National Research University Higher School of Economics

as a manuscript

Zyuzin Alexander Vladimirovich

RUSSIAN ENTERPRISE PERFORMANCE: INDUSTRIAL AND REGIONAL AGGLOMERATION EFFECTS

PhD Dissertation Summary

for the purpose of obtaining academy degree

Doctor of Philosophy in Economics

Academic supervisor:

Doctor of Science in Economics

Demidova Olga Anatolyevna

JEL: C21, C51, C53, R11, R12, R15

Moscow – 2023

Introduction and main definitions

Dissertation thesis is dedicated to the issues of modelling of agglomeration effects for private real sector Russian companies[1]. Classical approach classifies agglomerations by the way of their composition onto *localization* (or equally *concentration / industry cluster*) economies and *urban economies* (*diversity economies*). Concentration economies consist of firms that belong to the homogenous industries and locate in the same geographical region. Urban economies are mostly related to diversified multi-industry economy of the cities and sometimes larger objects (regions, states, small countries).

In the present work the following terminology is used: agglomeration externalities of localization economies are called *concentration effects* (the degree of *concentration* is expressed in terms of distribution uniformity of industries' labor across regions) and urban externalities are considered from the angle of benefits for the firms that arise due to the regional economy diversification – *diversification effects*. The major geographical units in this study are regions rather than cities, which is sometimes justified as a balance between data availability and the ability to find sensitive responses of firms' indicators to diversification externalities. *Agglomeration effects* in the context of this thesis include jointly *concentration* and *diversification* externalities.

The thesis analyzes an important relation among the agglomerations and the firms' performance. More precisely, an attempt is made to numerically estimate the influence of concentration and diversification effects on the sales margin (SM) and net profit margin (NPM) indicators of the real sector private companies.

The present study also includes a discussion of the additional questions that allow to discover more details about the relation between the agglomeration effects and the enterprises' operational efficiency. For example, the influence of the agglomeration externalities on the firms' performance is compared for the entities that differ by their

[1] A limitation of the analysis to consideration of only private real sector companies is done due to the following reasons: (a) financial entities' statements differ a lot in terms of content and analytical approaches; (b) non-commercial and governmental-held companies (or entities, where the government owns a significant share of capital) might have different from classical profit maximization goals (e.g. realization of socially important programs and projects, support for social and government initiatives, funding of public goods etc.).

size, industry, initial performance, or geographic location (inside or outside of the industry cluster).

To estimate the degree of industry concentration the Ellison-Glaeser (or EG) index (Ellison, Glaeser, 1997) is used, and to get estimates of the degree of regional economy diversification the Herfindahl-Hirschman index (or HHI) is applied. The calculation of the indices is based on the information about the number of employees. To preserve the benefits of working with the population of entities any missing values in the input variable (employees) were recovered with the estimations, obtained using the available characteristics of the company, including productivity, industry and business scale.

The relevance of the work

The so-called "*tradable*" industries are most sensitive to the implementation of regulation (Porter, 2003). Tradable industries include those, located in one geographical location (region) and can carry out activities (compete, sell goods and services) in other regions. Local industries are opposite to tradable as they can sell their goods and services only within their city or region of residence. Among the industry groups compiled in the present thesis two ("IT and telecommunication" and "Manufacturing industries") are the most fully consistent with the concept of "tradable" industries. Other industry groups can be called *partially tradable*, as they include both tradable and local industries: "Agriculture, fishing and forestry", "Mining", "Transport", "Construction", " Wholesale and retail trade, food service", "Services".

The practical application of the research results is largely determined by the answer to the question – where the policy of regional specialization or industry concentration is the most efficient. Which regions are the best for introduction of stimulating regimes and which industries to stimulate or subsidize? It can be assumed that the study is primarily applicable for such tradable industries, which are steadily influenced by the agglomeration effects.

The study is also relevant for partially tradable industries, but with a number of limitations that require further study. For example, it is important to find industries

(according to their 2-digit OKVED codes, so that are parts of industry group) that are the most affected by the concentration and diversification effects. It can be assumed that, in some cases, stimulating policy may also be useful for the following partially tradable industry groups: “Agriculture, fisheries and forestry” (fishing companies are tradable), “Transport” (excluding urban transport, the industry is considered to be tradable) and “Mining” (service and servicing companies, that are also parts of this industry group, are tradable).

Regional and sectoral policy in Russia is implemented using an important tool of preferential regimes. The Accounts Chamber of the Russian Federation reports[2] about three effects arising from the establishment of a special regime: an increase in profits for resident companies in a region, where a preferential regime is established; an increase of certain regional socio-economic indicators (it is assumed that in such a region, due to the stimulating industry policy, the degree of concentration of enterprises that belong to the subsidized industry will increase); positive external spatial effects for neighboring regions.

It is important to provide some details on how Russia implements a stimulating policy, aimed at the development of industries and regions. Such a policy includes a set of regulatory methods: subsidies, tax breaks, simplification of administrative and bureaucratic procedures, customs benefits, advisory support and a number of other measures. The mechanism for implementing such a stimulating policy is realized via the creation of special and particular economic zones (SEZ and PEZ) and territories of advanced socio-economic development (TASD). The difference between them lies in the range of subsidies provided, regulatory mechanisms and status.

“Now there are 43 particular economic zones in Russia and their significant share is aimed at stimulating industrial production (24 PEZs), and the high-tech and IT sectors (7 PEZs) (Federal Law of July 22, 2005 N116-FL "On the Special Economic Zones"). Among all the PEZs 17 were introduced between 2019 and 2022, which indicates the

[2] Report "Analysis of the mechanism of establishing and functioning of preferential regimes as a tool for socio-economic development and foreign economic policy", approved by the Board of the Accounts Chamber of the Russian Federation on December 21, 2021 (URL: https://ach.gov.ru/checks/prefregime).

great importance of stimulating regional policies. There are currently 117 of TASD introduced in Russia, which makes them also a widespread regulatory tool. In total, there are more than 160 preferential regimes operate in Russia.

Covering notes to the corresponding laws (concerning preferential regimes) state direct goals of implementation PEZ, SEZ and TASD programs (including job creation, infrastructure development, exploration of mineral extraction sites), while externalities (in particular - agglomeration effects) arising from the growth in one or more industries in certain regions are neglected. However, agglomeration externalities can be a significant part of the general success of an enterprise in a PEZ, SEZ or TASD.

Stimulating regulation of industries and certain territories, based on the introduction of preferential regimes, is a widespread global practice. The Accounts Chamber of the Russian Federation analyzed 36 countries of Europe, Asia, North and South America and found out that the introduction of analogues to Russian SEZ, PEZ and TASD is a common practice. However, the preferential regimes in foreign countries differ much in terms of operational functioning, conditions for granting subsidies and regulatory approaches. At the same time, all preferential regimes despite of the way of their regulation, provide more favorable environment for business entities and investors compared to unsubsidized territories. European Union rarely applies preferential regimes, while in China and India this is the most widely used tool of regional and industry regulation (more than 55% of all preferential regimes in the world). In general, developing countries with emerging economies use preferential regimes far more frequently than developed countries, where existing stimulating programs are maintained, but the process of introducing new ones is fading. The stimulating mechanisms, established for residents of territories, where preferential regimes are introduced in Russia, are similar to those most often applied globally - tax incentives and co-financing of large investment projects aimed at infrastructure development.

An important is to note that only predictable agglomeration effects with stable appearance, direction of influence and well-studied behavior can be taken into account when planning an introduction of some preferential regime. That is, key important is to check whether the agglomeration effects are a stable phenomenon – appear each period

during a significant range of time. Secondly, the direction of influence of such effects should be predictable and stay unchanged over long time periods. Thirdly, it would also be good to track the dynamics of the magnitude of the agglomeration effects over time - whether it increases, decreases, or remains statistically unchanged.

Considering that an integral part of this work is the analysis of the stability of agglomeration effects, all the above makes the study and its results relevant and practically applicable.» (Zyuzin, Demidova, 2022). The results obtained in the work allow, assuming certain limitations, to make predictions about the effectiveness of stimulating regional and industry economic policy. Owing to versatile assessments (by company scale, industry and initial performance) of the impact of concentration and diversification effects on the firms' performance, as well as to the wide coverage of considered industries allows to use obtained results as a starting point or benchmark in solving more global problem – to describe an optimal distribution of industries by region.

The estimates of industry concentration and regional economy diversification are based on the information about the average number of employees per annum. These estimates are done using the general population of Russian private real sector enterprises (or its close approximating sample). Missing values (if any) in such data were filled in with the estimates obtained using the information about the enterprise, which includes available data about the effectiveness, content and scale of firms' business.

A number of other direct and indirect reasons justify the novelty, relevance and importance of the research question. First, alike studies of agglomeration effects have often been performed for a number of foreign economies. As to the Russian economy and enterprises – only a few major works are known so far.

Secondly, there is a lot of freedom (a lot of different approaches are applied in the works devoted to agglomerations) in choosing methodologies for defining, evaluating and comparing agglomerations. This situation creates an opportunity to consider operational efficiency as a dependent variable and propose new approaches of modeling

the impact of agglomerations on it in this work. The idea of linking agglomeration effects with the operational efficiency is quite popular (Lu, Tao, Yu, 2012; Davidson, Mariev, 2015). However, in the context of Russian local market the study that links SM, NPM and agglomeration effects, is new and may be of interest.

Thirdly, Russia is a very good subject for research on the agglomeration economy. Its' regions significantly differ from each other by the number of economical and natural reasons – an access to the sea, administrative or financial resources, an availability of the minerals, different natural and climatic conditions (soils, forests, rivers, borders etc.) create prerequisites for the formation of industry clusters in different regions. Thus, there are processes of orientation of some regions to a certain type of economic activity, which would provide them relative production advantages. For example, an access to the sea provides employment in the fishing or aquaculture industry; good soils provide incentives for farming, and so on. Other regions develop uniformly and become urban industrial centers, having more diversified economies with highly developed manufacturing, trade, R&D and services. First of all, these are federal cities and their neighboring regions. These mechanisms stay behind the processes of the regional differentiation by the degree of the diversification of their economy. A natural question arises - how this difference in the economic structure of regions affect the performance of enterprises in it? In the present thesis such differences between regions are controlled via the introduction of variables that are proxy characteristics of regional investment climate (Table 4).

Finally, the distribution of economic activity in Russia was much influenced by the USSR planning economy. However, the modern research found major structural shifts in the distribution of Russian companies – in the Soviet period concentration of companies was artificially decreased (randomization of the distribution of economic activity) and in contrast during the period from 1989 to 2013 years the major part of inefficiently located companies relocated or closed. By 2013 the concentration of economic activity in Russia became alike to the leading market economies (like Germany, Japan or UK) (Mikhailova, 2016).

Now, the distribution of economic activity across Russian regions is not random. Depending on the regions' specifical traits and peculiarities, there might appear some industry clusters. For example, the industry of fishing is the most developed and largely represented in the regions with the direct access to the sea (or open waters). High-tech industries need relatively large labor market, which makes it possible to quickly find highly qualified specialists and develop along with the growth of the R&D investments (in Russia high-tech and IT companies are mostly concentrated in Moscow, Moscow region, St. Petersburg and Leningrad region). Higher degree of the industry concentration determines the presence or absence of the industry clusters, which are the source of the agglomeration externalities of concentration. These effects have not been previously assessed for Russian enterprises and the task of doing that seems to be relevant and interesting now.

Brief literature review

The problem of agglomeration economies was widely discussed in the literature and there was suggested a detailed theoretical description of the origin and causes of various agglomeration effects. Also, there were collected a lot of empirical evidence either supporting or contradicting to the theoretical models and concepts. The overall result, however, is that in the most of papers there was confirmed the existence of certain agglomeration effects. A distinctive feature of the empirical papers on the issue of the agglomeration economies is vastly varying research methodology, including the freedom in the choice of modeling techniques, models' specifications and dependent variables, which can significantly differ even in the papers, where the common goal had been declared. Basically, such studies are devoted to the European and the USA cases (references are given below in the text). The similar studies were also conducted for the other countries but much less frequently.

The theoretical framework for concentration economies is usually associated with Alfred Marshall's *Principles of Economics* (Marshall, 1920), where he in fact illustrated the possibility of the existence of positive agglomeration externality effects for companies located in the industry cluster. The theory was developed during the XX

century and enriched by a number of works, and the main ideas were summarized and formalized (Rosenthal, Strange, 2004; 2006).

The positive agglomeration externalities that an enterprise can get working within a geographic cluster arise due to the potential cost reduction. There are three main sources of savings from geographical location:

(a) Better access to both suppliers and consumers results in a decrease of logistics costs and economic disintegration, and thus the firms also gain benefits from this externality (Cohen, Morrison Paul, 2009; Holmes, 1999);

(b) Labor pooling results in the opportunity for the firms to hire employees at lower costs. The existence of this channel has been shown in a number of works (Baumgardner, 1988; Cohen, Morrison Paul, 2009; Nakamura, 1985);

(c) Knowledge spillover effects boost technical progress due to enhanced and intense information and knowledge exchange, and cooperation in knowledge production (Audretsch, 1998; Audretsch, Feldman, 1996; Morrison, Siegel, 1998; Wheaton, Lewis, 2002).

Agglomeration effects that appear via (a)-(c) channels are also caller MAR effects (Marshall-Arrow-Romer effects). The channels (a) and (b) have quite straight and simple mechanisms, while (c) might seem a little confusing taking into account the modern rates of technological (including telecommunication) progress. The first, who criticized channel (c) was Paul Krugman (Krugman, 1991), who supposed that the knowledge can be transferred by distance through modern remote communication channels. In further studies, however, there was demonstrated the expediency of separating the concepts of *knowledge* and *information*, so that only the information can be transmitted remotely, while personal communication is often needed for the effective knowledge exchange (Audretsch, 1998; Fischer et al., 2006; Glaeser et al., 1992; von Hippel, 1998).

It was also found out that the recipients of the information assimilate the distributed information better in a local cluster rather than at a physically long distance. That is,

personal contact increases the effectiveness of communication (Audretsch, 1998). Thus, all three channels matter even in modern realities[3].

It should be noted that the agglomeration effects might also be negative for the enterprises. Basically, this impact is related to the problem of growing competition in the cluster. For Chinese manufacturing companies it was demonstrated that the companies' market markups[4] are shrinking in large agglomerations as a result of increased competition (Lu, Tao, Yu, 2012).

Nevertheless, the natural assumption is that the existence of negative agglomeration effects will not always be confirmed and not for the enterprises of all sectors of the economy. The negative effects might appear if the enterprise has only a limited local market for sale of its' goods and services (for example, a local furniture factory or a small retail). Once this condition is violated, the negative effects will most likely not be found. This can be best demonstrated for easily scalable IT & telecommunication companies. Preserving all the advantages of the agglomeration, these companies compete among each other in a rather limited extent. Their potential consumers (individuals or entities) might locate far beyond the companies' cluster (sometimes potential demand might appear anywhere in the world), and due to the high rate of technological progress the range of products, that these companies offer, is rapidly expanding (product differentiation processes). Considering constraint production capacities and the inability (with rare exceptions) of a single company to meet the entire global demand for a product, in line with the fact that the supply of products is constantly expanding with new innovative goods and services, it can be considered that the residual demand is large enough to eliminate price pressuring effects of competition, which lead to the consequent negative agglomeration effects for the enterprise.

[3] Despite the development of digital channels of transmitting information, the role of personal communication cannot be ruled out. First, remote channels are more expensive, implying the payments of the costs of communication. Secondly, personal communication is more effective and can cover a wider range of issues to discuss than those that can be touched remotely. For example, when deciding whether to invest in a company, investors prefer face-to-face meetings, and discussing the results of laboratory experiments one may often require a personal look at the external conditions, created for conducting the experiment.

[4] The ratio of price to marginal costs. To estimate the markup authors follow the methodology, suggested in De Loecker, J. and F. Warzynski. 2012. "Markups and Firm-Level Export Status", American Economic Review

Moreover, given the global geography of sales of products and services, even if competitive pressure appears, its' reason may differ from the agglomeration.

The basic principles of urbanization economies, and the externalities that arise in connection with the formation of urban agglomerations, were described by Jane Jacobs for the New York (Jacobs, 1961). As already mentioned, urbanization economies may be considered beyond the boundaries of a single city. Externalities of urbanization economies differ significantly from MAR concentration effects. Institutional effects play the major role in urban economies (while for MAR, the effects are operational): less exposure to industry shocks; incentives to innovate and offer a wider range of products and services; lower unemployment (including lower terms for finding a job when considering the case of the natural unemployment); higher quality of life of the population, including the development of the social infrastructure, transport, and a high level of education and medicine.

There are several basic directions of the empirical research in agglomeration economics. The most popular are the attempts to relate the innovative activity of enterprises to its' geographical location (inside or outside the cluster), and characteristics of the cluster, where the enterprise works (industry or urban cluster; are there regional or country boundaries between clusters, etc.). There are studies of the cluster impact on the other characteristics of the enterprise, including export activity, market markup, enterprise growth rates, or transportation costs. Urban agglomerations (cities, regions, states, countries) are often considered in terms of the dynamics of their own development and growth. Such research usually demonstrate that the presence of a cluster accelerates the rate of growth of the gross city or regional product, and the agglomeration is growing faster than the country on average (Ketels, Protsiv, 2020; Quigley, 1998; McCann, van Oort, 2019; Tao, Huang, Tao, 2020). The results of the main empirical studies of industry and urban agglomeration effects are summarized in Table 1.

Table 1

Summarized results of the main empirical research of the concentration and urbanization agglomeration effects (based on the foreign data sets)

Research (authors, year)	**Dependent variable**	**Key explanatory variable(s)**	**Results**	**Notes**
Innovative activity and agglomerations				
Baptista, Swann (1998) *Do Firms in Clusters Innovate More?*	Number of innovations.	Industry HHI with shares of employees, working in each studied region, as weights.	Propensity to innovate increases for the companies inside of their industry cluster.	The results of innovative activity of 248 manufacturing companies studied over the period from 1975 to 1982.
Bell (2005) *Clusters, Networks, and Firm Innovativeness*	Estimation of innovation activity was based on the 3 indicators: new products, new services, implementation of new production technologies.	The results were based on the interview answers received from the industry experts.	Cluster and coordinated management positively affect the innovative activity.	Innovative activity was estimated by the industry experts using the scale from 1 to 5.
Hervas-Oliver et al. (2018) *Agglomerations and Firm Performance: Who Benefits and How Much?*	Number of new products or patents and/or the firm.	Localization coefficient, based on the regional and industry employment.	Localization increases firms' innovativeness.	Authors studied 6.7 thousand of private Spanish companies in 2001.
Рост фирм в кластере				
Li, Lu, Wu (2012) *Industrial Agglomeration and Firm Size: Evidence from China*	Company size calculated as the log of the employees.	Localization of the firms' neighborhood expressed through the log of the employees laboring for the firm's neighbors.	Localization increases firms' growth rate. The larger are neighbors the higher is the firms' growth rate.	Chinese manufacturing companies, data set for the period 1998-2005.
Agglomerations and the firms' performance				

Ciccone, Hall (1996) *Productivity and the Density of Economic Activity*	Revenue per a unit of labor.	Spatial concentrations expressed through the industry labor density in each studied USA state.	Double spatial concentration leads to a 6% increase of per capita productivity on average.	Observations are states (51 in total). The authors relate the average density of labor to the average level of per capita productivity in the industry in the state.
Knoben et al. (2016) *Agglomeration and Firm Performance: One Firm's Medicine Is Another Firm's Poison*	Operational income per a unit of labor.	Localization is measured using the data about the density of distribution of industry employment across regions.	Cluster effects can be positive or negative depending on certain industry group or regional characteristics.	The research was conducted for the Netherlands' companies.
Martin, Mayer, Mayneris (2011) *Spatial Concentration and Plant-Level Productivity in France*	TFP (total factor productivity)	Industry localization and city economy diversification.	No evidence of relation between city economy diversity and TFP was found. Was found positive industry localization effect.	Study of French manufacturing companies in the period 1996-2004. In total 126.7 thousands of firms.
Other aspects of agglomeration effects				
Malmberg, Malmberg, Lundequist (2000) *Agglomeration and Firm Performance: Economies of Scale, Localization, and Urbanization among Swedish Export Firms*	Gross monetary volume of export.	Industry localization and the degree of urbanization of firms' actual geographical location.	Localization and urbanization externalities both increase firms' export activity.	The research is based on the data set that includes 10 thousand of Swedish exporters in 1994.

Holmes (1999) *Localization of Industry and Vertical Disintegration*	The degree of the vertical disintegration (expressed as a ratio between factor costs and revenue).	Localization is measured as a density of employees in the surrounding companies.	Localization increases the degree of the vertical disintegration (the larger are production chains)	Author studied private USA companies in 1987. In total 368 thousands of observations.

Source: composed by the author

Empirical research in Russia

The problems of agglomeration economies in Russia discussed a lot in the early 2000s, when Russian statistics had been accumulated a data about the functioning of the economy in the post-Soviet period. The heated discussions continued until the mid-2010s. The researches were primarily interested in the consequences of the transition from a command to a market economy for the organization of economic activity in the country.

First of all, the changes could occur in the distribution of resources between industries: some industries could shrink relative to others - relatively increased. For example, it was demonstrated that despite the high costs of redistributing resources between regions, the spatial structure of production in Russia has tendencies to change. The formation of production centers can occur both around the industry leaders of the Soviet era, as well as in the regions that were not significant at the dawn of a market economy in Russia (Golovanova, 2008).

Secondly, changes could occur in the geography of the distribution of industries across the country regions. As noted in the work of Golovanova S.V. and Kadochnikov S.M. from 2011, the Soviet model of economic development assumed the widespread use of the Territorial Production Complexes (TPCs), which is well applicable only to the resource extracting industries and for the regions with low population density and weak development of the manufacturing. Market economy created the prerequisites for changing the behavior of the economic agents and incentivized the development of the new concept for the formation of clusters and the distribution of industries across Russian regions (Golovanova, Kadochnikov, 2011).

There are other works that have studied the principles of the formation of agglomerations in Russia. There was demonstrated an existence of the rapidly growing manufacturing industry clusters developing around the large urbanization centers (Mikhailova, 2016). In another work (Kolomak, 2015), there were considered the principles of clustering and the issues of “choosing” geographical places for cluster formation. Thus, often the dynamics of development and firms’ migration of the large

agglomerations of the USSR, formed in the period of a command economy (often in the form of a TPCs), is compared to the period after the establishment of a market economy in Russia.

Compared to the foreign literature, it can be said that there are relatively few studies based on Russian data, however, one can find works dedicated to the classical areas in the economy of agglomerations: urbanization and concentration. The main limitations in the analysis of agglomeration effects in Russia are short time series and panels, which in combination with quite a limited number of considered industries makes it difficult to conduct a comparative cross-industry analysis.

«A study of Russian regions revealed that regional specialization (that is, low diversification of the regional economy), measured with the Krugman Specialization Index, is effective only if there is a resource base, otherwise specialization will be economically inefficient (Rastvortseva, Chentsova, 2015). The opposite result was obtained in the article (Kolomak, 2015), where there was discovered that regional specialization leads to an increase in the GRP.» (Zyuzin, Demidova, Dolgopyatova, 2020).

N. Davidson and O. Mariev in their paper (Davidson, Mariev, 2015) considered both MAR (similar to the concentration effects in this paper, although the modeling is significantly different) and urban effects. The authors studied the performance of Russian manufacturing firms for the period from 2002 to 2008. (7 111 enterprises in total) and found a U-shaped relation between the logarithm of revenue and the degree of the localization effect. The localization was measured as the ratio of the total revenue of the entire industry in the constituent entity of the Russian Federation to the revenue of a particular firm in this industry in the region. The authors also revealed the presence of positive diversification effects for the firms' revenue. Previously, similar results were obtained by a team of authors for manufacturing firms between 2001 and 2004 (Vorobyev et al., 2014).

Another paper (Skhvediani, Sosnovskikh, 2020) examined spillover effects in the Russian economy, measured as the density of implemented innovations per worker. The authors studied the relationship between companies' innovation activity and relative degree of the concentration (in terms of the number of employees, revenue, investments and wages) and found that localization positively affects the density of usage of the innovations, and the regions with a high density of innovations locates in the neighborhood to each other. The study used panel data covering 83 Russian high-tech enterprises from 2009 to 2018. (830 observations in total).

In other research, conducted at different times, where different modeling techniques were applied, there was discovered the existence of the positive agglomeration MAR effects for Russian enterprises (Bessonova, Gonchar, 2017; Gonchar, 2009). A positive impact of urban agglomerations (diversification effects) on the output of enterprises was found for manufacturing enterprises[5] (Gonchar, Ratnikova, 2014).

Despite the empirical results pointing to the significant role of agglomeration effects for enterprises and regions, a survey of business and government representatives revealed that concentration effects are noticeable not for all industries. The authors stated two types of interaction within the cluster: "cooperative" and "interconnective"[6] and recorded the opinions of the representatives of the engineering and IT industries (as well as government officials) about their assessment of the degree of cooperation and interconnection among firms within the cluster in which their enterprise operates. It turned out that IT companies are more inclined towards cluster cooperation and are ready to invest in the development and strengthening of cluster relationships, unlike vehicle manufacturing enterprises. At the same time, there is a potential for an increase in the degree of influence of agglomeration effects, since cluster relationships in Russia were much less developed than in Europe at least when the research was being

5 An interview of 957 firms' representatives, conducted in 2009 in 357 Russian cities located in 48 regions.

6 This classification does not contradict those used in the thesis and differs only in that it characterizes the relationship between firms in a cluster, and does not address to any specific agglomeration effects. So, if the spillover effects are more typical for "cooperative" relations, the effect of logistic costs reduction is a demonstration of the benefits of the interconnections. At the same time, the classic agglomeration externality of the contraction of labor markets applies to both types of relations. At the same time, the classic labor pooling agglomeration externality applies to both types of relations. For example, some enterprises agree to invest jointly in the training of future personnel, then this is an example of collaborative relations, and the migration of trained specialists to a large cluster or the implementation of state educational programs in the regions, where the industry is concentrated, is better characterized by the concept of interconnective relations within the cluster.

conducted (Golovanova, Avdasheva, Kadochnikov, 2010). Thus, it is very important to consider agglomeration effects in the industry studies.

Thus, after reviewing the existing literature, it is possible to identify several important gaps that this study could fill.

Table 2

The uncovered gaps in the existing literature and the contributions to be done

Was done before	Contribution
There are estimations of the influence of the concentration and diversification effects on the Russian companies' innovative activity (Skhvediani, Sosnovskikh, 2020). The studies based on the foreign data also consider the influence of localization and diversification on the exporting volumes (Malmberg, Malmberg, Lundequist, 2000), firms' growth (Li, Lu, Wu, 2012), market markups (Lu, Tao, Yu, 2012), trade and investments (Knoben et al., 2016; Ciccone, Hall, 1996).	Quantify the influence of the concentration and diversification effects on the Russian enterprise performance.
The studies were conducted for a few separate industries and were based on samples of data. Typical sample size was about 7-10 th. of observations (including papers, mentioned above).	Offer and apply an approach of working with the population of private real sector companies, which allows precise estimation of concentration and diversification indices.
Business scale was not accounted directly but only via indirect firms' characteristics; no cross-industry analysis conducted; factor of initial difference in the performance was not considered.	Consider business scale, industry and initial performance when estimating the influence of the concentration and diversification effects on Russian companies' performance.
Dynamics was analyzed for cities, regions, countries (Ketels, Protsiv, 2020; Quigley, 1998; McCann, van Oort, 2019; Tao, Huang, Tao, 2020), but not for the enterprises. Firms were analyzed using relatively small	Study the dynamics of the concentration effects and identify increasing or decreasing trends, stability of sign and significance.

samples with the help of the regression models including FE and RE (which is poorly applicable for bigger samples or the population, when the panel becomes strongly disbalanced).	
In previous works the clusters were identified, described and measured (in terms of their location, borders or growth or the reasons of formation (Aleksandrova et al., 2019)), but the influence of the degree of concentration was not compared for the firms inside and outside of the cluster.	Study the difference in the influence of the concentration effects inside and outside of the industry cluster.

Source: composed by the author

Based on the above, the purpose was set and research objectives were formulated.

Goals and objectives of the study

The main goal of the thesis is to conduct a comparative analysis of influence of the industry concentration and regional economy diversification agglomeration effects on the enterprises' (belonging to different real sector industries) performance.

To achieve the goal, the following objectives are necessary to complete:

1. Collect and process the data set (including filling the missing values in the «*Employees*» variable and calculating indices of industry concentration and regional economy diversification), which includes the information from the balance sheet statement, profit and loss statement and the number of employees for each company of the population of Russian private real sector companies;
2. Estimate the influence of the concentration and diversification agglomeration effects on different business scale firms' performance;
3. Obtain the estimates of the influence of concentration and diversification agglomeration effects on companies depending on their industry;
4. Estimate quantile regression models to test for the differences in the degree of influence of the agglomeration effects on companies with varying initial performance.;
5. Study of over time stability and dynamics of changes in agglomeration effects;

6. Identification of clusters and comparison of the results from p. 5 for enterprises located inside and outside their clusters.

Methodology

The thesis consists of 4 chapters. The first one covers general methodological issues related to all subsequent parts of the study (data and work with missing values, determination of firms' performance, formalization of the agglomeration effects, aggregation of the industries into groups). In the remaining chapters the stated problems are being solved, obtained results discussed and applied approaches presented. This section provides a brief description of the methodology used in the work.

The data

During the work on the thesis, new interesting questions arose, and to answer them it was required more data than had been stated at the beginning. Therefore, two sets of data are used sequentially in the work. Thus, all calculations (pp. 1-6 of objectives section) were conducted using the most up-to-date data about Russian enterprises available (data for only 2017 were used for points 2-4, and for points 5-6, the data set for 2011-2018 was used).

The first sample is cross-sectional data for 2017 and includes information about the balance sheet, income statement, average annual number of employees, and some institutional characteristics of Russian regions as the variables of control. The observation is an enterprise (firm-level data). Total number of observations, included in the data set after processing the data was 647,697. Since only shell companies and obvious errors were removed from the data set, and the missing values in the 'Employee' variable were restored, this data set can be considered an approximate estimate of the general population of Russian private real sector firms. The core of the data base was kindly shared by the Center of Macroeconomic Analysis and Short-Term Forecasting (CMASTF) and then expanded with Rosstat (Federal Department of Statistics) and Federal Tax Service open databases. Regional characteristics were taken

from the periodical review "Regions of Russia" and electronic databases prepared by Rosstat.

To answer questions about the dynamics and stability of MAR effects, the database was significantly expanded. The main source of the data was Ruslana Bureau van Dijk (BvDR) database, which made it possible to form panel data set up with for 8 years duration from 2011 to 2018 (the most up-to-date data at that moment). The setup was the same: information about the financial and economic conditions of Russian enterprises (including revenue, operating costs, debt, profit, average annual number of employees, statistical identifiers, etc.). After processing raw data, the sample consisted of 1.3 million observations for the period from 2011 to 2018. The number of observed companies differs in each reporting year, ranging from 83 to 237 thousands of observations.

Aggregation into industry groups

Industries that are close to each other by the essence of their economic activity were aggregated into industry groups. The main idea of the aggregation is to enhance the convenience when working with the general population. Considering each 2-digit branch of OKVED separately would lead to the use of 4-digit OKVED codes to calculate concentration effects. Thus, the number of models under consideration would increase up to 70, and taking into account the use of 2 dependent variables it would be necessary to discuss 140 models. The number of models would increase along with the complication of the research objectives, which would be especially noticeable, when estimating quantile regressions to estimate the agglomeration effects for enterprises with different performance. The computational complexity of this problem is the estimation of 1 260 models (140*9, where 9 is the number of quantiles). It would be problematic to compare such a number of models and not very representative as well, so it was decided aggregate industries despite the availability of information even despite of the availability of 6-digit OKVED codes.

An aggregation of the industries into groups took place in accordance with the NACE methodology. The aggregation resulted in the formation of 8 industry groups. Previously, a similar method of aggregation was used in the work (Zyuzin, Demidova, Dolgopyatova, 2020). All the analysis was conducted using this classification (Table 3).

Table 3

Aggregated industrial groups: principles of formation

№ Group ID	Group name	Statistic codes (OKVED) included in the group
№1 (A)	Agriculture, fishing and forestry	01+02+03
№2 (B)	Mining	05+06+07+08+09
№3 (C)	Manufacturing	10+11+12+13+14+15+16+17+18+19+20+21+22+23+24+25+26+27+28+29+30+31+32+33+35+36+37+38+39
№4 (D)	Retail, wholesale and food services	45+46+47+55+56
№5 (E)	Construction	41+42+43+68
№6 (F)	Transport	49+50+51+52
№7 (G)	IT & telecommunication	53+58+59+60+61+62+63
№8 (H)	Services and other minor industries	69+71+73+77+78+79+81+82+95+96

Source: composed by the author

Dealing with the missing values

The share of missing values in the Employees variable varied from 5% to 12% depending on the year (2011-2018). For the largest companies (with assets exceeding several billions of rubles) missing values were restored manually, for the rest there were used estimates based on the available information about the company. Here follows the general approach of missing values estimation:

«

1) All the considered companies were aggregated into the industry groups (in accordance with the Table 3).

2) Missing values were estimated for each year groupwise separately using the linear regression model (OLS[7]) as a forecasting tool:

$Y_{jt}^{obs} = X_{jt}\beta_{jt} + \epsilon_{jt}$, where

Y_{jt}^{obs} – *observed part of the average annual number of employees of the company (for the industry j at the year t);*

X_{jt}^{obs} – *part of the matrix of the explanatory variables, that characterizes firm (including: revenue, capital, inventories, debts of j industry in t year and business scale dummies), which relates to* Y_{jt}^{obs} *in terms of the length and us used to get* $\hat{\beta}_{jt}$ *estimates.*

Obtained estimates $\hat{\beta}_{jt}$ were applied to estimate missing values in the average annual number of employees: $\hat{Y}_{jt}^{mis} = X_{jt}^{mis}\hat{\beta}_{jt}$., *where* – X_{jt}^{mis} part of the matrix with the explanatory variables that relates with the firms included in Y_{jt}^{mis}.» (Zyuzin, Demidova, 2022).

Concentration and diversification measures

Geographic concentration indices are the main tools of quantifying the degree of industry concentration. In the present thesis Ellison-Glaeser (EG) index is used (1) (Ellison, Glaeser, 1997). The EG index was chosen due to a number of advantages over other metrics of distribution inequality, such as the Gini index, Theil index, HHI index, MS index, or Krugman index. One of such advantages is that EG not determines the unevenness of the distribution of a particular industry, but also compares it to other industries in each geographic region, allowing to identify industry clusters. Moreover, estimates of industry concentration obtained using EG for one country are comparable with similar estimates for another. Ellison-Glaeser index also makes it possible to compare concentration effects among different industries (Ellison, Glaeser, 1997; Feser, 2000).

[7] To be sure in the robustness of the estimates, there were additionally applied multiple approaches and algorithms, such as Random Forest, MICE pmm (multiple imputations by chained equations predictive mean matching), and E-M imputation algorithm. The results were compared using a standard methodology with the division of the dataset on the training part and test subsample and then using the MAPE (Mean average predictive error) criterion. It turned out that linear regression is almost as good as the estimates obtained with the random forest, and when the share of missing values is small it can give even better results. Since the number of omissions was relatively small – 12% at maximum, so the forecasts were made using the OLS linear regression model.

«Ellison-Glaeser (EG) index of concentration is calculated as follows:

$$\gamma_j = \frac{\sum_i (s_{ij} - x_i)^2 - (1 - \sum_i x_i^2) \sum_j z_j^2}{(1 - \sum_i x_i^2)(1 - \sum_j z_j^2)} \quad (1),$$

where $j = 1 \ldots J$ – are indexed industries (J in total);

$i = 1 \ldots 82$ – indexed regions;

s_{ij} – share of employment in industry j out of total region i employment;

x_i – share of region i employment out of total national employment;

z_j – share of employment in industry j of total national employment.

The index is not standardized. Lower values of EG signal about uniform distribution of industry employment among regions and higher values reversely mean that the industry is concentrated in certain region(s).» (Zyuzin, Demidova, 2022).

«Another measure of the interest is the degree of the regional economy diversification. Simple and the most widely used metric for this purpose is the Herfindahl-Hirschman index (HHI, (2)). There are other possibilities of taking into account regional diversification, e.g., Theil or Gini indices, but HHI indicator was chosen due to computational simplicity and stability, which is achieved owing to the large number of observations.

$$HHI_i = \sum_j (s_{ij})^2 \quad (2)$$

where $j = 1 \ldots J$ – indexed industries (J in total);

$i = 1 \ldots 82$ – indexed regions;

s_{ij} – share of employment in industry j out of total region i employment.

HHI index varies in the range $HHI_i \in [\frac{1}{J}; 1]$. High values correspond to the situation of high regional specialization and low values mean that regional economy is diversified.» (Zyuzin, Demidova, Dolgopyatova, 2020).

Enterprise performance

In the present thesis *sales margin* is used as the main indicator of the enterprise performance (SM (3)). On the earlier stages of the study for the purposes of control there was simultaneously applied another measure – *net profit margin* (NPM (4)).

$$SM = \frac{revenue - costs\ of\ sales}{revenue} \quad (3)$$

$$NPM = \frac{net\ profit}{revenue} \quad (4)$$

Modeling

Throughout the whole study the general modeling approach was preserved and applied with some modifications to different issues. This is convenient for comparing and interpreting the results. To solve objectives 2, 3 ,4 ,5 linear regression models with three groups of variables were applied:

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \epsilon, \quad (5)$$

where β_k – regression coefficients $k = 1 \dots 3$; β_0 – constant; X_1 – firm specific regressors; X_2 – region specific regressors; X_3 – matrix of explanatory variables related to industry localization and regional economic diversification; ϵ – vector of random errors.

SM and NPM were used as dependent variables *Y*. This form of the model (5) can be used under different approaches to the enterprise grouping: by scale or by the companies' industry. Moreover, this form of the model is applicable for both samples, thus was used while studying dynamics issues. When working with the objectives 5 and 6, equation (5) was estimated sequentially for each industry group in each considered year.

Model estimates were obtained using least squares. An assumption about the normal distribution of errors was made (due to the large number of observations, we can assume an asymptotically normal distribution of errors). Formal tests revealed the heteroscedasticity of the regression errors, so standard errors in White's form were used. A Ramsey test was performed to verify the correct specification for each model.

Additionally, to compare the influence of the agglomeration effects on the enterprises with different initial performance (objective 4), quantile regression models were estimated. The general form of the quantile regression (QR) equation is the following:

$$Q_\tau(Y) = \beta_0(\tau) + X_1\beta_1(\tau) + X_2\beta_2(\tau) + X_3\beta_3(\tau) + \epsilon \quad (6)$$

where τ – corresponding quantile (the step τ is 0.1); $\beta_k(\tau)$ – regression coefficients for quantile τ ($k = 1 \dots 3, \beta_0$ – constant); X_1 – firm specific regressors; X_2 – region specific regressors; X_3 – matrix of explanatory variables related to industry localization and regional economic diversification; ϵ – vector of random errors.

For each model, the hypotheses about the equality of β coefficients of different τ were tested. They were consistently rejected for all industry groups, therefore, along with linear regression models, quantile regressions were also sensible.

Quantile regression models (QR) were estimated using R package «quantreg» (Koenker 2017).

All the variables that were included in the models are summarized in the Table 4.

Table 4

Variables that were included in the regression models

Variable	**Variable description**
Dependent variables	
NPM	Net profit margin.
SM	Sales margin.
Enterprise specific variables (X_1 regressors)	
logNetAss	Log of the firm's net assets. Net assets equals capital + future receivables – financial assets.
logEmployees	Log of enterprise annual weighted average employment.
NDEBIT	Net Debt/ EBIT multiplier. Net debt is the sum of long-term and short-term debts subtracting cash and its equivalents. EBIT is calculated as before tax profit + interest paid – interest received. EBIT measures on operational cash flow available for the use of the company.
logweffect	Log of the ratio of the revenue to the annual weighted average number of the employees. This is the proxy measure of the labor efficiency.
*Age dummy:** *age_old* *age_middle*	Dummy variables indicating company age: older companies (***age_old***) – operate for more than 20 years; middle aged companies (***age_middle***) – 10-20 years on the marker; younger companies (***age_short***) – 5-10 years on the market and the newcomers (***age_newcomers***) – operate less than 5 years.

age_short *age_newcomers*	
*Industry Dummies / Scale dummies***	Sets of dummies to indicate industry group (Table 3) or business scale depending on the model setup. Controlling industry group is № 8 (H – Services and other minor industries) and controlling business scale – micro firms.
Region specific variables (X_2 regressors)	
logAverMonthWage	Log of the average accrued wage in a region.
logRND	Log of R&D investments.
Unempl	Regional unemployment rate
Unemplsq	Squared regional unemployment rate
UnprofitWeight	Share of companies with losses in a region
DiffInvest	Percentage increase of investments in a region compared to the previous year
FirmBirthRate	Ratio of the difference between created and closed companies in considered year to the total number of firms in a region at the beginning of considered year
CriminalRate	Number of registered crimes per 100 people in a region
Industry concentration and regional economy diversification (X_3 regressors)	
Ellison-Glaeser	Ellison-Glaeser industry concentration index
HHI	Regional diversification measured by the simple Herfindahl-Hirschman index
logcore	Log of the ratio of the firm's industry employment to total regional employment (industry share in a region)
Explanatory variables included in the extended model specifications*	
Cluster*	Dummy indicator whether the company belongs to the geographical cluster*** (takes the value of 1 if it is and 0 otherwise).
EG*Cluster*	Cross-effect to distinguish the influence of concentration effects on SM for the firms inside and outside of the cluster.

Source: composed by the author

* These variables were implemented during the study of dynamics and sustainability of concentration effects.

** The choice of the set of dummies depends on the modeling approach: estimation of models for different business scales (industries dummified) or for different industry groups (business scale dummified).

*** The company belongs to the geographical cluster if: (a) its' own industry is highly concentrated (EG > 0.05) (Ellison and Glaeser 1997; Feser 2000); (b) the firm locates in top 3 regions, where the industry has the most of its' employment.

Variables *SM* and *NPM* are dependent variables. The expediency of their inclusion in the model follows from the objectives of the work. Variables *Ellison-Glaeser, HHI, Cluster, EG*Cluster* are the ones of interest (detailed justification is given in the thesis). *Age dummies* and *Industry dummies* were introduced into the model to take into account

the effects of experience (i.e., the difference in the efficiency of old and newly entered firms) and industry (on the stage before aggregation into industry groups).

The variables *logNetAss*, *logEmployees* reflect the labor- and capital-intensiveness of the enterprise. The basic expectation was that these factors may have a negative impact on the dependent variables (large enterprises are less efficient). By introducing the *logcore* variable the position of the enterprises' industry in the region was described (Davidson, Mariev, 2015). It was expected that the direction of influence of *logcore* would coincide with the direction of influence of HHI.

The *NDEBIT* variable indicates credit leverage of the company: this indicator is often controlled by banks, thus large values can mean trust to the company as only the most effective companies can manage paying high interest (as a consequence of large debt burden), and on the other hand, a high interest payment reduces net profit. The variable *logweffect* characterizes the degree of labor efficiency in the enterprise: for labor-intensive industries, a positive impact on SM and NPM is expected, and for capital-intensive industries, the effect may be insignificant.

The inclusion of the X_2 block in the model (thus, the variables that it consists of) is justified by the need to control the results of the study by the differences in the regional investment climate. The most common measure of the regions' investment attractiveness is the conclusions of rating agencies. The selection of variables in the thesis is based on the Expert RA rating agency methodology. This methodology identifies 64 main and 20 additional factors that affect investment attractiveness. The inclusion of all 84 variables in the model is impractical, since some blocks of variables taken into account by Expert RA[8] are redundant for the purposes of quantitative analysis (qualitative variables) and some other blocks might arise risks of multicollinearity. Among the five blocks of variables that assess *infrastructure* (group weight 40%), *social*, *economic* and *financial* resources, and the *environment*, for the thesis were selected only *social*, *economic* and *financial* factors in equal proportions.

[8] Expert RA report: https://www.raexpert.ru/researches/regions/regions_invest_2022/, and methodology: https://www.raexpert.ru/ratings/methods/current (both in Russian).

From the *social* block, which includes 18 variables, 3 factors were selected for the model (those giving the most general and versatile assessments): *logAverMonthWage*, *Unempl*, *CriminalRate*. The remaining variables from this block are either derived from the indicated values, or correlate to a large extent with them. Ramsey specification test revealed the necessity of inclusion of the square of the unemployment rate in the model (*Unemplsq*). There was expected a positive impact for SM and NPM from the decrease in *Unempl* and *CriminalRate* indicators, since the lower values of both of them indicate a better regional socio-economic climate and conditions for doing business in such locations. The influence of the *logAverMonthWage* variable can be ambiguous for SM and NPM, since higher average wages might mean higher labor efficiency and at the same time this factor increases the company's costs.

The economic and financial blocks of Expert RA's regional assessment guide include 8 and 6 factors respectively. For the objectives stated in thesis 4 indicators *logRND*, *FirmBirthRate*, *DiffInvest*, *UnprofitWeight* were selected to be included into the models as they characterize in the most capable way the overall enterprises' activity in the region. A priori, positive impact of *logRND* (new technologies add marginality), *FirmBirthRate* (if new companies enter the market, it means that entrepreneurs highly appreciate the probability of making a profit, which may follow from the observed high marginality of existing enterprises) and *DiffInvest* (investors will select the most promising regions from the point of view of business profitability for investments) indicators was expected. For the *UnprofitWeight* variable, the opposite logic applies - the higher is the share of unprofitable enterprises in the region, the more saturated is the market and less likely is the rise of SM and NPM for the existing enterprises.

The novelty

In this thesis the objectives (1)-(6) stated in the respective section are being solved. These issues have either not been discussed in the literature at all, or have been paid a little attention. Making an attempt to answer the questions above, there were collected significant databases (83-237 thousands of observations depending on the studied year) with up-to-date information about Russian enterprises of different industries (excluding

finance, insurance, government founded and non-commercial entities), new approaches and techniques were applied, new concepts were proposed in modeling the agglomeration effects. Being more specific, the following distinctive features of this study contribute to the Russian scientific literature (Table 2 describe the gap in the literature):

1.1. For the first time there were obtained quantitative estimates of the impact of concentration and diversification effects on the performance of different groups of Russian firms' (aggregation was based on the business scale into 4 groups: "large firms", "medium firms", "small firms" and "micro firms"). This approach allows you to compare the groups with each other;

1.2. Companies were aggregated into the industry groups in accordance to their economic activity (the aggregation follows the NACE[9] principles). The novelty is that for each industry group were estimated agglomeration effects of concentration and diversification.

The review of Russian studies on agglomeration economies and externalities contains the detailed description of the results obtained for Russian enterprises, however, suggested formats of conducting the research, including aggregation and approaches to the analysis of the results are new. Among foreign papers, there were found some attempts to compare agglomeration effects for different in terms of industry or scale enterprises, as well as to use larger samples that describe local general populations (for cities, states, regions, urban agglomerations and co-agglomerations, details are in the Table 1). However, to the best of our knowledge, there are no works that studied the impact of agglomerations directly on SM and NPM, and no cases of analyzing general populations of a national scale were found.

In this thesis, new methodological techniques are used, which were not found among Russian and foreign sources:

[9] The European Classification of Economic Activities: NACE Rev. 2 from 2008. High-level SNA/ISIC aggregation A*10/11.

2.1. A new approach of measuring and comparing the effects of concentration and diversification that appear for the enterprises with initially different performance was applied – using quantile regressions the agglomeration effects were compared for the best and the worst performing companies. Such analysis was held for each industry group separately;

2.2. For the first time, the stability of the direction of influence of concentration effects was tested in the period from 2011 to 2018 (for example, if in 2011 concentration effect was significant and positive for the industry group, then it was considered stable if it remained significant and positive throughout the entire considered period). Such an analysis was hold for each industry group;

2.3. An analysis of the dynamics of changes in the MAR industry concentration effects was conducted, which was never done before for Russian enterprises. Dynamics of change is a series of tests, demonstrating whether there is a significant difference between concentration effects estimates obtained in different studied years. For example, if from 2011 to 2012 for a certain industry group there was found an increase in the degree of influence of industry concentration on the firms' performance, then it was checked whether such a change was significant or accidental. The results were presented separately for each industry group;

2.4. A novel approach is to compare the magnitude of concentration effects for the firms located inside and outside of their geographic industry clusters.

An additional important distinguishing feature of the work is the use of all the advantages of working with the general population. An application of the new approaches to data processing allows to retain population and obtain estimates without data loss and significant distortions. Among the most popular ways to restore the missing values, was chosen one that allows to preserve the order of industries ranked according to their degree of concentration.

These main ideas addressed in the thesis and summarized above are unique and new for studies dedicated to this topic.

The results submitted for the thesis defense:

1. It was shown that the agglomeration effects of concentration and diversification differ for the enterprises of various business scales. An increase of the Ellison-Glaeser concentration index (or EG) (Ellison, Glaeser, 1997) by 0.1 results in 0.7% increase of SM of micro-firms and 7.5% increase for medium and large representatives of the industry, and an increase of the NPM varies from 0.2% for micro-firms to 1.2 % for large companies. This result does not contradict those shown in the works (Baptista, Swann, 1998; Holmes, 1999; Martin, Mayer, Mayneris, 2011; Davidson, Mariev, 2015) but is opposite to the findings of the other paper (Lu, Tao, and Yu, 2012). A change of the Herfindahl-Hirschman index (HHI), which is a measure of the diversification of the regional economy, by 0.1 causes a change in the sales margin by 1.5% for micro-firms, however, for larger enterprises, the significance of diversification effects has not been confirmed. In general, this result is in line with the theoretical arguments of Jacobs (Jacobs, 1961; Jacobs, 1969) in that part that the diversification of the economy within certain regions can have a very limited effect on the enterprises performance: (a) regional market is limited; (b) regions are generally open systems; (c) Russian regions are much larger than the cities in terms of their scale);
2. It was shown that EG has a positive impact on the majority of studied industries, including "Agriculture, fishing and forestry", "Mining"[10], "Manufacturing", "Wholesale and retail trade, food service", "Transport" and "IT and telecommunications". For these industries, each 0.1 change in the value of the EG leads to an increase of the SM by the value varying in the range 1.2% - 138.6% depending on the industry and to an increase in the NPM by the value from 0.4% to 16.0%, depending on the industry as well. These findings in support of positive concentration effects are consistent with the results published earlier at different times (Davidson, Mariev, 2015; Holmes, 1999; Martin et al., 2011). The

[10] Concentration effects for the “Mining” group are significant during the interval from 2011 to 2018 except 2017, which is represent in the Table 5.

diversification effects are weaker and they have a negative impact on all of the industries except “IT and telecommunications”. With a 0.1 change of the HHI, values of dependent variables SM and NPM decrease by 1.2%-6.5% and 0.1%-0.5%, respectively, depending on the industry. The exception is the “Construction” industry, where both effects are stronger, up to a sevenfold change in SM in response to 0.1 change in the EG. Following the “Construction” industry, concentration and diversification have the greatest impact on enterprises of “Agriculture, fishing and forestry” and “Transport” industries, while for the “Mining” industry, no significant effects of concentration and diversification were found;

3. Higher-performing firms were, on average, more likely to be adversely affected by the increased industry concentration comparing to the lower-performing firms. Thus, the initial level of business efficiency is a significant factor, when determining, what impact the agglomeration effects will have. This situation is consistent with widely accepted predictions and ideas (Audretsch, Feldman, 1996; Jacobs, 1961; Shaver, Flyer, 2000) that not very efficient firms benefit from the proximity to the industry leaders owing to the spillover effects and spreading information;
4. Concentration effect steadily appears each studied year and is statistically significant for the majority of industries both inside and outside of the cluster. Throughout the reviewed period (2011-2018), there were found consistently positive concentration effects for the “Agricultural”, “Mining”, “Transport” and “IT & Telecommunications” industries, which is consistent with the previous results (Davidson, Mariev 2015; Zyuzin, 2021). Only one industry was affected negatively by the concentration effect during the whole period - “Services” sector in case, when the firms are located within an industry cluster. Some peculiarities were found for the "Manufacturing" and "Services" industries, for which 5-6 years periods, when the concentration effects were negative alternated with 2-3 years periods, when this agglomeration effect had become positive. The power of the

concentration effects varied for most industries having a stronger or less intense impact on SM depending on the year.

The described above results are summarized in the Table 5, that contains coefficient estimates for the main variables of interest. The modeling principles applied to obtain the results demonstrated in the Table 5 are described in details in the next section.

Table 5.

Summary characteristics of the coefficients of the main variables of interest (EG and HHI) that indicate the influence of agglomeration effects of concentration and diversification on SM and NPM (modeling techniques are describe in the «Methodology remarks» section)

For companies different by their scale of business								
	Large firms		Medium firms		Small firms		Micro firms	
LR*_SM: EG	0.6579*** (0.1865)		0.7500*** (0.2067)		0.4170*** (0.1116)		0.0726' (0.0372)	
RL_SM: HHI	0.2408 (0.2051)		0.2661 (0.1646)		0.1580' (0.0818)		0.1690*** (0.0356)	
LR_NPM: EG	0.1256** (0.0437)		0.0787** (0.0290)		0.0275* (0.0122)		0.0201*** (0.0054)	
LR_NPM: HHI	-0.0317 (0.0200)		-0.0185 (0.0130)		-0.0059 (0.0068)		0.0275*** (0.0053)	
For different industry groups								
	Agriculture, Fishing, and Forestry	Mining	Manufacturing	Wholesale, Retail and Food Services	Construction	Transport	IT and telecommunication	Services and other minor industries
LR_SM: EG	5.7067*** (0.3923)	0.1803 (0.1703)	1.8520*** (0.4003)	5.4161*** (0.7154)	–70.7048*** (2.3383)	13.8603*** (2.0790)	0.1266*** (0.0379)	–0.9954*** (0.1049)
LR_SM: HHI	0.6518* (0.2579)	0.6276 (0.5211)	0.0032 (0.0750)	0.1295** (0.0467)	0.2302*** (0.0697)	0.3675* (0.1464)	–0.0212 (0.1803)	0.2221* (0.0977)
QR*_SM: EG**	+ / ↓ / ***	+ - / ↔ / x	+ - / ↓ / ***	+ / ∪ / ***	- / ∪ / ***	+ / ↑ / ***	+ / ↑ / ***	- / ∪ / ***
QR_SM: HHI	+ / ↔ / ***	+ / ↑ / *	- / ∪/ x	+ / ∩ / ***	+ / ↓ / ***	+ - / ∩ / *	+ - / ∩ / x	+ - / ↓ / *
LR_NPM: EG	0.3304** (0.1010)	0.0273 (0.0318)	–0.0736 (0.0539)	0.5796*** (0.1037)	–13.4015*** (0.4195)	1.6030*** (0.3111)	0.0428*** (0.0053)	–0.1151*** (0.0165)
LR_NPM: HHI	–0.1425** (0.0445)	–0.0784 (0.1230)	0.0233* (0.0099)	0.0169** (0.0057)	0.0335** (0.0117)	0.0154 (0.0167)	0.0562* (0.0282)	0.0257 (0.0168)
QR_NPM: EG	+ / ↑ / ***	+ / ∩ / *	+ - / ↓ / **	+ / ↑ / ***	+ - / ↓ / ***	+ / ↑ / ***	+ / ↑ / ***	+ - /

								↓ / ***
QR_NPM: HHI	- / ↓ / **	+ - / ↔ / x	+ / ↑ / **	+ / ↑ / *	+ / ↑ / ***	+ / ↑ / x	+ / ∩ / *	+ / ↑ / *
Time consistency (SM)	High	High	Low	NA	NA	High	Moderate	Moderate
Effects monotonicity (SM)	High	Low	Low	NA	NA	Low	Low	Moderate
Effects significance (SM)***	High	Moderate	Moderate	NA	NA	High	Moderate	High

**LR – linear regression model, QR – quantile regression model.*

***Coefficient signs that appear the most often or indication that the sign varies / coefficient behavior when the quantile of the initial performance grows (↑ / ↓ - monotone growth or decrease, ↔ - no difference between quantiles, ∪ / ∩ - nonlinear behavior of the respective form) / the most often significance rate or its' absence [x]).*

****Identified for each industry in general model (without discriminating companies' location relative to a cluster) – High-concentration was significant during the whole studied period, Moderate-number of years, when the concentration effect was not significant is less than a half (3 or less years), Low-concentration effects were not significant in half or more studied years.*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{'}p < 0.1$

Source: composed by the author.

Discussion of the research

Preliminary results were discussed at all the stages of the research and were presented at seminars, Russian national and international conferences, and also published in indexed reviewed scientific journals.

Conferences

April conference XXI (27.05.2020). Section L «Firms and markets». Report: «Concentration and diversification of Russian economy: regional and industrial aspects».

April conference XXII (16.04.2021). Section L «Firms and markets». Report: «Concentration and diversification of Russian economy: an analysis of Russian enterprise performance 2011-2018».

4-th workshop «Applied econometrics» associated to XXIII April International Academic Conference on Economic and Social Development (15.04.2022). Report: «An agglomeration effect of the industry concentration in Russia: cross-industry study of stability over time (from 2011 to 2018)».

Seminar report to the Department of applied economics HSE (03.06.2021) entitled "Cross-Industry Analysis of Russian Enterprise Performance: Do Concentration and Diversification Matter?".

Seminar report to HSE Centre for Spatial Econometrics in Applied Macroeconomic Research (04.03.2021) entitled: « Concentration and diversification of Russian economy: an analysis of Russian enterprise performance 2011-2018».

3-rd workshop «Applied econometrics» associated to XXII April International Academic Conference on Economic and Social Development (14.04.2021). Report: «Estimation of distortions in compiled variables (indices) that arise from missing values in the dataset. Challenges, possible solutions and additional options».

ERSA 2022, 61th Congress “Disparities in a Digitalising (Post-Covid) world – Networks, Entrepreneurship and Regional Development” (23.08.2022). Report:

«Clusters and Russian private firms' performance: cross-industry analysis» (co-author O.A. Demidova).

Report at the Scientific-research seminars of HSE Doctoral School of Economics (20.02.2019, 06.03.2020). Reports entitles: «Concentration and diversification of Russian economy: regional and industry peculiarities» (content changed along with the progress of the study).

List of author's original articles

Zyuzin A.V., Demidova O.A., Dolgopyatova T.G. Localization and Diversification of Russian Economy: Regions' and Industries' Peculiarities //. Prostranstvennaya Ekonomika = Spatial Economics. 2020, vol. 16, no 2. pp. 39-69. DOI:10.14530/se.2020.2.039-069 (in Russian).

Zyuzin A.V. Cross-Industry Analysis of Russian Enterprise Performance: Do Concentration and Diversification Matter? // Economic policy. Vol. 16. № 4. Pp. 42-83. DOI: 10.18288/1994-5124-2021-4-42-83.

Zyuzin A.V., Demidova O.A. An influence of the industry clusters on the Russien private enterprise performance: cross-industry analysis (2011-2018) // Voprosy Ekonomiki. 2022, vol. 11, pp. 90-116. DOI: 10.32609/0042-8736-2022-11-90-116 (in Russian).

Policy implication and theoretical value of the research

Linking the operational enterprise performance (which owners often treat as one of the main indicators of the company's performance) and the estimates of agglomeration effects of industry concentration and regional economy diversification, allows to extend the opportunities for planning regional and industry policy including fine-tune and targeting. In particular, one can make assumptions about the effectiveness of economic policies aimed at developing special economic zones or stimulating the development certain industries.

The study has the theoretical value as it makes another step towards describing the concept of the optimal distribution of industries across regions. It becomes possible to estimate marginal spillover effects and offer such redistribution of employment within an industry between regions that the average operating results of enterprises of the whole industry would increase.

References

Audretsch D.B. (1998). Agglomeration and the Location of Innovative Activity. // *Oxford Review of Economic Policy*, Vol. 14, No 2, pp. 18–29. https://doi.org/10.1093/oxrep/14.2.18

Audretsch D.B., Feldman M. (1996) R&D Spillovers and the Geography of Innovation and Production // *American Economic Review*. 1996. Vol. 86. Issue 3. Pp. 630–640.

Baptista R., Swann P. (1998). Do Firms in Clusters Innovate More? *Research Policy*, Vol. 27, pp. 525–540. https://doi.org/10.1016/S0048-7333(98)00065-1

Baumgardner J.R. (1988) Physicians' Services and the Division of Labor Across Local Markets Author // *Journal of Political Economy*. Vol. 96. Issue 5. Pp. 948–982. https://doi.org/10.1086/261571

Bell G. (2005). Clusters, Networks, and Firm Innovativeness. // *Strategic Management Journal*, Vol. 26, No. 3, pp. 287–295. https://doi.org/10.1002/smj.448

Bessonova E., Gonchar K. (2017). Incentives to Innovate in Response to Competition: The Role of Agency Costs. // *Economic Systems*, Vol. 41, pp. 26–40. http://dx.doi.org/10.1016/j.ecosys.2016.09.002

Ciccone A., Hall R. (1996). Productivity and the Density of Economic Activity. // *The American Economic Review*, Vol. 86, No. 1, pp. 54–70.

Cohen J. P., Morrison Paul C.J. (2009). Agglomeration, Productivity and Regional Growth: Production Theory Approaches. *Handbook of Regional Growth and Development Theories*, pp. 101–117.

Davidson N., Mariev O. (2015). The Impact of Spatial Concentration on Enterprise Performance. // *Economy of Region*, No. 4, pp. 95–105. https://doi.org/10.17059/2015-4-8

Ellison G., Glaeser E. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. // *Journal of Political Economy*, Vol. 105, No. 5, pp. 889–927.

Feser E. J. (2000). On the Ellison-Glaeser Geographic Concentration Index. *University of North Carolina*, pp. 1-19. Open access by URL: http://works.bepress.com/edwardfeser/28/

Fischer M., Scherngell T., Jansenberger E. (2006). The Geography of Knowledge Spillovers Between High-Technology Firms in Europe: Evidence from a Spatial Interaction Modeling Perspective. *Geographical Analysis*, Vol. 38, pp. 288–309. https://doi.org/10.1111/j.1538-4632.2006.00687.x

Glaeser E.L., Kallal H.D., Scheinkman J.A., Shleifer A. (1992) Growth in Cities // Journal of Political Economy. Vol. 100. Issue 6. Pp. 1126–1152. URL: https://www.jstor.org/stable/2138829

Golovanova S.V., Avdasheva S.B., Kadochnikov S.M. (2010) Inter-firm Cooperation: Analysis of Clusters' Development in Russia // Russian Management Journal. vol. 8, no 1, pp. 41-66 (in Russia)

Golovanova S.V. Industry Product Distribution Structure Changes in Russia During the Period of Economic Growth in 1997-2004 (2008) // *Journal of Modern Competition*. no 5, pp. 58-72 (in Russian).

Golovanova S.V., Kadochnikov S. M. Cluster Relations in Russia and Regional Policy of Developing Clusters // Book: XI international scientific congress on developing economy and society: Book 3 / Editors.: E. G. Yasin. Moscow.: HSE, 2011. pp. 383-391 (in Russian)

Gonchar K.R., Ratnikova T.A. Estimating and explaining urban agglomeration effects for Russian manufacturing companies // Book: XIV April International Academic Conference on Economic and Social Development: in 4 books. Book 3 / Eitor: E. G. Yasin. Moscow.: HSE, 2014. pp. 55-64 (in Russian)

Gonchar K. (2009). Urban Agglomeration Effects and Company Productivity in Russia: Empirical Evidence Based on Manufacturing Industry Survey. *14th Annual Conference of The International Society for New Institutional Economics.*

Hervas-Oliver J.-L., Sempere-Ripoll F., Alvarado R., Estelles-Miguel S. (2018). Agglomerations and Firm Performance: Who Benefits and How Much? *Regional Studies*, Vol. 52, No. 3, pp. 338–349. https://doi.org/10.1080/00343404.2017.1297895

Holmes T. (1999). Localization of Industry and Vertical Disintegration. *The Review of Economics and Statistics*, Vol. 81. No. 2, pp. 314–325.

Huang Y., Hong T., Ma T. (2020). Urban Network Externalities, Agglomeration Economies and Urban Economic Growth. *The International Journal of Urban Policy and Planning*, Vol. 107, pp. 1-15. https://doi.org/10.1016/j.cities.2020.102882

Jacobs. J. (1961). The Death and Life of Great American Cities. *New York (NY): Random House*.

Ketels C., Protsiv S. (2021). Cluster Presence and Economic Performance: a New Look Based on European Data. *Regional Studies*, Vol. 55, No. 2, pp. 208-220. https://doi.org/10.1080/00343404.2020.1792435

Knoben J., Arikan A.T., van Oort F., Raspe O. (2016). Agglomeration and Firm Performance: One Firm's Medicine Is Another Firm's Poison. *Environment and Planning A*, Vol. 48, No. 1, pp. 132–153. https://doi.org/10.1177/0308518X15602898

Kolomak E.A. (2015) Evolution of Spatial Distribution of Economic Activity in Russia // Regional Research of Russia. Vol. 5. Issue 3. Pp. 236–242. https://doi.org/10.1134/s2079970515030065

Krugman P. (1991) Increasing Returns and Economic Geography // Journal of Political Economy. Vol. 99. Issue 3. Pp. 483–499. https://doi.org/10.1086/261763

Li D., Lu Y., Wu M. (2012). Industrial Agglomeration and Firm Size: Evidence from China. *Regional Science and Urban Economics*, Vol. 42, pp. 135–143. https://doi.org/10.1016/j.regsciurbeco.2011.07.003

Lu Y., Tao Z., Yu L. (2012). Agglomeration and Markup. *University Library of Munich, MPRA Paper*, No. 38974. https://doi.org/10.2139/ssrn.2065303

Malmberg A., Malmberg B., Lundequist P. (2000). Agglomeration and Firm Performance: Economies of Scale, Localization, and Urbanization among Swedish Export Firms. *Environment and Planning A*, Vol. 32, pp. 305–321. https://doi.org/10.1068/a31202

Marshall A. (1920). Principles of Economics. *London: Macmillan and Co.*

Martin P., Mayer T., Mayneris F. (2011). Spatial Concentration and Plant-Level Productivity in France. *Journal of Urban Economics*, Vol. 69, pp. 182–195. https://doi.org/10.1016/j.jue.2010.09.002

McCann P., van Oort F. (2019). Theories of Agglomeration and Regional Economic Growth: a Historical Review. *Handbook of Regional Growth and Development Theories*, Capello R., Nijkamp P. (eds.). *Massachusetts, Edward Elgar Publishing Inc*, pp. 6-23. https://doi.org/10.4337/9781788970020

Mikhailova T.N. (2016). Transformation of Industrial Geography in Russian Federation // *Journal of Russian Entrepreneurship*. vol. 17, no 3, pp. 351-358. http://dx.doi.org/10.18334/rp.17.3.2223 (in Russian)

Morrison, C. J., & Siegel, D. (1998) Knowledge Capital and Cost Structure in the U.S. Food and Fiber Industries // *American Journal of Agricultural Economics*. Vol. 80. Issue 1. Pp. 30–45. doi:10.2307/3180266

Nakamura R. (1985) Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities // Journal of Urban Economics. Vol. 17. Issue 1. Pp. 108–124. https://doi.org/10.1016/0094-1190(85)90040-3

Porter M. (2003) The economic performance of regions. *Regional studies*. Vol. 37. No. 6–7. Pp. 549–578. https://doi.org/10.1080/0034340032000108688.

Quigley J. M. (1998). Urban Diversity and Economic Growth. *Journal of Economic*

Perspectives, Vol. 12, No. 2., pp. 127-138.

Rastvortseva S.N., Chentsova A.S. (2015) Regional Specialization and Geographical Concentration of Industry in Russia // Regional Science Inquiry. Vol. 7. Issue 2. Pp. 97–106

Rosenthal S.S., Strange W.C. (2006). The Micro-Empirics of Agglomeration Economies. *A Companion to Urban Economics*, Arnott R. J., McMillen D.P. (eds.). Hoboken, NJ, Wiley-Blackwell. https://doi.org/10.1002/9780470996225.

Rosenthal S.S., Strange W.C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. *Handbook of Regional and Urban Economics: Cities and Geography*, Henderson J.V., Thisse J. (eds.). North Holland, Elsevier Inc., Vol. 4, pp. 2119-2171.

Skhvediani A., Sosnovskikh S. (2020). What Agglomeration Externalities Impact the Development of the Hi-tech Industry Sector? Evidence from the Russian Regions. *International Journal of Technology*, Vol. 11, No. 6, pp. 1091-1102. https://doi.org/10.14716/ijtech.v11i6.4423.

von Hippel E. (1998) Economics of Product Development by Users : The Impact of " Sticky " Local Information // *Management Science*. Vol. 44. Issue 5. Pp. 629–644. https://doi.org/10.1287/mnsc.44.5.629

Vorobyev P.V., Davidson N.B., Kislyak N.V., Kuznetsov P.D. (2014) Industrial diversity and concentration in the Russian cities as factors of economic efficiency // *Vestnik URFU*. Economics and management series. No. 6, pp. 4-18 (in Russian)

Wheaton W.C., Lewis M.J. (2002) Urban Wages and Labor Market Agglomeration // *Journal of Urban Economics*. Vol. 51. Issue 3. Pp. 542–562. https://doi.org/10.1006/juec.2001.2257

Zyuzin A. (2021). Cross-Industry Analysis of Russian Enterprise Performance: Do Concentration and Diversification Matter? // *Economic Policy*, Vol. 16, No. 4, pp. 42-83. https://doi.org/10.18288/1994-5124-2021-4-42-83.

Zyuzin A.V., Demidova O.A. (2022) An influence of the industry clusters on the Russian private enterprise performance: cross-industry analysis (2011-2018) // *Voprosy Ekonomiki*. Vol. 11, pp. 90-116. DOI: 10.32609/0042-8736-2022-11-90-116 (in Russian).

Zyuzin A.V., Demidova O.A., Dolgopyatova T.G. (2020) Localization and Diversification of Russian Economy: Regions' and Industries' Peculiarities //. *Prostranstvennaya Ekonomika = Spatial Economics*. vol. 16, no 2. pp. 39-69. DOI:10.14530/se.2020.2.039-069 (in Russian)