VILNIUS UNIVERSITY

Rima
RUBČINSKAITĖ

The Impact of Clusters' on Economy and Innovation in the Baltic States

SUMMARY OF DOCTORAL DISSERTATION

Social Sciences
Economics S 004

VILNIUS 2019

This dissertation was written between 2014 and 2018 at Vilnius University Faculty of Economics and Business Adminsitration.

Academic supervisor -
Prof. Dr. Gindrutė Kasnauskienė (Vilnius University, Social Sciences, Economics, S 004).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:
Chairman – Prof. Dr. Algirdas Miškinis (Vilnius University, Social Sciences, Economics, S 004).
Members:
Prof. Dr. Tomas Baležentis (Vilnius University, Social Sciences, Economics, S 004);
Prof. Dr. Vincentas Rolandas Giedraitis (Vilnius University, Social Sciences, Economics, S004);
Dr. Peter Huber (Austrian Institute of Economic Research, Social Sciences, Economics, S 004);
Prof. Habil. Dr. Borisas Melnikas (Vilnius Gediminas Technical University, Social Sciences, Economics, S 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 14:00 hour on October 30, 2019 in Room 403 of the Faculty of Economics and Business Administration, Vilnius University. Address: 9 Saulėtekio ave, room No. 403, Vilnius, Lithuania, phone No.: +37052366126; e-mail: evaf@evaf.vu.lt.

The text of this dissertation can be accessed through the library of Vilnius University as well as on the website of Vilnius University: *https://www.vu.lt/naujienos/ivykiu-kalendorius*.

VILNIAUS UNIVERSITETAS

Rima
RUBČINSKAITĖ

Klasterių poveikis Baltijos šalių ekonomikai ir inovacijoms

DAKTARO DISERTACIJOS SANTRAUKA

Socialiniai mokslai
Ekonomika S 004

VILNIUS 2019

Disertacija rengta 2014– 2018 metais Vilniaus universiteto Ekonomikos ir verslo administravimo fakultete.

Mokslinė vadovė – prof. dr. Gindrutė Kasnauskienė (Vilniaus universitetas, socialiniai mokslai, ekonomika, S 004).

Gynimo taryba:
Pirmininkas – **prof. dr. Algirdas Miškinis** (Vilniaus universitetas, socialiniai mokslai, ekonomika, S 004).
Nariai:
prof. dr. Tomas Baležentis (Vilniaus universitetas, socialiniai mokslai, ekonomika, S 004);
prof. dr. Vincentas Rolandas Giedraitis (Vilniaus universitetas, socialiniai mokslai, ekonomika, S 004);
dr. Peter Huber (Vienos ekonomikos tyrimų institutas, socialiniai mokslai, ekonomika, S 004);
prof. habil. dr. Borisas Melnikas (Vilniaus Gedimino technikos universitetas, socialiniai mokslai, ekonomika, S 004).

Disertacija ginama viešame gynimo tarybos posėdyje 2019 m. spalio mėn. 30 d. 14 val. Vilniaus universiteto Ekonomikos ir verslo administravimo fakulteto 403 posėdžių salėje. Adresas: (Saulėtekio al. 9, II rūmai, Vilnius, Lietuva), tel. +37052366126; el. paštas evaf@evaf.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: *https://www.vu.lt/naujienos/ivykiu-kalendorius*.

CONTENTS

SUMMARY

1. INTRODUCTION

Relevance and importance. The formation of local and global networks has a major impact on the changing structure of the modern economy. There is evidence of the impact of strong clusters on regions' economy. Clusters and networks are catalysts for accelerating industrial transformation and developing new regional competitive advantages. Clusters accelerate the growth of job places & new enterprises, thus contributing to growth and prosperity.

Research problem. Cluster research has been conducted in developed countries (USA, Germany, Canada), while no cluster identification research has been done in the Baltic States (Estonia, Latvia, Lithuania). Cluster monitoring in Europe is tailored to US-identified clusters and more closely reflects the EU countries' specialization in terms of the business sector's economic activities. Researchers in the Baltic States have addressed the topic of clusters, but there is a lack of research on clustering in the Baltic region.

Clusters are associated with export and innovation. However, the identification of clusters is mainly based on labor factor features. The productivity is associated with export development and innovation. Productivity is also linked to long-term economic development and the identification of strong cluster structures. However, there is a lack of research that would exploit the productivity factor together with labor factors while identifying clusters.

In order to fill the identified research gaps, the research problem of the dissertation was formulated using the following questions: "Do the business sector's economic activities tend to cluster in the Baltic region, what is the nature of such clustering, and how does it influence regional economy and innovation?"

Research object: the clusters of Baltic States (Estonia, Latvia, Lithuania) and their composition.

Aim of the thesis: to identify clusters in the Baltic region and to evaluate the impact of these clusters on the economy and innovation within the Baltic region.

Research tasks:

1. To analyze and systematize the theoretical assumptions of cluster research;

2. To analyze the latest cluster identification methods, compare them, and adapt them for clustering research in the Baltic region;

3. To analyze the research methods of clusters' impact on the economy and innovation as well as select and adapt the appropriate methods for for the evaluation ofclusters' impact on the economy and innovation;

4. To identify and analyze clusters and their composition;

5. To analyze the impact of clusters on the economy within the Baltic region;

6. To analyze the impact of clusters on innovation within the Baltic region;

7. To interpret the results of empirical research, present a summary, and develop recommendations.

Scientific Novelty of the Thesis

The cluster concept has been used more actively since the 20th century. However, the term *cluster* is used across different studies with different definitions. As a consequence, researchers use different methods for identifying clusters in different geographic coverage. The cluster maps made in Europe were mainly based on data of the economy of the United States of America and were not verified for another geographic scope. The main indicator to for identifying clusters is employment, despite the assumptions that clusters are the source of economic growth, labor productivity, and export growth in a region. In this work, there was an attempt to design the cluster identification algorithm, which can be applied in different geographical coverage territories, by testing it with panel methods. According to the knowledge of the author, it is the first time that productivity indicators were used for cluster identification together with employment factor. After the identification of the clusters, the identified clusters' impact on the economy and innovation was analyzed. According to the knowledge of the author, cluster identification was performed for the first time in the Baltic States (Estonia, Latvia, Lithuania) region as a whole by applying cluster analysis.

Thesis structure. The thesis is structured in three main parts: the first part focuses on the analysis of theoretical approaches in relation with clusters & regional economy; the second part is dedicated to analyze previous methods & develop a methodological approach for the empirical research; and the third one is dedicated to an empirical analysis, conclusions, and recommendations. The first part of the work analyzes theoretical approaches and assumptions in cluster research and the concept of the cluster in different theoretical concepts, its relevance in the context of regional economic development, and the clusters' impact on the development of the economy and innovation in regions and countries. In order to design the methodological approach for the empirical research, cluster identification methods exploited by other researchers were analyzed in detail in the second part of thesis. The research methods that allow us to assess the impact that clusters have on the economy and innovation are also explored in the second part. The empirical research results were presented in the third part of the work. The impact of identified clusters on the economy using panel models is performed in the third part. The impact on innovation, due to a lack of data, was investigated using a simple linear regression model. Main conclusions and recommendations are presented in the end of the dissertation.

2. ANALYSIS OF THEORETICAL APPROACHES AND ASSUMPTIONS IN CLUSTER RESEARCH

Contemporary research on regional economic development emphasizes that regions are successful when they have an agglomeration of economic activities (Storper, 1997; Lesage and Fischer, 2008; Woodward, 2011; Liviano and Arauzo - Carod, 2011; Mukim, 2012; Felipe and McCombie, 2012; Ketels et al., 2013). In the 21st century, clusters and networks are associated with regional economic development and innovation (Porter, 2003; Das and Finne, 2008; Spenser et al., 2010; Boschma, 2017; Crescenzi and Iammarino, 2017), the agglomeration of economic activities in specific geographical areas, urbanization processes, and the fourth industrial revolution.

Although the concept of a cluster was first introduced by Michael Porter (Porter, 1998, 2003), other researchers (Woodward, 2011; Arauzo, 2008) emphasize that the cluster as an agglomeration of certain types of economic activities was first introduced and analyzed by Alfred Marshall (1890). External factors that determine the agglomeration of economic activities include intensive supplier relationships, knowledge transfer, and workforce concentration (Woodward, 2011). The main factors of the agglomeration process mentioned by Marshall (1890):

- Intensive relationships between suppliers and buyers, enabling greater productivity through vertical disintegration and specialization;
- The ability to obtain specific knowledge of an economic activity through enhanced relationships between units of economic activity (Mukim, 2012; Martynovich and Lundquist, 2016; Hervas-Oliver et al., 2017);
- The labor market concentration, where agglomeration leads to higher productivity as a result of increasing employee compliance in business (Woodward, 2011; Felipe and McCombie, 2012; Hervas-Oliver et al., 2017).

Researchers who represent a strand of new economic geography and explore spatial distributions of economic activities exploit the concept of a cluster as a phenomenon of the agglomeration of economic activities. Thanks to the concentration of industries in a given geographical area positive economic effects are at the core of the new concept of economic geography (Krugman, 1991, 2010; Venables, 1996; Hanson, 1996; Woodward, 2011; Spenser et al., 2010; Garretsen and Martin, 2010). Recent research on agglomeration and clusters reveals that concentration of firms in regions leads to higher wages (Porter, 2003; Mion, Naticchioni, 2009; Ketels, 2013).

Several key schools of the cluster concept can be distinguished in economic theory: the new economic geography and regional competitiveness. The regional competitiveness trend is represented by M. Porter, author of the concept of competitiveness, and his followers. The neoclassical direction of economic theory is presented by researchers from economic geography and the distribution of economic activities.The principal scholars are A. Marshall, P. Krugman, M. Fujita, G. Ottaviano, T. Tabuchi, J.F. Thisse. In the following sections, the concept of a cluster is explored in the context of regional economic growth/development and competitiveness theories.

The Concept of a Cluster in the Context of Regional Economic Development and Growth Theories

The branch of economic geography analyzes space as a factor determining the strengths and weaknesses of an economic system and explores how that economic system works in a given space or geographical area. The geographical area of an economic activity determines the resources of the local factors of production, the advantages or disadvantages of the economic activity in terms of geography, and the natural resources of the chosen local area. Economic geography argues that the geographical distribution of production resources (raw materials, natural advantages) is only partly determined by external factors (Capello, 2011). Historical factors have an impact on the distribution of resources geographically, such as human capital, social capital, land, labor productivity, and the availability of (proximity to) local resources (Capello, 2011; Krugman, 1998, 2010; Arauzo-Carod, 2008; Garretsen and Martin, 2010). All these factors are the source of an increasing productivity in a particular region.

The direction of economic geography that explores the significance of the region analyzes:

(a) The different territorial distribution of different types of production, which are defined by the factors of the choice of economic activity;

b) The factors determining differences in the spatial distribution of the market between producers and the functional spatial distribution of activities (Capello, 2011; Garretsen and Martin, 2010; Liviano and Arauzo-Carod, 2011; Felipe and McCombie, 2012).

The Weber and Hoteling models, together with the central location theory, were aimed at explaining why enterprises cluster in a particular location. Some local economic growth models induced the emergence of new economic development theories, such as Solow's economic growth theory (Hagemann, 2009). Regional development and growth theories distinguish hypothetical assumptions about transport costs, agglomeration, and uneven distribution of production systems and resources. The models explore factors that determine business clustering in a location, which are transportation costs, distance to major markets, and local resources. The Hoteling model proves that there is price competition for identical goods in a regional market. This model proves that an internal economy of scale in a local space or regional economy of scale can be explained by the economy of scale of transport costs. This may be a factor explaining the assumption that the three Baltic countries comprise a single region.

Cumbers and MacKinnon (2004) discussed the added value of clusters as a theoretical model. The researchers proved the advantages of clusters in the context of the knowledge economy. Researchers highlighted external links that are critical to the development of clusters, be it new sources of information, a skilled workforce, entrepreneurs, or ideas. This is especially true in the peripheral regions, which are characterized by a lack of urban economies.

Most studies conducted under the influence of new economic geography have been country-specific (Italy - Mion, 2003; Spain - Pablo-Marti et al., 2011; Sweden - Nefke et al., 2009; USA - Desmet and Vernon Henderson, 2014; United Kingdom - Martin et al., 2015; Poland - Nazarczuk and Krajewska, 2018; Vietnam - Hoang et al., 2019) and rarely done on a regional scale (Catalonia-Liviano et al., 2011). At European Union level, the regions are classified according to the NUTS-2 system, and this system is the most widely used in various regional development studies. However, as the EU regions are very different in geographical locations, with differences in political systems and histories, with different levels of economic development, the definition of what a region is crucial in regional development research. Regions are defined on the basis of different approaches: administrative, geographical, economic, social, historical (Dawkins, 2003; Dzemydaitė, 2016). According to the report *The State of the Region. The Top of Europe – Striving for Direction in a Complex Environment* (Ketels et al., 2015), the Baltic States – Lithuania, Latvia, Estonia – are allocated to the Baltic Sea Region. The Baltic Sea Region also includes the following countries: Northern European countries (Denmark, Finland, Iceland, Norway, and Sweden), the northern regions of Germany (Hansestadt Hamburg, Mecklenburg-Vorpommern and Schleswig-Holstein), the northern regions of Poland (Pomorskie, Warminsko-Mazurskie and Zachodnio-Pomorskie), and a large part of the Northwest Federal District of Russia (excluding the four most remote regions, i.e., the Republic of Komi, Komi-Arkhangelsk region, Nenetsky, and Vologodskaya regions). Other researchers allocate the Baltic States to the region of Central and Eastern Europe, which includes the Czech Republic, Poland, Romania, Slovakia, and Hungary (Blajer-Golebiewska, 2014). These countries and their regions differ significantly in their economic development, history, and institutional structure. The three Baltic countries are similar in their histories, political systems, and economic development as well in their institutional setups. Thus, the assumption that the Baltic States comprise a single region is applied in this thesis.

In conclusion, economies of scale can be considered as an external factor in the selected region. One of the key issues in exploring agglomeration phenomena are the patterns of labor distribution in the Baltic region. In the context of the new economic geography, the cluster is a concentration of certain economic activities in a specific geographical area – a region.

The Concept of a Cluster in the Context of Regional Competitiveness

The concept of a cluster in the context of competitiveness was developed by M. Porter (1998; 2003), author of the concept of competitiveness. The concept of regional competitiveness is actively used by the European Commission in designing its regional development policy. There are a few main definitions in relation to the research of the transformation of an industrial economy into an economy of clusters and knowledge clusters (Cooke, 2001, 2012; Breschi and Malerba, 2001; Ergazakis et al., 2004; Bathelt et al., 2004; Delgado et al., 2011; Ketels, 2012, 2013; Karlsson and Rouchy, 2014; Belick Manzini and Di Serio, 2017) and which are related to the definition of a cluster: clusters, cluster organizations, and networks. It is important to consider the following characteristics while exploring these definitions: geography, industry, or sector, nature of relationship, purpose.

Clusters are traditionally defined as a geographic place-based partnership from linked economic activities (Porter, 1990). Collaboration may or may not occur.

Cluster organizations focus on specific geography, but this geography can be driven more by policy than by the boundaries of economic areas (Sölvell et al., 2008). Cluster organizations are always focused on a group of related economic activities (the so-called *cluster category*) and provide a framework for collaboration.

Enterprise networks (Ketels, 2012) may or may not be related to specific geography and economic activity groups. By definition, they are designed specifically for active collaboration. This cooperation may be unlimited in time or focused on the execution of a specific project. According to the given definition of an enterprise network, a cluster organization is a type of specific network.

The cluster definition used in the regional competitiveness policy of Europe is defined according to the Porter's definition, i.e., the cluster is a group of related economic activities in a given geographical area. However, both Porter and his followers argue that the cluster is not just a cluster of business activities in a particular geographical area. In addition, both Porter and his followers exploit only the bussiness sector for identifying clusters, although the importance of both academic institutions and the public sector is recognized. For example, Porter (2003), Delgado (2013), and Titze and Brachert (2016) used an input-output approach in their research to identify clusters. Later, however, Delgado (2013) acknowledged that an input-output relationship analysis does not identify clusters qualitatively, especially when agglomerated data are used instead of the enterprise level data. The main disadvantage of the input-output method is that it is not linked with the geographic area. Thus, the author of this thesis suggests that the concept of a cluster has several uncertainties:

- With regard to the links between economic activities, the variety of clusters' identification methods prove the uncertainty of the definition; also, there is no clear link between these methods and the geographical area, and it is unclear when and which method of cluster identification is appropriate;

- Mostly, only data on business economic activities are analyzed, although the concept of a cluster is broader in the scientific literature, i.e., it includes public and research institutions;

- The concept has a geographic uncertainty, namely regarding the administrative areas are most commonly used and it is not clear which geographical coverage region would be most appropriate for clusters' identification.

The definition of a cluster in two different theoretical concepts is similar in terms that it examines the dispertion or concentration of economic activities in a particular region. In the context of competitiveness theory, much attention is paid to the relationships between economic activities in the value added chain, and the main methods for identifying the relationships are a correlation between economic activities in a specific geographic area and the analysis of input-output links. However, the latter method has the drawbacks of not linking economic activities to a specific geographical area unless enterprise level data are used. Definition of what we consider a region is very important for researchers representing new economic geography strand. If clusters are related economic activities in a specific geographical region, the concept of a region is often interpreted differently in different contexts.

In this thesis, the author applied Porter's definition of a cluster: a cluster is a group of related business economic activities in a particular region, in this case – the Baltic States (Estonia, Latvia, Lithuania). A business economic activity is a four-digit NACE Rev. 2 economic activity of the business sector. For the first time, geographic coverage is different than in other regional and country studies – i.e., all three Baltic States (Estonia, Latvia, and Lithuania) were treated as one region in this work.

A Comparison of Clusters Identification Methods

Quantitative clustering studies that investigate economic agglomeration phenomena typically use industry data from national or

EU industry statistical databases. Porter (2003), and later Delgado et al. (2013), used both LQ (Location Quotient) and input-output analysis as well as agglomeration and collocation methods to improve cluster identification and used data of different levels. Three cluster identification methods, which have been applied in the USA, Germany, and Canada, were analyzed and compared in this work.

Delgado and Porter's Clusters Identification Method

The approach developed by Porter and his followers for identifying clusters is exploited by the European Commission to monitor clustering process in the European Union (Ketels and Protsiv, 2014a; 2014b; 2016). Delgado et al. (2013) used the database of the 2009 Service, Manufacturing and Industry 675 (Six-Digit US Industry Classification System) to improve the clusters identification method suggested by Porter. Researchers argued that Porter's suggested cluster identification algorithm is of good quality because it identifies different types of interdependencies. Criteria used for improving the method is based on the assumption that the group of linked economic activities must also include the demand and supply side of the industries and also technologies, knowledge, and skills. Porter (2003) identified clusters mainly by exploiting the colocation patterns of service and manufacturing economic activities. An important feature of clusters as proposed by Porter was that one economic activity was assigned only to one cluster. M. Delgado et al. (2013) developed an algorithm for identifying clusters further using a national input-output database, the Occupational Employment Statistics (OES). Delgado et al. (2013) also made an assumption about agglomeration as an important condition for cluster development (Spenser et al., 2010; Delgado et al., 2013; Felipe and McCombie, 2012; Arauzo-Carod, 2010; Garretsen and Martin, 2010; Chasco et al., 2012), the concentration of EAs was also explored and included. The colocation patterns of EAs were studied by exploring the correlation coefficients of employment of economic activities in a geographical area. Porter

(2003) and Delgado et al. (2013) have assumed that this approach can identify various interrelationships between different economic activities, such as technology, demand links, and workforce skills (Delgado et al. 2013). Delgado et al. (2013) compared the identified clusters to those identified by Porter and results were similar, i.e. similar groups of EAs in the clusters. Although researchers mainly exploited the six-digit level data of economic activities of the US, some cluster similarity matrices were constructed at four-digit level. Likewise, Porter's (2003) clustering study had the same limitations, i.e., the study was mainly conducted at the level of 4 digit EAs. However, some calculations were done using two or three digit level data. Another disadvantage of the study conducted by Delgado et al. (2013) was that the employment by occupation data were not geographically linked.

Brachert's et al. Clusters Identification Method

Brachert et al. (2011) further developed a method proposed by Titze et al. (2011). The method reflected the sectoral and spatial interrelationships between industry clusters. The three-step approach for cluster identification proposed by German researchers has been applied in Germany. They combined the input-output method with the spatial concentration using the *Gi* indicator, analyzing the vertical relationships of clusters within the region and with neighboring regions. However, because the input-output matrix was transformed into a qualitative input-output matrix, some information was lost. The fact that the study was based on assumptions, such as the similarity between intersectoral relationships at the national and regional levels, it did not allow the determination of genuine buyer-supplier relationships. Likewise, the application and interpretation of the Stenberg and Litzenberger (2004) cluster index, which was used in the approach, depends on the choice of its threshold value. Another drawback is that input-output linkage reflects only the interrelationships between industries or economic activities, but does not reflect the flow of knowledge or cooperation or innovation. Significantly, the assumption of productivity uniformity in all regions of the country was very rigid and did not consider geographic location as a fundamental factor for the patterns of unique business structures.

The clustering methods proposed by Porter and Delgado (US) and Titze (2011) and Brachert (2011) are similar in their methodological approach, i.e., that both exploited input-output relationships and concentration methods. The identification of clusters as done by Canadian researchers (Spenser et al. 2010) used an alternative method. The biggest difference from the previous cluster identification approaches was that Spenser et al. (2010) analyzed labor factors by labor market areas that overlap with urban regions in the Canadian statistical system.

Spenser's et al. ClustersIdentification Method

Canadian researchers used the 2001 Census data of Canada, since this data have a clear geographical distribution and contain information on population income and other important labor force characteristics. These data were enriched by employment data for 300 industries across 140 urban regions. One of the major disadvantages of using the cluster identification algorithm proposed by Porter or the Delgado group is that it was adapted to one of the world's largest economies with its own unique structure and characteristics. Even Porter (2003) emphasized that this method is not applicable in many parts of the world. Ketels (2014b) recognizes that the structure and composition of German clusters is different from that of the US.

Canadian researchers made similar assumptions as were used in the US. The assumptions were made regarding industry employment specialization, collocation patterns, cluster scale, and critical mass (Spenser et al., 2010).

The cluster identification method proposed by Canadian researchers was conducted in four steps:

1) By exploring the concentration of industries exploiting employment data from the 2001 Census and estimating the LQs of 300 industries across 140 regions of Canada;

2) By exploring industrial collocation patterns using collocation matrices, as well as by analyzing the LQs of 218 industries, it was determined how much the same pair of industries tend to cluster in different regions; if industries tend to concentrate in more than 50% of time, then a tendency to cluster in the same geographical area was observed;

3) The identification of regions where clusters tended to cluster; for this employment data from the Canada 2001 Census was used, and three criteria were defined for size (more than 1000 employed), specialization (the LQ values of industries included in the cluster within the region is 1), and scale (the cluster is defined according the rule that more than half of the 4-digit economic activities, of which the LQ is 1, are included) (Spenser et al., 2010). A comparison of the approaches described above is presented in Table 1.

Table 1. Comparison of clusters‘ identification approaches.

Charac-teristics	**Delgado, Porter, Stern's approach* (USA)**	**Brachert, Titze, Kubis's approach** (Germany)**	**Spenser's et al.*** approach (Canada)**
Geogra-phical level	Economic Areas (*n* = 172)	NUTS 3 (*n* = 430)	Census data (*n* = 140)
Econo-mic activity level	2007 North American Industrial Classification System (NAICS), (6-digits level, 675 economic activities)	2003 Germany input-output database (71 industries)	1990 Standard Industrial Classification System (SIC, 4 digits level, 879 economic activities)
Cluster identi-	5-step approach of exploiting	3-step multidimensio-	4-step approach of

Charac-teristics	Delgado, Porter, Stern's approach* (USA)	Brachert, Titze, Kubis's approach** (Germany)	Spenser's et al.*** approach (Canada)
fication ap-proach	similarity matrixes, the coaglomeration index, input-output linkages, the analysis of employment by economic activities, clustering function, the clusters' within links indicator, the clusters' between links indicator	nal approach of exploiting Sternberg and Litzenberger's clusterization index, *Gi* statistics, and the qualitative input-output matrix	exploiting LQ, collocation matrix, specialization, & agglomeration criteria

Source: compiled by the author, * Delgado et al., 2013; ** Brachert et al., 2011; *** Spenser et al., 2010.

In summarizing the overview of cluster determination methods, it can be concluded that cluster research is based on concentration, spatial distribution, and linkages of research methods. There is no consensus on which cluster identification approach is the best. It could be assumed that although clusters identified in the US are used as the basis in other geographic areas, the US economy is unique, and other countries and regions do not necessarily have the same structure of clusters.

Approach to Evaluating the Impact of Clusters on the Economy

The impact of clusters and their composition on regional economies was proved in developed countries (the US - Porter, 2003; Delgado et al., 2008; Germany - Brachert et al., 2011; Europe - *DG Enterprise and Industry Report on Innovation Clusters in Europe*, 2010). One of the most important factors in exploring the impact of clusters on the economic development of a region are labor force parameters. Clusters' effects on the economy are measured by the concentration of a labor force in a particular economic activity in a particular region as well as wage rates in specific economic activities in specific regions (Porter 2003; Delgado et al., 2008; Brachert et al., 2011). The wage rate indicator is treated as an element of gross value added.

In economic geography, the spatial distribution of economic activities was explored using Krugman's concept of new economic geography (Arauzo-Carod et al., 2010). The most comprehensive studies on the impact of clusters on the economy and innovation were carried out by US researchers Porter (2003), Delgado et al. (2010, 2013). Ch. Ketels (2007), M. P. Feldmann (2010), J. Koo, K.-R. Cho (2011), J. Leibovitz (2004), R. Teigland, and G. Lindqvist (2005). Researchers have exploited descriptive statistics, linear or log-log regression analyses, LQs, and the *Gini* index (Porter, 2003; Delgado et al., 2010; Spencer et al., 2010; Ketels and Protsiv, 2010, 2016) to explore the impact of clusters' on the economy. Using Porter's cluster definition, the links between specific types of narrow economic activities and their impact on regional employment and economic performance were analyzed (Porter, 2003; Delgado et al., 2011; Kubis et al., 2010, p. 217; Brachert et al., 2011; Pires et al., 2013). Research on the economic impact of different clusters varies across geographic areas. While strong clusters in the US are found to have an impact on the economy regardless of cluster specialization (Porter, 2003), studies in other countries show that specific clusters in a given location are a specific feature of that location (Hausman et al., 2012; Lin, 2011; Ketels, 2013; Antonioli et al., 2015).

Sölvell et al. were the first to use Porter's cluster definition to identify clusters in the Baltic States by analyzing patterns of economic activity concentration in the new EU member states – Cyprus, Slovenia, Malta, Hungary, the Czech Republic, Slovakia, Poland, Estonia, Latvia, and Lithuania (Sölvell et al., 2008). The main finding of their study is that regional concentration in these countries is significantly lower than that in the US and slightly lower than in the old EU member states (Sölvell et al., 2008). The correlation of employment in economic activities across geographies has been a key method for identifying clusters and cluster category (Porter, 2003; Sölvell et al., 2008). Ketels and Protsiv conducted an EU-wide study to investigate whether the existence of clusters can lead to higher welfare aspirations using the European concept of a new development path policy (Ketels and Protsiv, 2013). The results of this study confirmed the results of Porter's empirical study, i.e., that the presence of strong clusters (a factor of agglomeration of different economic activities) has a positive and significant impact on average wages in a particular region. When comparing the cluster studies in the Baltic States (Sölvell et al., 2008; Ketels and Protsiv, 2014; Ketels and Protsiv, 2016), it is important to note that the research objectives and the methodology of cluster identification were different in all three cases. In addition, the adapted cluster definition proposed by Delgado et al. (2013) may not have been suitable for small countries, whose economies are not strong.

In conclusion, the main researches on the economic impact of clusters in Europe, including the Baltic countries, were conducted using the concept of regional competitiveness. The main factors in these studies are labor force parameters. In these studies, the Baltic States were analyzed from a country perspective and the assumption of a single region was not applied.

Approach to Exploring the Impact of Clusters on Innovation

There are several trends in approaches used to explore clusters' impact on innovation:

- when the main object of research are cluster companies, and in this case the results of business development or innovation activities as well as the factors that influenced those results were explored;

- when the main object are clusters and their impact on innovation in a given geographical region;
- studies focusing on cluster or other business development policy support measures, which affect the innovative performance of a particular region.

An analysis of the more recent studies (2012–2018) on the impact of clusters' on business innovation performance was done to identify the factors that are important for successful business innovation performance (Terstriep and Lüthje, 2018; Kržeij et al., 2016; Braune et al., 2016; Cook et al., 2013; Li and Geng, 2012). The impact of clusters as an organizational structure on enterprise innovation depends on internal cluster resources (human, infrastructure, R&D infrastructure, quality of intra-cluster collaboration, involvement of cluster companies in global value chains – i.e., when multinational companies belong to the cluster) and external factors (market potential, the involvement of international networks, etc.). Although most studies find that clusters, as an organizational structure, have a positive influence on the innovation performance of firms, some studies do not confirm a significant difference between clustered and non-clustered firms (Kržeij et al., 2016). However, the results of the studies are difficult to compare because of the different approaches, the different ways of collecting data, and the different time periods.

The impact of US clusters on the innovative potential of regions is demonstrated by Porter (2003) by exploiting patent number data. However, the USA's high-value-added economy has such sectors as biotechnology or information technology, for which patenting is important. The economies of the Baltic States or Germany, for example, are based on industries for which R&D activities are more important than patents. Segarra-Blasco et al. (2018), while discussing the region as a factor of a specific geographical area that can lead to the agglomeration of manufacturing sectors and / or innovative activities, discussed several models of clusters:

(a) the concentration of a *specific industry* in the region (Marshall-Arrow-Romer, hereinafter referred to as the MAR type),

(b) the Jacobs type, when *different types* of businesses are concentrated in a particular region,

(c) the Porter type, when the concentration of economic activities is a result of competition within the same business sector.

However, in the opinion of the author of this thesis, the Porter type cluster is more closely related to the Jacobs type, because Porter's hypothesis included not only core economic activities but demand and suppy sides as well.

Other studies of the 2012–2019 period explored the impact of clusters on regional innovation (Jia et al., 2015; Anokhin et al., 2019; Belso-Martínez et al., 2017; Brachert et al., 2016; Gallié et al., 2013). Studies on the impact of clusters on regional innovation confirm the importance of agglomeration, labor force and its quality, capital, R&D resources, and R&D infrastructure.

The new economic geography theory assumes the concentration of economic activity as a result of higher productivity. This led to research proving that the spatial dimension of innovation activity has an impact on productivity (Audretsch et al., 2003; Mukim 2012; Martynovich and Lundquist, 2016; Hervas- Oliver et al., 2017). As in studies that analyze the impact of policy measures on clusters, the studies that explore the impact of clusters on innovation are usually case studies at the levels of a country, region, or cluster. The impact of the identified clusters on the economy is assessed by different researchers depending on the theoretical model they use. For example, by comparing cluster's employment dynamic with the dynamic in regional or national employment, or by examining wages in region-specific clusters. However, wages may reflect the effects of supply and demand forces and do not reflect the productivity change, which is important in the long term (Krugman, 1994). The impact of clusters on innovation is examined at different levels (company, economic activity, or industry) and using a variety of models and factors: patents, number of graduates, R&D expenditure, etc. While much of the research finds that clusters have a positive impact on business innovation performance (Porter, 2003; Delgado et al., 2010; Spencer et al., 2010; Ketels and Protsiv, 2013, 2016; Belso-Martizez et al., 2017; He et al., 2015), some studies do not detect a significant difference between cluster and non-cluster firms (Rodriguez-Pose and Comptour, 2012; Krželj et al., 2016). Most of the studies analyzing the impact of clusters on innovation have been conducted in countries with different economic development than the Baltic States. In conclusion, there is a lack of research on the impact of clusters on innovation both at the European level and in the Baltic region. In addition, research on the impact of clusters on innovation is largely based on patenting rates and does not always reflect other factors that may have an impact on innovation.

3. METHODOLOGICAL APPROACH OF EMPIRICAL RESEARCH

The development of the empirical research model took into account the agglomeration forces as one of the conditions for entities clustering in a region. The assumption that the Baltic States can be treated as a single region was made with respect to the EU or with respect to the Baltic Sea region or with respect to Central and Eastern Europe. Thus, the geographical coverage of the study encompasses the Baltic States – Estonia, Latvia, and Lithuania. The empirical research consists of three main parts:

(a) a cluster identification and validation algorithm, which is aimed to group economic activities of the business sector with similar characteristics;

b) research of the impact of clusters on the economy;

c) research of the impact of clusters on innovation.

The Cluster Identification and Validation Algorithm was aimed to explore economic activities of the business sector that tend to cluster in a region and explore whether the region has a unique business sector structure. It has been assumed that the economic activities in a given cluster are linked due to geographical factors.

Approach to Identifying Clusters in the Baltic States

Cluster similarity characteristics have largely reflected the regions of specialization, agglomeration, and collocation in the above discussed cluster identification approaches, which have been applied in the USA and Europe. it is important that. Clusters of USA identified by Delgado et al. (2013) were transferred to the European Union Cluster Monitoring Instrument and rely primarily on employment indicators. The identification of strong clusters is done on the basis of employment, which is essentially the result of agglomeration. The impact of clusters on the economy is measured by higher wages. However, while discussing the wage factor, it can be argued that its size may depend on the ratio of labor supply and demand in a particular region or in a specific economic activity or the overall state of a particular sector and economy in that region. The cluster identification algorithm applied by Canadian researchers relates only to agglomeration in urban areas and was applied only to the economic activities of more than 1000 employed persons. Thus, this could not be applied by the author of this research, as, for example, 75 of the 1333 observations equalled to zero, 150 were less than 100, 461 (35% of all observations) less than 1000, and 331 had no data. The observation in this research is a data of characteristic of one of the economic activities of a business sector in one of the Baltic States in *t* year. As the cluster definition is based on a set of different economic activities within a defined geographical area, the economic activities of the business sector of all Baltic countries were considered together, assuming that it is a single region. The cluster identification method was developed with the aim of identifying groups of related economic activity in the Baltic region, taking into account the possible regional specialization and the collocation of business sector activities.

The first step in identifying clusters is to identify the clusters of business EAs (business economic activities) using the *k*-mean method. Business EAs were grouped by exploiting the similarity of characteristics in the beginning (2008) and in the end (2016) of the empirical research period using the *k*-mean method. Because of the large differences in the data analyzed, all data used for cluster analysis were standardized. When using the *k*-mean method, the expected number of clusters should be set. Because the number and set of observations for each year of the period considered were different, the number of clusters with a *k* value between 20 and 90 was explored. Calinski-Hrabasz's pseudo-*F* index was used to validate the identified number of clusters. Studies by Everit et al. (2011) have identified the Calinski-Haabasz and Duda-Hart indicators as one of the most appropriate indexes for validation (Stata, 2017). The Calinski-Harabasz pseudo-F index, for the number of *g* groups and the number of n observations (in this case, the business EAs), is calculated as follows:

$$\frac{(B)/(g-1)}{(W)/(N-g)}, \qquad (1)$$

where *B* is the sum of the squares sum of the squares and the vector product, and *W* is the sum of the squares of the distances in the cluster and the vector product. High values of the Calinski-Harabasz pseudo-*F* index indicate discrete cluster structures. After the analysis of the *k*-means, the preliminary composition of EAs in clusters was analyzed by employment, labor productivity, and value added factors.

The second step in cluster identification is to compare the results of cluster analyses at the beginning and end of the period and identify the clusters of similar business EAs. If the same group of EAs falls into the cluster at the beginning and end of the period, it is assumed that these groups are similar and treated as a cluster of the business EAs.

Following the assumption that the Baltic States were treated as a single region, the collocation patterns of clusters were examined, i.e., the third step of cluster identification was performed. A nonparametric correlation analysis of the employment of EAs in clusters was done. The Kendal τ coefficient was chosen, eliminating large differences between the values of employment and having in consideration the short observation period, which was less than 10 years. After these three steps, the groups of EAs were considered as clusters. An analysis of the composition of the identified clusters was conducted to determine which type of cluster was identified: MAR, Jacobs, or Porter. Assuming that the identified clusters are heterogeneous, a panel-based approach was selected for further analysis of clusters‘ impact on the economy.

Empirical Research on the Impact of Clusters’ on the Economy

The Cobb-Douglas production function model was first considered. However, cluster analysis and the validation of clusters of economic activity classes revealed that clusters are heterogeneous and that a simple multiple regression approach may not be appropriate. According to Baltagi (2005), the panel model is more appropriate as it takes into account the heterogeneity of individuals (in this case, clusters) and provides "more information, more variability, less collinearity between variables, more degrees of freedom and greater efficiency" (Baltagi, 2005, p. 5). Therefore, panel models were applied to explore the possible impact of clusters on the economy. Following the production function framework, it is assumed that value added is a dependent variable and that independent variables are factors of the labor force – i.e., the number of persons employed and labor productivity. Since the capital factor is not used in this research, a linear expression of the panel model was tested first:

$$Y_{it} = \alpha_i + \beta_1 L_{it} + \beta_2 P_{it} + u_{it}, i = 1, \dots, N;\ t = 1, \dots, T, \quad (2)$$

where Y_{it} is the value added of cluster *i* (EUR million) in year *t*, L_{it} is the number of full time equivalents of employment of cluster *i* in that year, P_{it} is the labor productivity of cluster *i* in year *t*, u_{it} is the error term. For the selection of a panel model, the analysis was performed in the following order:

- application of a constant coefficients model (CCM);
- application of a fixed effects model (FE);
- application of a random effect model (RE);
- application of the Breusch-Pagan Lagrange test;
- application of the Hausman test to select either a fixed or random effects model.

The CCM model ignores the fact of panel data and is the most restricted cross-sectional data model. Individual specific effects models are the fixed effects model (FE) and the random effects model (RE). The assumption is that there is heterogeneity between individuals or groups, which is represented by the α_i parameter. The question is whether individual effects correlate with independent variables. The fixed effects model is applied if correlation exists. In case of no correlation, the RE is applied. In a fixed effects model, individual-specific effects of α_i may correlate with an independent variable; α_i is included as a free member in the equation. Each individual will have a different α_i and the same slope factor.

Approach to Exploring the Impact of Clusters on Innovation

After exploring the methods used to evaluate clusters' impact on innovation and due to the limitations of accessing detailed data on innovation such as R&D expenditure by four-digit NACE or employment data at this level, the assumption that labor productivity could have an impact on total business R&D expenditure was considered. The assumption was considered since OECD experts (McGowan et al., 2015) associate productivity with corporate innovation activity. Thus, the following regression model for the effect on innovation was constructed:

$$y_{berd\ t} = \beta_0 + \beta_1 DN_{aver\ cls\ gr\ it} + \varepsilon, \quad (3)$$

where $y_{berd\ t}$ is the total R&D expenditure of the business sector in year t (EUR million) and $DN_{aver\ cls\ gr\ it}$ is the average productivity of the clusters' group in percentage in period t. It was assumed that the amount of R&D expenditure may vary depending on the size of labor productivity. Therefore, clusters were grouped by labor productivity into the following groups:

- A clusters' group with $200 < DN_{aver}$;
- A clusters' group with $150 \leq DN_{aver} \leq 200$;
- A clusters' group with $DN_{aver} < 150$.

Variables and Data Sources

Clusters are associated with export growth and productivity gains (Porter, 2003; Delgado et al., 2010). However, to the best of the author's knowledge, no research has been carried out to apply the factor of productivity on cluster identification. An analysis of the theoretical models of regional development and growth proves that clusters are the result of agglomeration forces. Agglomeration is linked with an increase in productivity (Woodward, 2011; Felipe and McCombie, 2012; Hervas-Oliver et al., 2017). Productivity is also related to export growth. According to Ketels (2008), it is important for clusters that reflect regional specialization to be productive. The clusters identified in studies of the US, Germany, and Canada were based mainly on the employment factor (Porter, 2003; Delgado et al., 2013; Spenser et al., 2010; Brachert et al., 2011). Productivity is also associated with innovation and investment in knowledge capital, i.e., R&D, company-specific capabilities, databases, design, and other forms of intellectual property (McGowan et al., 2015). Niţoi and Pochea (2016), who examined productivity dynamics in Central and Eastern European countries during the period of 1995–2014, allocated the Baltic States to the same productivity's dynamicgroup. This fact was considered to have proven the assumption that the Baltic States are a single region. McGowan et al. (2015) and Niţoi and Pochea (2016) argued that not only the total number of employees but also the number of hours worked is important for the productivity factor. Niţoi and Pochea (2016) mainly used an indicator of the ratio between real value added and total hours worked in their research on productivity growth and convergence. In addition to the labor force variable, i.e., the number of persons employed, several productivity variables (labor productivity and value added per FTE) were selected for cluster identification (Table 2).

Table 2. Variables for cluster identification in the Baltic States

Variable	**Definition**
Employment	Persons employed – number
Value Added (VA)	Value added at factor cost – million euro
Full time equivalent (FTE)	Employees in full time equivalent units – number
Labor productivity	Wage-adjusted labor productivity (Apparent labor productivity by average personnel costs) – percentage

Source: compiled by the author.

The employment factor may reflect the agglomeration of economic activities or specialization in the region, while FTE and labor productivity variables are related to the productivity of economic activity. The Eurostat Structural Business Statistics include sectors of industry, construction, trade, and services (from B to N and S95).

Research limitations and data availability. There was a lack of data for all economic activities of the business sector in all Baltic countries (Estonia, Latvia, Lithuania) for the selected period (2008–2016). The EAs that have data for at least half of the period were selected for cluster analysis. Data on EAs the mean of annual employment of which were less than 100 persons during the period of analysis were also dropped. The period was chosen as Nace Rev. 2 was introduced in 2008. The cluster analysis was done with different sets of EAs. For the cluster analysis, nominal value added data were exploited because the real value added data are not available at such a detailed level. In the absence of production volume data at constant prices, the focus of research was on the regional-sectoral dimension (250-300 business classes × 3 countries) rather than on time (9 points). Also, cluster analysis relied on employment and productivity characteristics rather than similar dynamics when the use of real data has a significant impact. The value added was not deflated in the panel models under the NACE Rev. 2 price index, since it could be assumed that prices in the two-digit category reflect the general trend of the industry, and in some cases (for example, the Construction Sector) even the trend of the whole section. Also, due to a lack of data on innovation at such a detailed level, this study was led to a general model of clusters' impact on innovations and was based on the productivity factor associated with total business R&D expenditures.

Thesis Statements:

1. Business sectors' economic activities of medium and high productivity cluster in the Baltic States (Estonia, Latvia, Lithuania).

2. The suggested cluster identification algorithm can use both employment and productivity indicators.

3. The lack of strong cluster structures in the Baltic region could limit the region's economic development and innovation activities.

4. MAIN FINDINGS OF EMPIRICAL RESEARCH

Results of Cluster Analysis

The first step of cluster analysis was done by applying the *k*-mean method for business sector EAs at the beginning (2008) and at the end (2016) of the period. The pseudo-*F*index was used to select the number of clusters of *k*-mean method. The higher the pseudo-*F* index value, the more appropriate is the number of clusters. Results were obtained using STATA15‘ software. The highest concentration of pseudo-*F* values was in the interval of

$$20 \leq k \leq 24.$$

Thus, the highest value was for a number of identified clusters, i.e., 24. The results of the *k*-mean cluster analysis were analyzed by exploiting the variables of employment, labor productivity, and value added. Additionaly, the composition of clusters with regard to sectors & distribution across countries was analyzed. The second step of analysis was done by comparing the groups of EAs in 2008 and 2016 clusters. If a group of EAs stays together in clusters of 2008 and 2016, it is assumed to have similar characteristics and thus could be a cluster. To confirm the collocation pattern, a correlation analysis of EAs' groups was performed. A group of EAs in which EAs are correlated with at least of the half of the EAs by $\tau \geq 0.3$ in the group was confirmed to have a collocation pattern by employment.

The medium and high labor productivity clusters, i.e., $150 \leq DN_{Aver}$ were analyzed further. The change of clusters' productivity and in the number of EAs is presented in Table 3.

Table 3. High and medium clusters change in productivity and the number of EAs (lp – low productivity, mp – medium productivity, hp – high productivity).

Clusters	**Productivity change**		**Quantity of economic activities (EA) in cluster**	**Number of correlated EAs in cluster**
1_19	lp>mp	↑	12	10
1_23	lp>mp	↑	10	3
2_13	mp>mp	=	3	1
2_17	mp>lp	↓	4	3
2_19	mp>lp	↓	3	3
2_20	mp>mp	=	4	3
2_23	mp>mp	=	12	6
4_4	hp>hp	=	4	4
4_7	hp>lp	↓	6	3
4_13	hp>mp	↓	9	2
4_20	hp>mp	↓	12	8
4_23	hp>mp	↓	5	5
5_13	lp>mp	↑	11	3
7_7	hp>lp	↓	12	4
7_13	hp>mp	↓	5	3
7_15	hp>lp	↓	7	5
9_12	hp>mp	↓	5	5
9_14	hp>lp	↓	3	2
9_16	hp>mp	↓	4	1
9_22	hp>hp	=	3	2
10_13	lp>mp	↑	6	4
12_20	lp>mp	↑	3	2
14_23	lp>mp	↑	14	8
15_4	mp>hp	↑	7	7
15_7	mp>lp	↓	15	2
15_9	mp>lp	↓	5	4

Clusters	**Productivity change**		**Quantity of economic activities (EA) in cluster**	**Number of correlated EAs in cluster**
15_12	mp>mp	=	5	1
15_13	mp>mp	=	21	11
15_14	mp>lp	↓	4	4
15_15	mp>lp	↓	18	13
16_10	hp>hp	=	3	2
16_22	hp>hp	=	4	3
19_8	mp>lp	↓	3	2
19_6	mp>lp	↓	8	7
20_4	hp>hp	=	5	
20_12	hp>mp	↓	7	
20_13	hp>mp	↓	3	2
20_16	hp>mp	↓	8	8
21_12	mp>mp	=	6	6
21_14	mp>lp	↓	7	3
21_16	mp>mp	=	5	4
21_20	mp>mp	=	7	7
23_4	lp>hp	↑	5	
23_13	lp>mp	↑	4	4
23_20	lp>mp	↑	3	2
24_20	lp>mp	↑	5	
24_23	lp>mp	↑	3	3

Source: Compiled by the author.

Main cluster identification results:

(a) five high-productivity and four average-productivity clusters out of the 47 examined remained within the same productivity group at the beginning and the end of the period;

b) four high-productivity clusters fell into the lp clusters group and eight into the mp group;

c) Nine from the lp groups change the productivity to the mp level, one from the mp group changes productivity to the hp group;

d) not all identified clusters are linked through input-output linkages – for example, cluster “4_4” consists of E3812lv – collection of hazardous waste, G4633lt - Wholesale of dairyproducts, eggs and edible oils and fats, G4638lt - Wholesale of other food, including fish, crustaceans and molluscs, G4672lt - Wholesale of metals and metal ores while cluster „9_22“ has two EAs „G4671ee - Wholesale of solid, liquid and gaseous fuels and related products“ ir „G4675lt - Wholesale of chemical products“,

e) The identified high productivity clusters do not reflect the Baltic clusters declared in the European Cluster Observatory; Lithuanian EAs dominated in clusters.

Analysis of the Composition of Clusters of Medium and High Productivity

The highest productivity clusters the average productivity of which did not change during the period are “4_4,” “16_10,” and “16_22.” Cluster “4_4” included economic activities from different sectors, i.e., E3812lv, G4633lt, G4638lt, G4672lt. Three of the business classes included in the cluster belong to the G sector from Lithuania, one of the economic activities belong to the Estonian E sector. Although the cluster remained in the same productivity cluster at the beginning of the period under review, it can can be classified as a Jacobs-type cluster. Cluster “16_10” consisted of two business classes from Lithuania – D3513lt and H5222lt – and was also classified as a Jacobs-type cluster. Cluster 16_22 included F4110lt, G4621lt, N7739lv, and only two of those could be related through an input-output link and both are from different countries. Because labor productivity did not decrease in these three clusters, it can be concluded that they are only locally related. The medium labor productivity clusters are: “2_20,” “2_23,” “15_13,” “2_12,” “21_16,” “21_20.” Two EAs are just the same, one being from Lithuania and the other from Latvia. This may indicate regional specialization in the M7120 EA. One cluster included EAs from only sector C - Manufacturing.

Eighteen of the 40 analyzed clusters can be classified as Porter-type clusters, while others as Jacobs-type. The clusters that can match the Porter cluster type are: “1_23,” “2_19,” “2_20,” “2_23,” “4_20,” “7_13,” “7_15,” “9_14,” “9_22,” “12_20,” “15_7,” “15_9,” “15_15,” “19_6,” “21_14,” “21_16,” “21_20,” “23_20.” In order to confirm that these clusters belong to the Porter type, additional input-output research could be conducted. In some clusters, the same EA from different countries came together, which could mean a specialization of the region. Such activities include the M7120 - Technical testing and analysis, H494 - Freight transport by road1, I5610 - Restaurants and mobile food service activities, H5223 - Service activities incidental to air transportation. However, this could be seen not as a regional cluster specialization but as a specialization in specific EAs. These activities belonged to the cluster of medium to low labor productivity. It is important to mention that most Porter clusters consisted of a small number of business classes, i.e., two to three EAs.

In conclusion, the cluster identification algorithm applied in this research properly grouped EAs by employment and productivity characteristics. Jacobs and Porter type clusters were identified. Most of the identified clusters belonged to the medium and low labor productivity groups.

Impact of Clusters’ on the Economy of the Baltic Region

The analysis of the scatter diagrams of clusters by value added and the employment factors revealed a division between several groups. Therefore, the assumption was made that there could be different mechanisms of impact to value added. Thus, clusters were grouped accordingly by the full time equivalent (FTE) variable:

- 1000 < FTE;
- 1000 ≤ FTE ≤ 5000;
- FTE ≤ 9000.

A panel data analysis was performed following this linear expression:

$$\boldsymbol{Y_{i,t} = \alpha_i + \beta_1 L_{i,t} + \beta_2 P_{i,t} + u_{i,t}}, \qquad (4)$$

where $Y_{i,t}$ – value added, $L_{i,t}$ – number of positions in cluster i in year t, $P_{i,t}$ – productivity in cluster i in year t, i – number of clusters, t – year of the reference period. Only two groups were explored using the panel models, i.e., “1000 < FTE” and “1000 ≤ FTE ≤ 5000”. The group of FTE ≤ 9000 included only two clusters. Thus, there was no possibility to apply panel models. Three panel models were tested with each of the clusters’ group: constant coeficients model (CCM), fixed effects model (FE), and random effects model (RE). The constant coeficients model is presented hereafter:

$$y_{it} = \alpha + \beta_1 L_{it} + \beta_2 P_{it} + (\alpha_i - \alpha + e_{it}) \qquad (5)$$

The CCM ignore the panel data model. If this model is a true model, then the independent variables do not correlate with the error values. As the CCM ignores the potential heterogeneity of clusters, a graphical analysis was performed to check for heterogeneity (Fig. 1).

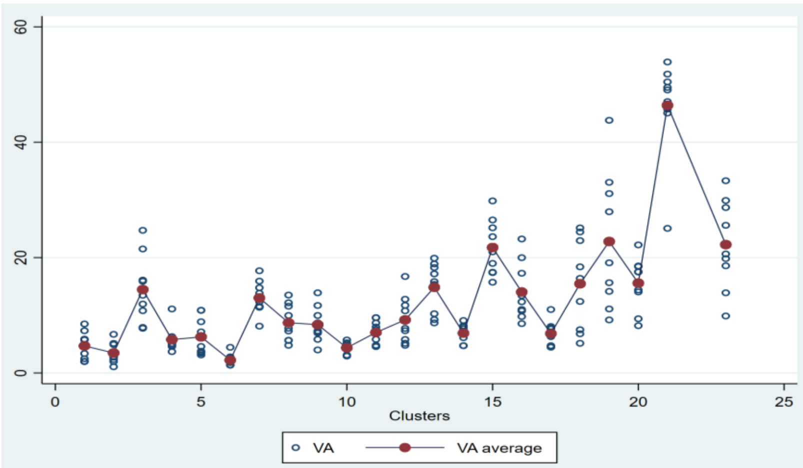

Figure 1. The heterogeneity of cluster groups’ “1000 < FTE” clusters by value added.

Source: compiled by the author.

Thus, fixed-effects and random-effects models were applied. Fixed effects model:

$$y_{it} = \alpha + \beta_1 L_{it} + \beta_2 P_{it} + v_i + \epsilon_{it}. \quad (6)$$

In FE, the least squares method is applied for the average of a dependant variable based on the average values of the independent variables over the period, i.e.:

$$y_{it} - \overline{y_t} = (L_{it} - \overline{L_i})'^{\beta_1} + (P_{it} - \overline{P_i})'^{\beta_2} + (e_{it} - \overline{e_i}). \quad (7)$$

In this model, the number of observations equals NT and the individual effect α_i is removed, since only the individual effect mean α remains. Hypothesis: $H_0: \alpha_1 = \alpha_2 = \cdots \alpha_{22}$, H$H_A: at\ least\ one\ \alpha_s \neq \alpha_j$, F test statistic is applied. The F-statistic for the model as a whole and the independent variables is statistically significant (p> F = 0.0000). Hypothesis for t statistic: $H_0: \beta_1 \neq 0, \beta_2 \neq 0$; $H_A: \beta_1 = 0, \beta_2 = 0$. H_0 is rejected because p <0.05. The correlation between the errors and the independent variables is weak (0.1832), the high value of the Rho coefficient ρ (0.8082) demonstrates that 80.8% of the dispertion could be explained by the heterogeneity of the panel model groups (Table 4). The Rho factor is calculated using the formula:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}, \quad (8)$$

where σ_u is the standard deviation of the panel model, and σ_e is the standard deviation of ϵ_{it}.

Table 4. Results of panel models for the clusters' group "1000 < FTE."

	CCM	**FE**	**RE**
Number of observations (number of clusters)	198 (22)	198 (22)	198 (22)
F(2,195) p>F	-	161,49 0,000	-
Wald χ^2 (2)	362,74	-	357,56

	CCM	FE	RE
Prob > χ^2	0,0000		0,0000
R^2	-	0,6359	0,6360
within	-	0,6499	0,6499
between	-	0,6370	0,6371
α	-11,86	-11,48	-11,84
β_1	0,0258	0,0254	0,0258
β_2	0,0583	0,0572	0,0583
σ_u	-	6,0184	6,0566
σ_e	-	2,9320	2,9320
ρ	-	0,8082	0,8101

Source: compiled by the author.

Overall all models (CCM, FE, RE) appear fine and indicate the relationship between variables. An overall R^2 (0.6359) indicates that the model will explain 63.6% of the change of the dependent variable. The total value of R^2 is close to the value of R^2 between groups (0.6370), whereas the value of R^2 within the group is slightly higher (Table 4). For further analysis, a random effect model was constructed with the following general formula:

$$y_{it} = \alpha + \beta_1 L_{it} + \beta_2 P_{it} + v_{it} + \varepsilon_{it}, \qquad (9)$$

where y_{it} is the value added of cluster group *i* in year *t*, L_{it} is the FTE of cluster group *i* in year *t*, $P_{it}$is the productivity of cluster group *i* in year *t*, $\beta_1$and β_2 are coefficients of independent variables, $v_{it}$is the error between cluster groups, $\varepsilon_{it}$is the error within the cluster group. If the random effect method is used, the assumption is made that errors are not correlated with independent variables, in which case time-independent variables can be included in the model and used to predict the independent variable. In the considered random effects model, the individual effect of groups α_i is within error. Comparing the values of R^2 in the fixed and random effects models suggests that the values of R^2 within the group, between groups, and overall values are almost the same. Within the R^2 group, the same values as in the fixed effect model and statistically significant (p> F, p> t), Rho coefficient, coefficients for independent variables were slightly higher. Differences between the clusters do not correlate with independent variables (corr (u_i, X) = 0.0000). The random effect model tests the assumption that cluster groups are not similar and their differences, are random, and vary during the 2008–2016 period. The coefficients for this model were calculated using the generalized least squares method (GLS). To test which model is the most appropriate for the "1000 < FTE" clusters' group, the Breusch-Pagan Lagrange (BPL) and Hausman tests were exploited. The BPL test is designed to test, for the random effect model, whether σ_u^2, i.e., $cor(u_{it}u_{is})$ is significantly different from 0. If the test is statistically significant, the random effect method is used. The BPL test follows the equation:

$$PV_{klasteris,t} = Xb + u_{klasteris} + e_{klasteris,t} \tag{10}$$

where $PV_{klasteris,t}$ is the value added of the i cluster, Xb – independent variables, $u_{klasteris}$ is the error between the cluster groups, $e_{klasteris,t}$ is the error within the cluster. The BPL test uses the least squares method and ir χ^2 statistics. As $p > \chi^2$, BPL is statistically significant. This confirms that the RE model could be more efficient. To choose between FE and RE models, the Hausman test was performed by applying

$$H = (\widehat{\beta_{RE}} - \widehat{\beta_{FE}})'(V(\widehat{\beta_{RE}}) - V(\widehat{\beta_{FE}}))(\widehat{\beta_{RE}} - \widehat{\beta_{FE}}), \tag{11}$$

where V is the matrix of the covariates of dispersion.

If $\widehat{\beta_{RE}} - \widehat{\beta_{FE}}$ are close, then the difference will be close to zero; if significantly different, the value will far from zero. The Hausman test hypothesis:

H_0: coefficient differences are not systematic,

H_A: systematic differences in coefficients:

$$\chi^2 = (b_{FE} - b_{RE})'[(V_{FE} - V_{RE})^\wedge(-1)](b_{FE} - b_{RE}) \tag{12}$$

Since χ^2 estimated (0.5945) greater than 0.05, the Hausman test is statistically insignificant and the H_0 hypothesis is rejected, i.e., the differences in the coefficients are systematic. It could be stated that the RE model for the clusters' group "1000 < FTE" is more appropriate.

All three models (CCM, FE, RE) were applied to the clusters' group "1000 ≤ FTE ≤ 5000." Results of the models are presented in Table 5. All models are fine overall; however, after a graphical analysis (Fig. 2) of the clusters' group of "1000 ≤ FTE ≤ 5000", heterogeneity was confirmed.

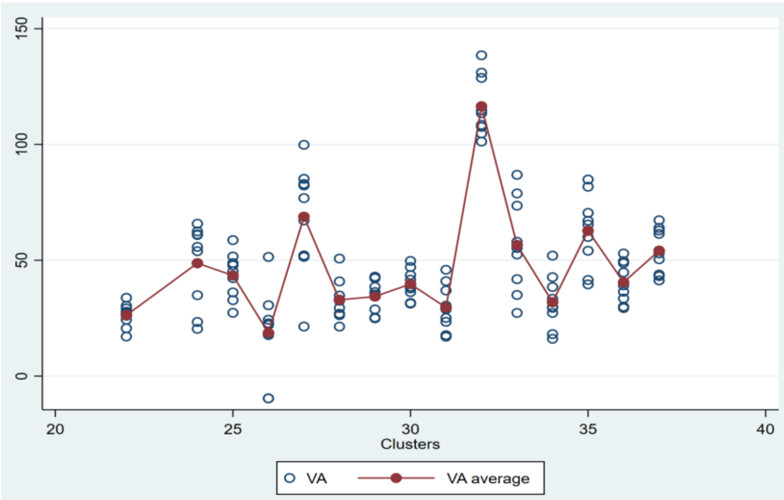

Fig 2. The heterogeneity of the clusters' groups "1000 ≤ FTE ≤ 5000" clusters by value added.
Source: compiled by the author.

The fixed effect model hypothesis: H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_{22}$, H_A: at least one $\alpha_s \neq \alpha_j$, F test statistic applied. The F-statistic for the model as a whole and the independent variables is statistically significant (p> F = 0.0000). The correlation between the errors and the independent variables is weak (-0.1756), and the high value of the ρ coefficient (0.8029) confirms the heterogeneity of the clusters and the validity of the panel model analysis. The highest overall value of R^2 is within the group (0.6231), and there is a slight difference between the groups and the overall R^2. A comparison of the values of R^2 in the fixed and random effects models suggests that the values of R^2 within the group, between groups, and overall values are nearly the same. After the Hausman test (H_0: coefficient differences non-systemic, H_A: coefficient differences systemic), $H_0$was confirmed (p>χ^2). Thus, the FE model was deemed more appropriate is for this clusters' group.

Table 5. Results of panel models for the clusters' group "1000 ≤ FTE ≤ 5000."

	CCM	**FE**	**RE**
Number of observations (number of clusters)	135 (15)	135 (15)	135 (15)
F(2,195) p>F	-	97,54 0,0000	-
Wald χ^2 (2) Prob > χ^2	214,36 0,0000	-	208,33 0,0000
R^2	-	0,5353	0,5775
within	-	0,6231	0,6191
between	-	0,5123	0,5660
α	-32,58	-34,41	-32,21
β_1	0,0253	0,0269	0,0249
β_2	0,1527	0,1452	0,1544
σ_u	-	16,95	13,49
σ_e	-	8,3983	8,3983
ρ	-	0,8029	0,7205

Source: compiled by the author.

The cluster identification and validation algorithm in the Baltic region was validated by exploring cluster impact on the economy for the value added variable and by exploiting panel models that confirmed the heterogeneity of the identified clusters. The analysis confirmed that the RE model is more apropriate for the clusters' group

“1000 < FTE,” although it is not clear what the reasons are in this case. The clusters’ group “1000 < FTE” includes 22 clusters, only four of which had no change in labor productivity. Based on the overall slope of the independent variables, it can be argued that the change in labor productivity in this group may have a greater impact on the change in value added than the change in the FTE.

The change of labor productivity in clusters’ group of “1000 ≤ FTE ≤ 5000” has a bigger impact on value added, which is even greater than in the previous group (the slope coefficient in this group is 0.15 as compared to 0.058). Meanwhile, the impact of the FTE is very similar (0.025 and 0.026). In the group of clusters “1000 ≤ FTE ≤ 5000” there were 15 clusters, six of which did not change labor productivity during the analyzed period, and 7 of which were assumed to be Porter-type clusters. The other clusters’ group of “1000 <FTE” had 9 Porter-type clusters, and the productivity of the eleven clusters decreased. In general, it could be stated that EAs of the business sector cluster in the Baltic States. The impact of clusters on the economy was proved, and the labor productivity has bigger effect on the economy. The impact of labor productivity in the clusters’ group of higher labor productivity could indicate stronger cluster structures in this group.

Impact of Clusters on Innovation in the Baltic States

First, to check the possible function of regression, an analysis of the scatter diagrams by cluster groups “Annual average R&D expenditures of the Baltic business sector and the average annual productivity of cluster groups” was performed. After this analysis, an assumption was made that linear dependence could be tested. The results of the analysis are presented in Table 6. The results of the analysis confirmed the assumption that two clusters’ groups (“200 <DN_{Avrg}“ and “DN_{Avrg} <150”) has linear relationship between the business sector’s annual R&D expenditures and labor productivity. The assumptions of normality, heteroskedacity and autocorrelation were tested and held for two clusters' groups ("200 <DN_{Avrg}" and "DN_{Avrg} <150").

Table 6. Results of the regression analysis in different clusters' productivity groups.

	200 < DN_{Avrg}	**150 ≤ DN_{Avrg} ≤200**	**DN_{Avrg} < 150**
R^2_{Adj}	0,41	0,23	0,61
β_0	-45,63	-79,54	-30,99
β_1	0,49	1,25	0,90
p (0,05) >F	0,0380	0,1151	0,0079
p (0,05) >t, β_0	0,4190	0,426	0,3790
p (0,05) >t, β_1	0,0380	0,115	0,008

Source: compiled by the author.

The two-equation result from regression analysis:

$$\hat{y}_{MTEP150} = -30{,}99 + 0{,}9DN_{150}.$$
$$\hat{y}_{MTEP200} = -45{,}63 + 0{,}49DN_{200}.$$

According to the first, if a clusters' group's labor productivity will increase by 1%, the business sector's R&D expenditures will increase by 0.9 million EUR. According to the second, a labor productivity increased by 1% will result the business sector R&D expenditures increaseby 0.49 million EUR. The interesting fact is that the change of labor productivity of the cluster group with lower labor productivity has a greater impact on the business sector's R&D expenditures. According to McGowan et al. (2015), both capital and labor resources have an impact on productivity as well as innovation factors. If a higher productivity indicates a higher quality of the workforce in the cluster or the uptake of new technologies, this could explain the differences in results between higher and lower productivity clusters. Other assumption could be made that the clusters' group with medium labor productivity could have a different mechanism of impact on

R&D expenditures. Although the models applied do not explain the causal link between labor productivity and R&D expenditure, it clearly indicates that there is a relationship between business R&D expenditures and labor productivity. The different process of the impact in different labor productivity clusters' groups could indicate different the labor quality in these groups, which was the main assumption in the model. The clusters' group with "DN_{Avrg} <150" included 10 clusters 4 of which belong to Porter type clusters. There were 8 clusters in the higher productivity group with only two of the Porter type. Although no further research has been carried out, it is possible that Porter-type clusters, which are related not only to the geographical factor but also through the potential input-output links, have a greater potential for innovation. However, this group of clusters also had the lowest productivity. In conclusion, this study confirmed the link between productivity and innovation activity, which was revealed by other researchers.

5. CONCLUSIONS

The analyzed theoretical models and conducted research by previous authors confirmed that clusters are a relevant scientific and practical topic in the context of EU competitiveness and industrial policy-making. The following main conclusions are made:

1. The analysis of different theoretical approaches related to the cluster concept confirmed:

a. The clustering of economic activities in the region is associated with the phenomena of agglomeration. Regional and location development and growth theories treat clusters as the concentration economic activities in the region. Agglomeration is treated as an internal factor for the region and an external factor for the enterprise.

b. Regional development and growth theories tend to explore the patterns of labor force distribution, from uniform dispersion to full agglomeration.

c. The uncertainty of determination of the region is obvious in the literature. This uncertainty is not usually addressed because data are collected according established administrative units. Marshall and the New Economic Geography representatives treat clusters as a phenomena of the agglomeration of economic activities in the region. The distinctive feature of Porter’s cluster definition is that it assumes the diversity of economic activities in a cluster, i.e., that clusters could reflect not only core industry activity but also supply and demand sides. Porter has linked the economic activities of the cluster to both geographical coverage and input-output relationships. Some researchers argue that the input-output relationship could be important in the context of low labor mobility (the European case).

d. While theories of regional economic development and growth tend to explain the phenomena of agglomeration in the regions, clustering studies aim to identify patterns of uneven production systems and resource distribution in the regions.

e. Different theoretical assumptions of the cluster concept in cluster research determine the uncertainties of the cluster concept in terms of geographical coverage, linkage of economic activities, and composition.

f. Porter's definition of clusters was applied in the European Union. However, despite that, the definition itself was not aimed to reflect a narrow specialization, the European map of clusters is basically based on indicators reflecting national specialization at the EU level. Also, there is a drawback of such application of clusters identified in the USA, as the US economy is one of the largest, more integrated, and has a unique structure. This implies ignoring the unique business structure of European countries and regions and raises doubts about the relevance of USA's clusters on EU scale.

g. The analysis of the limited cluster studies conducted in the Baltic States reveals that the conducted research is mostly related to sector-specific (manufacturing, information technology) research in a particular Baltic country, and that the research itself is more related to the concept of competitiveness. According to the best of the Thesis author's knowledge, there is no research conducted with an aim to identify clusters in the Baltic States, nor the impact of these clusters on the economy and innovation. Thus, there is a lack of research on clustering and the economic impact of clusters in the Baltic region.

2. The analysis of scientific publications on the impact of clusters on the regional economy revealed that research was carried out at different levels and in different geographical areas: from the enterprise level to cluster and industries or sectors, and from city to region and country. Cluster research in large countries (such as the USA, Germany, France, or Spain) confirmed the assumptions that clusters are associated with agglomeration and specialization. The most commonly used independent variable for assessing the impact of clusters is the annual average wage, which may reflect the interaction between demand and supply forces at the regional and national level, but not necessarily the impact of clusters on regional development.

3. Results of the analysis of the research on the impact of clusters on innovation are difficult to compare due to the diversity of levels of the analyzed objects and the application of different models, different data collection methods, and other factors. However, the analysis revealed a mostly positive effect of clusters on innovation at the regional level. There is also a lack of research on the impact of clusters on innovation in the Baltic States.

4. The analysis of clusters identification methods revealed that:

a. Depending on the level of detail and quality of the data and the purposes of the research, various methods or combinations of cluster identification methods are exploited. The main variables for identifying clusters are employment and number of companies. The analysis of identification methods revealed that an input-output relationship analysis did not qualitatively identify the clusters and did not link them to the geographical area and was not therefore considered. Even if clusters are associated with export and productivity growth, productivity variables are not used in cluster identification.

b. Clustering methods employ methods of identification related to the agglomeration phenomenon, regional specialization, and collocation.

5. The analysis of the impact of clusters on economy and innovation revealed that:

a. Research on the impact of clusters usually explores labor and wage developments, which may not necessarily be due to clustering.

b. Research on the impact of clusters on innovation is mainly based on exploiting the knowledge production function, and the results do not always show positive effects, especially when analyzed at the enterprise level.

6. The identification of clusters in the Baltic States (Estonia, Latvia, Lithuania) revealed that:

a. The clustering algorithm exploited in the Thesis can be adapted to identify clusters of different geographic coverage if a single region assumption is possible to apply.

b. It can be concluded that the Porter and Jacobs type clusters were distributed almost equally in the Baltic region, most of which belonged to the group of medium labor productivity. Thus, it could be assumed that the region lacks strong cluster structures.

c. Only few economic activities from the different Baltic States were grouped in the same cluster, suggesting that this may reflect a regional specialization in these business activities.

d. On the cause of why unrelated business EAs dominate the region, an assumption could be made that there are some common factors that have impact on employment and productivity in the region.

7. The research of cluster impact on economy has revealed that:

a. The mechanisms of cluster impact on value added differ in different employment groups.

b. Changes in labor productivity in higher employment clusters may lead to greater changes in value added compared to lower employment clusters.

c. The insufficient economic development of the region can be explained by the dominance of a weak structure of medium labor productivity clusters in the region.

8. The research of cluster impact on innovation revealed that:

a. There may be other than labor productivity factors that could have impact on innovation in the region, especially in the cluster group of medium labor productivity.

b. The assumption that labor productivity and R&D expenditure are positively related was confirmed. The change of labor productivity in the lower cluster productivity group had greater impact on R&D expenditures. This may be explained by the fact that the higher labor productivity cluster group has already undergone a major change in labor productivity due to the higher quality of its labor force, and therefore the impact on R&D expenditure may be less significant.

c. Although it can be argued that cluster labor productivity may be positively related to innovation, the dominance of medium labor productivity clusters in the region may lead to lower innovation performance.

9. The proposed cluster identification algorithm based on productivity and employment was sufficient enough to identify the different types of clusters in the region by employment and productivity; therefore, it can be developed further taking into account other possible characteristics of the different types of clusters.

LIST OF PUBLICATIONS

Rubčinskaitė, R., Kasnauskienė, G. (2017). The Role of Economic Activities Clusters in Gross Value Added Generation in the Baltic State. Proceedings of the International Central Bohemia University Conference "Innovations in Science and Education 2017," http://ojs.journals.cz/index.php/CBUIC/article/view/957

Rubčinskaitė, R., Kasnauskienė, G. (2016). Identification of Clusters Contribution to Economic Growth in the Baltic States (2016). Development in Economics: Theory and Practice. The 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.pdf

Rubčinskaitė, R. (2014). Regional Dimension of Baltic Countries Clusters' Impact on Economy: First Insights. Proceedings of the 4th International Scientific Conference "Practice and Research in the Private and Public Sector 2014" ISSN (online) 2029-7978. http://prpps.mruni.eu/wp-content/uploads/2014/06/PRPPS-2014-PROCEEDINGS.pdf

PERSONAL DETAILS

In 2010, Rima Rubčinskaitė graduated from the Faculty of Economics and Business Administration, Vilnius University with a Master's degree in economics. She started her research career as a junior researcher in the research grant project "Research of Attracting Investments into the Lithuanian Capital Market" (2012–2014) funded by the Research Council of Lithuania. She also worked as an expert in innovation and technology transfer in various projects, such as contract R&D projects "Pilot Application of Higher Education Policy Research Methodology" (2012–2014), "SMEs Consultation for the Preparation of Project Applications of International Programs Services" (2013–2014), as an expert in the project "InoSpurtas" (2018–2021), coordinated by the Agency of Science, Innovation and Technology, and as a researcher in the research grant project "Reassessment of the Optimum Currency Area in the Persistently Heterogeneous European Union" (2018–2022), funded also by the Research Council of Lithuania. She is also a a member of the international association Regional Study Association since 2017. R. Rubčinskaitė attended workshops on networks analysis, panel data analysis, participated in eight international scientific conferences, where she presented her research results on higher education policy, innovation, clusters, agglomeration, and resilience. In 2014, R. Rubčinskaitė was admitted into doctoral studies at the Faculty of Economics and Business Administration, Vilnius University.

Vilniaus universiteto leidykla
Universiteto g. 1, LT-01513 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 15 egz.