THE ROLE OF ECONOMIC ACTIVITIES' CLUSTERS IN GROSS VALUE ADDED GENERATION IN THE BALTIC STATES

Rima Rubčinskaitė,[1] Gindrutė Kasnauskienė[2]

Abstract: Economic globalization affects regional development through different economic value creation chains. The structure of regional economic activities' groups and size of clusters could also influence the structure and size of gross value added in a particular region. This paper examines the impact of the main labor force indicators on the generated gross value added in diverse economic activities in the Baltic States and the neighbor regions of Finland and Poland in the period of 2000-2013 according to Eurostat data. The research has shown that the structure of economic activities clusters in the Baltic States (Estonia, Latvia, Lithuania) significantly differs from the neighboring regions of Poland and Finland. We also found that the labour force input has a significant impact on the following economic activities' clusters in the Baltic States: "Industry (except construction)", "Wholesale and retail trade, transport, accommodation and food service activities", "Construction", "Professional, scientific and technical activities; administrative and support service activities", "Financial and insurance activities", "Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organisations and bodies". The impact of the labor force on generated gross value added differs in diverse economic activities. The authors believe that the results of this study could be useful for policy makers in building more progressive national and regional economic development strategies.

Introduction

The empirical research results indicate that the composition of clusters as groups of specific economic activities has a significant impact on value creation in the studied regions. The structure of regional economic activity groups and the size of clusters could also influence the structure and the size of gross value added in a particular region. The questions we discuss in this paper are of considerable importance to the Baltic States: Does location (i.e. region) matters in the Baltic States? How much do the Baltic States vary or are similar in terms of the distribution of labor force across economic activities? How much do the Baltic States vary in the structure of gross value added of economic activities? How much does the input of the labor force matter to the generation of gross value added in a specific economic activity?

There are two main strands of research used to analyze clusters and the agglomeration impact on regional development that are briefly presented in the paper. The Porter's cluster theory is widely accepted by policy makers at the regional, national and the European Union levels. The clusters and networks became part of the EU strategy "Europe 2020" and the industrial policy "An Integrated Industrial Policy for the Globalisation Era Putting Competitiveness and Sustainability at Centre Stage."

This study aims to evaluate the role of economic activities' clusters in gross value added generation in the Baltic States (Estonia, Latvia, Lithuania) and neighbor regions of Finland and Poland. The originality of the paper lies in the fact that it has been for the first time, to the authors knowledge, that a few small open economies of the same geographical location were analysed as one region and compared with the neighbour regions of other countries in terms of labour distribution across clusters of economic activities and its impact on gross value added.

The following part of the paper reviews the modern regional development theories. Further, the methods used to investigate the subject are described. Then follows the discussion on a geographical scale of the research. The next part highlights the analysis of labor and gross value added structure of the selected regions and the main results of a regression analysis on the impact of labour force on gross value added in selected clusters and regions. The last section discusses the main findings and conclusions.

Literature review: Cluster Theory and New Economic Geography

Two different research strands on regional economic development were developed in the end of the 20th century and the beginning of the 21st century: a cluster theory by Michael Porter and a new economic geography by Paul Krugman. Both of them highlight the location being quite little under

---

[1] Faculty of Economics of Vilnius University, rima.putiene@gmail.com

[2] Faculty of Economics of Vilnius University, gindra.kasnauskiene@ef.vu.lt

consideration in economics earlier. M. Porter and his followers argue that the location defines the specific structure of interrelated economic activities and the patterns of such collocation can be called clusters (Porter, 2003; Brachert, Titze, Kubis, 2011; Delgado, Porter, Stern, 2012). The analysis of the USA County Business Patterns (CBP) data allowed M. Porter to classify clusters in the following categories: traded, local, and resource-dependent (Porter, 2003). According to Porter's cluster theory, clusters have an impact on the economy and innovation of the region and use the concentration, correlation and regression analysis to prove this (Porter, 2003; Sölvell, Ketels, Lindqvist, 2008; Brachert, Titze, Kubis, 2011; Delgado, Porter, Stern, 2012; Pires et al., 2013; Ketels, Ch., Protsiv, S., 2013). The main variables used are employment, establishments, wage size, and patent indicators (Porter, M., 2003; Brachert, Titze, Kubis, 2011; Pires et al., 2013; Delgado, Porter, Stern, 2013).

Paul Krugman focuses attention on the relationship between the location and trade (Krugman, 2010). According to Martin and Sunley (1996, p. 6), Krugman's theory "combines the models of imperfect competition and scale economies used in a new trade theory with the location theory's. It highlights the emphasis on the significance of transport costs and "the interaction of external economies of scale with transport costs" to be the key to his explanation of regional industrial concentration and the formation of regional "centers" and "peripheries." Others add that new economic geography explains "market expansion effect" (i.e. the increased local demand for upstream activities when a new company starts production in a certain location) and "market crowding effect" (i.e. a local supply for downstream activities) (Ottavianno, 2002).

While discussing regional economics, there are two main groups of theories: a location theory that analyses "economic mechanisms that distribute activities in space" and a regional growth (or development) theory where research objects are "spatial aspects of economic growth and the territorial distribution of income" (Capello, 2011). New Economic Geography (NEG) as a location theory analyses the diversity of locations, determinants of local growth using quantitative approach. The cumulative causation model (Myrdal, 1957; Kaldor, 1970) and the endogeneous growth model (Luca, 1988; Romer, 1987) belong to the same local growth theories where growth is treated as territorial competitiveness (Capello, 2011). It should be added that Porter's cluster theory also deals with competitiveness as the main source of inspiration for a cluster concept in the region.

While both new theories deal with spatial distribution as such, they both received critique and were widely discussed and compared. While Porter's cluster theory received much critique on the empirical basis (Woodward, 2012) and the geographical scale of cluster (Martin and Sunley, 2003; Brachert, Titze, Kubis, 2011) , this means that the definition of a region matters. Krugman's new economic geography was criticised because of focusing "on forces and processes that were important a century ago but much less relevant today" (Krugman, 2010, p.3). Both theories are still under the scholarly focus, and both are exploited in regional economics studies.

Methodology and geographical scale of the research

Most of the research carried out into clusters mainly measures cluster strength exploiting spatial agglomeration of labor force such as location quotient, Gini coefficient. Input-output methods, network analysis, discriminant analysis methods are used for defining clusters (Hill and Brennan, 2000; Porter, 2003; Delgado, Porter, Stern, 2013; Kies, 2008). The multidimensional approach developed by Titze et al. (2011) exploits the combination of spatial concentration and input-output methods (Brachert, Titze, Kubis, 2011). This combination allows to identify horizontal or vertical dimensions of clusters, where horizontal dimensions mean "clusters of firms producing similar output" and compete in the same market, and where the vertical dimension means that a cluster includes "firms linked through input-output relations" (Malmberg and Maskell, 2002; Brachert, Titze, Kubis, 2011).

It is argued that cluster measures should be compared with agglomeration measures as industry employment and firm's concentration and should be analyzed together with regional policy's key variables - productivity, employment growth and innovation (Woodward, 2013).

The school of NEG applies economic models with the focus on the economy of scale and agglomeration of firms and labor (Martin and Sunley, 1996; Cumbers and MacKinnon, 2004). Diverse approaches to analyze the spatial distribution of clusters of economic activities are exploited. Most research exploits administrative units of analysis, for example counties, districts, regions, etc… (Pablo-Marti and Arauzo-Carod, 2011).

This research exploits the percentage of individuals in employment and gross value added to avoid the problem of the regional size. For example, the population size of Lithuania was more than 3 million during the given period, while in Latvia – 2.5 million and 1.0-1.5 million in other selected regions. The population size impacts the size of the number of persons employed in the region. The distribution analysis was done by exploiting labor percentage share through clusters of economic activities in the selected regions. To understand which region is homogeneous and has the same economic function, i.e. the input of labor impacts the outcome of gross value added, a regression analysis of the selected regions and the selected clusters of economic activities was exploited.

To define geographical unit of analysis one should have in mind how much regions are similar in economy, social and political development, cultural values. This study focuses on the geographical scale of the Baltic States – Estonia, Latvia and Lithuania (NUTS 2 classification) – and their neighboring regions in the North (Finland regions - Länsi-Suomi, Helsinki-Uusimaa, Etelä-Suomi) and in the South (Poland regions - Warminsko-Mazurskie, Podlaskie). The Baltic States have similar history, similar political systems and comparable economic development as well as similar local resources. According to the report "The State of the Region. The Top of Europe - Striving for Direction in a Complex Environment" (Ketels and Pedersen, 2015), the Baltic States belong to Europe's macro-region – the Baltic Sea Region which includes Denmark, Finland, Iceland, Norway, Sweden, North Germany regions (Hansestadt Hamburg, Mecklenburg-Vorpommern and Schleswig-Holstein), North Poland regions (Pomorskie, Warminsko-Mazurskie and Zachodnio-Pomorskie) and Russian regions (Archangelsk Komi Republic, Nenetski AO, Vologodsk district). Other researchers see the Baltic States as part of the region of Central and Eastern Europe which also includes Czech Republic, Poland, Romania, Slovenia, Hungary (Blajer-Golebiewska, 2014). To confirm that the Baltic States are similar in economic outcomes, their gross value added (GVA) structure was analyzed. The neighbor regions were added to examine whether there is some relationship between the neighbor regions and the Baltic States in labour and GVA structure.

The main aim of the research was to analyse the impact of labour input on gross value added in the main clusters of economic activities in the Baltic States and the neighbour regions (Estonia (EE), Latvia (LV), Lithuania (LT), Länsi-Suomi (FI19), Helsinki-Uusimaa (FI1B), Etelä-Suomi (FI1C), Podlaskie (PL34), Warminsko-Mazurskie (PL62)). The regions were classified according to the EU NUTS2 classification. The Eurostat data of economic activities' clusters of 2000-2013 period was used. The cluster research exploits labour indicators as the main measure of agglomeration in the region and as the main indicator of the impact on regional economy. Thus, the employment indicator (i.e. the number persons employed) was exploited in this study which was assumed as the independent variable in further analysis. The dependent variable was assumed as the gross value added in clusters of economic activities. The following clusters of economic activities according to NACE Rev. 2 data were analysed (AFF) "Agriculture, forestry and fishing", (IND) "Industry" (includes "Mining and quarrying"), (MNF) "Manufacturing" (includes "Manufacturing", "Electricity, gas, steam and air conditioning supply", "Water supply; sewerage, waste management and remediation activities"), (CONS) "Construction", (WHRT) "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities", (ICT) " Information and communication", (FINS) "Financial and insurance activities", (RE) " Real estate activities", (PRFSC) "Professional, scientific and technical activities, Administrative and support service activities", (PAEDU) "Public administration and defence; compulsory social security; Education; Human health and social work activities", (ARTS&) "Arts, entertainment and recreation; Other service activities; Activities of households as employers; undifferentiated goods-and services-producing activities of households for own use; Activities of extraterritorial organisations and bodies".

Main results and discussion

Firstly, the structure of GVA and employment was analyzed. Then the regression analysis was done using panel data. The data of cluster of economic activities "Public administration and defense; compulsory social security; Education; Human health and social work activities" was used only for the calculations of the percentage share. One limitation of the research was that the EUROSTAT data did not have gross value added data of all the above listed clusters of economic activities of Poland's regions. The main research question was what share of gross value added of a cluster could be explained by a share of the persons employed. The shares of indicators were calculated in percentages.

The economic activity clusters were grouped by the mean of the percentage share of GVA and the persons employed of a 5- year period (Table 1). This analysis aimed to have the first insight on how similar or diverse the locations were. The main aim was to decide if we could assume that the Baltic States are similar in GVA and labor structure.

Table 1: The mean of the percentage GVA share and the mean of percentage share of persons employed of economic activities clusters in the Baltic States and NUTS2 neighbor regions of 2009-2013 period

	Indicator / Region	FI19	FI1B	FI1C	EE	LV	LT	PL34	PL62
>1	GVA, %	FINS	AFF						
	EMPL, %	RE	AFF	RE			RE	ICT, RE	ICT, RE
≤1-5<	GVA, %	AFF, ICT, ARTS&	FINS, ARTS&	AFF, ICT, FINS, ARTS&	AFF, ICT, FINS, ARTS&	AFF, ICT, FINS, ARTS&	AFF, ICT, FINS, PRFSC, ARTS&	FINS, ARTS&	FINS, ARTS&
	EMPL %	ICT, FINS, ARTS&	FINS, RE, PRFSC, ARTS&	AFF, ICT, FINS, ARTS&	AFF, ICT, FINS, RE, ARTS&	ICT, FINS, RE, ARTS&	ICT, FINS, ARTS&	FINS, PRFSC, ARTS&	FINS, PRFSC, ARTS&
≤5-10<	GVA, %	CONS, RE, PRFSC	CONS, ICT, RE	CONS, RE, PRFS	CONS, RE, PRFSC	CONS, RE, PRFSC	CONS, RE, PRFSC	AFF, CONS	AFF, CONS
	EMPL, %	AFF, CONS, PRFSC	MNF, CONS, ICT	CONS, PRFSC	CONS, PRFSC	AFF, CONS, PRFSC	AFF, CONS, PRFSC	CONS	CONS
≤10-15<	GVA, %	WHRT	MNF	WHRT	MNF, PAEDU	MNF, PAEDU	PAEDU		
	EMPL, %		IND, PRFSC	MNF		MNF	MNF	MNF	AFF
≤15-20<	GVA, %	MNF, PAEDU	IND, WHRT, PAEDU	MNF, PAEDU	IND, WHRT	IND	IND, MNF		
	EMPL, %	IND, MNF, WHRT		IND, WHRT	IND, MNF, PAEDU	IND, PAEDU	IND	IND, WHRT, PAEDU	IND, MNF, WHRT, PAEDU
≤20	GVA, %	IND		IND		WHRT	WHRT	IND, PAEDU	IND, PAEDU
	EMPL %	PAEDU	WHRT, PAEDU	PAEDU	WHRT	WHRT	WHRT, PAEDU	AFF	

Source: Authors

As the main variable for the spatial distribution and the agglomeration analysis is labor, the indicator of share of employment was the main argument for further grouping of locations. Thus, for further research the following clusters and regions were selected:

- "Construction" – all regions (FI19, FI1C, EE, LV, LT, PL34, PL62);
- "Industry" – FI19, FI1C, EE, LV, LT, PL62;
- "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities," (ICT) "Information and communication" – EE, LV, LT;
- "Information and communication" - FI19, FI1C, EE, LV, LT;
- "Financial and insurance activities" - FI1C, EE, LV, LT;

- ▪ "Professional, scientific and technical activities, Administrative and support service activities" - FI19, FI1C, EE, LV, LT;
- ▪ "Arts, entertainment and recreation; Other service activities; Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use; Activities of extraterritorial organizations and bodies" - FI19, FI1C, EE, LV, LT.

While grouping, the main focus was laid on the homogeneity criteria of the Baltic States, i.e the percentage share of the cluster's persons employed in the three Baltic States should fall into one group. Other regions were analyzed in the same group in case the share of employment of these regions belonged to the same percentage share group.

Further research was done by exploiting a classical linear regression model (CLRM). For this purpose, the data was analyzed using „MS Excel" and „Gretl" software. First, scatter diagrams were exploited where the independent variable was the percentage share of persons employed in a cluster and the dependent variable was the percentage share of GVA of a cluster. The assumptions of CLRM for normality, autocorrelation and homoscedasticity were tested using Jarque-Bera test, Durbin-Watson statistics and the White test.

The regression models of the selected clusters in the selected regions which vary in composition follow hereafter. There was no possibility to exploit the analysis with dummy variables for the regions because of the small number of observations. There was no possibility to use the multiple regression analysis because the variables were calculated as the percentage share inside the selected region. The applied level of significance was 0.05 (α =0,05) in all the applied tests.

The "Construction" cluster was analyzed using two different region scales (Table 1 in Annex 1), one of which was the Baltic States and the other – the Baltic States and the Northern neighboring regions and the Southern neighboring regions. Both the regional models are statistically significant according to F statistics. The interesting findings are that the larger scale regional regression model meets all the CLRM model assumptions for normality, autocorrelation and homoscedasticity and could predict the average value of the "Construction" cluster percentage share of GVA by the percentage share of persons employed, i.e. if one increase percentage share of persons employed by 1%, the percentage share of GVA of this cluster will increase by 0.91%. To evaluate why the assumption for normality in the EE, LV, LT region was not fulfilled, we used a scaterdiagram (Figure 1, Annex 2). The normality assumption was employed because it is assumed that the number of independent variables in the model is optimal and there is no influence of omitted variables is small. Thus, in the case of the EE, LV, LT region, we could assume that there could be other independent variables that have the influence. The "Industry" cluster was analyzed using two different region scales (see Table 3 below), one of which was the Baltic States and the other – the Baltic States and the Northern neighbor regions FI19, FI1C and the Southern neighbor region PL62. Both the regional models are statistically significant according to F statistics. Both the models do not meet the normality asumption even though the adjusted determination coefficient is high in both cases.

The analysis of the scatter diagrams of the "Industry" cluster (Figures 2, 3, Annex 2) demonstrates that the linear model is not suitable for predicting a stochastic relationship. The analysis of the "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities" cluster was done only for the Baltic States (EE, LV, LT). Even though the linear model is statistically significant and has high determination coefficient, it does not fulfill the normality criteria. Thus, it could not be used for prediction. The analysis of the "Professional, scientific and technical activities, Administrative and support service activities" cluster was done for two regions (Table 4 in Annex 1): one - FI19, FI1C, EE, LV, LT and the other was the Baltic States. We could note that these models also do not meet the normality criteria even though the determination coefficients are high and F statistics is significant. The cluster "Information and communication" was analyzed in two regions scales (see Table 5 in Annex 1): one of FI19, FI1C, EE, LV, LT and the other for the EE, LV, LT. The models are statistically significant. The determination coefficient is lower than in the previous models. Both the models do not meet the normality criteria. The second model (EE, LV, LT) also does not meet the homoscedasticity criteria. Both the models could not be used for prediction of a stochastic relationship. It could be noted that in earlier research, a negative correlation between the labor variable and the GVA was identified in the LT region (Rubčinskaitė and Kasnauskienė, 2016). The cluster

"Financial and insurance activities" was analysed in two different composition regions (see Table 6, Annex 1). The models could not be used for prediction because they do not meet the normality criteria. The "ARTS&" cluster was analysed in two different regional scales (see Table 7, Annex 1). Though both the models are statistically significant, they do not meet the normality criteria. Though, it could be noted that the determination coeficients and estimators are different in both models.

We could argue that the labour input impact on the generated GVA is similar in the following clusters "Construction," "Industry," "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities." It should be also noted that the concentration of labor force in these clusters is high and falls into the groups of „≤10-15<" and „≤15-20<". There are also different means of the determination coefficient in clusters. While the determination coefficient in clusters "Industry" and "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities" is high and comparable, the determination coefficients of the regression models of clusters "Information and communication", "Financial and insurance activities", "Arts, entertainment and recreation; Other service activities; Activities of households as employers; undifferentiated goods-and services-producing activities of households for own use; Activities of extraterritorial organisations and bodies" are very different and lower.

The authors argue that all the models require improvements on including more variables and possible employment of non-linear regression models. The different scales of regions proved that the location scale matters. Even the structure of a cluster composition is similar, the relationship model is different on the regional scale and in a cluster.

A preliminary analysis of the economic activity composition in the selected regions confirms that "center-periphery" forces could matter. An example could be Finland Helsinki-Uusimaa (FI1B) which has very a different cluster composition compared to other Finland's regions under analysis. This could also be explained by a higher level of urbanization in the region. This fact confirms the assumption that regions matter. The specificity of the regions and agglomeration is important in both the theories presented earlier which confirms their relevance in our research. This is proved by a different composition of activities of the neighboring regions of the Baltic States. The neighboring regions under analysis, except for Helsinki-Uusimaa, are periphery regions of Poland and Finland and the identified differences in labor and the GVA structure could suggest that center-periphery forces work well on the national or state level.

Conclusion

According to the author's knowledge, this is the first attempt to analyze the Baltic States as the whole region and compare it with the neighboring regions that belong to other nations and states. Also, this is an attempt to refuse the usual Cobb-Douglas production function while analyzing the labor input impact on GVA. This research was also an attempt to evaluate critically and assess the labor as the main indicator of agglomeration in cluster research by analyzing its possible impact on gross value added of the selected sector. The impact of labour force on generated gross value added is statistically significant in the selected clusters and regions. The impact is very different in the selected clusters and this implies that we should look for other than labor indicators when it comes to assessing the impact of clusters or industries agglomeration. Thus, we could accept Woodwards suggestion to explore more deeply the industries' agglomeration variables to assess the impact of clusters and industries on the economic development of regions.

The research results could be useful for regional development policy makers. A more in-depth analysis of regions' structure of economic activities' clusters could be recommended and the regional development policy should take into account the center – periphery forces and to invest smartly in regions.

References

Blajer-Golebiewska, A. (2014), "Economic Activity-Based Cluster Analysis of European Union Countries", World Review of Business Research, Vol. 4, No. 1, March 2014 Issue, pp. 48 -61

Brachert, M., Titze, M., Kubis, A. (2011). Identifying industrial clusters from a multidimensional perspective: Methodical aspects with an application to Germany. Papers in Regional , Science, Vol. 90, No 2, pp. 419-440

Capello, R. (2011). Location, Regional Growth and Local Development Theories. AESTIMUM 58, Giugno, pp. 1-25

Cumbers, A., MacKinnon, D. (2004). Introduction: Clusters in Urban and Regional Development.Urban Studies, Vol. 41, Nos 5/6, pp. 959–969

Delgado, M., Porter, M.E., Stern, S. (2012). Clusters, Convergence, and Economic Performance. NBER Working Paper Series, No. 18250, retrieved from http://190.242.114.8:8081/jspui/bitstream/11146/402/1/1680-DELGADO_2011_CLUSTERS_CONVE.PDF

1Pablo-Martí, F., Arauzo-Carod, J.-M. (2011). Spatial distribution of economic activities. Published in New challenges for European regions and urban areas in a globalised world : 51st ERSA Congress : 51st European Congress of the Regional Science Association International : 37th Spanish Regional Science Association Conference : 30th August - 3rd September 2011, Barcelona, Spain, hdl:10419/119944

Gianmarco I.P. Ottaviano. (2002). Regional Policy in the Global Economy: Insights from New Economic Geography. HWWA Hamburg Institute of International Economics Discussion Paper 211. ISSN 1616-4814, retrieved from https://www.econstor.eu/bitstream/10419/19392/1/211.pdf

Hill, E.W., Brennan, J.F. (2000). A methodology for identifying the drivers of industrial clusters: The foundation of regional competitive advantage. Economic Development Quarterly 14, pp. 65–96, doi: 10.1177/089124240001400109

Kaldor, N. (1970). The Case of Regional Policies. Scottish Journal of Political Economy, n. 3, pp. 337-348

Ketels, Ch. H.M., Protsiv, S. (2013). Clusters and the New Growth Path for Europe. Work Package 301. MS47 "Research paper on the role of clusters for the new growth path", Working Paper no 14, retrieved from http://www.foreurope.eu/fileadmin/documents/pdf/Workingpapers/WWWforEurope_WPS_no014_MS47.pdf

Ketels, Ch., Pedersen, H. (2015). The State of the Region. The Top of Europe - Striving for Direction in a Complex Environment. Retrieved from https://www.bdforum.org/wp-content/uploads/2015/11/SORR_report_web.pdf

Kies, U., Mrosek, T., Schulte, A. (2008). A statistics-based method for cluster analysis of the forest sector at the national and subnational level in Germany. Scandinavian Journal of Forest Research, 23 23, pp. 445-457

Krugman, P. (2010). The New Economic Geography, Now Middle-Aged. Prepared for presentation to the Association of American Geographers, April 16, 2010, retrieved from https://www.princeton.edu/~pkrugman/aag.pdf

Lucas, R. (1988). On the Mechanics of Economic Development. Journal of Monetary Economics, vol. 22, pp. 3-42

Martin, R., Sunley P. (1996). Paul Krugman's Geographical Economics and Its Implication for Regional Development Theory: A Critical Assessment. Economic Geography, Vol. 72, No. 3, 259-292

Martin, R., Sunley, P. (2003). Deconstructing clusters: Chaotic concept or policy panacea? Journal of Economic Geography 3, pp 5–35, http://www.cbr.cam.ac.uk/fileadmin/user_upload/centre-for-business-research/downloads/working-papers/wp244.pdf

Myrdal G. (1957), Economic Theory of Under-developed Regions, General Duckworth & Co., London

Porter, M.E. (2003). The Economic Performance of Regions. Regional Studies, Vol. 37, pp. 549-578.

Pires, J.C., Cravo, T., Lodato, S., Piza, C. (2013). Industrial Clusters and Economic Performance in Brazil, Inter-American Development Bank. Working Paper Series No. IDB-WP-475, retrieved from http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=38216320

Romer, P. (1987). Growth Based on Increasing Returns due to Specialisation. American Economic Review, Papers and Proceedings, vol. 77, pp. 56-67

Rubčinskaitė, R., Kasnauskienė, G. (2016). Identification of Clusters Contribution to Economic Growth in the Baltic States (2016. Development in Economics: Theory and Practice. 12th International Prof. Vladas Gronskas Young Researchers Scientific Conference. Reviewed Selected Papers. http://www.khf.vu.lt/dokumentai/failai/konferenciju/gronsko2016/GRONSKO_KONFERENCIJOS_LEIDINYS_2015.p

Sölvell, O., Ketels, Ch., Lindqvist, G. (2008). Industrial specialization and regional clusters in the ten new EU member states. Competitiveness Review: An International Business Journal, Vol. 18 Iss ½, pp. 104 – 130, retrieved from http://dx.doi.org/10.1108/10595420810874637

Titze, M., Brachert, M., Kubis, A. (2011). The identification of regional industrial clusters using qualitative input–output analysis (QIOA). Regional Studies 45, pp. 89–102

Woodward, D. P. (2012). Industry Location, Economic Development Incentives, and Clusters. The Review of Regional Studies, 42, pp. 5-23.

Annex 1

Table 1: Regression model of "Construction" cluster

NACE Rev.2 / region	Regression function	R^2	F, p
"Construction" –FI19, FI1C, EE, LV, LT, PL34, PL62	$Y_{gva} = 2,84 + 0,91\ X_{empl} + \varepsilon_i$	$R^2 = 0,8391$ $R^2_{Adj} = 0,8376$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ $F(1, 110) = 573,7053$ $p = 1,89e\text{-}45$
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 728,957; p= 5,11696e-159; **H_0 not rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistics = 1.98118; p = 0,712735; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 0,802214; $P(\chi^2 > 0,802214) = 0,669578$; **$H_0$ not rejected**		
"Construction" – EE, LV, LT	$Y_{gva} = 2,81 + 0,816\ X_{empl} + \varepsilon_i$	$R^2 = 0,6856$ $R^2_{Adj} = 0,6777$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ $F(1, 40) = 87.24417$ $p = 1.33e\text{-}11$
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 353,312; p= 1,90202e-077; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 2,67397; p = 0,992969; **H_0 not rejected** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 3,94938; $P(\chi^2 > 3,94938) = 0,138804$; **$H_0$ not rejected**		
Source: Authors			

Table 2: Regression model of "Industry" cluster

NACE Rev.2 / region	Regression function	R^2	F, p
"Industry" – FI19, FI1C, EE, LV, LT, PL62	$Y_{gva} = 1,15 + 0,930\ X_{empl} + \varepsilon_i$	$R^2 = 0,9567$ $R^2_{Adj} = 0,9562$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ $F(1, 82) = 1863,95$ $p = 1.07e\text{-}57$
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 9815.8, with p-value 0; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1.98942; p-value = 0.682165; **H_0 not rejected** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 1,26759; p = $P(\chi^2 > 1.26759) = 0,530575$; **$H_0$ not rejected**		
"Industry" –EE, LV, LT	$Y_{gva} = -0,946 + 0,976\ X_{empl} + \varepsilon_i$	$R^2 = 0,942$ $R^2_{Adj} = 0,940$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ $F(1, 40) = 652,98$ $p = 2.17e\text{-}26$
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 1217,14, p= 5.02534e-265; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1.98942; p-value = 0.682165; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White testas: LM = 1,29682; $P(\chi^2 > 1,29682) = 0,522875$; **$H_0$ not rejected.**		
Source: Authors			

Table 3: Regression model of "Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities" cluster

NACE Rev.2 / region	Regression function	R^2	F, p
"Wholesale and retail trade; repair of motor vehicles and motorcycles, Transportation and storage, Accommodation and food service activities" – EE, LV, LT	$Y_{gva} = 0,0198 + 1,01\ X_{empl} + \varepsilon_i$	$R^2 = 0,991$ $R^2_{Adj} = 0,990$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ F(1, 40) = 4459.363 p = 1.20e-42
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 1,08515; p= 0,00136877; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1.88; p = 0. 413435; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 1,14565; p ($\chi^2 > 1,14565$) = 0,563931; **H_0 not rejected**		
Source: Authors			

Table 4: Regression model of "Professional, scientific and technical activities, Administrative and support service activities" cluster

NACE Rev.2 / region	Regression function	R^2	F, p
"Professional, scientific and technical activities, Administrative and support service activities" - FI19, FI1C, EE, LV, LT	$Y_{gva} = 2,41 + 0,881\ X_{empl} + \varepsilon_i$	$R^2 = 0,8395$ $R^2_{Adj} = 0,8372$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ F(1, 68) = 355,9227 p = 9.96e-29
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 1727,75; p= 0; **H_0 rejected**. Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,85122; p-value = 0,408194; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White testas: LM = 0,0082853; p ($\chi^2 > 0,0082853$) = 0,995866; **H_0 not rejected**		
"Professional, scientific and technical activities, Administrative and support service activities" - EE, LV, LT	$Y_{gva} = 0,865 + 0,9996\ X_{empl} + \varepsilon_i$	$R^2 = 0,9651$ $R^2_{Adj} = 0,9642$	H_0: β_1=0; H_A: $\beta_1 \neq 0$ F(1, 40) = 1107,242 p = 8,99e-31
	$u_i \sim N\ (0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 2169,44; p= 0; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,95544; p = 0,510532; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 2,13751; p ($\chi^2 > 0,0082853$) = 0,343436; **H_0 not rejected.**		
Source: Authors			

Table 5: Regression model of "Information and communication" cluster

NACE Rev.2 / region	Regression function	R^2	F, p
"Information and communication" - FI19, FI1C, EE, LV, LT	$Y_{gva} = 7,98 + 0,969\ X_{empl} + \varepsilon_i$	$R^2 = 0,5562$ $R^2_{Adj} = 0,5496$	H_0: $\beta_1=0$; H_A: $\beta_1\neq0$ $F(1, 68) = 85,21121$ $p = 1,29e-13$
	$u_i \sim N(0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 96,5006; p= 1,1096e-021; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,62273; p = 0,117841; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 4,078571; p $(\chi^2 > 4,078571) = 0,130122$; **$H_0$ not rejected**		
"Information and communication" - EE, LV, LT	$Y_{gva} = 7,86 + 0,773\ X_{empl} + \varepsilon_i$	$R^2 = 0,4284$ $R^2_{Adj} = 0,4142$	H_0: $\beta_1=0$; H_A: $\beta_1\neq0$ $F(1, 40) = 29,99118$ $p = 2.57e-06$
	$u_i \sim N(0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 26,1672; p= 2,0791e-006; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,62273; p = 0,117841; **H_0 not rejected** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 6,24994; p $(\chi^2 > 6,24994) = 0,0439382$; **$H_0$ rejected**.		

Source: Authors

Table 6: Regression model of "Financial and insurance activities" cluster

| NACE Rev.2 / region	Regression function	R^2	F, p
"Financial and insurance activities" - FI1C, EE, LV, LT	$Y_{gva} = 7,47 + 0,999\ X_{empl} + \varepsilon_i$	$R^2 = 0,6202$ $R^2_{Adj} = 0,6131$	H_0: $\beta_1=0$; H_A: $\beta_1\neq0$ $F(1, 54) = 88,18038$ $p = 6,04e-13$
	$u_i \sim N(0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 70,0082; p= 6,27933e-016; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,78327; p = 0,30921; **H_0 not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 0,863751; p $(\chi^2 > 0,863751) = 0,64929$; **$H_0$ not rejected**		
"Financial and insurance activities" - EE, LV, LT	$Y_{gva} = 4,81 + 0,889\ X_{empl} + \varepsilon_i$	$R^2 = 0,5548$ $R^2_{Adj} = 0,5437$	H_0: $\beta_1=0$; H_A: $\beta_1\neq0$ $F(1, 40) = 49,85956$ $p = 1,54e-08$
	$u_i \sim N(0, \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 73,4859; p= 1,1034e-016; **H_0 rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,02603; p = 0,000728527; **H_0 not rejected**. Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 3,72791; p $(\chi^2 > 3,72791) = 0,155058$; **$H_0$ not rejected**		

Source: Authors

Table 7: Regression model of "Arts, entertainment and recreation; Other service activities; Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use; Activities of extraterritorial organisations and bodies" cluster

| NACE Rev.2 / region	Regression function	R^2	F, p
"ARTS&" - FI19, FI1C, EE, LV, LT	$Y_{gva} = 6,03 + 0,65\ X_{empl} + \varepsilon_i$	$R^2 = 0,5252$ $R^2_{Adj} = 0,5182$	$H_0: \beta_1=0;\ H_A: \beta_1 \neq 0$ $F(1, 68) = 75,2325$ $p = 1,31e-12$
	$u_i \sim N\ (0,\ \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 18,813; p= 8,219e-005; **H₀ rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 1,64205; p = 0,131577; **H₀ not rejected** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 0,711541; p ($\chi^2 > 0,711541$) = 0,700633; **H₀ not rejected**		
"ARTS&" - EE, LV, LT	$Y_{gva} = 7,37 + 0,732X_{empl} + \varepsilon_i$	$R^2 = 0,4699$ $R^2_{Adj} = 0,4566$	$H_0: \beta_1=0;\ H_A: \beta_1 \neq 0$ $F(1, 40) = 35,46396$ $p = 5,46e-07$
	$u_i \sim N\ (0,\ \sigma^2)$ H_0: errors are normally distributed; H_A: errors are not normally distributed. Jarque-Bera test = 39,4895; p= 2,66048e-009; **H₀ rejected.** Errors do not correlate: cov $(X_i, u_i) = 0$. H_0 : no first order autocorrelation; H_1 : first order correlation exists. Durbin-Watson statistic = 0,990543; p = 0,00045779; **H₀ not rejected.** Homoscedasticity H_0 : Heteroscedasticity does not exist; H_1 : Heteroscedasticity exists. White test: LM = 6,30044; p ($\chi^2 > 10,0584$) = 0,0428427; **H₀ rejected.**		

Source: Authors

Annex 2

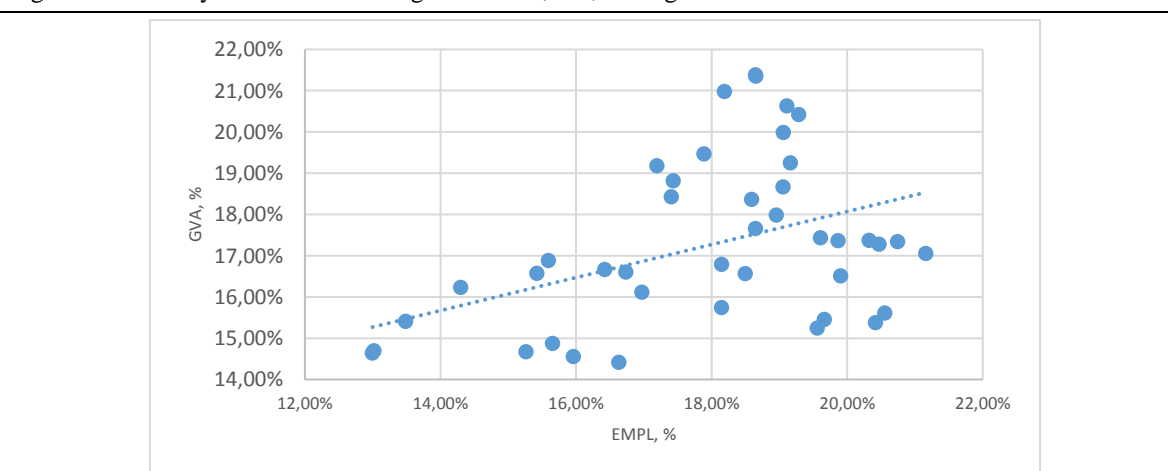Figure 1: "Construction" cluster scater diagram for EE, LV, LT region
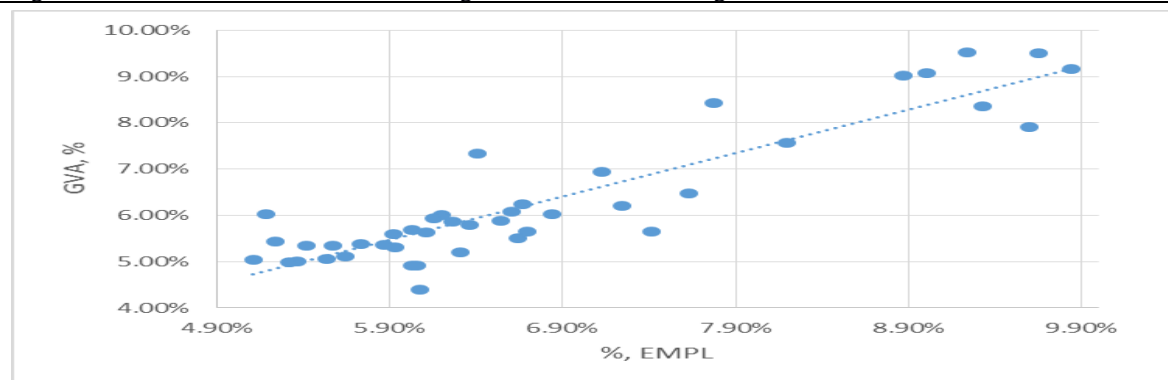
Source: Authors

Figure 2: "Industry" cluster scater diagram for FI19, FI1C, EE, LV, LT, PL62 region
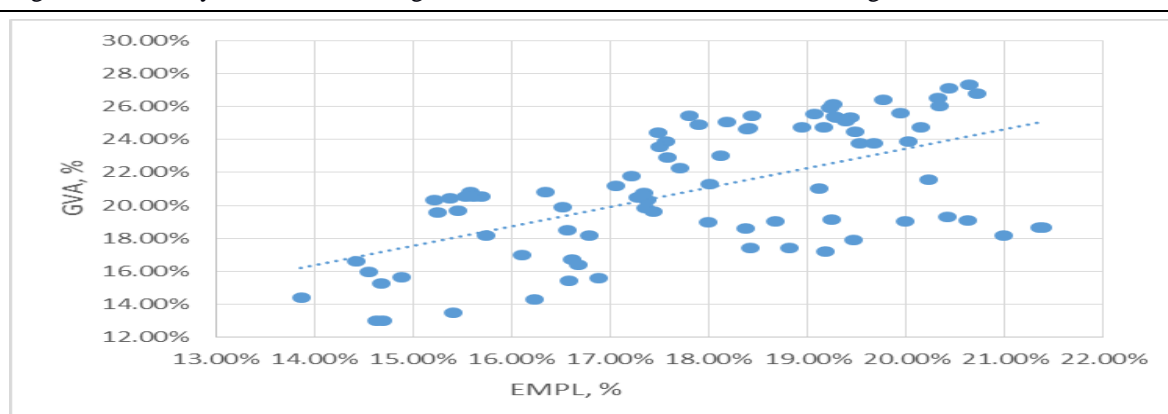
Source: Authors

Figure 3: "Industry" cluster scater diagram for EE, LV, LT region

Source: Authors