DO UKRAINIAN MANUFACTURING FIRMS BENEFIT FROM AGGLOMERATION?

by

Kostiantyn Zahorulko

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2018

Thesis Supervisor: ______________________ Professor Volodymyr Vakhitov

Approved by ____
Head of the KSE Defense Committee, Professor Tymofiy Mylovanov

____

____

____

Date ______________________________

Kyiv School of Economics

Abstract

DO UKRAINIAN MANUFACTURING
FIRMS BENEFIT FROM
AGGLOMERATION?

by Kostiantyn Zahorulko

Thesis Supervisor: Professor Volodymyr Vakhitov

Why do Ukrainian firms make sense to cluster near one another? We test Marshall's theories of industrial agglomeration by examining four manufacturing industries. We constructed intra-industry and potential market`s effects, which consider the difference in location, industry, year and cluster`s size. These effects represent linkages inside each industry separately and the whole manufacturing market together. The agglomeration effects showed significant influence on firm's productivity. So, Ukrainian firms could benefit from these effects by cooperating with the local authority to reduce transfer cost of goods, ideas, and labor. Nevertheless, the overall effect of the increasing distance between firms inside one industry or within all manufacturing firms could be controversial. For the small town or villages, the increase of distance would show that these small location produces for the big regional market and has a benefit from this.

To My Mother

For Your Endless Love, Support, and Encouragement

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION ... 1

CHAPTER 2 LITERATURE REVIEW ... 5

2.1. Production function estimation ... 6

2.2. Empirics of agglomeration economies ... 7

2.3. Clustering essentials ... 9

CHAPTER 3. Methodology ... 11

CHAPTER 4. Data description ... 17

CHAPTER 5. Estimation results ... 20

5.1. Urbanization effect ... 22

5.2. Intra-industry effect ... 24

5.3. Potential market effect ... 28

CHAPTER 6. Conclusion ... 32

WORKS CITED ... 34

APPENDIX A ... 37

APPENDIX B ... 38

APPENDIX C ... 38

APPENDIX D ... 41

APPENDIX E ... 42

APPENDIX F ... 43

APPENDIX G ... 45

LIST OF FIGURES

Figure 1. Histogram of firm`s age. 17

Figure 2. Distribution of the urbanization effect at different clusters. 22

Figure 3. The urbanization effect coefficient at 50, 100, 200km radius 23

Figure 4. The intra-industry effect coefficient at 50, 100, 200km radius 25

Figure 5. The intra-industry effect coefficient at 50, 100, 200km radius 26

Figure 6. The intra-industry effect coefficient at 50, 100, 200km radius 27

Figure 7. The potential market effect coefficient at 50, 100, 200km radius 29

Figure 8. The urbanization effect coefficient at 50, 100, 200km radius 30

Figure 9. The potential market effect coefficient at 50, 100, 200km radius 31

Figure 10. The illustration for DJ (Metallurgical production and production of finished metal products) industry the clusters with the 100km radius 38

Figure 11. The illustration for DL (Manufacture of electric, electronic and optical equipment) industry the clusters with the 100km radius 38

Figure 12. The illustration for DM (Production of vehicles and equipment) industry the clusters with the 100km radius 39

Figure 13. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with the 100km radius 39

Figure 14. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with the 50km radius 40

Figure 15. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with 200km radius 40

LIST OF TABLES

Table 1. The mean variables descriptive statistics 19

Table 2. The mean of the urbanization effect coefficient 23

Table 3. Distribution of the intra-industry effect for DK (Manufacture of machinery and equipment) industry 24

Table 4. The mean of the intra-industry effect coefficient 25

Table 5. The mean of the urbanization effect coefficient 26

Table 6. The mean of the intra-industry effect coefficient 27

Table 7. Distribution of the potential market effect for DK (Manufacture of machinery and equipment) industry 28

Table 8. The mean of the potential market effect coefficient 29

Table 9. The mean of urbanization effect coefficient 30

Table 10. The mean of the potential market effect coefficient 31

Table 11. The codes of the economic activities 37

Table 12. The main variables descriptive statistics in 2001 41

Table 13. The main variables descriptive statistics in 2013 41

Table 14. The input-output table for 2014 42

Table 15. Regression results of TFP on the urbanization effect for DK (Manufacture of machinery and equipment) industry 43

Table 16. Regression results of TFP on the intra-industry effect for DK (Manufacture of machinery and equipment) industry 43

LIST OF TABLES - Continued

Table 17. Regression results of TFP on the intra-industry effect for DK (Manufacture of machinery and equipment) industry 44

Table 18. Summary statistic of Agglomeration effects at DK industry in Kyiv cluster in 50, 100, 200km radius 45

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to my thesis supervisor prof. Volodymyr Vakhitov for his insightful suggestions, guidance, and support, which were extremely helpful.

I am also grateful to all research workshop professors for their thorough reviews and valuable comments.

Special thanks to my friend Polovnikov Ilia for fruitful discussions and readiness to help.

I very appreciate Victoria Strakhova for your support. Your generosity has inspired me to help others and give back to the KSE community.

Finally, I am immensely grateful to my mother for her ceaseless support and patience, and to my girlfriend Liliia for her timely help that allowed me to complete this research.

GLOSSARY

Word

UKRSTAT – State Statistics Service of Ukraine

KOATUU – classification of objects of the administrative-territorial system of Ukraine

KVED – the Statistical Classification of Economic Activities in Ukraine

TFP – total factor productivity

AGNES – Agglomerative Nesting

IOT – the input-output table

C h a p t e r 1

INTRODUCTION

Industries are geographically concentrated. This concentration is too great to be explained by exogenous spatial differences in natural advantage. Why does this concentration occur?

Ellison G., Glaeser, E. L. and Kerr, W. R., 2010

The most important reason for the firm's concentration is cutting the transportation cost. Marshall (1890) separated three different types of the transportation cost – the costs of moving goods, people and ideas. The agglomeration theory is trying to investigate these reasons. This empirical work aimed at three agglomeration effects – urbanization effect, intra-industry effect and input-output linkages effect.

In general, the agglomeration economies are the concept that describes any effects, which lead to firms` and workers` income increasing as a consequence of the growing local economy according to Combes and Gobillon (2015). The most ambitious regional project in the U.S. history, which well-studied example of the growing local economy, is the Tennessee Valley Authority (TVA). The TVA project had been started more than 100 years ago and had all kinds of the agglomeration effects.

The main goal of organizing TVA was to make people live better through agriculture development. Nevertheless, the TVA led to the substantial increase in the agriculture employment, the real outcome was increasing industrial growth. New working dams in that area lead to cheaper electro energy and higher workers` wages. When program's subsidies to farmers ended, people switched from agriculture to manufacturing. This is an evidence of the positive outcomes variety, nevertheless, the main goal of TVA was to develop the agriculture in the specific place.

Ukraine has a great magnitude from all three kinds of enterprises: agriculture, manufacturing and IT. Furthermore, McCann and Folta (2011), Pe'er and Keil (2013), Ferragina and Mazzotta (2015), Basile and Pittiglio (2017) give the evidence of positive influence the agglomeration increasing both the number of startups and firms` lifetime.

Before the effects calculation, we need to make sure that our estimation would be unbiased. For that purpose, we incorporated the differences within industries, regions sizes, spatial locations and years. Ellison G., Glaeser, E. L. (2010) stated that the agglomeration effects depended on the method of shaping the regions, which has some concentration. Some industries concentrate within three regions, other concentrates only at one region. To conclude the differences between industries and firm's locations, we performed the clusterization. The approach used in the paper is the agglomerative clustering, which is the type of hierarchical clustering, known as Agglomerative Nesting (AGNES). To incorporate the differences in location size, as our firms locate at cities, towns, villages or part of cities, we produced the clusterization based on the weighted distances. We chose the population in the location as the weighted factor.

After the regions or clusters defined, we separated them into different subsets and produced calculation on each subset respectively. At this point, it is necessary to describe in more details our three agglomeration effects.

The urbanization effect is the first effect. According to Combes and Gobillon (2015), it is calculated as a ratio of population in a location on location`s area. And to make it unbiased, we subtracted the number of people employed at the particular firm from the population of the location. In general, the next two effects are the weighted distances.

The intra-industry effect is the weighted distance of the particular firm in the particular industry to all other firms in the same region and the same industry. Firstly, we calculated the Euclidian distances within all locations using coordinates. Then, our weighted mechanism is following: we aggregated the total output of the same industry firms (in our case it is sales per year) in the same region at the particular year for each location and for the whole region and divided the aggregated output of each location on the total output of the region. And the same procedure we performed for each of the eleventh chosen industries, the thirteen available years and the nine selected regions.

The input-output linkages effect is based on the previous effect. We used the input-output table to come from each intra-industry weighted distance to the weighted distance of the whole manufacturing firms. As the input-output linkages effect based on intra-industry effect, in the sense that almost all industry used up to 90% of the same industry products, we investigate these two distance effects with urbanization separately and compare the magnitude inside industries and the intersection of all manufacturing industries.

In this study, we test the hypothesis that the rise of the distance between firms leads to decreasing agglomerations effects. Although, we want to check the significance of agglomerations effects in Ukraine and compare their magnitude on productivity. Before an estimation, we need to separate the total factor productivity (TFP) from the total sales. We used Levinsohn and Petrin (2003) approach as they were working with the manufacturing firms too. After the TFP calculation, we produced three linear regressions to get the coefficient on our effects respectively and another two regressions to estimate the urbanization effects with each distant effect separately.

This work proceeds with the review of the relevant literature, covered in Chapter 2. The following chapters present the data description, methodology, and estimation results. Finally, the conclusions are described in Chapter 6.

Chapter 2

LITERATURE REVIEW

Considering Marshall (1890), industrial agglomeration reduces three types of transport cost – the cost of moving goods, people and ideas. According to localization economies, there are internal effects to the shared industry and location, which seem to be external to the firms.

The most recent study, Ellison, Glaeser, Kerr (2010), compare effects of all three Marshallian theories and the importance of shared natural advantages. They estimated that shared natural advantages were more important than any single Marshallian factor but not as important as the cumulative effect of the three Marshallian factors. Another important result, which they gained was that customer-supplier relationships have the strongest effect among the Marshallian factors. The authors used coagglomeration index to compare the magnitude of effects.

In the current study, we are going to separate the effects of labor, capital, material cost and agglomeration effects on firms` productivity.  Therefore, the literature review of agglomeration effects impact on firms` productivity is arranged in the following way:

- first, we summarize the findings on the production function estimation;
- second, we consider empirics of agglomeration economies;
- next, we review the literature findings on clustering algorithms, considering the spatial data clusterization techniques.

2.1. Production function estimation

We start our brief review of the literature on estimation of production functions with the summarizing paper by Ackerberg, Caves, Frazer (2015 henceforth ACF). This paper briefly explained why there is a need to struggle with TFP estimation beyond usual FE or even OLS. For illustration purposes, consider a simple Cobb–Douglas production function in logs: $y_{it} = f(k_{it}, l_{it}, \omega_{it})$, where k_{it} is capital, l_{it} is labor and ω_{it} represent "productivity" shocks that are potentially observed or predictable by firms when they make input decisions.

The crux of the identification problem inherent in estimating such a production function is that the inputs such as capital and labor are chosen by firms (ACF mentioned that). This means that if the econometric unobservable ωit is observed (or partially observed) by the firm prior to choosing kit and lit, then these choices will likely depend on ω_{it}, creating correlation between (k_{it}, l_{it}) and ω_{it}, and rendering OLS estimates of βk and βl inconsistent (ACF mentioned that underlining this endogeneity problem dates back to Marschak and Andrews (1944).

ACF focused on papers by Olley and Pakes (1996), Levinsohn and Petrin (2003), suggested an alternative approach for producing more consistent estimates considering the current data generating processes. Collard-Wexler and De Loecker (2016) introduced an estimator that gave an opportunity to avoid measurement error in the capital stock, relying on the hybrid IV-control function approach.

The most recent study was undertaken by Mollisi and Rovigatti (2017), who considered all papers mentioned before such as Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015) and proposed a new estimation procedure based on Wooldridge (2009) with using dynamic panel instruments. The last showed how a consistent estimator could be produced by using a single step GMM framework only, while others used two-steps estimation.

Considering the literature on the production function estimation, we still can highlight three main approaches: Olley and Pakes (OP), Levinsohn and Petrin (LP) and GMM. There are two already mentioned papers, which provided more advanced techniques based on OP and LP: Ackerberg, Caves, Frazer (2015) and Collard-Wexler, De Loecker (2016). The LP technique is the most appropriate approach in the current investigation as we deal with manufacturing firms as it was done in the initial study. More advanced techniques are appropriate, but our main focus of this thesis is the exploration of the agglomeration effects and incorporating the capital shocks or other kinds of shocks are above our discussion.

2.2. Empirics of agglomeration economies

Combes and Gobillon (2015) underlined that the estimation of the overall local characteristic impact, mostly used the ratio of local employment density on local productivity, there is no way for separating such effects as sharing, matching, or learning from each other, so all of them act simultaneously. Policy implication directly depends on the choice of local determinants, which affect local productivity and the first one is the economy of density.

Combes and Gobillon (2015) suggested that the local density of economic activities should lead to increasing the productivity of firms and workers. The same paper suggests that the common model should be a regression: the logarithm of regional wage or TFP on the logarithm of current employment or population density with the use of aggregate data. As emphasized by Melo et al. (2009), the estimations vary because of different countries, industries, periods of time. So, the elasticity estimates vary from 0.04 to 0.07, which implies 3-5% productivity gain if density double.

Considering specific estimates, there are a number of studies. Ciccone and Hall (1996), Rosenthal and Strange (2008) with usage aggregate data for the United States. Ciccone (2002), Ciccone and Hall (1996) with usage NUTS 3 regions for France, Germany, Italy, Spain, and the United Kingdom. Brulhart and Mathys (2008) with usage 245 NUTS 2 regions for 20 western and eastern European countries.

There are other estimations, which used almost the same estimation method as those mentioned above. The results differ significantly depending on the adopted empirical strategy. The main finding as summarized by Combes and Gobillon (2015) is that quite large agglomeration gains are present, a long-run elasticity of productivity with respect to the density is up to 0.13, the strength of agglomeration effects increased over time.

The results of initial studies, which estimate agglomeration economies for separate countries using wages or TFP aggregated by region, were provided by Rosenthal and Strange (2004). The findings of recent articles, which used the individual level of datasets, including workers' or firms' precise location, were covered by Combes and Gobillon (2015). Summarizing the economy of the density topic, Combes and Gobillon (2015) also underlined that it was more expedient to study TFP rather than wages because it showed a direct productivity measure at the firm level.

Lafourcade and Mion (2005) showed contradictory findings that large plants tend to cluster within narrow geographical units (concentration) and small establishments preferred to co-locate within wider distance-based clusters (agglomeration). There are also studies analyzing the connection between the profit maximization firms` behavior and agglomeration economies by Crozeta, Mayerb, Mucchielli (2004); Combes and Lafourcade (2011). All three papers used the framework proposed by Ellison and Glaeser (1997), which removes the plant size from localization. Such a framework is not used in this investigation because of the assumption of spatial and geographic independence. The size of the region and mutual location do not seem to matter for the index value. So, the local employment density is to be used.

Since the majority of studies, including the case of Ukraine itself presented by Vakhitov and Bollinger (2010), suggested that the firm-level approach, compared to the aggregate approach, showed the most stable results, such an approach gained credence as the most appropriate one.

2.3. Clustering essentials

"*The agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It's also known as Agglomerative Nesting. The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram*."

(Kassambara, 2017, p. 67)

As was mentioned, the agglomeration effects depended on the method of shaping the regions. The industry`s concentration does not always suit perfectly within the official region` boundaries. To conclude the differences between industries, firm's locations, we need to do the clusterization analysis.

There are several clustering methods: K-Means, Density-based Clustering Algorithm (DBSCAN) and a Multidirectional Optimum Ecotope-Based Algorithm (AMOEBA) and Agglomerative Nesting (AGNES).

Before starting build clusters, we have to choose such method, that has an ability to set the radius in km, to weight the distances between locations (towns, villages, cities or city` parts) using its population, to make consistent interpretation of the clustering mechanism. The AGNES has all mentioned above strengths comparing to other approaches. The agglomerative clustering is the type of hierarchical clustering. It produces the dendrogram, which visualizes the hierarchy of the locations: there are list of locations on the x-axis and y-axis is a measure of closeness of either individual locations points or clusters. To make sure that the AGNES is the best choice, we need to consider the pros and cons of other methods.

The computation difference lays above our discussion, so let me focus on the pros and cons of each method. The first approach has main control parameter – the number of clusters. The DBSCAN technique has two main control parameters: a minimum point in a cluster and an "epsilon", which is the distance between points inside the cluster (Ester M. and 1996). The main advantage of both algorithms is high predictive power in accordance with their advanced technique of discovering noises. The main disadvantage is relatively complicated interpretation. The interpretation of epsilon parameter of the DBSCAN technique is still under discussion.

The last-mentioned approach AMOEBA uses polygon-based data (Aldstadt, J., and A. Getis, 2006). It is both main pros and cons. This algorithm simply joins smaller regions into a bigger cluster based on some underlying variable. Usage of polygon-based data provides better data on the area of the locations, clusters. For the purpose of that study, we can imagine the locations as the points with different weights, which are the population of the locations. So, usage of the AMOEBA approach could be the further development of the topic.

Chapter 3

METHODOLOGY

Firstly, we need to estimate the production function to identify the agglomeration spillovers. To set up a model of the production function we use a simple model developed by Rosenthal and Strange (2004) was used:

$$y_i = g(A_i) * f(x_i) \tag{1}$$

where A_i is the agglomerations effects, x_i is factor influence on firm`s output such as labor, capital, material cost.

Another component of the first equation is given by the LP approach:

$$f(x_i) = \beta_l l_t + \phi_t(k_t, m_t) + \eta_t \tag{2}$$

where t is a period of time, l_t is employment at the firm, $\phi_t(k_t, m_t)$ is the instrumental variable, which corrected the classical capital share using the material cost. The total factor productivity component (TFP) η_t includes the potentially observed effects excluding magnitude of the capital, material and labor costs.

We modified the agglomeration component in the following ways:

$$A_i = Urbanization_i * Intra\_industry_i \tag{3a}$$

$$A_i = Urbanization_i * \ Potential\ market_i \tag{3b}$$

where i imply that the agglomeration effect is calculated for each firm.

Proceeding the first tree equations, we can derive TFP as function of the urbanization and intra-industry, or urbanization and potential market in log terms:

$$\eta_t = f(x_i) - \beta_l l_t - \phi_t(k_t, m_t) \quad [4]$$

$$\eta_t = \beta_U * Urbanization_i \quad [5]$$

$$\eta_t = \beta_I * Intra\_industry_i \quad [6a]$$

$$\eta_t = \beta_P * \ Potential\ market_i \quad [6b]$$

$$\eta_t = \beta_U * Urbanization_i + \beta_I * Intra\_industry_i \quad [7a]$$

$$\eta_t = \beta_U * Urbanization_i + \ \beta_P * \ Potential\ market_i \quad [7b]$$

As was mentioned, the measurement of the agglomeration effects depends on the method of shaping regions. So, before we would proceed with the agglomeration effects calculation, we need to shape our regions considering the differences in the industry, spatial location of each region and size of region. We selected for our purpose the AGNES clusterization approach with weighting on the population of locations (means city, village, town, etc.). Intra-industry and potential market effects depend on the radius of the corresponding industrial cluster. Using classification of objects of the administrative-territorial system of Ukraine (KOATUU) based approach to shape industrial clusters weekly follows some economic logic.

Müllner (2017) puts the general scheme of the agglomerative clustering procedure, which we are going to use, as follows:

1. Start with N singleton clusters (nodes) labeled −1, . . . , −N, which represents the input points.
2. Find a pair of nodes with minimal distance among all pairwise distances.
3. Join the two nodes into a new node and remove the two old nodes. The new nodes are labeled consecutively 1, 2, . . .
4. The distances from the new node to all other nodes is determined by the method parameter (see below).

5. Repeat N − 1 times from step 2, until there is one big node, which contains all original input points.

After we proceed all above mentioned steps, we would have the dendrograms for each industry. The dendrogram is a tree of our locations, which we could observer cutting a tree at the different threshold. Here threshold could be interpreted as a radius of the cluster. Such interpretation became possible because we used one of the most basic methods called 'average'.

Müllner (2017) continues to describe how this method works: this 'method' attribute of the first input parameter d. This specifies which metric was used in the distance method, which generated the first argument. The parameter method specifies which clustering scheme to use. The clustering scheme determines the distance from a new node to the other nodes. Denote the dissimilarities by d, the nodes to be joined by I, J, the new node by K and any other node by L. The symbol |I| denotes the size of the cluster I.

$$d(K, L) = \frac{|I| \cdot d(I, L) + |J| \cdot d(J, L)}{|I| + |J|} \qquad [8]$$

The distance between two clusters A, B is the average distance between the midpoints in the two clusters:

$$d(A, B) = \frac{1}{|A||B|} \sum_{a \in A, b \in B} d(a, b) \qquad [9]$$

There are markets inside each cluster. The size of the market depends on the size of the cluster. So here we should separate local market`s effects within the different size of clusters. Ellison, Glaeser, Kerr (2010) compare all three Marshallian effects at 250, 500, 1000 miles thresholds. In Ukraine, we have the more concentrated spatial organization. So, we chose the cluster`s radiuses at 200km, 100km, and 50km value for current study.

Our main assumption is that clusters with the same radius look differently for each industry. So, we reproduce cluster analysis for each industry. There is at APPENDIX C the illustration for chosen four industries the clusters with the 100km radius.

The first component in equation 3 – "urbanization" is coming from the following equation:

$$Urbanization_i = \frac{population_{\,l} - employment_i}{area_{\,l}} \quad [10]$$

where $population_{\,l}$ is a population of location, where firm located; $employment_i$ is labor employed by the particular firm; $area_{\,l}$ is the area of location (city, town, village, city`s part), where a particular firm is located.

"Intra-industry" is an effect inside some industry. This effect was developed to investigate the industry-specific magnitude on the firm`s productivity. It is computed as:

$$intra\_industry_l^t = \sum_{l=1}^{N} d_l^t w_l, \text{ where } w_l = \frac{\sum_{i=1}^{F} output_i^t}{\sum_{l=1}^{N} output_l^t} = \frac{\sum_{i=1}^{F} output_i^t}{\sum_{l=1}^{N}\sum_{i=1}^{F} output_i^t} \quad [11]$$

where each parameter means:

- w_l is the weight of location;
- t – year, when the output was reported;
- l – location (city, town, village, city`s part) of each firm;
- N is a total number of location inside the region;
- i – a number of firms inside selected location;

- F is a total number of firms inside the location;
- d_l^t – distance to from the selected location to others inside the region, which was shaped by clustering using selected max radius.

At this point, it is necessary to underline that the clusterization was performed for each industry separately.

The potential market effect is the index, which was constructed to look at the firms` input-output linkages within the manufacturing sector. Basically, this index could be interpreted as the weighted distance to the potential markets inside the manufacturing sector. The weighting factor is relative weight from the input-output table given by UKRSTAT. The OECD gave the definition of an input-output table (IOT) as the sale and purchase relationships between producers and consumers within an economy. "Potential market" effect calculated as:

$$potential\_market_i = \sum_{ind=}^{I} \sum_{j=1}^{M} \left[\sum_{i=1}^{N} \frac{output_i^j}{\sum_{l=1}^{N} output_i^j} * distance_i \right] * w_j \qquad [12]$$

where we summarize the weighted distance using the relative weight of sub-industry in IOT.

Let`s summarize all steps of calculation from the moment, when we got raw data till the moment when we got the coefficients for each effect according to equations 5 – 7b:

1. Omitting all firms that enter or exit over the sample period.
2. TFP estimation using traditional approach (Olley and Pakes, 1996).
3. Clusterization using weighted distance and AGNES algorithm.
4. Aggregation population by years, locations and calculation total output at each cluster.
5. Calculation all three agglomeration effects.
6. Running all five regressions.

Beginning from the step 3, we calculated for each industry and for each radius. After the third step, additionally, we calculated for a cluster.

The purpose of the whole procedure is to investigate the existence of the agglomeration effects. The analysis of the significant effects would serve as the evidence for the potential benefits from the concentration of the manufacturing firms within different industries and spatial organizations.

C h a p t e r 4

DATA DESCRIPTION

In our empirical research, we make use of the data of Ukrainian firms in 2001-2013. The database came from several statistical records submitted annually to the State Statistics Service of Ukraine (Ukrstat). The original dataset contains the statistical information about firms from all industries. Considering Henderson (2003) specifications, Vakhitov (2010) empirical study and overall tendency in the agglomeration study to use the manufacturing data, the final dataset consists of only manufacturing firm's observation. This fact corresponds to using the subsector "D" according to the system of economic activities. The whole structure is given in APPENDIX A.

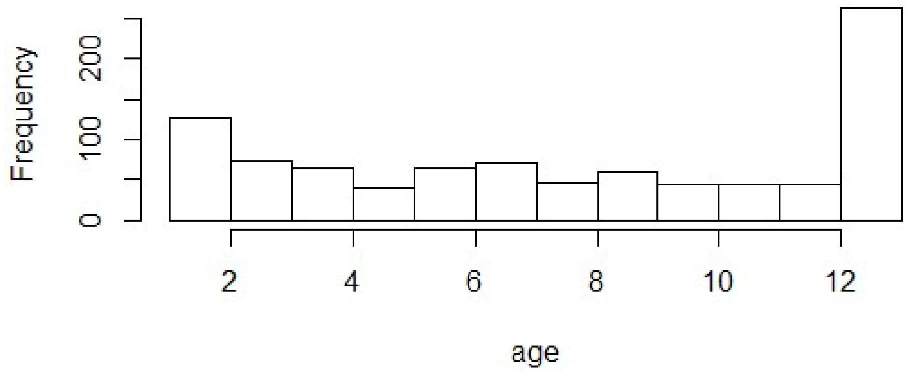


Figure 1. Histogram of firm`s age.

The final dataset consists of 64 825 firms, 303 643 firm-year observation overall. So, there are the panel data with the unique code for each firm. We used traditional approach (Olley and Pakes, 1996): entry and exit of firms are accounted for in TFP estimation by constructing a balanced panel; i.e. by omitting all firms that enter or exit over the sample period.

We focused on all manufacturing industries. To join the industry list with the input-output table, given by UKRSTAT, we need to modify the industry list with. Based on the KVED version 2005, we united the sub-sections DB+DC and DD+DE into DBC and DDE respectively. So, we have sub-sections DBC and DDE names are "Textile production; manufacture of clothes, fur, and fur products; Manufacture of leather, leather goods and other materials" and "Wood processing and production of wood products, except furniture; Pulp and paper production; publishing activity" respectively. The full range of industries is provided in APPENDIX A.

As we found the pattern of the firm`s locations. The whole picture of firm`s locations patterns from different industries is given in APPENDIX B.

The dataset includes sales as such output, material cost, capital at the end of the year given in thousand UAH and the number of people employed at the firms. All of them were corrected considering Producer price indices (to basic year), taken from the UKRSTAT official website.

To calculate the localization effect, we took the data on the areas and population of each city, town, village, where firms are located, from the UKRSTAT on based on KOATU classification. Totally we have 399 locations.

We picked nine main cities, based on firm`s locations patterns from different industries: Kyiv, Dnipro, Zaporozhye, Kharkiv, Donetsk, Lugansk, Mykolaev, Odesa and Lviv. It was done for separating important clusters, which includes nine listed cities. After every cluster analysis using different cluster radius (50km, 100km and 200km), we would take into consideration only important clusters.

The input-output table is the newest available (for 2014) and comes from Ukrstat as well. The oldest table is available for 2009, the table does not differ significantly by years, so for simplicity of calculation 2014 input-output table was used for all years.

Table 1. The mean variables descriptive statistics

Year	Output	Empl.	Capital	Material cost	Area of city/town	Population of city/town	Quantity of firms
2001	4662	100	2773	3095	241	728591	23522
2002	5217	90	2624	3311	247	747104	25114
2003	7178	85	2726	4586	251	758557	25669
2004	11058	87	2947	6945	261	792938	25359
2005	12384	87	3309	8068	267	812247	25190
2006	14220	82	3781	9097	277	840787	25818
2007	19110	82	5120	12497	274	833186	25510
2008	24462	82	6729	16144	280	849407	24671
2009	21011	72	7598	13630	281	853807	23528
2010	28969	71	8151	19096	216	656929	19709
2011	33712	69	8871	22536	216	658568	20595
2012	34925	68	10896	21355	216	655690	20315
2013	35052	69	12369	20972	214	651998	18643

To calculate the proposed index, we modified both KVED and input-output table. As column names in input-output table proposed by Ukrstat aggregated DB with DC and DD with DE. In the other way, some columns in input-output table needed to be aggregated to get subsection in KVED. The main variables descriptive statistics in 2001 and 2013 are given in APPENDIX D. The input-output table is given in APPENDIX E.

Chapter 5

ESTIMATION RESULTS

In this study, we try to find more evidence that the rise of the distance between firms leads to decreasing agglomerations effects. To find some evidence on the first hypothesis, we produced the estimation of the intra-industry and the potential markets effect at regions with different radiuses 50, 100, 200 km.

Firstly, we would consider the distributions of each effect. It`s the necessary step in order to compare the differences, which occur through the different industries and locations. Secondly, we would analyze the coefficients on the each separately. Then, we would look at the coefficients at the regressions with two agglomeration effects.

The last step is checking the significance of agglomerations effects in Ukraine and compare their magnitude on TFP. The main expectation is the negative sign of the coefficient, as decreasing of the distances leads to decreasing the transportation cost. The last thing should cause the increase of productivity.

Before the analysis of the effects, it wouldn`t be a bad idea to speculate on how the predetermined cluster` centers unite depending on the different industries. At 50km radius, the predetermined centers don`t unite. At 100km radius, the industries (the names of industries are given at APPENDIX A):

- DF (Production of coke, refined products, and nuclear materials), DG (Chemical production) and DK (Manufacture of machinery and equipment) unite Dnipro, Zaporozhye cities with the surrounding smaller locations;

- DA, DBC, DG, DL unite Donetsk, Lugansk cities with the surrounding smaller locations;
- DJ (Metallurgical production and production of finished metal products) unites Nikolaev, Lviv cities with the surrounding smaller locations.

At 200km radius, the industries:

- DBC, DF, DG, DH, DI, DJ, DL, DM unite Dnipro, Zaporozhye cities with the surrounding smaller locations;
- DA (Manufacture of food products, drinks, and tobacco products) unites Dnipro, Zaporozhye, Kharkov cities with the surrounding smaller locations;
- DDE (Wood processing and production of wood products, except furniture; Pulp and paper production; publishing activity) and DK (Manufacture of machinery and equipment) unites Dnipro, Zaporozhye, Nikolaev cities with the surrounding smaller locations;
- all industries unite Donetsk, Lugansk cities with the surrounding smaller locations;
- DI (Manufacture of other non-metallic mineral products) and DJ (Metallurgical production and production of finished metal products) unite Nikolaev, Lviv cities with the surrounding smaller locations;
- DBC, DF, DG, DH, DL, DM unite Nikolaev, Odesa cities with the surrounding smaller locations.

The interpretation of such difference in clusterization lays under discussion of the current paper, but it is interesting direction for further investigation. The main idea of underlining these differences in clusterization is to show that the administrative boundaries do not fit perfectly the manufacturing organization boundaries.

The distribution of all three agglomeration effects was given below for the DK (Manufacture of machinery and equipment) industry as an example. We chose for the analysis the DK industry`s distribution as it represented other industries pretty well. The whole range of the distributions, which vary by radius region and industry could be found at the Google Drive, which contains all files related to the thesis (folder "density").

5.1. Urbanization effect

Urbanization effect doesn`t vary many troughs the different industries and doesn`t depend on the different ways of shaping the region. The mean is laying from 7.5 to 8. The distribution of the effect for DK (Manufacture of machinery and equipment) industry is given on figure 2.

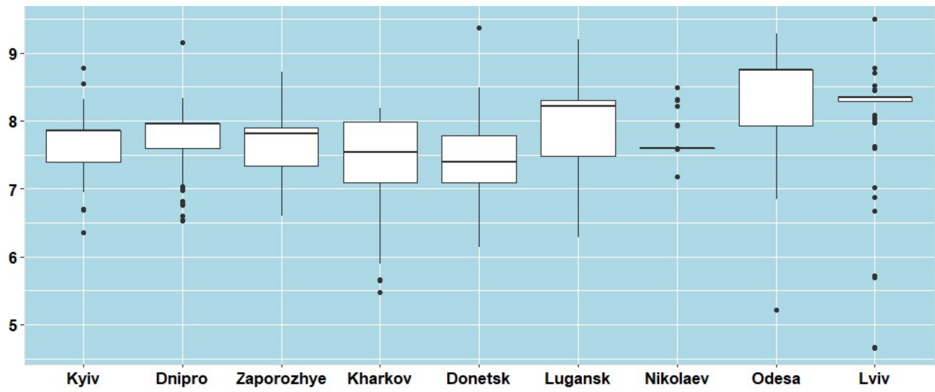


Figure 2. Distribution of the urbanization effect at different clusters.

The coefficient of equation 5 is presented below. Using the sparklines, we try to visualize the dynamics, which comes from the radius differences. The fracture of the line at the sparklines showed that at some point the manufacturing firms switch from using the spatial widespread labor market to the labor market, which is only cover big cities.

Table 2. The mean of the urbanization effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Dnipro	0,17	0,21	0,15	0,72	0,15	0,36	0,15	0,10	0,04	0,11	0,00
Donetsk	0,06	0,21	0,06	0,24	0,12	0,26	0,22	0,20	0,12	0,30	0,46
Kharkov	0,12	0,00	0,12	0,03	0,04	0,06	0,14	0,09	0,12	0,08	0,04
Kyiv	0,22	0,18	0,22	0,66	0,08	-0,02	0,12	0,00	0,28	0,22	0,07
Lugansk	0,03	0,11	0,03	-0,41	0,06	0,19	0,09	0,16	0,11	0,27	0,13
Lviv	0,11	0,01	0,11	-0,62	0,02	0,06	0,11	0,03	0,12	0,09	-0,01
Nikolaev	0,25	0,25	0,25	-0,90	0,25	0,25	0,07	0,08	0,12	0,24	-0,12
Odesa	0,04	0,02	0,04	0,03	-0,04	-0,03	0,12	-0,02	-0,04	0,06	-0,05
Zaporizhzhia	0,12	0,11	0,12	0,93	0,07	0,21	0,11	0,07	-0,01	0,09	0,12

Notes: the coefficients were calculated using equation 5

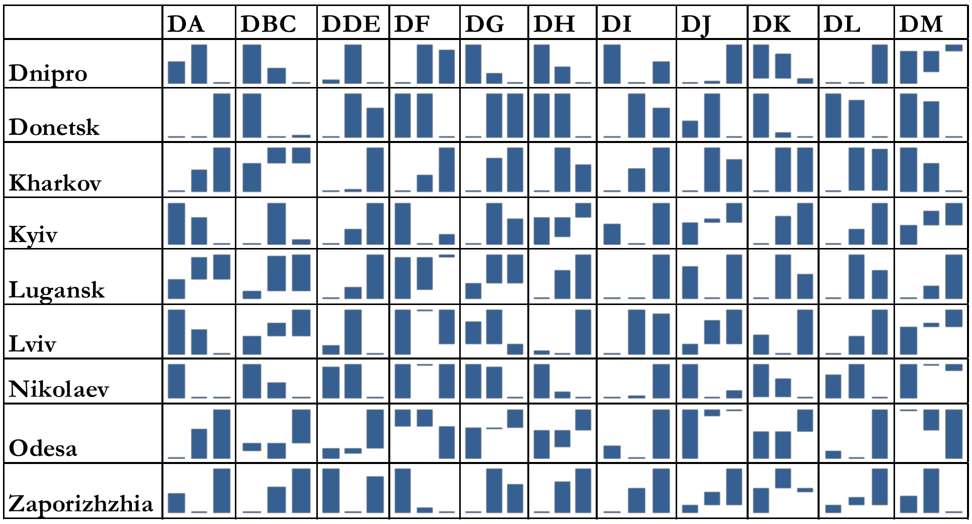


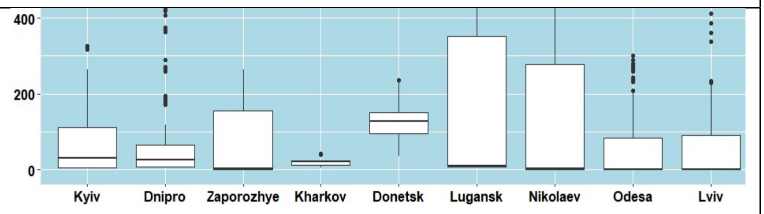
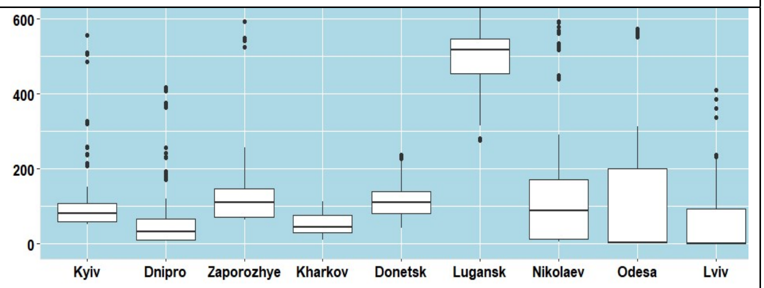
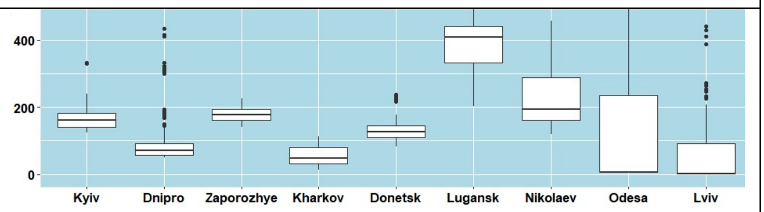
Figure 3. The urbanization effect coefficient at 50, 100, 200km radius

The selected cell in the table above represents the coefficient rise with increasing of the region radius. It shows that all industries, among all regions, have the urbanization effects stronger when the region covers more territory. So, it implies that smaller locations have mutual benefits with bigger locations by sharing labor markets. But some industries in such regions as Dnipro, Donetsk, Lviv, Nikolayev showed that there was a drop in the value of the coefficient on the urbanization effect with increasing the of the radius. Such cases demonstrated that the urbanization effects became stronger at the center of the region. So, in these four regions, the big cities would become even bigger, small even smaller.

5.2. Intra-industry effect

The intra-industry effect is the agglomeration effect, calculated using equation 11. In order to incorporate the share of each location (town, a city, village, part of city, etc.) in the total output of the cluster, we used the algorithm described in methodology. The distribution of the intra-industry effect is given below.

Table 3. Distribution of the intra-industry effect for DK (Manufacture of machinery and equipment) industry

Mean, km	Radius, km	Distribution of the industry effect, km (on the x-axis there are the region centers)
~ 20	50	
~ 70	100	
~ 140	200	

The first column of table 3 is a mean of the intra-industry effect coefficient at 50, 100, 200km radius, which calculated excluding Lugansk and Nikolaev as outliers. We could see that the mean value increases substantially from the 50km to 100km radius, and continue after 100km radius. There is huge variance at Lugansk and Nikolaev. At radiuses 100km and 200km the mean at that regions became a few times more than at first five regions. Nikolaev region and Lviv united at these radiuses, which affected more Nikolaev than Lviv. At 200km radius Donetsk and Lugansk also united, which decreased mean at Lugansk. It could explain the fact that the distances were weighted and at these regions this approach produced outliers. These moments is not discussed in this paper and should be considered in the further investigation.

Table 4. The mean of the intra-industry effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	-0,0058	-0,0064	-0,0079	-0,0179	-0,0056	-0,0037	-0,0037	-0,0044	-0,0082	-0,0054	-0,0057
Dnipro	0,0008	0,0004	-0,0003	-0,0004	-0,0002	0,0005	0,0000	-0,0007	-0,0006	-0,0003	-0,0001
Zaporizhzhia	0,0009	0,0005	0,0001	0,0025	-0,0001	0,0008	0,0004	-0,0005	-0,0007	-0,0003	0,0001
Kharkov	0,0010	0,0007	0,0008	-0,0075	0,0017	0,0015	0,0017	0,0013	0,0019	0,0007	0,0009
Donetsk	0,0003	0,0009	0,0009	0,0054	0,0012	0,0011	0,0006	0,0006	0,0009	0,0011	0,0011

Notes: the coefficients were calculated using equation 6a

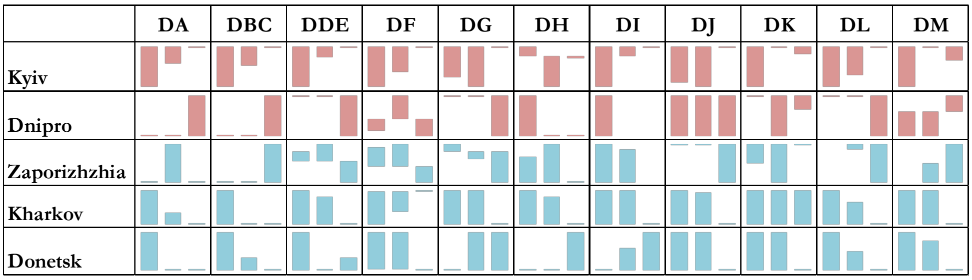


Figure 4. The intra-industry effect coefficient at 50, 100, 200km radius

The value inside cells of table 4 is a mean of the intra-industry effect coefficient at 50, 100, 200km radius. And the histogram inside each cell visualizes the dynamics of the coefficients at each value of radius respectively. From the distributions in table 2, we could conclude that at table 3, there should be a huge difference between the value of coefficients at 100km and 200km radiuses. At table 3, there are second and third bars presents these values at mentioned radiuses respectfully. The assumption, about the differences at 100km and 200km radiuses, is true only about Kyiv region. And Kyiv is only one region, which has all negative coefficients. This implies that decreasing the distances to the other firms in the same industry would benefit the manufacturing firms. And the biggest magnitude of the coefficient appears at the 50km radius.

Overall, we could see that the intra-industry effect has bigger magnitude at 50km radius than at 100km, and in most cases, the value of the effect becomes almost zero at the 200km radius.

Positive values of coefficient could be explained as the industries and locations, where the intra-industry linkages could benefit all firm within some region and industry. From another perspective, we could observe the intra-industry effect,

when running the regression using equation 7a. We would focus on the analysis of the first five regions for the intra-industry and potential market effects, as they showed most consistent distribution.

Table 5. The mean of the urbanization effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	0,0822	-0,0262	0,0816	0,0318	-0,1195	-0,0638	0,0707	-0,2736	0,0638	0,0908	-0,1256
Dnipro	0,1338	0,1886	0,1689	0,7070	0,1585	0,3544	0,1511	0,1113	0,0666	0,1089	-0,0093
Zaporizhzhia	0,0627	0,0540	0,1237	0,9048	0,0743	0,1910	0,0881	0,0871	0,0206	0,0864	0,1246
Kharkov	0,1132	-0,0054	0,2033	-0,0139	0,0411	0,0460	0,1288	0,0766	0,1168	0,0827	0,0136
Donetsk	0,0485	0,1687	0,1674	0,2296	0,0651	0,1899	0,2103	0,1823	0,0955	0,2637	0,3807

Notes: the coefficients were calculated using equation 7a

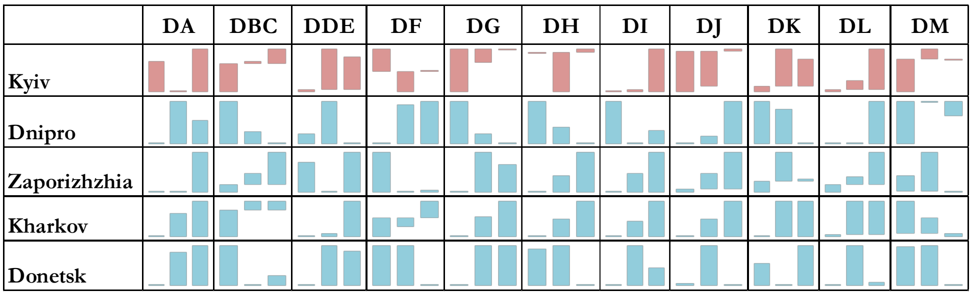


Figure 5. The intra-industry effect coefficient at 50, 100, 200km radius

Comparing values of the respective coefficients, we could observe that there is a drop of the coefficient on the urbanization effect almost at all industries and region.

Nevertheless, the values of the intra-industry effect coefficients have not changed much at table 6 comparing to table 4.

Table 6. The mean of the intra-industry effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	-0,0049	-0,0076	-0,0077	-0,0141	-0,0067	-0,0045	-0,0029	-0,0093	-0,0075	-0,0050	-0,0080
Dnipro	0,0007	0,0003	-0,0004	-0,0023	-0,0003	0,0005	-0,0001	-0,0007	-0,0007	-0,0003	-0,0001
Zaporizhzhia	0,0009	0,0005	0,0001	-0,0009	-0,0002	0,0006	0,0004	-0,0005	-0,0007	-0,0003	0,0000
Kharkov	0,0010	0,0007	0,0008	-0,0108	0,0017	0,0015	0,0017	0,0013	0,0019	0,0007	0,0009
Donetsk	0,0003	0,0006	0,0007	0,0052	0,0011	0,0008	0,0002	0,0004	0,0007	0,0007	0,0004

Notes: the coefficients were calculated using equation 7a

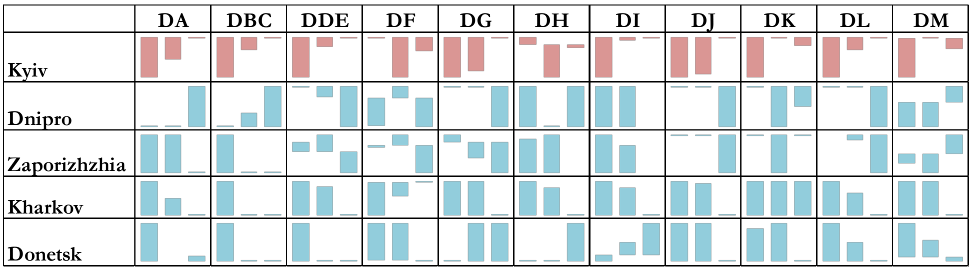


Figure 6. The intra-industry effect coefficient at 50, 100, 200km radius

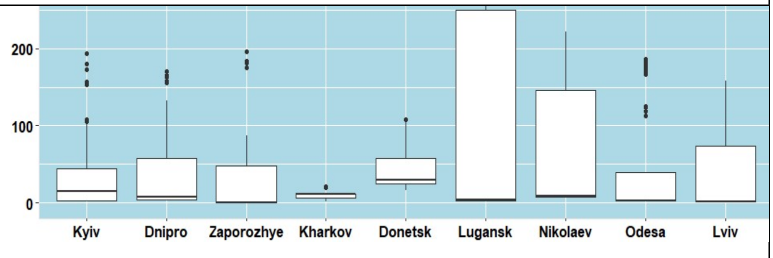
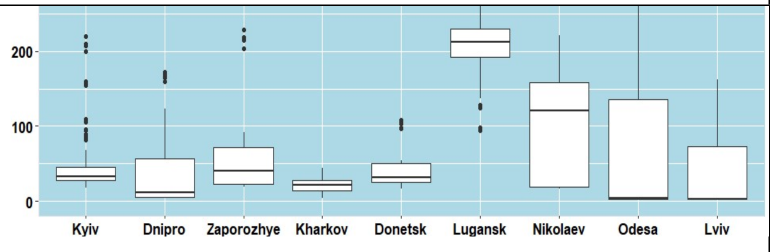
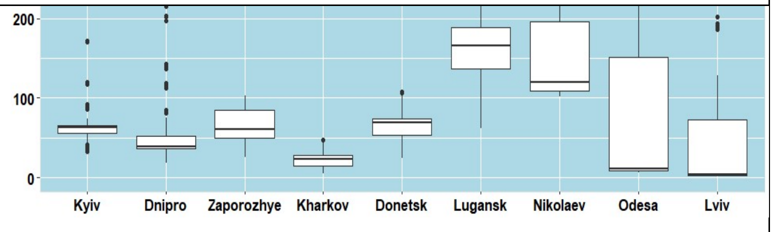
At this stage, we could compare the magnitude of the urbanization effect and the intra-industry effect. The mean of the urbanization effect is about 7.75 people per sq. km and the mean of the intra-industry effect is about 100km at 200km radius. The value of coefficient on the urbanization effect at Kyiv region with 200km radius in DK industry (Manufacture of machinery and equipment) is equaled to 0.0883. The value of coefficient on the intra-industry is equaled to -0.0064. So, increasing the number of people per sq. km on 10% would bring about 6.8% increase in log of output (0.775*0.0883 = 0.0684325). The decreasing the weighted distance of all firms inside DK industry on 10% would bring about 8.96% increase in log of output (14*0.0064 = 0.0896). In this particular case, the intra-industry effect is higher. Furthermore, there are many cases, when the intra-industry effect would bring more benefits than the urbanization. At this example, we could see, that the intra-industry effect is significant in sense of the coefficient value. The results of F-statistics, in that case, give us significance level at p=0.99 for both coefficients.

5.3. Potential market effect

The potential market effect is the agglomeration effect, calculated using equation 12. We could see that the outlier's problem appeared at the intra-industry effect. It was smoother than at the previous agglomeration effect and the IOT weighting method could be the reason for such results.

The first column of table 7 is a mean of the potential market effect value at 50, 100, 200km radius, which calculated excluding Lugansk and Nikolaev as outliers. We could see that the mean value continues double at 100km and 200km radius.

Table 7. Distribution of the potential market effect for DK (Manufacture of machinery and equipment) industry

Mean, km	Radius, km	Distribution of the potential market effect, km (on the x-axis there are the region centers)
~ 30	50	
~ 60	100	
~ 120	200	

The value inside cells of table 8 is a mean of the potential market effect coefficient at 50, 100, 200km radius. And the histogram inside each cell visualizes the dynamics of the coefficients at each value of radius respectively. From the distributions in table 7, we could conclude that at table 8, there should be about 20% difference between the value of coefficients at 100km and 200km radiuses. At table 8, there are second and third bars presents these values at mentioned radiuses respectfully.

Table 8. The mean of the potential market effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	-0,0072	-0,0132	-0,0129	-0,0192	-0,0082	-0,0037	-0,0068	-0,0043	-0,0159	-0,0084	-0,0098
Dnipro	0,0010	0,0007	-0,0003	0,0015	-0,0001	0,0015	0,0001	-0,0006	-0,0015	-0,0003	-0,0002
Zaporizhzhia	0,0012	0,0010	0,0002	0,0051	0,0001	0,0017	0,0010	-0,0004	-0,0017	-0,0003	0,0003
Kharkov	0,0013	0,0015	0,0017	0,0040	0,0022	0,0025	0,0036	0,0016	0,0035	0,0011	0,0015
Donetsk	0,0004	0,0019	0,0014	0,0051	0,0016	0,0018	0,0007	0,0008	0,0023	0,0014	0,0016

Notes: the coefficients were calculated using equation 6b

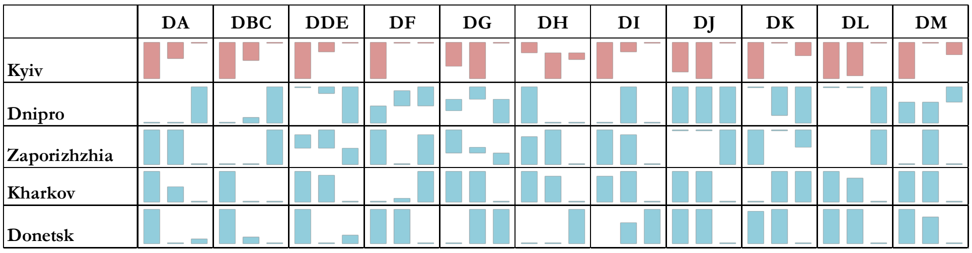


Figure 7. The potential market effect coefficient at 50, 100, 200km radius

The assumption, about the differences at 100km and 200km radiuses, is true only about Kyiv region. And Kyiv is only one region, which has all negative coefficients. This implies that decreasing the distances to the other firms in the same industry would benefit the manufacturing firms. And the biggest magnitude of the coefficient appears at the 50km radius. DJ and DK industries have negative sign coefficient at Kyiv, Dnipro, and Zaporizhzhia, which implies the existence of intra-region linkages benefit the local manufacturing producers and increase their productivity.

Overall, we could see that the intra-industry effect has bigger magnitude at 50km radius than at 100km, and in most cases, the value of the effect becomes almost zero at the 200km radius.

Positive values of coefficient could be explained as the industries and locations, where the input-output linkages could benefit all firm within some region and industry. From another perspective, we could observe the potential market effect when running the regression using equation 7b.

Table 9. The mean of urbanization effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	0,0825	-0,0173	0,0759	0,2390	-0,1276	-0,0594	0,0875	-0,2637	0,1399	0,0835	-0,1339
Dnipro	0,1340	0,1966	0,1682	0,6997	0,1565	0,3253	0,1479	0,1138	0,0784	0,1129	-0,0197
Zaporizhzhia	0,0630	0,0742	0,1197	0,8528	0,0698	0,1585	0,0806	0,0853	0,0351	0,0880	0,1114
Kharkov	0,1132	-0,0070	0,1993	0,0311	0,0415	0,0413	0,1232	0,0698	0,1050	0,0762	0,0121
Donetsk	0,0484	0,1686	0,1703	0,2335	0,0654	0,1869	0,2188	0,1702	0,0715	0,2635	0,4260

Notes: the coefficients were calculated using equation 7b

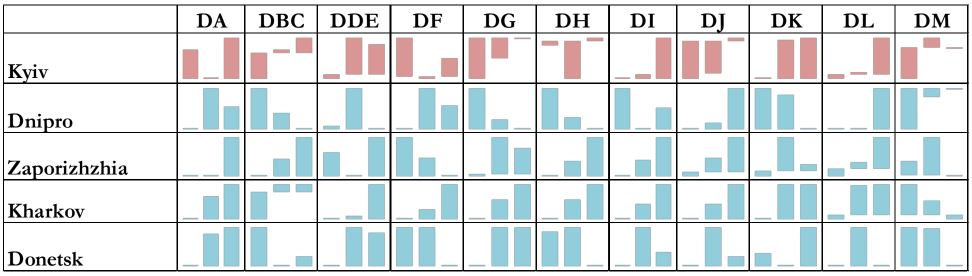


Figure 8. The urbanization effect coefficient at 50, 100, 200km radius

Comparing values of the respective coefficients, we could observe that there is a drop of the coefficient on the urbanization effect at first three regions. Kharkov and Donets have bigger values of the coefficient on the urbanization effect. It implies that the potential market effect takes a part of the urbanization effect. Nevertheless, the urbanization effect compensates for the negative values of the intra-industry effect. At this point, the coefficients at table 9 have changed much compared to table 4.

Table 10. The mean of the potential market effect coefficient

	DA	DBC	DDE	DF	DG	DH	DI	DJ	DK	DL	DM
Kyiv	-0,0061	-0,0157	-0,0127	-0,0148	-0,0099	-0,0048	-0,0052	-0,0088	-0,0128	-0,0083	-0,0139
Dnipro	0,0009	0,0005	-0,0004	-0,0011	-0,0002	0,0012	-0,0001	-0,0007	-0,0016	-0,0003	-0,0001
Zaporizhzhia	0,0011	0,0009	0,0001	0,0012	0,0000	0,0016	0,0009	-0,0004	-0,0018	-0,0002	0,0002
Kharkov	0,0013	0,0016	0,0018	0,0040	0,0022	0,0024	0,0035	0,0015	0,0034	0,0012	0,0015
Donetsk	0,0004	0,0013	0,0010	0,0050	0,0014	0,0013	0,0000	0,0006	0,0020	0,0009	0,0002

Notes: the coefficients were calculated using equation 7b

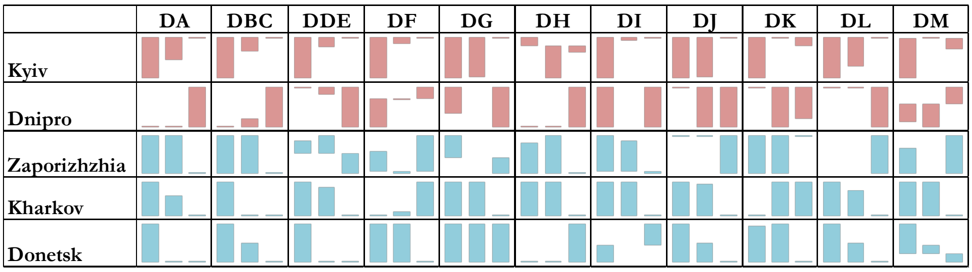


Figure 9. The potential market effect coefficient at 50, 100, 200km radius

At this stage, we could compare the magnitude of the urbanization effect and the potential market effect. The mean of the urbanization effect is about 7.75 people per sq. km and the mean of the potential market effect is about 50km at the 200km radius.

The value of coefficient on the urbanization effect at Kyiv region with 200km radius in DK industry (Manufacture of machinery and equipment) is equaled to 0.1974. The value of coefficient on the potential market is equaled to -0.0113. So, increasing the number of people per sq. km on 10% would bring about 15.29% increase in log of output (0.775*0.1974 = 0.152985). The decreasing the weighted distance of all firms to the all other firms through the all manufacturing industries on 10% would bring about 13.56% increase in the log of output (12*0.0113 = 0.1356). In this particular case the effect of the urbanization is higher, but we could find the cases, when the potential market effect would bring more benefits than the urbanization. At this example, we could see, that the potential market effect is significant in sense of the coefficient value. The results of F-statistics, in that case, give us significance level at p=0.99 for both coefficients.

Chapter 6

CONCLUSION

In this study, we found more evidence that the rise of the distance between firms leads to decreasing agglomerations effects. Nevertheless, overall effect of the increasing distance between firms inside one industry or within all manufacturing firms could be controversial. For the small town or villages, the increase of distance would show that these small location produces for the big regional market and has a benefit from this.

Among Localization, Potential market`s and Sub-industry effects, second had the biggest influence on TFP according to betta coefficient. Looking at coefficient in APPENDIX F and summary statistic in APPENDIX G, we can clearly see that Agglomeration effects became less significant through the distance. The above-mentioned findings go in line with coagglomeration patterns developed in Ellison, Glaeser, Kerr (2010).

Considering the distribution of the intra-industry effect, we noticed that the industry-specific linkages value is pretty the same up to the 100km radius and increases twice at 200km. The potential market effect value increases twice at the 100km radius and has the relatively small increase after.

The results of the regression on these two effects – the coefficient have negative sign mostly at Kiev region and positive at all other regions. It shows that in Kyiv region the decreasing the transportation cost would benefit all producers. Other regions have benefited from the neighborhood region`s firms and markets. More deep analysis of coefficients could show the main regional roads, which benefit all manufacturing firms.

All three agglomeration effects showed significant influence on firm's productivity. So, Ukrainian firms could benefit from these effects by cooperating with the local authority to reduce transfer cost of goods, ideas, and labor. Nevertheless, the overall conclusion is that only Kiev region could put this on agenda. As only there almost all coefficient for the intra-industry and potential market effects the have negative sign. Other regions have controversial results, which shows that there only between regions linkages affect productivity. Moreover, we could find some bundle of industries, which operates at two – tree regions and has the negative sign on coefficients. So that improving of some particular infrastructure object could increase the manufacturing productivity at these two-three regions, where the improvement would decrease the distance index (means the value of the intra-industry or potential effects).

WORKS CITED

Ackerberg, D. A., K. Caves, and G. Frazer. 2015. "Identification properties of recent production function estimators". *Econometrica*, 83(6), 2411-2451.

Aldstadt, J., and A. Getis. 2006. "Using AMOEBA to create a spatial weights matrix and identify spatial clusters". *Geographical Analysis*, 38(4), 327-343.

Basile, R., R. Pittiglio, and F. Reganati. 2017. "Do agglomeration externalities affect firm survival? Regional Studies", 51(4), 548-562.

Brülhart, M., and N. A. Mathys. 2008. "Sectoral agglomeration economies in a panel of European regions". *Regional Science and Urban Economics*, 38(4), 348-362.

Ciccone, A. 2002. "Agglomeration effects in Europe". *European Economic Review*, 46(2), 213-227.

Collard-Wexler, A., and J. De Loecker. 2016. "Production function estimation with measurement error in inputs". *National Bureau of Economic Research*. No. w22437.

Combes, P. P., and L. Gobillon. 2015. "The empirics of agglomeration economies". *Handbook of regional and urban economics*. Elsevier. Vol. 5, pp. 247-348.

Combes, P. P., and M. Lafourcade. 2011. "Competition, market access and economic geography: Structural estimation and predictions for France". *Regional Science and Urban Economics*, 41(6), 508-524.

Crozet, M., T. Mayer, and J. L. Mucchielli. 2004. "How do firms agglomerate? A study of FDI in France". *Regional science and urban economics*, 34(1), 27-54.

Duranton, G., and D. Puga. 2004. "Micro-foundations of urban agglomeration economies". *Handbook of regional and urban economics*. Elsevier. Vol. 4, pp. 2063-2117.

Eberts, R. W., and D. P. McMillen. 1999. "Agglomeration economies and urban public infrastructure". *Handbook of regional and urban economics*, 3, 1455-1495.

Eid, J., H. G. Overman, D. Puga, and M. A. Turner. 2008. "Fat city: Questioning the relationship between urban sprawl and obesity". *Journal of Urban Economics*, 63(2), 385-404.

Ellison, G., E. L. Glaeser, and W. R. Kerr. 2010. "What causes industry agglomeration? Evidence from coagglomeration patterns". *American Economic Review*, 100(3), 1195-1213.

Ellison, G., and E. L. Glaeser. 1997. "Geographic concentration in US manufacturing industries: a dartboard approach". *Journal of political economy*, 105(5), 889-927.

Ester, M., H. P. Kriegel, J. Sander, and X. Xu. 1996. "A density-based algorithm for discovering clusters in large spatial databases with noise". *In Kdd*. Vol. 96, No. 34, pp. 226-231.

Ferragina, A. M., and F. Mazzotta. 2015. "Agglomeration economies in Italy: impact on heterogeneous firms' exit in a multilevel framework". *Economia e Politica Industriale*, 42(4), 395-440.

Fujita, M., and J. F. Thisse. 2002. "Economics of agglomeration: Cities, industrial locations, and regional growth".

Hall, R. E., and C. I. Jones. 1996. "The productivity of nations". *National Bureau of Economic Research*. No. w5812.

Hsieh, C. T., and P. J. Klenow. 2009. "Misallocation and manufacturing TFP in China and India". *The Quarterly journal of economics*, 124(4), 1403-1448.

Jacobs, J. 1969. "The Life of Cities. Random House".

Kassambara, A. 2017. "Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning". (Vol. 1). *STHDA*

Kline, P., and E. Moretti. 2013. "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority". *The Quarterly Journal of Economics*, 129(1), 275-331.

Lafourcade, M., and G. Mion. 2005. "Concentration, spatial clustering and the size of plants".

Levinsohn, J., and A. Petrin. 2003. "Estimating production functions using inputs to control for unobservable". *The Review of Economic Studies*, 70(2), 317-341.

Marshall, A. 1890. "Principles of political economy". *Maxmillan*, New York.

McCann, B. T., and T. B. Folta. 2011. “Performance differentials within geographic clusters”. *Journal of Business Venturing*, 26(1), 104-123.

Melo, M. T., S. Nickel, and F. Saldanha-Da-Gama. 2009. “Facility location and supply chain management–A review”. *European journal of operational research*, 196(2), 401-412.

Mollisi, V., and G. Rovigatti. 2017. “Theory and Practice of TFP Estimation: The Control Function Approach Using Stata”.

Müllner, D. 2017. “The fastcluster package: User’s manual”.

Olley, G. S., and A. Pakes. 1992. “The dynamics of productivity in the telecommunications equipment industry”. *National Bureau of Economic Research*. No. w3977

Pe'er, A., and T. Keil. 2013. “Are all startups affected similarly by clusters? Agglomeration, competition, firm heterogeneity, and survival”. *Journal of Business Venturing*, 28(3), 354-372.

Rosenthal, S. S., and W. C. Strange. 2003. “Geography, industrial organization, and agglomeration”. *The review of Economics and Statistics*, 85(2), 377-393.

Rosenthal, S. S., and W. C. Strange. 2004. “Evidence on the nature and sources of agglomeration economies”. *Handbook of regional and urban economics*. Elsevier .Vol. 4, pp. 2119-2171.

Syverson, C. 2004. “Product substitutability and productivity dispersion”. *Review of Economics and Statistics*, 86(2), 534-550.

Syverson, C. 2011. “What determines productivity?” *Journal of Economic literature*, 49(2), 326-65.

Vakhitov, V., and C. Bollinger. 2010. “Effects of ownership on agglomeration economies: evidence from Ukrainian firm level data”. Journal of Urban Economics, 5, 1-13.

Wooldridge, M. 2009. “An introduction to multiagent systems”. John Wiley and Sons.

APPENDIX A

Table 11. The codes of the economic activities

Subsection	KVED
DA Manufacture of food products, drinks and tobacco products	15 Manufacture of food products and drinks 16 Manufacture of tobacco products
DBC Textile production; manufacture of clothes, fur and fur products; Manufacture of leather, leather goods and other materials (DB+DC)	17 Textile productions 18 Clothing production; production of fur and fur products 19 Manufacture of leather, leather goods and other materials
DDE Wood processing and production of wood products, except furniture; Pulp and paper production; publishing activity (DD+DE)	20 Wood processing and production of wood products, except furniture 21 Production of paper mass, paper, cardboard and products from them 22 Publishing and printing activity, reproduction of recorded media of information
DF Production of coke, refined products and nuclear materials	23 Production of coke, refined products and nuclear materials
DG Chemical production	24 Chemical production
DH Manufacture of rubber and plastic products	25 Manufacture of rubber and plastic products
DI Manufacture of other non-metallic mineral products	26 Manufacture of other non-metallic mineral products
DJ Metallurgical production and production of finished metal products	27 Metallurgical production 28 Manufacture of fabricated metal products
DK Manufacture of machinery and equipment	29 Manufacture of machinery and equipment
DL Manufacture of electric, electronic and optical equipment	30 Manufacture of office equipment and computers 31 Manufacture of electrical machinery and equipment 32 Manufacture of radio, television and communication equipment 33 Manufacture of medical equipment, measuring instruments, optical instruments and devices, watches
DM Production of vehicles and equipment	34 Manufacture of cars, trailers and semitrailers 35 Manufacture of other transport equipment

Notes: there were used slightly modified KVED-2005

APPENDIX B

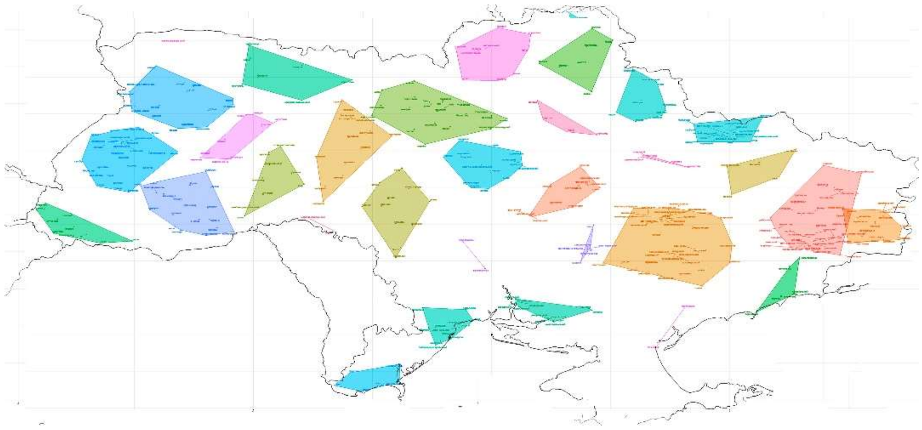


Figure 10. The illustration for DJ (Metallurgical production and production of finished metal products) industry the clusters with the 100km radius

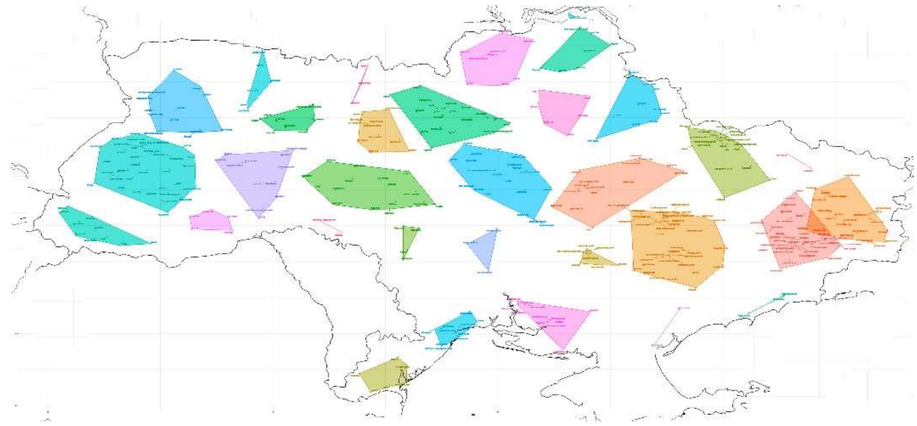


Figure 11. The illustration for DL (Manufacture of electric, electronic and optical equipment) industry the clusters with the 100km radius

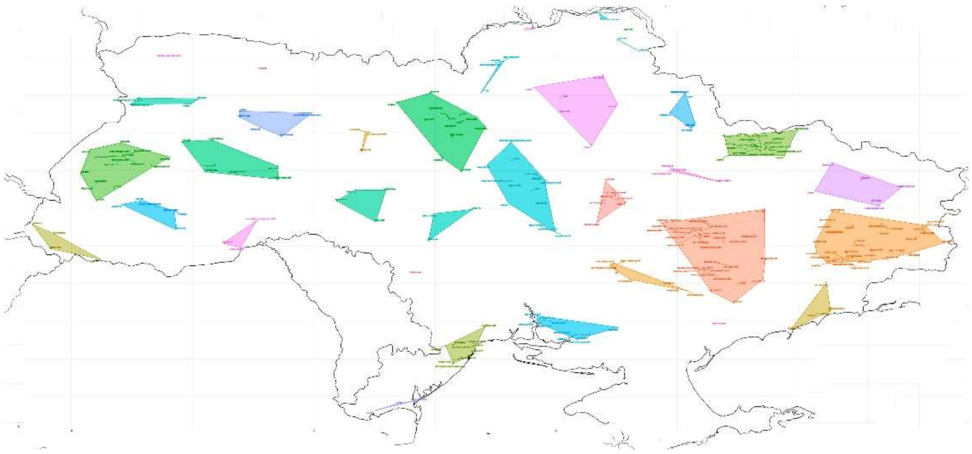


Figure 12. The illustration for DM (Production of vehicles and equipment) industry the clusters with the 100km radius

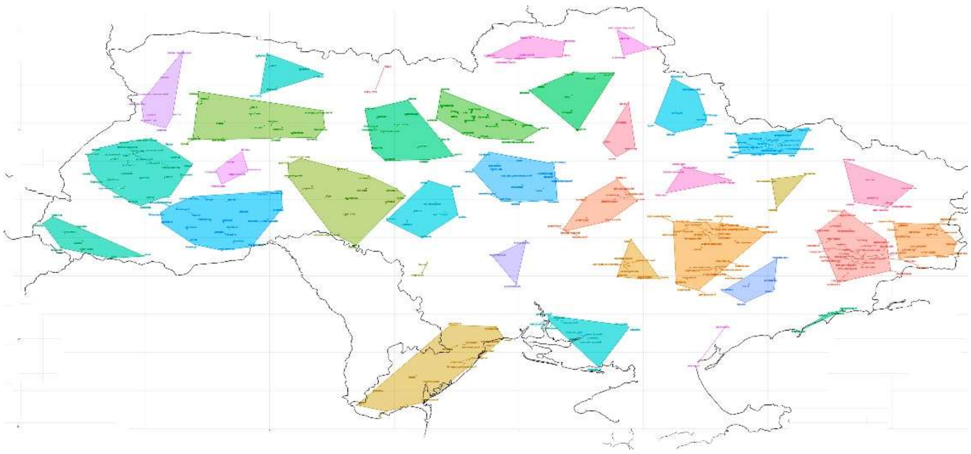


Figure 13. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with the 100km radius

APPENDIX C

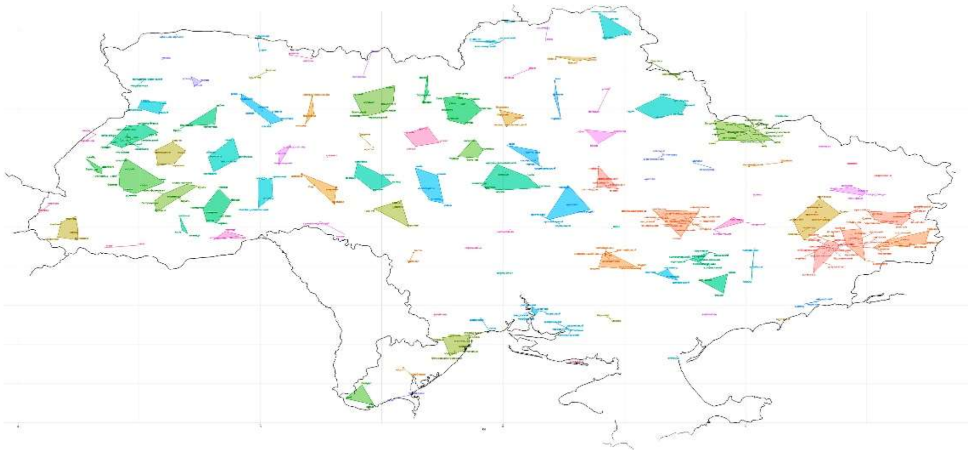


Figure 14. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with the 50km radius

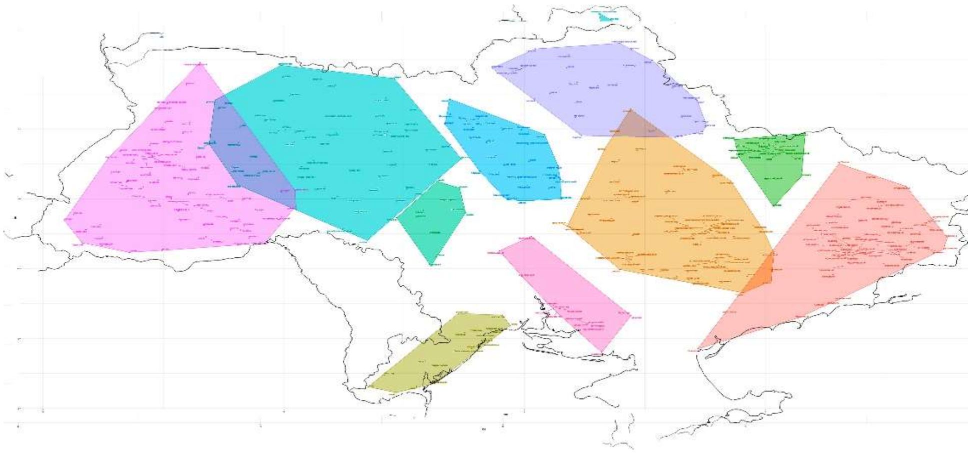


Figure 15. The illustration for industry "DK Manufacture of machinery and equipment" the clusters with 200km radius

APPENDIX D

Table 12. The main variables descriptive statistics in 2001

Statistic	N	Mean	St. Dev.	Min	Max
output	23,522	4,662	64,316	0.10	5,421,435
empl	23,522	100	721	0	55,625
tasset_eoy	23,522	2,773	31,228	0.000	2,680,333
matcost	23,522	3,095	47,238	0	3,931,230
area	23,522	241	272	2.7	848
population	23,522	728,591	869,926	3,907	2,611,327

Table 13. The main variables descriptive statistics in 2013

Statistic	N	Mean	St. Dev.	Min	Max
output	18,643	35,052	439,198	0.10	28,272,460
empl	18,643	69.3	520	0	31,917
tasset_eoy	18,643	12,369	260,926	0	30,398,920
matcost	18,643	20,972	320,113	0.000	22,103,893
area	18,643	214	330	2.700	848
population	18,643	651,998	1,020,989	3,907	2,611,327

APPENDIX E

Table 14. The input-output table for 2014

industry	DA	DBC	DDE	DF	DG	DH	DI	DJ	DL	DK	DM	DN
DA	79%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DBC	4%	34%	2%	1%	2%	0%	1%	4%	0%	1%	2%	10%
DDE	29%	0%	31%	0%	3%	1%	2%	1%	1%	0%	0%	3%
DF	7%	0%	1%	3%	4%	0%	2%	81%	1%	0%	0%	0%
DG	15%	1%	8%	3%	42%	8%	3%	3%	1%	1%	0%	1%
DH	25%	1%	3%	0%	3%	21%	2%	4%	3%	2%	2%	2%
DI	10%	0%	1%	0%	1%	1%	18%	21%	1%	1%	0%	1%
DJ	5%	0%	1%	0%	1%	2%	2%	69%	7%	13%	9%	6%
DL	8%	0%	1%	0%	1%	0%	2%	7%	44%	5%	9%	11%
DK	5%	0%	1%	1%	1%	0%	2%	10%	6%	13%	13%	5%
DM	6%	0%	1%	0%	1%	0%	1%	6%	1%	1%	45%	9%
DN	2%	0%	1%	0%	1%	0%	1%	5%	1%	1%	1%	13%

APPENDIX F

Table 15. Regression results of TFP on the urbanization effect for DK (Manufacture of machinery and equipment) industry

	50km	100km	200km
Urbanization	0.297***	0.346***	0.331***
	(0.066)	(0.053)	(0.032)
Constant	−2.368***	−2.748***	−2.616***
	(0.525)	(0.420)	(0.257)
Observations	6,177	6,311	7,284
R^2	0.003	0.007	0.014
Adjusted R^2	0.003	0.007	0.014
Residual Std. Error	1.259 (df = 6175)	1.255 (df = 6309)	1.252 (df = 7282)
F Statistic	20.361*** (df = 1; 6175)	42.802*** (df = 1; 6309)	103.819*** (df = 1; 7282)

Note: *p<0.1, **p<0.05, ***p<0.01

Table 16. Regression results of TFP on the intra-industry effect for DK (Manufacture of machinery and equipment) industry

	50km	100km	200km
Intra-industry	−0.014***	−0.013***	−0.005***
	(0.002)	(0.001)	(0.0003)
Constant	0.099***	0.126***	0.192***
	(0.019)	(0.020)	(0.019)
Observations	6,177	6,311	7,284
R^2	0.013	0.018	0.033
Adjusted R^2	0.012	0.018	0.033
Residual Std. Error	1.253 (df = 6175)	1.248 (df = 6309)	1.240 (df = 7282)
F Statistic	78.980*** (df = 1; 6175)	116.237*** (df = 1; 6309)	252.167*** (df = 1; 7282)

Note: *p<0.1, **p<0.05, ***p<0.01

Table 17. Regression results of TFP on the intra-industry effect for DK (Manufacture of machinery and equipment) industry

	50km	100km	200km
Potential market	-0.031^{***}	-0.030^{***}	-0.009^{***}
	(0.003)	(0.003)	(0.001)
Constant	0.113^{***}	0.136^{***}	0.182^{***}
	(0.020)	(0.020)	(0.020)
Observations	6,177	6,311	7,284
R^2	0.013	0.017	0.022
Adjusted R^2	0.013	0.017	0.022
Residual Std. Error	1.252 (df = 6175)	1.249 (df = 6309)	1.248 (df = 7282)
F Statistic	84.463^{***} (df = 1; 6175)	109.117^{***} (df = 1; 6309)	162.295^{***} (df = 1; 7282

Note: *p<0.1, **p<0.05, ***p<0.01

APPENDIX G

Table 18. Summary statistic of Agglomeration effects at DK industry in Kyiv cluster in 50, 100, 200km radius

Radius	Effects	N	Mean	St. Dev.	Min	Max
50km	Urbanization Intra-industry Potential markets	50km	184 628 168 666 168 666	8.01 52.17 37.70	4.64 0.03 0.53	9.50 847.23 893.08
100km	Urbanization Intra-industry Potential markets	100km	217 345 196 111 196 111	7.92 83.33 62.07	4.64 0.0001 1.22	9.50 836.03 771.86
200km	Urbanization Intra-industry Potential markets	200km	217 345 196 111 196 111	7.92 83.33 62.07	4.64 0.0001 1.22	9.50 836.03 771.86