagged behind compared to the rest of the Eurozone. This makes sense since they are small countries that were under the Soviet Union's reign until 1990-1991 and once they exited the Soviet Union they had to establish free market institutions, as well as their sovereignty from scratch. Therefore, they were in a totally different historical phase, compared with all the rest.

The graphs below indicate significant differences between the groups, validating the fact that the clustering technique "wisely" separated them from the rest of the Eurozone. Some of the most important differences are briefly presented below:

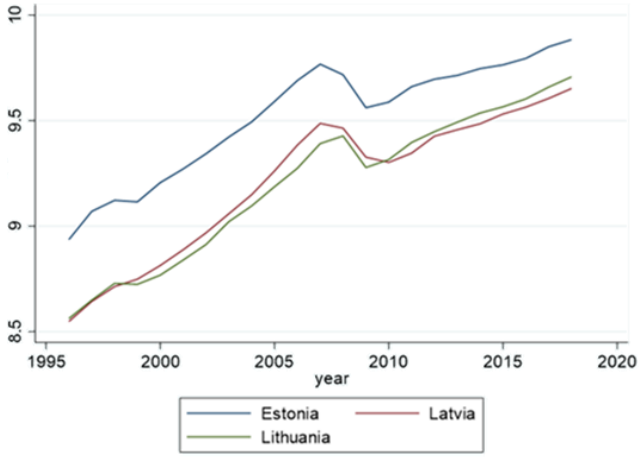

Figure 5.1.2.A: Natural logarithm of GDP per capita: Group A for the time period 1996-2018

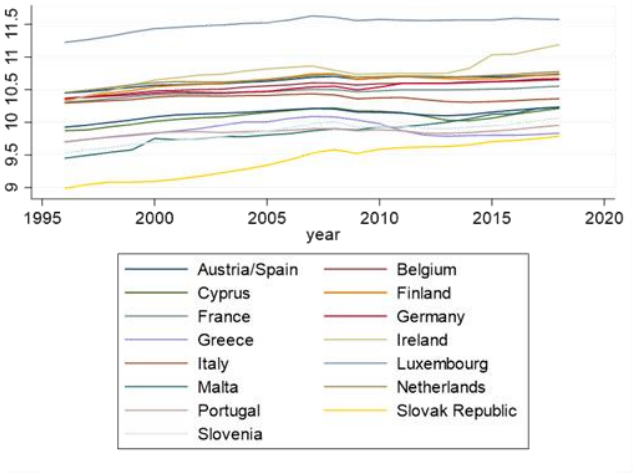

Figure 5.1.2.B: Natural logarithm of GDP per capita: Group B for the time period 1996-2018

It is noticed in the figures 5.1.2.A and 5.1.2.B that ex-Soviet Union countries (group A) have differences with the group of 16 countries (group B). More specifically, in 1996, the lngdp of group B varied from 9.5 to 10.5 with two exceptions, Slovak Republic and Luxembourg, while group A did not exceed 9. From 1996 till 2006 there was a strong continuous rise in lngdp in group A, but during the period of 2006-2009, a small reduction of lngdp is noticed. Concerning group B there was a continuous and moderate growth of lngdp until 2009 followed by a very small fall of the variable. From 2009 onwards there is an increase in lngdp in both groups. From

the graph of group A it is observed that Estonia finally converges with Lithuania and Latvia, as at the beginning of the period its level of GDP was less close to the level of GDP of Latvia and Lithuania. The overall steeper increase in the level of GDP of group A can be attributed to the fact that those countries are weaker starting from a much lower level compared to group B countries, so they had the potential to grow faster in their new environment. It is also worth noting that the countries of group B also converged.

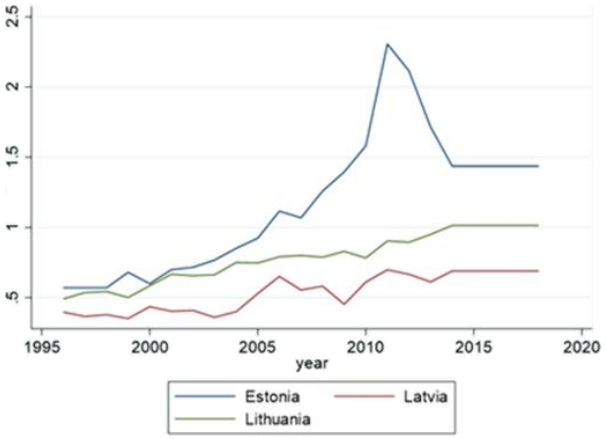

Figure 5.1.3.A: Research and development as a percentage of GDP: Group A for the time period 1996-2018

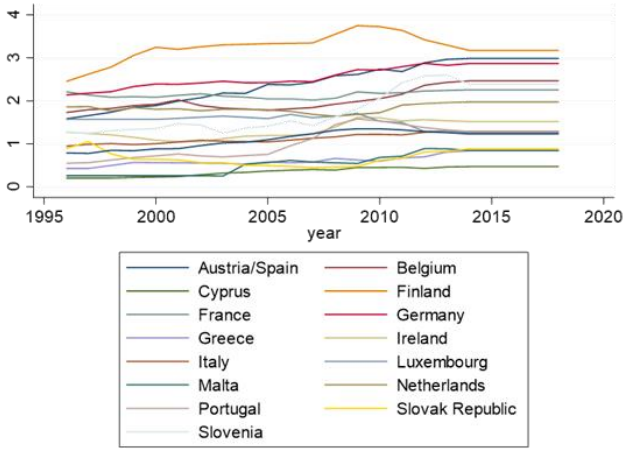

Figure 5.1.3.B: Research and development as a percentage of GDP: Group B for the time period 1996-2018

Significant differences should also be considered concerning r&d among the two groups. In figure 5.1.3.A, in 1996, group A commenced with low research and development expenditure as a percentage of their GDP near 0.5, while countries of group B ranged from 0 to 2.5 with their majority exceeding 1% of GDP in the same year. From 2013 onwards countries of group B have stabilized their research and development expenditure as a percentage of their GDP, while this non-volatility is observed in group A from 2014 onwards. Another noticeable fact is that Estonia (figure 5.1.3.B) during 2012-2013 had almost a 1.5 percent decline in its research and development as a percentage of its GDP.

It makes sense that at the beginning of the sample period, research and development expenditure was at low levels as countries were under total reform (i.e. they could not fund such

projects and most probably even if they could fund them there was no demand from their domestic market and the education sector).

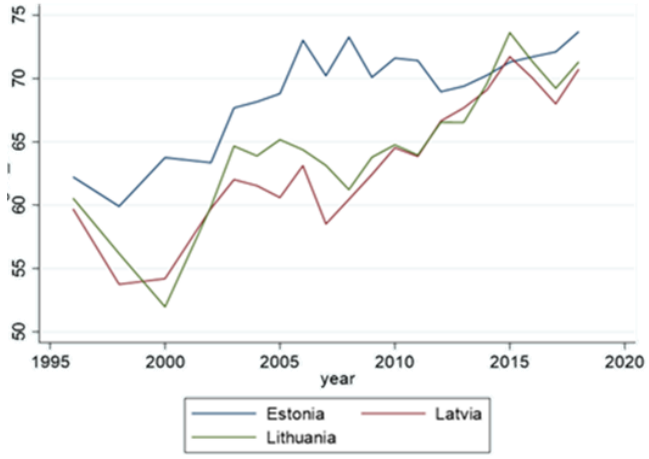

Figure 5.1.4.A: Government effectiveness: Group A for the time period 1996-2018

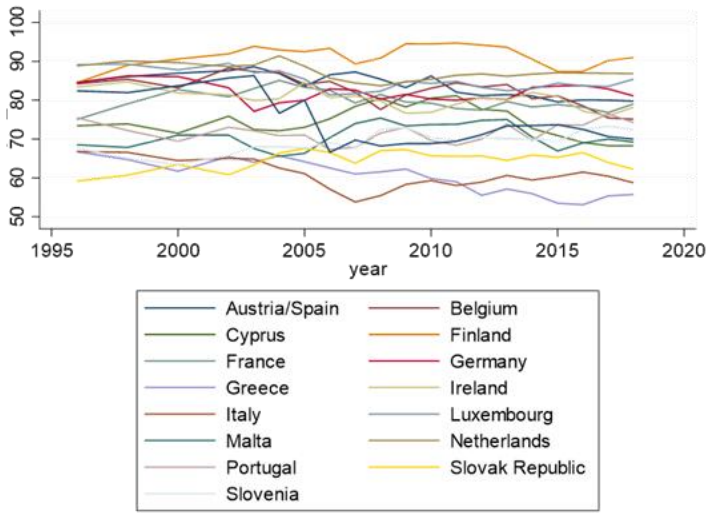

Figure 5.1.4.B: Government effectiveness: Group B for the time period 1996-2018

In 1996 countries of group A (figure 5.1.4.A) started with a level of government effectiveness near 60, while group B countries (graph 3(b)) varied in the range of 70 to 85 with the exception of Slovak Republic, Finland and Cyprus. Countries belonging to group A increased their government effectiveness reaching 70-75 with Estonia having the highest government effectiveness. Nonetheless, countries in group B generally remain in quite stable levels of effectiveness, with the exceptions of mainly Malta and Greece where government effectiveness is somewhat deteriorating.

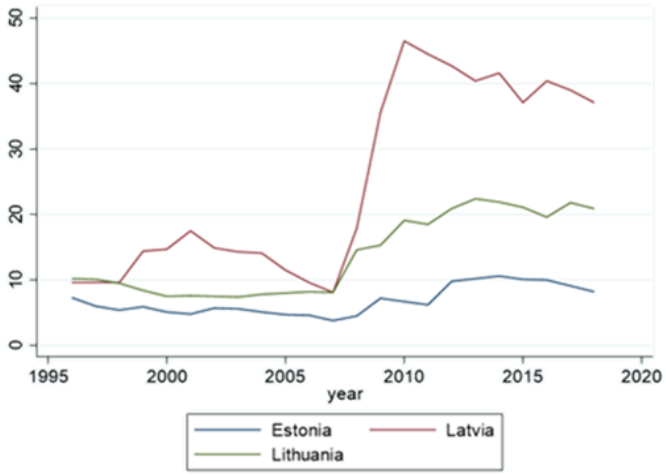

Figure 5.1.5.A: Debt as a percentage of GDP: Group A for the time period 1996-2018

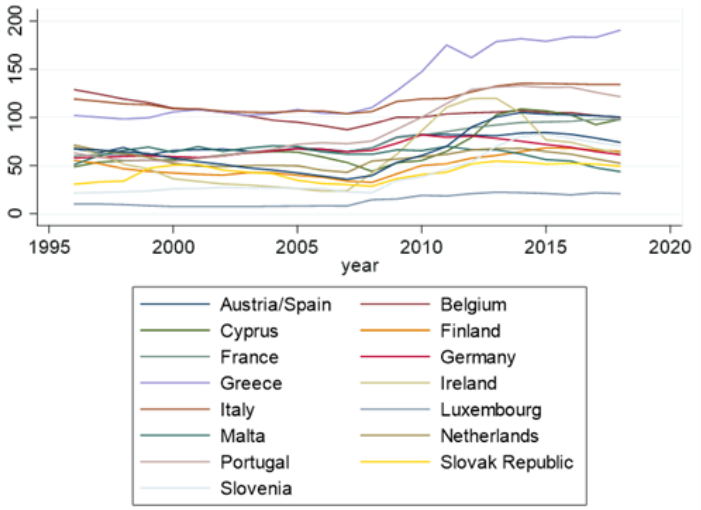

Figure 5.1.5.B: Debt as a percentage of GDP: Group B for the time period 1996-2018

According to figure 5.1.5.A, during the period 1996-2007, the ex-Soviet countries had debt to GDP below 20%, while the majority of countries in group B (figure 5.1.5.B) had debt to GDP ratio of 50% and higher, except Slovenia and Luxembourg. Additionally, after 2007, there was a significant increase in the debt-to-GDP ratio for the ex-Soviet countries, but continued to be lower than group B's debt to GDP.

It is noteworthy that almost all debt burdens from the Soviet era were alleviated for Latvia, Estonia, and Lithuania (Owen & Robinson, 2003). Therefore, the three ex-Soviet Baltic states started off their careers as sovereign states essentially debt-free.

Additionally:

"Public expenditures had sharply increased during the boom years. With the crisis, this proved out of line with the tax base, which was likely to be permanently reduced following the crisis. In structural terms, fiscal imbalances going into the crisis at end-2008 were high across all three countries." (Purfield & Rosenberg, 2010)

All in all, the debt of GDP ratio for those three countries remains significantly low compared to group B.

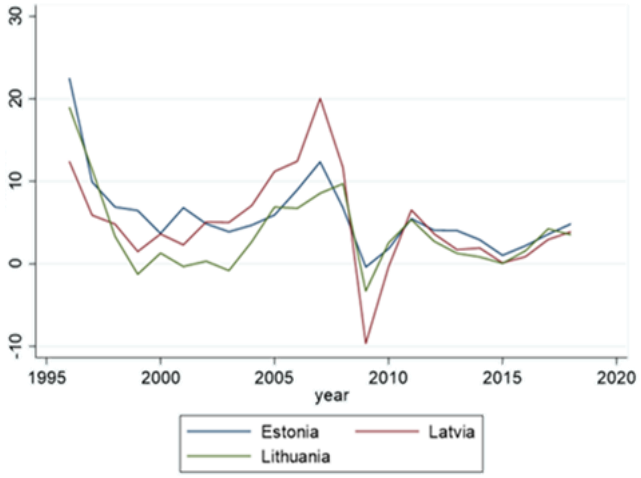

Figure 5.1.6.A: Inflation: Group A for the time period 1996-2018

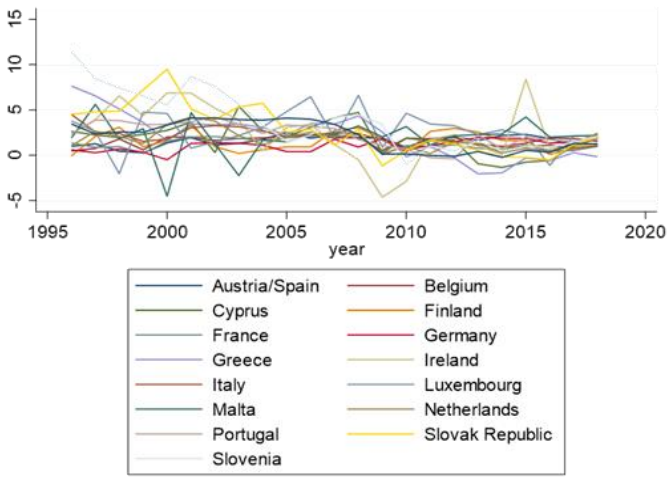

Figure 5.1.6.B: Inflation: Group B for the time period 1996-2018

The countries of the former Soviet Union in the year 1996 started with a higher level of inflation than the majority of countries of Group B with the exception of Slovenia which was also at the same level of inflation as the former Soviet countries in the same year. Nevertheless, during the period of 1997-2002, the level of inflation was approximately the same level for the two groups. From 2006 to 2009 there was a significant decline in inflation in the former Soviet countries (as a matter of fact, this period is characterized mainly by deflation), while in group B this trend was also present, but not so intense, i.e. the deceleration of inflation was more gradual during the period 2006-2009 for group B. From 2011 onwards, inflation for group A fluctuated around the same levels as in group B.

It is worth noting that the cophenetic correlation coefficient is 0.826, close to 1. This means that the cluster maintains the pairwise actual distances (Euclidean distances) to the extent of 82.6%. So, the cluster can be characterized as decently fit.

This clustering clearly indicates the difference between the two groups (group A and group B). In the current research, the clustering led to distinguishing the atypical countries (Estonia, Latvia, Lithuania) from the rest of the Eurozone. So, the following econometric analysis concerns all the countries of the Eurozone with the exemption of the three countries.

5.2 Regression Results

In this section, the results of the econometric model applied for all Eurozone countries, except Estonia, Latvia and Lithuania, (16 in total) are presented below. These 16 countries have been chosen based on the clustering result:

Table 1 Fixed Effects Results: Regression of grgdp on gfcf, , r&d, gremployees, dummydebt, dummydebtsq for the time period 1996-2018

Variables	grgdp
gfcf	0.036
	(0.043)
r&d	0.069
	(0.567)
gremployees	0.493**
	(0.083)
dummydebt	-1.334**
	(0.296)
dummydebtsq	-0.012*
	(0.004)
Constant	3.447*
	(1.403)
Observations	368
R^2	0.565

Note: The model is estimated with both country and time-fixed effects. Robust standard errors are in parentheses placed below the coefficients. * $p < 0.05$, ** $p < 0.01$

Out of the 5 coefficients included in the model, the last three (gremployees dummydebt, dummydebtsq) are strongly statistically significant (high t-ratios) and bear the "right" sign as expected in theory. In other words, when growth rate of employment goes up by 1% the overall economy's growth rate is expected to go up as well by 0.49%. Higher debt-to-GDP ratios are negatively related to GDP real growth. The debt to GDP ratio is a dummy (dummydebt). According to the model when debt to GDP ratio is less than 80%, there is a "flat" negative influence on growth rate of GDP (-1.334).

Over the 80% threshold for the debt to GDP ratio, there is an additional negative effect on real growth, which is not "flat"; it is proportional (-0.012) to the debt to GDP ratio squared. This is "small", but with the "correct"- negative- sign and still statistically significant. Thus, establishing a quadratic effect on debt to economic growth. I have tried different thresholds (75%, 85%, 90%) and the 80% threshold fits the data best (better t-ratios for both debt variables, better R square etc). After all the official threshold of the monetary union of the Eurozone area is 60%, but most eurozone countries have been hovering above that; according to our data the overall mean debt of GDP is 70.27%, over that period; therefore an 80% threshold makes sense.

The main mechanism through which debt influences growth negatively is interest payments. At low debt as a percentage of GDP, interest rates are usually low and interest payments are also low. However, at "high" levels of debt (presumably above the threshold) the markets attribute higher risk to the country and penalize its borrowing by higher rates; therefore countries with higher debt to GDP ratio are expected to find themselves in a borrowing spiral trap (having to borrow at higher rates to roll over their debt payments), therefore increasing the cost burden of servicing their debt.

So far 56.59% (overall R squared) of the dependent variable has been explained by three independent variables that are all statistically significant, plus two non-significant independent variables (gfcf and r&d). The latter are useful however, because they stabilize the statistical behavior of our econometric model.

The independent variables gfcf and r&d are theoretically expected to have some positive effect on growth. Nonetheless, in the model this effect, even though it still bears a positive sign, appears as statistically insignificant as it is "absorbed "over time by the gremployees variable. In other words, the model suggests through the t-ratios that gremployees as a variable "dominates" over capital formation (and r&d), as a variable determining growth.

In the current research, three tests were applied in order to detect the presence of cross-sectional dependence, autocorrelation and heteroskedasticity.

For the purpose of checking the presence of cross-sectional dependence the Pesaran's test of cross-sectional independence was used. The detection tool showed that panels are not correlated (Pr = 0.6441). For detecting serial correlation, the Wooldridge test for autocorrelation in panel data was applied. The test showed that there is absence of serial correlation as the null hypothesis is accepted (Prob > F = 0.1650). The test applied for heteroskedasticity is the Modified Wald test. The result of this test is the rejection of the null hypothesis, which indicates that heteroskedasticity is present (Prob>chi2 = 0.0000). Consequently, due to the problem of heteroskedasticity, I run the model with heteroskedasticity-robust standard errors (also called sandwich estimators).

Finally, the F test was used to compare the model with and without the presence of the two variables of debt. As shown from the F (F=72.980>F2, 365=1.6664), the presence of the two variables is vital for the model.

6. Conclusion

Based on the clustering two groups of countries were chosen: group A (Estonia, Lithuania, Latvia) and group B (Austria, Belgium, France, Netherlands, Finland, Germany, Ireland, Luxembourg, Malta, Greece, Slovak Republic, Slovenia, Cyprus, Italy, Portugal and Spain). This clustering separation is sound since Estonia, Lithuania and Latvia are small countries under the Soviet Union's reign until 1990-1991 and after their independence, they had to establish free market institutions, as well as their sovereignty from scratch. Therefore, they were in a totally different historical phase, compared to the other group of countries (group B). So, due to the fact that group A consists of atypical countries, further analysis was focused on group B.

Concerning the results of the econometric model for group B: The growth employment variable incorporates both trend and business cycle fluctuations. Therefore, it has an understandable significant influence on growth.

What is more interesting: it seems there is a statistically significant quadratic negative effect on growth for debt to GDP over 80%.

There is also a “flat” negative effect on growth below the 80% threshold for the countries of group B. In essence, this is the “aggregate”, (weighted average) effect of aggregate debt to GDP ratio of countries with less than the 80% threshold on their growth rates.

We expect that in less indebted countries the negative effect of their individual debt on their individual growth rates is proportional but “smaller”. However, the exact disaggregation of this negative effect for each country needs further econometric examination.

Therefore, the reduction of debt to GDP for all countries, especially the most burdened ones, is expected to enhance future growth rates. Besides growth enhancement, debt reduction will allow for faster convergence among all member states in group B (since group A countries do not have a debt problem).

References

Abdulkadir, A., Afriana, W., & Azis, H. (2020, October 29). THE EFFECTS OF R&D EXPENDITURES ON ECONOMIC GROWTH IN OECD COUNTRIES. *Airlangga Journal of Innovation Management*, *1*(2), 175. https://doi.org/10.20473/ajim.v1i2.21607

Afonso, A., & Alves, J. R. (2014). The Role of Government Debt in Economic Growth. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2468805

Amado, G. (2022). Revisiting the Debate on the Eurozone Crisis: Causes, Clustering Periphery and Core, and the Role of Interest Rate Convergence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4279444

Anastasiou, A., Apergis, N., & Zervoyianni, A. (2023, January 2). Convergence of public expenditures and revenues in EU28 during 2002–2019: Evidence from club-clustering analysis before and after the European debt crisis. *International Journal of Finance & Economics*. https://doi.org/10.1002/ijfe.2756

Barro, R. J. (1990, October). Government Spending in a Simple Model of Endogeneous Growth. *Journal of Political Economy*, *98*(5, Part 2), S103–S125. doi: https://doi.org/10.1086/261726

Carvalho, P. R., Munita, C. S., & Lapolli, A. L. (2019, February 20). Validity studies among hierarchical methods of cluster analysis using cophenetic correlation coefficient. *Brazilian Journal of Radiation Sciences*, *7*(2A). https://doi.org/10.15392/bjrs.v7i2a.668

Cavenaile, L., & Dubois, D. (2011, November). An empirical analysis of income convergence in the European Union. *Applied Economics Letters*, *18*(17), 1705–1708. https://doi.org/10.1080/13504851.2011.560104

Cecchetti, S. G., Mohanty, M. S., & Zampolli, F. (2011). *The real effects of debt* (Ser. Bis working paper, no. 352). SSRN. Retrieved October 30, 2023, from https://ssrn.com/abstract=1946170

Checherita-Westphal, C. D., & Rother, P. (2010). The Impact of High and Growing Government Debt on Economic Growth: An Empirical Investigation for the Euro Area. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1659559

Dias, J. (2017). Unemployment and sovereign debt crisis in the eurozone: a k -means- r analysis. *Physica a: Statistical Mechanics and Its Applications*, *482*, 108–117. https://doi.org/10.1016/j.physa.2017.04.022

Égert, B. (2013). The 90% Public Debt Threshold: The Rise & Fall of a Stylised Fact. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2304242

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NE.GDI.FTOT.ZS

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/GB.XPD.RSDV.GD.ZS

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.DEFL.KD.ZG

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.MKTP.KD.ZG

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/GC.TAX.TOTL.GD.ZS

Glossary | DataBank. (n.d.). https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.DEFL.KD.ZG

Glossary:government debt. Glossary:Government debt - Statistics Explained. (n.d.). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary%3AGovernment_debt

Guloglu, B., & Tekin, R. B. (2012). A panel causality analysis of the relationship among research and development, innovation, and economic growth in high-income oecd countries. Eurasian Economic Review, 2(1), 32–47. doi: https://doi.org/10.14208/BF03353831

History of the EU, EU pioneers | European Union. (n.d.). European Union. https://european-union.europa.eu/principles-countries-history/history-eu_en

Ignatov, I. (2023). Convergence Determinants and Club Formation in the EU over 1999-2021. *Икономически изследвания*, (3), 37-63.

Irac, D., & Lopez, J. (2015, October). Euro area structural convergence? A multi-criterion cluster analysis. *International Economics*, *143*, 1–22. https://doi.org/10.1016/j.inteco.2015.01.005

Kantar, E., Deviren, B., & Keskin, M. (2014, November). Hierarchical structure of the European countries based on debts as a percentage of GDP during the 2000–2011 period. *Physica A: Statistical Mechanics and Its Applications*, *414*, 95–107. https://doi.org/10.1016/j.physa.2014.07.001

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011, September). The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*, *3*(02), 220–246.doi: https://doi.org/10.1017/s1876404511200046

Lymonova, E. (2019). Estimation of the effect of taxes and gross fixed capital formation on economic growth of euro area. *Academy Review*, *1*(50), 5–14. doi:https://doi.org/10.32342/2074-5354-2019-1-50-1

Maestas, N., Mullen, K. J., & Powell, D. (2023, April 1). The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity. *American Economic Journal: Macroeconomics*, *15*(2), 306–332. doi: https://doi.org/10.1257/mac.20190196

Matesanz, D., & Ortega, G. J. (2015, October). Sovereign public debt crisis in Europe. A network analysis. *Physica A: Statistical Mechanics and Its Applications*, *436*, 756–766. https://doi.org/10.1016/j.physa.2015.05.052

Mazurek, J. (2014, January 25). Examination of European Union economic cohesion: A cluster analysis approach. *Business and Economic Horizons*, *9*(4), 8–17. https://doi.org/10.15208/beh.2013.18

Mehrara, M., & Musai, M. (2013, September). The Causality between Capital Formation and Economic Growth in MENA Region. *International Letters of Social and Humanistic Sciences*, *8*, 1–7.doi: https://doi.org/10.18052/www.scipress.com/ilshs.8.1

Ncanywa, T., & Makhenyane, L. (2016). *Can investment activities in the form of Capital Formation influence economic growth in South Africa?*.Retrieved from http://hdl.handle.net/10386/1653

OECD Publishing. (2015). *Frascati manual 2015 : guidelines for collecting and reporting data on research and experimental development* (Ser. The measurement of scientific,

technological and innovation activities).doi: https://doi.org/10.1787/9789264239012-en

Onuferová, E., Čabinová, V., & Matijová, M. (2020, September 1). Categorization of the EU Member States in the Context of Selected Multicriteria International Indices Using Cluster Analysis. *Review of Economic Perspectives*, *20*(3), 379–401. https://doi.org/10.2478/revecp-2020-0018

Owen, M. E. W., & Robinson, M. O. (2003, September 9). *Russia Rebounds*. International Monetary Fund.

Palacio-Niño, J. O., & Berzal, F. (2019). Evaluation metrics for unsupervised learning algorithms.doi: https://doi.org/10.48550/arXiv.1905.05667

Pełka, M. (2018). ANALYSIS OF INNOVATIONS IN THE EUROPEAN UNION VIA ENSEMBLE SYMBOLIC DENSITY CLUSTERING. *ECONOMETRICS*, *22*(3), 84–98. https://doi.org/10.15611/eada.2018.3.06

Proença, C., Neves, M., Dias, J. C., & Martins, P. (2021, July 27). Determinants of sovereign debt ratings in clusters of European countries – effects of the crisis. *Journal of Financial Economic Policy*, *14*(3), 403–427. https://doi.org/10.1108/jfep-01-2021-0017

Purfield, C., & Rosenberg, C. (2010). Adjustment Under a Currency Peg: Estonia, Latvia and Lithuania During the Global Financial Crisis 2008-09. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.1750701

Romer, P. M. (1986, October). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, *94*(5), 1002–1037.doi: https://doi.org/10.1086/261420

Saraçli, S., Doğan, N., & Doğan, S. (2013, April 23). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, *2013*(1). doi: https://doi.org/10.1186/1029-242x-2013-203

Soava, G., Mehedintu, A., Sterpu, M., & Raduteanu, M. (2020, December 4). Impact of Employed Labor Force, Investment, and Remittances on Economic Growth in EU Countries. *Sustainability*, *12*(23), 10141.doi: https://doi.org/10.3390/su122310141

Sodipe, O. A., & Ogunrinola, O. I. (2011). Employment and economic growth nexus in Nigeria. *International Journal of Business and Social Science*, *2*(11). Retrieved from https://core.ac.uk/download/pdf/12356253.pdf

Solow, R. M. (1956, February). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65. https://doi.org/10.2307/1884513

Sylwester, K. (2001, December). R&D and economic growth. *Knowledge, Technology & Policy*, *13*(4), 71–84.doi: https://doi.org/10.1007/bf02693991

Taylor, M. P., & Mankiw, N. G. (2014, January 1). *Economics*.

WGI-Documents. (n.d.). https://info.worldbank.org/governance/wgi/Home/Documents

Appendix

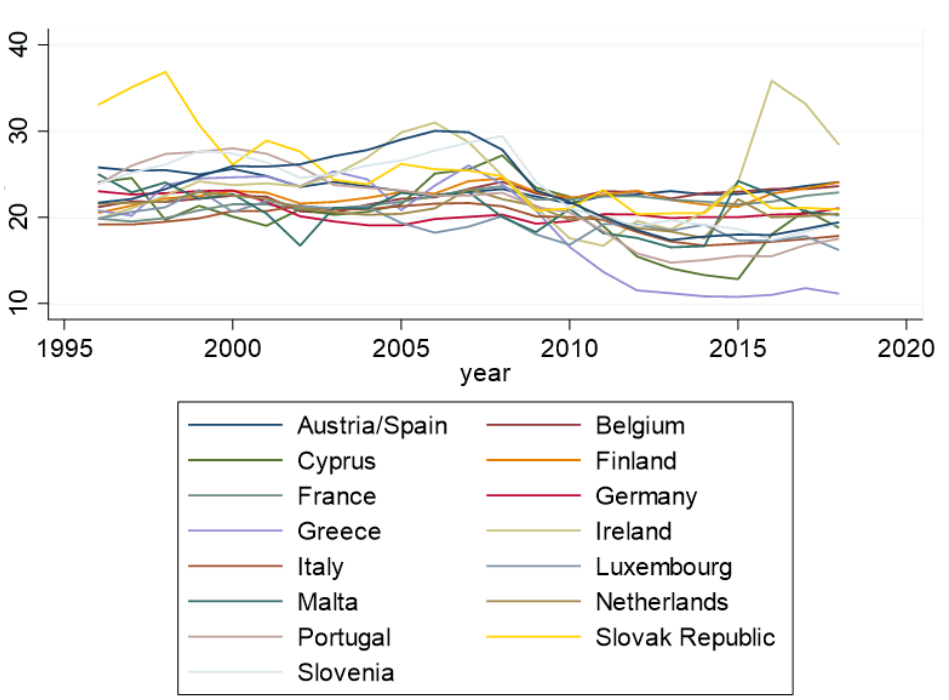

Figure A.1: Gross fixed capital formation as a percentage of GDP: Group B for the time period 1996-2018

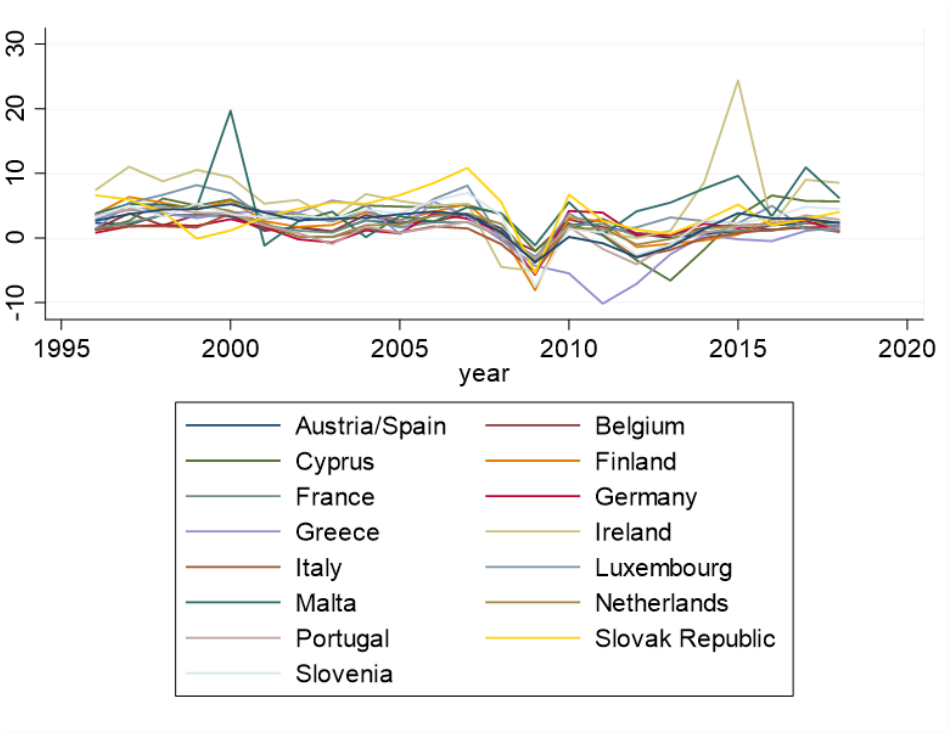

Figure A.2: Growth rate of GDP: Group B for the time period 1996-2018

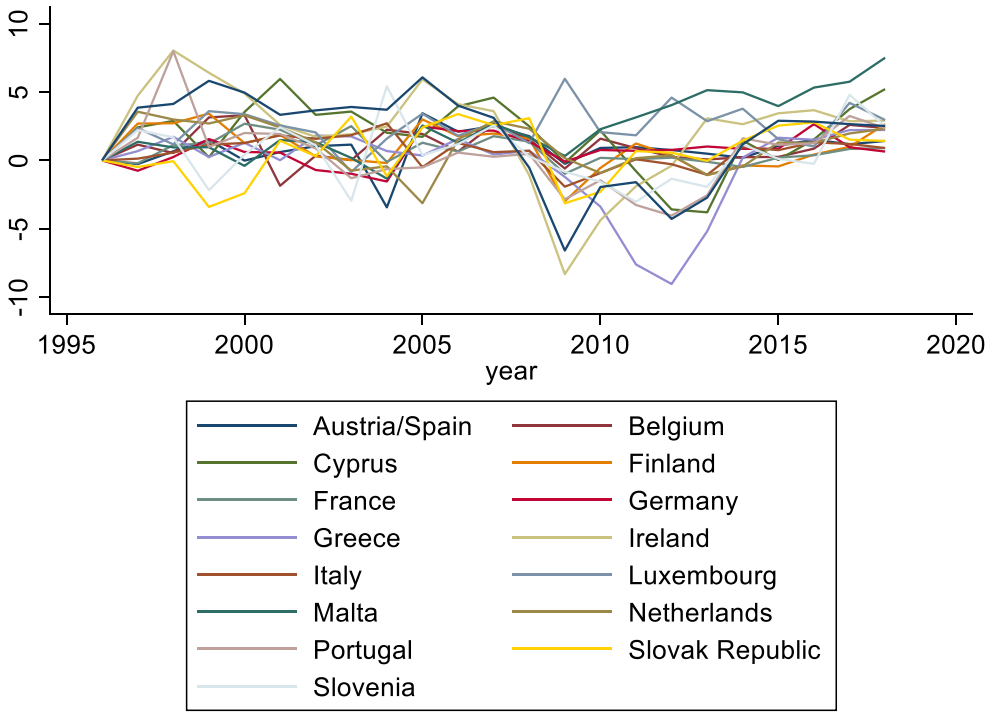

Figure A.3: Growth rate of employees: Group B for the time period 1996-2018

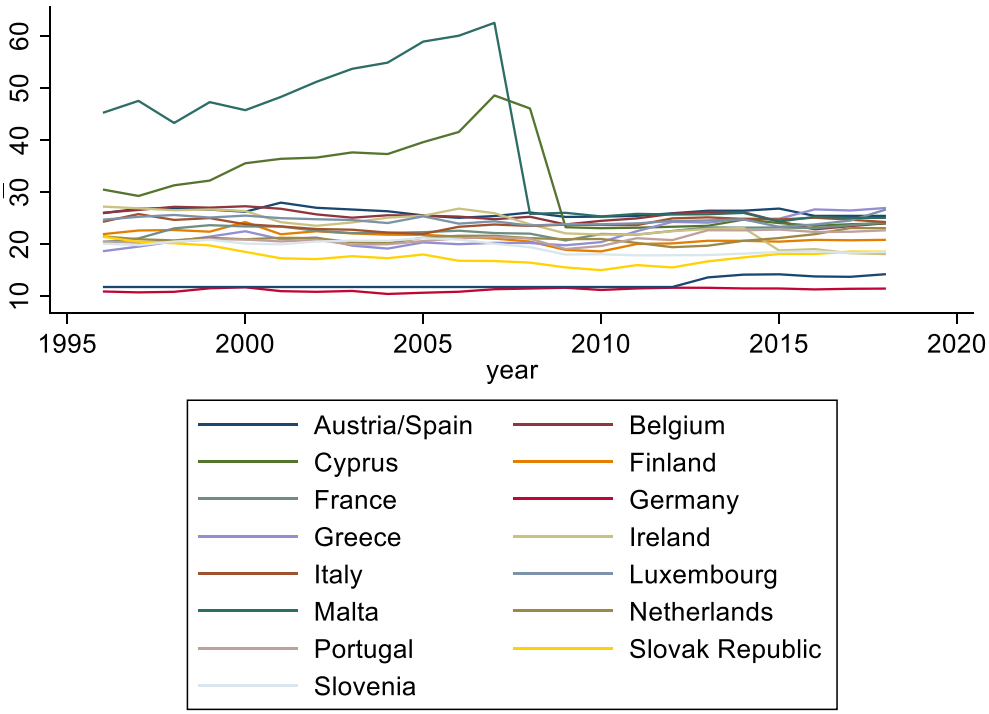

Figure A.4: Tax revenue as a percentage of GDP: Group B for the time period 1996-2018

Variable	Obs	Mean	Std. dev.	Min	Max
grgdp	368	2.463783	3.303384	-10.1493	24.3704
lngdp	368	10.32171	.5029224	8.99024	11.63
gfcf	368	21.80886	3.685005	10.77	36.8557
r_and_d	368	1.53034	.8341318	.20407	3.74883
gremployees	368	1.112155	2.241035	-9.04154	8.04461
goveff	368	76.35536	9.699293	53.1194	94.6996
debt	368	70.27065	34.31659	7.4	190.7
taxrev	368	22.65248	7.623978	10.4263	62.5025
inflation	368	2.071126	1.939395	-4.6246	11.4525
dummydebt	368	.6820652	.466308	0	1
dummydebtsq	368	40.90665	69.37747	0	363.6649

Table A.1: Summary statistics