Received: 5 July 2022 | Revised: 16 February 2023 | Accepted: 8 April 2023

DOI: 10.1111/rsp3.12668

ORIGINAL ARTICLE

Does complementarity matter for the emergence of new specialization industries in the regions of Russia?

Evgeniy Kutsenko[1] | Sabyasachi Tripathi[1] | Kirill Tyurchev[1,2]

[1]Institute for Statistical Studies and Economics of Knowledge, HSE University, Moscow, Russia

[2]Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic

Correspondence
Sabyasachi Tripathi, Institute for Statistical Studies and Economics of Knowledge, HSE University, Myasnitskaya Ulitsa, 11 Moscow, 101000, Russia.
Email: sabya.tripathi@gmail.com

Funding information
The article is based on a study funded by the Basic Research Program of the HSE University.

Abstract

The purpose of this article is to assess how internal and external complementarity linkages are responsible for the emergence of new specialization industries in the context of Russian regions in the period from 2005 to 2015, taking into account other factors that are significant for the emergence of such industries. The internal complementary linkages are measured by the share of new specialization industries related to existing specialization industries in the same region. External complementarity linkages are measured by the share of new specialization industries related to existing specialization industries in neighboring regions. Our data show that internal and external complementarity linkages have a strong positive effect on the emergence of new specialization industries, with a greater effect had by external ones. The number of specialization industries, the number of employees, market potential, gross domestic expenditure on research and development (R&D), domestic patent applications, working population with a higher education, and the distance from the administrative center of a region to the nearest million-plus city also influence the emergence of new specialization industries, while gross regional product (GRP) per worker, the share of companies using the internet, and special economic zones do not. In addition, we assess the models that predict beginning of the process of diversification in a region. Finally, some policy measures are suggested for the creation of new specialization industries for Russia.

KEYWORDS
agglomeration effect, complementary, market potential, new specialization industries, Russian regions

JEL CLASSIFICATION
L60, O10, P23

1 | INTRODUCTION

The emergence of new industries is a central phenomenon in innovation and economic development and is an important area in economic studies. The arrival of new industries creates structural changes that are the crucial elements for achieving sustainable development and increased well-being (Boschma, 2017; Hausmann & Hidalgo, 2011).

Entrepreneurial and innovative activities also lead to the possible emergence of new industries. In recent decades, entrepreneurial activity has grown due to a sharp decrease in the cost of IT startups, creating internet platforms such as Alibaba, Airbnb, and Uber. (Burtch et al., 2018). However, low entrepreneurial and innovative activity in Russia limits the development opportunities for new industries and new products (Zemtsov et al., 2019, 2022; Zemtsov & Kotsemir, 2019).

Industry relatedness is an important driving force behind industrial diversification. Recent studies show that industries are more likely to emerge and develop in a region when they are related to preexisting industries in the region (Neffke et al., 2011). In addition, new technologies are more likely to arise in regions with an already established presence of related technologies (Kogler et al., 2013; Rigby, 2015). Xiao et al. (2018) found that the development of new industry specialization is positively associated with the new industry's relatedness to the region's current industries. They also showed that though the innovation capacity of a region is important, relatedness is a more important driver of diversification in regions. In fact, recent evolutionary economic geography (EEG) argues that regional diversification emerges as a path development process, as a region often branches into related industries, whereas unrelated industries exit the region. This argument has been empirically tested and supported on the basis of Chinese data (He et al., 2017).

Recently, several studies (e.g., Balland et al., 2019; Balland & Boschma, 2021; Boschma & Iammarino, 2009) have emphasized the role of complementary capabilities and interregional linkages behind regional diversification. Boschma and Iammarino (2009) showed strong evidence that related variety contributes to regional economic growth for Italian regions. Regions that are well endowed with sectors that are complementary in terms of competences (i.e., that show related variety) perform better. Balland et al. (2019) found that relatedness has a positive effect upon technological diversification within regions. Diversifying into complex technologies is easier when such technologies are more closely related to the existing knowledge core of the region. Finally, regions tend to grow more if they specialize in complex technologies related to existing technologies in the region. Balland and Boschma (2021) find that having complementary interregional linkages significantly increases the probability of regions in Europe developing new technological specializations, calculated on the basis of patent data. They also find that peripheral regions tend to diversify less, but their capacity to diversify increases significantly when they are connected to regions with complementary capabilities.

Related diversification positively impacts employment, innovation potential, and the growth of the regional economy in the short run and reduces substantial costs (Boschma & Iammarino, 2009; Boschma et al., 2012; Falcıoğlu, 2011; Frenken et al., 2007; Hartog et al., 2012 ; Neffke et al., 2011). However, related diversification has some risks; for example, established specializations are more vulnerable to economic crises and external shocks. At the same time, long-term growth may not be possible in related industry diversification. On the contrary, though unrelated diversification has certain advantages (such as the growth of the economy in the long run), it is also associated with higher costs, diffusion of resources, and loss of competitive advantages (Bathelt & Boggs, 2003; Essletzbichler, 2015; Quatraro, 2010).

After the collapse of the Soviet Union, the Russian economy began its transformation, both in connection with the need to restore destroyed production chains and in connection with ongoing market reforms. Unlike the European and American regions, whose industry portfolio was formed under market laws, the Russian economy had long been planned, and therefore, the transition to a new logic of development should have led to a change in the concentration of certain industries on its territory (Kolomak, 2020).

Russia is a federal state consisting of 85 politically equal subjects of the federation—regions that are diverse in terms of their resource potential and level of well-being and characterized by an uneven distribution of productive forces. For example, the difference between the most populated region (Moscow) and the least populated (Nenets Autonomous Okrug) is 284 times, and in terms of gross regional product (GRP), this difference reaches 353 times (between Moscow and the Altai Republic). Inequality (measured by the Gini index) in GRP has increased from 0.517 in 1995 to 0.612 in 2012 (Rastvortseva & Chentsova, 2015). Kolomak (2021) argued that the increase in the heterogeneity of city development in Russia was due to the growth of large cities and a decrease in the population of small cities. Therefore, uneven regional development is very noticeable in Russia. This also has a significant impact upon the emergence of new specialization industries. According to our calculations, some regions have experienced the emergence of six new specialization industries during the period of 2005–2015 and some regions have none.

At the same time, the regions have the authority to conduct their own industrial and innovation policies and to create various tools for the development of their own economies. For example, the regions implement their own programs to support small- and medium-sized businesses, create special agencies to attract foreign investment together with the national government, and launch free economic zones in their territory to create new large-scale industries.

In the context of the Russian economy, Lyubimov et al. (2018) used the network approach to measure the level of economic complexity and the diversification opportunities of Russian regions. They found that the complexity of Russian regional economies varies substantially: relatively high in western and central regions, lower in southern and northern regions, and the lowest in eastern regions. Lyubimov (2019) argued that the current stock of know-how in Russia is relatively low and fragmented; thus, it does not allow Russia to diversify into a broad range of more complex products. Kadochnikov and Fedyunina (2013) indicated that it is not industry variety per se but the variety of related industries located relatively close to each other in the product space that significantly contributes to economic growth in Russian regions. So why is the Russian case important? First, it is another object for testing hypotheses about the impact of complementarity on the emergence of new industries. Until now, there have not been many studies carried out on this topic, and in Russia, there have been none at all. Second, Russia is a large federal country with a large number of regions that are diverse and highly differentiated in terms of the level of economic development. Third, it is a country that is part of BRICS and the post-Soviet space. There are even fewer studies about effects of complementarity for such countries.

The purpose of our work is to study changes in the specialization industries of the Russian regions in the period from 2005 to 2015, as well as to identify the factors that contributed to this, especially the complementarity of specialization industries.

In this paper, we will try to assess how internal and external complementarity linkages are responsible for the emergence of new specialization industries in the context of Russian regions. Therefore, we consider regional characteristics to assess how they are beneficial in the promotion of new specialization industries. Our dependent variable is the total number of new specialization industries that appear in each region from the period of 2005 to 2015. Meanwhile, the effect of complementary specialization industries is measured by two variables, the share of new specialization industries against the share of existing specialization industries in the region, and the share of new specialization industries against the existing specialization industries in the neighboring regions. To calculate robust regression results, we control several other important independent variables such as the size of the region, existing skills, innovation activities, wealth, infrastructure, and the presence of large cities. Therefore, we expect initial conditions to play an important role in industry specialization. To capture the initial conditions, the average values for 2005, 2006, and 2007 are considered. This smooths out the possible omissions and more robustly reflects the impact of a time lag for the new specialization industries.

The paper is organized as follows. Section 2 presents a review of the literature. Section 3 describes our empirical framework. The measurement of variables, data sources, and a description of the data are presented in Section 3.1. The results of the estimation are described in Section 4. Finally, the major conclusions and policy implications are made in Section 5.

2 | REGIONAL DIVERSIFICATION: WHY DO NEW SPECIALIZATION INDUSTRIES EMERGE?

The issues facing economic diversification have a special place on the agenda of economists who study factors of regional development (Boschma, 2017; Chen, 2018; Hausmann et al., 2014; Hausmann & Klinger, 2006). New economic activities lead to positive changes in the economy not only on a national scale, but also at the level of individual regions (Saviotti & Frenken, 2008). In contrast, an economy that does not increase the diversity of sectors over time may be subject to structural unemployment and stagnation (Pasinetti, 1993).

The diversification of a region's economy is usually based on an existing set of local opportunities (Neffke et al., 2011; Rigby, 2015). If new industries are not related to the existing industry structure, diversification may have a mixed effect on long-term economic development (Nooteboom, 2000). Therefore, it is common to distinguish between two types of diversification—activity related and unrelated to the existing industry portfolio.

Many researchers define the relatedness of industries through the prism of the export basket of countries (Hausmann & Klinger, 2006; Hidalgo et al., 2007). Countries are developing through export diversification and within their own product/industry space. In other words, the issues surrounding sectors' relatedness are becoming crucial for economic growth (Bathelt & Boggs, 2003; Quatraro, 2010). Researchers (e.g., Hausmann & Klinger, 2006) show that the more goods a country already exports that are related to a product that it does not yet export, the more likely it is to start exporting that product in the future (Content & Frenken, 2016). As a result, countries that specialize in a large number of industries have more opportunities to sustain economic growth than less affluent countries that have a limited industry portfolio (Saviotti & Frenken, 2008). To develop a new industry, it is necessary to have a certain related 'mix' of industries; otherwise, diversification will be unrelated and will involve large risks and high costs. High-income countries tend to have unrelated diversification, as opposed to low-income countries with related diversification (Boschma & Capone, 2015; Petralia et al., 2017). The level of economic development probably affects the nature of diversification: rich countries can afford to experiment with the development of new unrelated industries, while poor countries may face negative consequences for their economies if they pursue such a path. The same results were obtained for the regions of Russia (Eferin & Kutsenko, 2021). Other studies show that a region with a related industry portfolio will demonstrate higher employment growth rates than a region developing unrelated industries since the former does not need to create fundamentally new infrastructure (Frenken et al., 2007). In contrast, narrow specialization may make the region vulnerable to economic crises and external shocks associated with fluctuations in global prices or changes in population preferences (Krugman, 1993; Saviotti & Frenken, 2008).

Neighboring regions can also influence the ways of diversification and contribute to the formation of a new industry space. If the region does not have a critical mass or the necessary capital for the development of a certain industry, then the neighboring region can affect this development due to spillover effects (Content & Frenken, 2016; Hidalgo et al., 2007). Thus, a region has a high probability of developing a certain industry if a neighboring region specializes in it (Bahar et al., 2014; Boschma et al., 2017). The reasons behind the greater influence of the nearest territories include the 'simplicity' of the interactions between the two entities due to the presence of similar patterns and levels of economic development, a similar resource base, climatic conditions, the size of territories, and the presence of cultural or ethnic ties (Bahar et al., 2014; Giroud, 2013; Kerr, 2008; Stein & Daude, 2007). Other factors may include possible migration between the territories and the ease of labor movements due to geographical proximity (Andersen & Dalgaard, 2011), which lead to knowledge spillovers and knowledge diffusion between neighbors, and further to the emergence of similar or related industries (Bahar et al., 2014).

Boschma et al. (2017) show that a US state has a higher probability of developing a competitive advantage in a new industry if a neighboring state specializes in that industry. In the long term, this process may lead to convergence in the export structures of neighboring regions. Other studies (Hafner, 2018; Isaksen, 2015; Mao & He, 2019) also show that the emergence of a new industry in the region may depend upon similar external factors, and this applies primarily to export-oriented industries. For example, the variety of export specializations in China's prefectures has shifted from coastal regions to the neighboring ones over the last decade or so (Mao & He, 2019). Thus, neighboring regions can be a source of new industries whose specializations can be transposed in geographical proximity. Most importantly, recent studies (Balland et al., 2019; Balland & Boschma, 2021; Boschma & Iammarino, 2009) find that complementary interregional linkages are the main drivers behind the creation of new specialization industries.

On the basis of the previous discussion, we developed two main hypotheses:

Hypothesis 1. New specialization industries are more likely to occur in regions if related industries are located in one's own region.

Hypothesis 2. New specialization industries are more likely to occur in regions if related industries are located in neighboring regions.

There are other important factors that may affect the emergence of the new specialization industries. Traditionally, economic theory suggests that the emergence of new industries is determined by fundamental factors of production, such as the availability of labor, transport infrastructure, and a critical mass of capital in a country or region (Hausmann & Klinger, 2006). Therefore, regional development is essentially an endogenous process, as it combines the infrastructure, natural resources, institutional capabilities, knowledge, and skills of a region (Maskell & Malmberg, 1999). Zhao and Guan (2013), using the sample of 20 leading universities active in nanotechnology, showed that nanotechnology is currently a scientific-push rather than the market-pull industry, which has led to the limited creation of an industry. The UN (2018) stated that when income grows, demand initially changes from necessities to more sophisticated goods. If enough industrial products are supplied, the diversification triggers industrial development through the emergence of new industries. The expansion and consolidation of manufacturing industries to increase production efficiency and reduce prices allows for the broad diffusion of manufactured goods through mass markets. This again increases production efficiency and the purchasing power of all consumers by creating new disposable incomes, and keeps the circle turning.

3 | EMPIRICAL FRAMEWORK

To understand the economic determinants of the emergence of new specialization industries; we use the following reduced form equation:

$$NSA_i = \alpha_0 + \sum_{i=1}^{12} \alpha_i \mathbf{X}_i + e_i, \tag{1}$$

Where i indicates the region, NSA stands for number of new specialization industries; $\boldsymbol{X}$ indicates a set of explanatory variables; α_0 is a constant; and e represents a well-behaved error term.

Table 1 presents an explanation of the independent variables used in Equation (1). To measure the effect of complementary specialization industries, we considered the following two variables: (i) the share of new specialization industries against the existing specialization industries in the region and (ii) the share of new specialization industries against the existing specialization industries in the neighboring region. According to Hypotheses 1 and 2, we assume that these two variables have a positive impact upon the creation of new specialization industries.

To measure the impact of the size of the region, we consider the following three variables: (i) the number of specialization industries, (ii) the number of employees, and (iii) the market potential of the region. The impact of existing

TABLE 1 Details of the independent variables used in the regression model.

Abbreviation of independent variables	Explanation	Expected sign
Independent variables		
Effect of complementary specialization industries		
x_1	Share of new specialization industries related to existing specialization industries in one's own region	+
x_2	Share of new specialization industries related to existing specialization industries in the neighboring regions	+
Control variables		
Impact of region size		
x_3	Number of specialization industries	+
x_4	Number of employees	+
x_5	Market potential of the region	+
Impact of existing skills		
x_6	Share of the working population with a higher education	+
Influence of innovation activity		
x_7	Gross domestic expenditure on R&D (GERD) as a percentage of GRP	+
x_8	Number of domestic patent applications per 1 million workers aged 15–72 years	+
Impact of wealth		
x_9	GRP per worker employed in the region	+
Infrastructure effect		
x_{10}	Existence of special economic zones (SEZs)	+
x_{11}	Share of companies using the internet	+
Entrepreneurial activity		
x_{12}	Share of small companies	+
x_{13}	Volume of foreign direct investment	+
Effect of large cities		
x_{14}	Distance from between the administrative center of the region and the nearest million-plus city	−
x_{15}	Square distance between the administrative center of the region and the nearest million-plus city	+

Abbreviation: GRP, gross regional product.
Source: Author compilation.

skills is measured by the share of the working population with a higher education. More workers with a higher education increase the existing skills in the industry. Gross domestic expenditure on R&D as a percentage of GRP and the number of domestic patent applications per 1 million workers aged 15–72 years are used to measure the impact of innovation activities in the region. The impact of wealth, which is measured by GRP per worker employed in the region, is also considered. The existence of special economic zones and the share of companies using the internet are considered to investigate the impact of infrastructure on the number of new specialization industries. The share of small companies and the volume of foreign direct investment are used to measure the impact of entrepreneurial activities on the new specialization industries. On the basis of the literature review, we expect that all these variables have a positive effect upon the emergence of new specialization industries.

To measure the spillover effects of large cities, we consider the distance between the administrative center of the region and the nearest million-plus city, given that newer industries emerge in regions that neighbor regions where million-plus cities are located. However, the relationship could be nonlinear. Therefore, we consider the square of the distance between the administrative center of the region and the nearest million-plus city. On the basis of the literature review, we assume that a higher distance from a million-plus city likely reduces the probability of the emergence of a new specialization industry. Larger cities become primary magnets of economic activity and the greater distance to a larger city indicates lower market potential and higher transport costs. However, when the distance increases further, a new specialization industry may appear to serve the local market.

3.1 | Measurement of variables, data sources, and a description of the data

Currently, there are a number of methodological approaches to identifying specializations and relatedness among industries. Some of them are based on the co-export of products, the flow of labor among industries, or combined measures of input–output links and shared labor pools (Hidalgo et al., 2018). For example, scholars are currently investigating technology specializations and complementarities on the basis of patent data (Balland et al., 2019; Neffke et al., 2011; Rigby, 2015). At the same time, the localization coefficient is traditionally used to determine technological specializations. Relatedness, in turn, is measured using a co-occurrence analysis (Balland & Boschma, 2021).

To determine the economic specializations of the regions and the links between them, we used the clusters of related industries approach proposed by Delgado et al. (2016) and then developed by Ketels and Protsiv (2016). This approach involves grouping economic activities into clusters (i.e., industries). In fact, these clusters, when analyzing their distribution between regions, reflect its specialization (Ketels & Protsiv, 2021). Our use of this approach is due, firstly, to the fact that it allows us to identify links between industries and not technologies, and secondly, its connection with the use of similar data on employment.

The authors first separated traded economic activities from non-traded ones (Porter, 2003). Unlike the former, the latter are represented in all regions (schools, cinemas, local services) almost evenly and depend upon the distribution of the population, which immediately makes it pointless to talk about any geographical specialization. Then, only traded activities were identified for analysis. Subsequently, these economic activities were grouped into 51 clusters of related industries (i.e., industries) on the basis of the co-location of enterprises, the concentration of employment, input–output tables, and the common labor pool (Delgado et al., 2016).

In accordance with this approach, we used a classifier for 51 industries, each of which consists of a set of related economic activities. Although it was originally used on the US economy, it was subsequently transferred by Ketels and Protsiv (Ketels & Protsiv, 2016) to the economies of European Union countries and their NACE classifiers of economic activity. Since 2016, Russia has switched to a classifier that is fully synchronized with NACE, which allows for the comparison of European and Russian regions, including using common metrics and approaches; for example, identifying specialization industries. It should also be noted that despite the difference in the historical and institutional conditions for the development of the economies of the United States, the European Union, and Russia, they turn out to be very similar in terms of employment in the 51 industries studied. In the United States, employment in them is 36% of the national economy, in the EU and Russia, 44% and 47%, respectively. In all three countries, they account for about 50% of total wages (HSE University, 2021). This similarity indicates the possibility of using the method to analyze the economies of countries with different characteristics and development trajectories.

To identify specializations in the regions of the EU countries, Ketels and Protsiv (2016) used the same 51 industries and calculated four coefficients: concentration, localization, productivity, and growth.

Unlike the approach of Ketels and Protsiv (2016), we identify specialization industries for each region by calculating only the concentration and localization coefficients. Our use of these two coefficients is due to the fact that each of them separately reveals different types of regions. The concentration coefficient helps to identify strong industries nationwide (giving an advantage to large regions), while the localization coefficient better identifies the activities that dominate on a regional level (which is more appropriate for the regions with small economies). We have also excluded measures of growth and productivity because they characterize specialization industries (their growth rate or wage levels) rather than describing them in the economic space. Moreover, they largely depend on the general level of development of the region and its location, that is, on factors external to the industry.

The concentration coefficient is calculated by the following equation:

$$S_{ijt} = \frac{E_{ijt}}{\sum_{i=1}^{m} E_{ijt}}, \tag{2}$$

where i is a region ($i = 1 \dots 85$), j is an industry ($j = 1 \dots 50$), and E is the number of employees in year t.

The localization coefficient is calculated by the following equation:

$$LQ_{ijt} = \frac{\frac{E_{ijt}}{\sum_{i=1}^{m} E_{ijt}} \quad \sum_{j=1}^{n} E_{ijt}}{\sum_{i=1}^{m} \sum_{j=1}^{n} E_{ijt}}, \tag{3}$$

where i, j, and E are the same as in Equation (2).

To confirm that a certain industry is a specialization industry of a given region in a particular year, we used the following conditions proposed by Ketels and Protsiv for European regions:

- The region needs to be found among the top regions in terms of concentration that altogether make up 80% of the nationwide employment in the respective industry.
- The region needs to be found among the top 20% of regions in terms of localization for a certain industry.

Links between industries have been identified in Delgado et al. (2016) on the basis of the US data. The revealed connections reflect the spatial, economic, and skill connections between different industries on the scale of the world's largest, most diversified, and integrated economy, which in our opinion, is an appropriate natural experiment to observe inter-industry links. Depending on the values of indicators of connectivity between individual economic activities (related industries), the level of connection between industries was determined. For example, video production is strongly associated with communications and music, the automotive industry has strong links with metalworking and production technology, and tobacco production was not closely associated with any other industry.

For the purposes of our study, only the strong links between industries identified in their work were used as the most reliable for application in different countries. We do not extrapolate links between industries in the United States and Russia, but we believe that if the link has been confirmed in the United States, then the industries have the potential to interact, or in other words, are complementary to one another. We believe that the complementarity between industries is largely due to technological and production factors but not to certain national or cultural characteristics.

In our study, we propose a method for identifying new specialization industries in the regions. Our dependent variable—the number of new specialization industries—is measured by the total number of new specialization industries that emerged in each region during the period of 2005 to 2015. We compared the specialization industries of the regions in two periods: 2005–2007 and 2013–2015. An industry of specialization is considered new if it was not present during the period of 2005–2007, and it became such in 2013–2015. The distribution of new industries of specialization by regions of Russia is shown in Figure 1.

For the investigated decade, 76 new specializations appeared in the regions of Russia. Data show that the Vladimir Region produced the highest number (6) of new specialization industries among the 80 regions in Russia. Out of 80 regions, 34 regions did not have any new specialization industries. The new specialization industries of the Vladimir region are biopharmaceuticals; business services; leather and related products; marketing, design, and publishing; medical devices; and metalworking technology. Biopharmaceuticals has a strong relatedness to downstream chemicals, which is not a specialization of Vladimir region; therefore, this industry is not connected with the industry portfolio of the region. Downstream chemicals are present in the neighboring Yaroslavl and Nizhny Novgorod regions; therefore, the region's new specialization industry fits into the industry portfolio of its neighbors.

We noticed that in some regions, the emergence of new specialization industries was accompanied by a reduction in industry employment both at the national and regional levels. When this occurred, a new specialization could be appointed even if there was a decrease of industry employment in the region. To exclude it, we introduced an additional variable for the number of new specialization industries. It does not consider new specialization industries in those regions in which the growth in the number of employees in the period from 2005 to 2015 was less than 20%. With this tightening of the criterion, the number of new specialization industries in Russian regions is reduced

FIGURE 1 Map of new specialization industries in the regions of Russia: 2005–2015. *Source*: Author compilation.

from 76 to 53. When used, the number of regions with at least one new specialization industry is reduced from 42 to 34. Using this approach makes the dependent variable stricter and more conservative.

To measure the complementariness of the new specialization industries, we consider a dummy variable of the share of new specialization industries related to existing specialization industries in the region. We assign 1 to a region if the region has 50% or more of the new industries of the specialization related (having strong ties to other specialization industries, according to Delgado's approach) to its industry portfolio, that is, existing specialization industries; otherwise, we give a 0 to the region.

Similarly, the dummy of the share of new specialization industries related to existing specialization industries in the neighboring region is also measured. Since there may be several neighboring regions next to a region, a list of its neighbors was determined for each region and a single list of the specialization industries of these regions was compiled.

Control variables such as number of employed workers, GRP per worker, existing specialization industries, share of companies using the internet, share of small companies, volume of foreign direct investments, number of domestic patent applications, and expenditures on R&D are measured by considering the average values for the period of 2005 to 2007. The indicator of the population with higher education is taken for 2002, the closest available year to 2005 with qualitative data. The dummy variable for SEZs takes the value of 1 if a region has at least one SEZ; otherwise, it takes the value of 0.[1] Distance from the administrative center of the region to the nearest million-plus city is measured by road distance in kilometers. Finally, to measure the regional market potential, we consider the following equations:

$$\text{The market potential (MP) of region } r \text{ in year } t \text{ is } \mathrm{MP}_{rt} = \sum\nolimits_{r \neq s} \frac{\mathrm{GDP}_{st}}{\mathrm{d}_{rs}}, \tag{2}$$

where GDP_{st} is the GDP of one's own region and neighboring region s in year t and d_{rs} is the distance from the region r to the region s that is taken as the minimum path along the roads from one regional center to another. Data sources for the indicators used in the model are presented in Table A1.

It should also be mentioned that we use lagged explanatory variables; that is, we assume that the initial conditions influenced the emergence of new specialization industries in the period from 2005 to 2015. Where possible, we take averages over several periods to avoid a possible 'spike' in the values for some regions.

3.2 | Descriptive statistics

Table 2 details the means, standard deviations, minimum, maximum, and coefficient of variations for the variables used for the regression estimates. We consider a total of 80 regions of Russia for the analysis.

Table A2 presents the correlations among independent values as being less than 0.80, except for the correlation between the distance and distance square variables (i.e., 0.97). Young (2017) indicated that if the absolute value of the Pearson correlation coefficient is less than 0.8, collinearity is less likely to exist. This is also evidenced by the estimated variance inflation factors (VIF) that are presented in Table 2. On the contrary, the consideration of the squared factor is based on the theoretical background; therefore, we kept it in the regression model.[2] In this situation, we reconsider the transformed distance variable by subtracting the mean from the original variable. We then take the square of the transformed distance variable.[3] Still, the collinearity problem cannot be avoided for the distance variable. However, VIFs reduced significantly from 27.30 to 14.04 for the distance variable and 24.87 to 11.98 for the

[1]There are four types of special economic zones in Russia (industrial, technological, tourism, and logistics). In our analysis, we take into account only industrial and logistics ones that were created before 2010.
[2]https://www.researchgate.net/post/Is_it_necessary_to_correct_collinearity_when_square_terms_are_in_a_model.
[3]https://statisticalhorizons.com/multicollinearity/

TABLE 2 Descriptive statistics.

Variable	Obs.	Mean	Std. dev.	Minimum	Maximum	C.V.	VIF
Dependent variables							
Number of new specialization industries (v1)	80	0.95	1.101	0	6	115.9	
Dummy of new specialization industries (1/0) (v2)	80	0.575	0.497	0	1	86.5	
Measurement of complementary specialization industries							
Dummy of the share of new specialization industries related to existing specialization industries in the region (1 indicates if more than 50%, otherwise 0) (v3)	80	0.36	0.484	0	1	133.4	2.59
Dummy of the share of new specialization industries related to existing specialization industries in the neighboring region (1 indicates if more than 50%, otherwise 0) (v4)	80	0.51	0.503	0	1	98.1	2.15
Other control variables							
Number of specialization industries (v5)	80	7.025	4.575	0	15.7	65.1	3.38
Log of the total number of employees (2005–2007 average) (v6)	80	13.38	0.903	10.44	15.624	6.8	4.04
Log of market potential (2005–2007 average) (v7)	78	13.31	1.005	10.72	17.271	7.5	2.39
Log of the share of working population with higher education (2002) (v8)	80	2.91	0.17	2.38	3.58	5.84	1.85
Log of gross domestic expenditure on R&D (GERD) as a percentage of GRP (2005–2007 average) (v9)	80	−0.76	1.119	−3.53	1.522	−147.0	2.69
Log of number of domestic patent applications per 1 million workers aged 15–72 (2005–2007 average) (v10)	80	4.93	1.293	−0.37	7.7	26.3	2.68
Log of GRP per worker employed in the region (2005–2007 average) (v11)	80	5.3	0.505	4.28	7.22	9.5	2.11
Dummy of special economic zones (industrial and logistics) (v12)	80	0.09	0.284	0	1	325.0	1.20
Share of companies using the internet (2005–2007 average) (v13)	80	4.15	0.25	2.98	4.58	5.94	1.97
Share of small companies (2005–2007 average) (v14)	80	2.90	0.53	−0.40	3.53	18.32	1.48
Volume of foreign direct investment (2005–2007 average) (v15)	72	10.35	2.19	−1.10	15.82	21.19	1.70
Distance from the administrative center of the region to the nearest million-plus city (by road) (v16)	80	0.00	1,503	−852.63	5147.38		14.04
Square distance from the administrative center of the region to the nearest million-plus city (by road) (v17)	80	2,233,537	5,813,994	1064.39	26,500,000	260.30	11.98
Average VIF							3.75

Abbreviations: CV, coefficient of variation; GRP, gross regional product; Obs, observation; Std dev, standard deviation; VIF, variance inflation factor.
Source: Author calculations.

distance square variable. Finally, parsimonious regression models have been shown to avoid multicollinearity problems. We run separate regressions when considering the two distance variables together, or we consider one of the distance variables in each regression model with other variables.

Furthermore, we also noticed that there is a higher correlation between dependent and independent variables. The correlation between the number of new specialization industries (v1) and the share of new specialization industries related to existing specialization industries in the neighboring region (v4) is 0.69. The correlation between the dummy of new specialization industries (v2) and the share of new specialization industries related to existing specialization industries in the neighboring region (v4) is 0.88. In this case, the additional use of parsimonious regression is supported. On the contrary, it may be the case that our regression models suffer from endogeneity or reverse causality due to a variable included within residuals that is correlated with the dependent variable and one independent variable.[4] To handle this problem, we conducted a two-stage least squares (2SLS) regression analysis.

Table A2 presents the raw correlation coefficients of the variables. The correlation coefficient between the dummy of the new specialization industry and the dummy of the share of new specialization industries related to existing specialization industries in the region (or neighboring region) is high. Therefore, these two variables are dropped when we consider the dummy of new specialization industries as the dependent variable in Table 4.

4 | RESULTS OF CALCULATIONS: HOW TO INITIATE AND PROGRESS WITH REGIONAL DIVERSIFICATION

Table 3 presents the results of regression models of the determinants of the emergence of new specialization industries based on Equation (1) by employing the ordinary least squares (OLS) and 2SLS methods. The number of new specialization industries is used as a dependent variable in the calculation. The models are estimated with different specifications and by the number of observations. As per our discussion in Section 3.2, we consider the parsimonious regression models rather than the full models. Regressions 1–9 list the results of a parsimonious model, excluding controls that are not found to be statistically significant or matched with the expected sign of the regression parameters. To correct heteroskedasticity, all of the OLS regressions report results with robust standard errors. The significant values of F-statistics for regressions 1–9 (except regression 6) indicate that the overall model is statistically significant. The higher values of R^2 indicate that the regressions models explain a good percentage of the total variations in the dependent variable. As some of the independent variables are in logarithmic form and the dependent variable is in non-logarithmic form, we explain the results as a linear-log regression model whenever it is applicable.

Regression results show that the effect of complementary specialization industries measured by the dummy for the share of new specialization industries related to existing specialization industries in the region and the dummy for the share of new specialization industries related to existing specialization industries in a neighboring region all have a positive and statistically significant effect on the number of new specialization industries. Regression model 3 (or 1) indicates that if the share of new specialization industries related to existing specialization industries in the region (or neighboring region) is more than 50%, then a 1.083 (or 1.184) larger positive effect is expected on the emergence of new specialization industries. These results support our Hypotheses 1 and 2 and indicate that complementarity matters for the emergence of new specialization industries. It supports the findings of earlier studies (e.g., Bahar et al., 2014; Balland et al., 2019; Balland & Boschma, 2021; Boschma, 2017; Boschma et al., 2017; Boschma & Iammarino, 2009; Hausmann & Klinger, 2006; Neffke et al., 2011; Rigby, 2015; Saviotti & Frenken, 2008).

As we discussed in Section 3.2, the dummy variable of the share of new specialization industries in the neighboring region (v4) could be an endogenous variable, as it has a very high correlation with the dependent variable (v2). To

[4]https://www.statalist.org/forums/forum/general-stata-discussion/general/1347369-a-high-correlation-coefficient-between-the-dependent-variable-and-a-control-variable

TABLE 3 Estimated results from OLS and 2SLS regressions.

Variables	Dependent variable									
	Number of new specialization industries (NSI)									Dummy of NSI
	OLS									2SLS
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Effect of complementary specialization industries										
Dummy of the share of new specialization industries related to existing specialization industries in the region	0.348	0.386	1.083***							
	(0.308)	(0.311)	(0.226)							
Dummy of the share of new specialization industries related to existing specialization industries in the neighboring region	1.184***	1.161***								0.993***
	(0.270)	(0.264)								(0.102)
Other control variables										
Number of specialization industries	0.037	0.035			0.096***					
	(0.042)	(0.043)			(0.029)					
Log of the total number of employees	−0.261	−0.292						0.341***		
	(0.185)	(0.187)						(0.097)		
Log of market potential	0.140	0.117		0.182**						−0.009
	(0.128)	(0.122)		(0.083)						(0.036)
Log of the share of working population with higher education	−0.363	−0.304	−0.676	−0.381	−0.852	−1.037	−1.208**			0.007
	(0.437)	(0.432)	(0.565)	(0.571)	(0.656)	(0.782)	(0.535)			(0.241)
Log of gross domestic expenditure on R&D (GERD) as a percentage of GRP	0.039	0.051		0.088		0.255*	0.165			0.018
	(0.120)	(0.131)		(0.094)		(0.146)	(0.155)			(0.032)
Log of the number of domestic patent applications per 1 million workers	−0.042	−0.048					0.279***			
	(0.164)	(0.165)					(0.069)			
Log of GRP per worker employed in the region	−0.123	−0.094						−0.295		
	(0.220)	(0.210)						(0.201)		

TABLE 3 (Continued)

	Dependent variable									
	Number of new specialization industries (NSI)									Dummy of NSI
	OLS									2SLS
Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Dummy of special economic zones	0.037	0.0376		0.009	0.003		−0.033	0.062		0.094
	(0.298)	(0.302)		(0.297)	(0.398)		(0.360)	(0.355)		(0.110)
Log of the share of companies using the internet	0.253	0.269			0.049	0.315	−0.061			
	(0.694)	(0.695)			(0.355)	(0.459)	(0.429)			
Log of the share of small companies	0.716*	0.709*	0.875**							0.055
	(0.389)	(0.387)	(0.402)							(0.130)
Log of the volume of foreign direct investment	0.064	0.066	0.085**							−0.009
	(0.055)	(0.056)	(0.042)							(0.015)
Distance from the administrative center of the region to the nearest million-plus city	0.000								−0.0004**	
	0.000)								0.0002	
Square distance from the administrative center of the region to the nearest million-plus city		−0.000							7.78e-08*	
		(0.000)							4.36e-08	
Constant	−0.235	0.179	−0.928	−1.042	2.557	2.854	3.610*	−2.058	0.776***	0.119
	(2.794)	(2.795)	(1.666)	(1.682)	(1.697)	(2.156)	(2.095)	(1.523)	0.143	(0.768)
F-stat/Wald chi^2	5.90***	6.25***	6.73***	13.14***	2.84**	1.20	5.16***	4.51***	3.71**	110.37***
Mean VIF	2.27	2.25	1.07	1.42	1.22	1.45	1.13	1.08	8.19	
R-squared	0.541	0.540	0.323	0.522	0.141	0.053	0.119	0.075	0.0630	0.726
Observations	70	70	72	78	80	80	80	80	80	70

Note: Robust standard errors in parentheses.
Abbreviations: GRP, gross regional product; VIF, variance inflation factor.
***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$.

TABLE 4 Validity of the 2SLS results.

Tests of endogeneity		Value		
Durbin (score) chi^2 (1)		4.72309**		
Wu–Hausman F (1.71)		4.41363**		
First-stage regression summary statistics				
F (1.72)		36.55***		
	10%	15%	20%	25%
2SLS size of nominal 5% Wald test	16.38	8.96	6.66	5.53
LIML size of nominal 5% Wald test	16.38	8.96	6.66	5.53

***p < 0.01, and **p < 0.05.

check this, we used instrumental variable approach—the 2SLS regression model. Regression model 10 reports the estimated 2SLS regression results. The results also show that the share of new specialization industries related to existing specialization industries in the neighboring region (v4) has a positive and statistically significant effect on the dummy for the number of new specialization industries (v2). The coefficient 0.993 indicates that a 10% increase in the dummy variable of share of new specialization industries related to existing specialization industries in the neighboring region (v4) increases the dummy for new specialization industries (v2) by 9.93%. We used the dummy for the share of new specialization industries related to existing specialization industries in the region (v3) as an instrument. Table 4 presents the results of the 2SLS regression. The null hypothesis of the Durbin and Wu–Hausman tests is that the variable under consideration can be treated as exogenous. Here both test statistics are highly significant, so we reject the null of exogeneity; we must continue to treat the dummy for the share of new specialization industries related to existing specialization industries in the neighboring region (v4) as endogenous. At the same time, higher F-statistics than critical values reported by the first-stage regression summary suggest that our instrument is not weak.[5]

As we are dealing with a count data model with a dependent variable that is non-negative, a Poisson regression would be more appropriate if the conditional distribution-dependent variable follows a Poisson distribution (Maddala, 1983). Nevertheless, the Poisson regression is based on the strong assumption of variance–mean equality, which has been rejected in various applications. Therefore, it is possible use the Poisson quasi-maximum likelihood estimator (QMLE), which allows the variance–mean ratio to be any positive constant σ^2. Given that our sample has a high number of zero counts, it is expected to be overdispersed, and thus, the negative binomial estimation is preferable to the Poisson regression (Cameron & Trivedi, 1986). This is also supported by the histogram of the dependent variable presented in Figure A1.

Table 5 presents the negative binomial regression estimation. The results show that the two variables that measure the effect of complementary specialization industries have a positive and statistically significant effect on the number of new specialization industries. The results are consistent with our OLS regression results presented in Table 3.

To test the resulting models (and to prove their stability, which inspires more confidence), we propose tightening the criteria for our dependent variable. With this criterion, we consider the regression results that are presented in Table 6.

[5]We also did similar testing of the 2SLS regression results by considering number of new specialization industries (v1) as a dependent variable. However, though other tests met expectations, the Durbin and Wu–Hausman tests show that the dummy variable for the share of new specialization industries related to existing specialization industries in the neighboring region (v4) is not an endogenous variable.

TABLE 5 Estimated results from the negative binomial regression: dependent variable (number of new specialization industries).

Variables	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16	Model 17
Effect of complementary specialization industries							
Dummy for the share of new specialization industries related to existing specialization industries in the region	0.247	0.292	1.217***				
	(0.361)	(0.363)	(0.290)				
Dummy for the share of new specialization industries related to existing specialization industries in the neighboring region	1.782***	1.735***					
	(0.472)	(0.465)					
Other control variables							
Number of specialization industries	0.036	0.034					0.078***
	(0.051)	(0.052)					(0.029)
Log of the total number of employees	−0.234	−0.285	−0.536**				
	(0.333)	(0.330)	(0.238)				
Log of market potential	0.153	0.113				0.343**	0.270*
	(0.231)	(0.228)				(0.134)	(0.138)
Log of the share of working population with higher education	−0.742	−0.623		−1.262		−1.441	−1.605**
	(1.320)	(1.316)		(0.806)		(0.888)	(0.735)
Log of the gross domestic expenditure on R&D (GERD) as a percentage of GRP	0.000	0.009			0.069	0.225*	
	(0.161)	(0.167)			(0.129)	(0.132)	
Log of the number of domestic patents applications per 1 million workers	0.004	−0.011	0.217		0.346**		
	(0.243)	(0.245)	(0.164)		(0.150)		
Log of GRP per worker employed in the region	−0.418	−0.378			−0.256		
	(0.490)	(0.488)			(0.332)		
Dummy for special economic zones	0.131	0.119	0.060		−0.005		
	(0.443)	(0.444)	(0.400)		(0.434)		
Log of the share of companies using the internet	0.250	0.282			−0.118		
	(1.130)	(1.135)			(0.788)		

(Continues)

TABLE 5 (Continued)

Variables	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16	Model 17
Log of the share of small companies	0.677	0.617	0.828*				
	(0.610)	(0.605)	(0.469)				
Log of the volume of foreign direct investment	0.135	0.138	0.160**				
	(0.107)	(0.108)	(0.075)				
Distance from the administrative center of the region to the nearest million-plus city	0.000			−0.001**			
	(0.000)			(0.0002)			
Square distance from the administrative center of the region to the nearest million-plus city		0.000		1.26e-07@			
		(0.000)		(7.75e-08)			
Constant	−0.883	−0.005	1.298	3.535	0.0560	−0.326	0.700
	(5.654)	(5.514)	(2.675)	(2.477)	(2.741)	(2.675)	(2.394)
Log likelihood	−69.87	−70.024	−82.55	−100.75	−100.19	−98.08	−94.476
Pseudo R^2	0.271	0.269	0.156	0.043	0.048	0.050	0.085
LR chi^2	51.81***	51.50***	30.55***	9.02**	10.15*	10.41**	17.62***
Observations	70	70	72	80	80	78	78

Note: Robust standard errors in parentheses.
Abbreviation: GRP, gross regional product.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, and $^{@}p < 0.11$.

TABLE 6 Estimated results of the OLS regression: dependent variable (modified number of new specialization industries).

Variables	Model 18	Model 19	Model 20	Model 21	Model 22	Model 23
Effect of complementary specialization industries						
Dummy for the share of new specialization industries related to existing specialization industries in the region	−0.011		0.655***			
	(0.215)		(0.201)			
Dummy for the share of new specialization industries related to existing specialization industries in the neighboring region	1.076***					
	(0.198)					
Other control variables						
Number of specialization industries	0.000	0.060***				
	(0.027)	(0.022)				
Log of the total number of employees						0.153**
						(0.076)
Log of market potential	0.155*					
	(0.092)					
Log of the share of working population with higher education	0.092	−0.138				
	(0.680)	(0.438)				
Log of gross domestic expenditure on R&D (GERD) as a percentage of GRP	0.071		0.0569			
	(0.095)		(0.106)			
Log of the number of domestic patents applications per 1 million labor forces	−0.062			0.171**		
	(0.088)			(0.082)		
Log of GRP per worker employed in the region	−0.010	−0.192				−0.140
	(0.189)	(0.145)				(0.179)
Dummy for special economic zones	0.205	0.217				0.208
	(0.294)	(0.320)				(0.287)
Log of the share of companies using the internet	0.0263			−1.121		−0.216
	(0.408)			(0.757)		(0.457)
Log of the share of small companies			0.225	0.498		0.183
			(0.322)	(0.393)		(0.178)
Log of the volume of foreign direct investment			0.053*	0.069*		
			(0.029)	(0.036)		
Distance from the administrative center of the region to the nearest million-plus city					−0.0002*	
					(0.000)	
Square distance from the administrative center of the region to the nearest million-plus city					3.84e-08	
					(3.24e-08)	

(Continues)

TABLE 6 (Continued)

Variables	Model 18	Model 19	Model 20	Model 21	Model 22	Model 23
Constant	−1.915	1.642	−0.725	2.314	0.577	−0.299
	(2.089)	(1.447)	(1.050)	(2.934)	(0.106)	(1.703)
R^2	0.463	0.103	0.195	0.103	0.045	0.047
Observations	78	80	72	72	80	80

Note: Robust standard errors in parentheses,
Abbreviation: GRP, gross regional product.
$^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table 6 presents the OLS regression results with the modified dependent variable. Regression models 18–23 present the parsimonious model. The results show a similar effect exerted by complementary industries on the emergence of new specialization industries. A dummy for the share of new specialization industries related to existing specialization industries in the region and a dummy for the share of new specialization industries related to existing specialization industries in the neighboring region all have statistically significant effects on the dependent variable. For example, regression model 18 shows that a 10% increase in the dummy for the share of new specialization industries related to existing specialization industries in the neighboring region increases the number of new specialization industries by about 11%. This again bolsters our claim that complementarity matters for the emergence of new specialization industries in Russia.

We also estimated the probit models to determine which factors determine the beginning of the process of diversification of the region's economy and to explicitly identify the binary nature of our dependent variables. Findings are more or less similar. In our sample dataset of 80 regions, 34 regions did not report the emergency of new specialization industries from the period of 2005 to 2015. Therefore, the consideration of a binary set up is very much relevant. Table 7 presents the regression output of the probit models. The statistically significant values of the likelihood ratio (LR) chi-squares indicate that our model as a whole is statistically significant; that is, it fits significantly better than a model with no predictors.

Among the control variables, the number of specialization industries, the total number of employees, the number of domestic patent applications per 1 million workers, gross domestic expenditure on R&D (GERD) as a percentage of GRP, the share of small companies, and the volume of foreign direct investment have a positive and statistically significant impact upon the number of new specialization industries. The impact of entrepreneurial activities is measured by the volume of foreign direct investment and has an expected effect on the dependent variable. In contrast, the share of working population with a higher education has a negative effect on the emergence of new specialization industries. The distance between the administrative center of the region and the nearest million-plus city (by road) has a negative effect (as expected) on the emergence of new specialization industries. The square of the distance between the administrative center of the region and the nearest million-plus city (by road) has a positive effect on the dependent variable. Market potential has a positive and statistically significant effect on the emergence of new specialization industries. This indicates that demand plays an important role in increasing new specialization industries. However, the GRP per worker employed in the region, the share of companies using the internet, and the dummy for special economic zones do not have any impact upon the dependent variable.

The regression results in Tables 5 and 6 also demonstrate similar results for the number of specialization industries, market potential, the share of small companies, the volume of foreign direct investment, and domestic patent application. The regression results in Table 5 also show that education has a negative effect on the number of new specialization industries. Furthermore, the regression results in Table 6 show similar effects of distance on the dependent variable. Table 7 reports the probit regression estimation. The statistically significant positive effect of the number of specialization industries indicates that the initial diversified set of specialization industries plays an

TABLE 7 Estimated results from the probit regression: dependent variable (dummy for new specialization industries, 1/0).

Variables	Model 24	Model 25	Model 26	Model 27	Model 28
Number of specialization industries	0.130***				
	(0.045)				
Log of the total number of employees			0.600***		
			(0.213)		
Log of market potential	−0.183				
	(0.220)				
Log of the share of working population with higher education	−1.598				−1.716*
	(1.112)				(0.917)
Log of gross domestic expenditure on R&D (GERD) as a percentage of GRP		−0.098	0.075		0.312**
		(0.171)	(0.145)		(0.141)
Log of the number of domestic patents Applications per 1 million workers	0.737***				
	(0.283)				
Log of GRP per worker employed in the region	−0.301	−0.415	−0.391		−0.107
	(0.486)	(0.326)	(0.359)		(0.297)
Dummy for special economic zones	0.715	0.474	0.421		
	(0.677)	(0.536)	(0.535)		
Log of the share of companies using the internet	−0.710		−0.400		0.254
	(1.386)		(0.758)		(0.710)
Log of the share of small companies	0.531	0.0404			
	(0.772)	(0.291)			
Log of the volume of foreign direct investment	0.065				
	(0.097)				
Distance from the administrative center of the region to the nearest million-plus city				−0.0005*	
				(0.0003)	
Square distance from the administrative center of the region to the nearest million-plus city				8.49e-08	
				7.04e-08	
Constant	6.225	1.259	−4.074	0.0043	6.368*
	(5.731)	(1.732)	(3.197)	(0.2109)	(3.265)
LR chi^2	15.15*	12.90**	12.46**	6.30*	6.63*
Log likelihood	−38.605	−48.101	−48.317	−52.599	−51.235
Pseudo R^2	0.164	0.118	0.114	0.036	0.036
Observations	70	80	80	80	80

Note: Standard errors in parentheses.
Abbreviation: GRP, gross regional product.
***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$.

important role in the emergence of new specialization industries. The coefficient of regression model 24 shows that with a one-unit increase in the number of specialization industries, the *z*-score of the dependent variable increases by 0.130. The results show similar results for the level of employment, education, and R&D expenditures on the emergence of new industry specializations.

TABLE 8 The influence of different variables upon the beginning of the process of diversification and increase in the number of new specialization industries.

Variables	Launching of the diversification process	Number of new specialization industries
Share of new specialization industries related to existing specialization industries in the region	+++	+++
Share of new specialization industries related to existing specialization industries in the neighboring region	+++	+++
Control variables		
Number of specialization industries	++	+++
Total number of employees	++	++
Market potential	Not found	++
Gross domestic expenditure on R&D (GERD) as a percentage of GRP	+	Not found
Number of domestic patent applications per 1 million workers	+	+
GRP per worker employed in the region	Not found	Not found
Log of the share of companies using the internet	Not found	Not found
Log of the share of small companies	Not found	+
Log of the volume of foreign direct investment	Not found	++
Special economic zones	Not found	Not found
Distance from the administrative center of the region to the nearest million-plus city	Not found	−
Square distance from the administrative center of the region to the nearest million-plus city	Not found	+
Share of the working population with higher education	−	−

Note: +++ indicates a positive influence with a significance level of 1% and below, ++ indicates a positive influence with a significance level of 5%, + indicates a positive influence with a significance level of 10%, and − indicates a negative influence with the significance level of 10%.
Abbreviation: GRP, gross regional product.
Source: Author calculations.

5 | CONCLUSIONS AND POLICY IMPLICATIONS

Currently, the scientific literature continues to discuss the factors affecting the diversification of regional economies. A special place in this discussion is played by the issue of intra- and interregional complementarity. Not so long ago, Balland and Boschma (2021), on the basis of patent data, empirically confirmed the importance of interregional complementarity in the absence of new technologies in regions. Moreover, the link with complementary regions can compensate for the weakness of internal factors affecting diversification.

However, to determine the significance of connectivity for the regional economy, we used a different approach, based not on the technological component, but on the distribution of employment. To do this, we optimized the clusters of related industries approach proposed by Delgado et al. (2016) and Ketels and Protsiv (2016). Using it, we analyzed the change in the specialization industries of Russian regions in 2005–2015 and then proposed variables that made it possible to assess the connectivity of regional industry portfolios and the portfolios of neighboring regions. In contrast to the approach proposed by Balland and Boschma (2021), we consider the connectivity not with all possible regions, but solely with neighboring ones due to their spatial proximity.

In our work, we investigated the influence of complementary intraregional and interregional linkages upon the emergence of new specialization industries in the Russian regions on the basis of employment statistics. To more accurately identify the effect of complementarity, we consider other important variables such as skills, wealth, infrastructure, distance to million-plus cities, and demand. We take into account time lags and consider the OLS, the negative binomial regression, and the probit regression model for analysis. The probit regression allows us to determine the factors that influence the launch of the related diversification of the region's economy. OLS, in turn, allows us to focus on the factors that increase the number of new specialization industries in the region. The negative binomial regression analysis helps us model the overdispersed count variables. The influence of various factors upon the launch of the diversification process and the increase in the number of new specialization industries in the region are presented in Table 8.

The results obtained showed that there are a number of factors that simultaneously affect the process of launching the diversification of the regional economy and increasing the number of new specialization industries. Common to both cases are the connection with one's own and the neighboring region's industry portfolio, the number of a region's specialization industries, the number of employees, and patent activity. Nefedova and Treivish (2020) found that historically Russian industrial diversification formed on the basis of city differential, for example, textiles in Moscow and Vladimir; glass and pottery in Bryansk, Tver, and Vladimir; and metal and weapons in Tula. Similarly, Galtseva et al. (2022) found that the creation of mono-industrial characteristics in the northern economy of Russia was based on the production of gold. Polyachenko (2022) also suggested that different Russian cities are specialized in different industries such as oil, gas, mining, military, and nuclear production. Axenov et al. (2020) and Pilyasov and Molodtsova (2022) highlighted the dominance of the big and biggest industrial enterprises in cities. Our results bolster these findings and suggest that the number of specialization industries in a region is very important for creating new ones.

At the same time, another innovative indicator turned out to be important in launching the diversification process: the share of research and development costs in the GRP, but the distance between regions turned out to be insignificant. On the contrary, the distance between regions and market potential turned out to be significant precisely for increasing the number of new specialization industries. Market potential reflects the accessibility of external regional markets and is very important for value-added of Russian regions through industry specialization. The result supports the finding of Kolomak (2020).

The wealth of the region and the existing infrastructure turned out to be absolutely insignificant for both processes. Therefore, the use of the internet, which is used as a proxy for infrastructure, does not have an appreciable impact upon new branches of industry specialization. This result does not support the findings of Yu (2022) and Zhang et al. (2021). The probable reason could be that the interregional difference in GRP in Russia is very high, and high capita GRP does not necessarily mean inclusive development of the region (Mikheeva, 2020). The impact of the share of small companies on the new specialization industry does not support the findings of Hannah and Kay (1977). However, it supports the findings of Motała et al. (2018). The positive impact of foreign direct investment on the number of new specialization industries supports the finding of Mironko (2020), Castellani et al. (2020), and Wang et al. (2016). The impact of the distance variable supports previous research findings (e.g., Castaldi et al., 2015; Essletzbichler, 2015; Glaeser et al., 1992; Jacobs, 1969; Panne, 2004; Schumpeter, 1912). The positive effect of market potential replicates findings in earlier studies (e.g., Schmookler, 1966; UN, 2018; Zhao & Guan, 2013).

Our results confirm the conclusions made earlier by Balland and Boschma (2021) that being close to other wealthy regions is not the only factor impacting diversification, but the critical factor involves being connected to regions that provide complementary capabilities.

Our similar results were obtained on the basis of employment data in specialization industries in the regions of Russia rather than on patent data (technological relatedness) of regions in the European Union. This approach removes one of the limitations of patent data, which is that it is often poorly associated with medium- and low-tech industries and also weakly associated with the service sector.

As a result, the most important contribution of our models is the variables associated with the intraregional and interregional complementarity of specialization industries. For the regional authorities, it is important to understand the sectoral structure and industry links not only of one's own region, but also those of neighboring regions. Therefore, it is important that Governments must also realize that new specialization industries tend to appear adjacent to existing ones. Determining the priorities of regional economic specialization should consider not only one's own economic profile but also the industries of neighboring regions. However, this raises the problem of asymmetrical information when the authorities have knowledge only about their region and do not have a mandate to study their neighbors. In these conditions, authorities rely solely on information about their region, while businesses focus on supply and demand, which are not limited by administrative boundaries. A solution in this situation can come from the national government through interventions in the form of establishing uniform rules for determining the economic priorities of the regions and their public publication. However, the strategy for the Spatial Development Strategy of Russia, published in 2019, does not bring the necessary clarity to the connections between regional economies, and it does not focus on the need to take into account the sectoral profile of neighboring regions. In this regard, the creation of an information base on the current specializations of Russian regions, similar to the US Cluster Map Project and EU Smart Specialization Platform, looks reasonable. Our results confirm the need for a more thorough study of the country's spatial planning documents, which are still largely considered schemes for the development of infrastructure (transport, energy, etc.).

Our results speak in favor of tools aimed at developing cooperation between regions. The distribution of subsidies on a competitive basis for the implementation of projects and the factor of interregionality could be key initiatives, as well as compliance with specialization industries of the regions. For example, industrial clusters of the Ministry of Industry and Trade of Russia can be used as such a tool. Currently, within its framework, there is no restriction on the creation of interregional clusters, but now their share in the total number of industrial clusters does not exceed 10%. Such support instruments can stimulate the development of new specialization industries in regions located close to one another.

Finally, as our results showed, patent activity is significant for starting the process of creating new specialization industries and increasing their number. Meanwhile, Russia is one of the few countries where patent activity has declined in recent years. Since the creation of new industries, especially in high-tech sectors, is accompanied and supported by patent activity, one of the possible options is targeted support for regional companies involved in the development of technologies and their introduction into production.

ACKNOWLEDGMENTS

The article is based on a study funded by the Basic Research Program of HSE University. We are grateful to the three anonymous referees and to the editor of this journal for their detailed and constructive comments that have significantly improved the quality of the study.

ORCID

Sabyasachi Tripathi https://orcid.org/0000-0003-1980-5477

REFERENCES

Andersen, T. B., & Dalgaard, C. J. (2011). Flows of people, flows of ideas, and the inequality of nations. *Journal of Economic Growth*, *16*(1), 1–32. https://doi.org/10.1007/s10887-011-9060-7

Axenov, K., Timoshina, A., & Zemlyanova, A. (2020). Commercial redevelopment of industrial and residential periphery of Russian metropolis: St. Petersburg, 1989–2017. *Regional Science Policy & Practice*, *12*(4), 705–722. https://doi.org/10.1111/rsp3.12300

Bahar, D., Hausmann, R., & Hidalgo, C. A. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics*, *92*(1), 111–123. https://doi.org/10.1016/j.jinteco.2013.11.001

Balland, P. A., & Boschma, R. (2021). Complementary interregional linkages and smart specialisation: An empirical study on European regions. *Regional Studies*, *55*(6), 1059–1070. https://doi.org/10.1080/00343404.2020.1861240

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900

Bathelt, H., & Boggs, J. (2003). Toward a reconceptualization of regional development paths: Is Leipzig's media cluster a continuation of or a rupture with the past? *Economic Geography*, *79*(3), 265–293.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, *51*(3), 351–364. https://doi.org/10.1080/00343404.2016.1254767

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy*, *44*(10), 1902–1914. https://doi.org/10.1016/j.respol.2015.06.013

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, *85*(3), 289–311. https://doi.org/10.1111/j.1944-8287.2009.01034.x

Boschma, R., Martín, V., & Minondo, A. (2017). Neighbour regions as the source of new industries. *Papers in Regional Science*, *96*(2), 227–245. https://doi.org/10.1111/pirs.12215

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. *Papers in Regional Science*, *91*(2), 241–256.

Burtch, G., Carnahan, S., & Greenwood, B. (2018). Can you gig it? An empirical examination of the gig economy and entrepreneurial activity. *Management Science*, *64*(12), 5497–5520. https://doi.org/10.1287/mnsc.2017.2916

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29–53. https://doi.org/10.1002/jae.3950010104

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional Studies*, *49*(5), 767–781. https://doi.org/10.1080/00343404.2014.940305

Castellani, D., Marin, G., Montresor, S., & Zanfei, A. (2020). "Foreign direct investments and regional specialization in environmental technologies," SEEDS working papers 0620, SEEDS, sustainability environmental economics and dynamics studies.

Chen, J. (2018). Interpreting economic diversity as the presence of multiple specializations. RRI working paper 2018–02, Morgantown, WV: West Virginia University.

Content, J., & Frenken, K. (2016). Related variety and economic development: A literature review. *European Planning Studies*, *24*(12), 2097–2112. https://doi.org/10.1080/09654313.2016.1246517

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1–38. https://doi.org/10.1093/jeg/lbv017

Eferin, Y., & Kutsenko, E. (2021). Adjusting smart specialization concept for Russian regions. *Public Administration Issues*, *3*, 75–110.

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies*, *49*(5), 752–766. https://doi.org/10.1080/00343404.2013.806793

Falcıoğlu, P. (2011). Location and determinants of productivity: The case of the manufacturing industry in Turkey. *Emerging Markets Finance and Trade*, *47*(sup5), 86–96.

Frenken, K., Oort, V. F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296

Galtseva, N. V., Favstritskaya, O. S., & Sharypova, O. A. (2022). Prospects for modernization of regional economies in remote regions of far north-east Russia. *Regional Science Policy & Practice*, *14*(1), 27–47. https://doi.org/10.1111/rsp3.12288

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics*, *128*(2), 861–915. https://doi.org/10.1093/qje/qjs073

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152. https://doi.org/10.1086/261856

Hafner, K. A. (2018). Growth-instability frontier and industrial diversification: Evidence from European gross value added. *Papers in Regional Science*, *98*(2), 799–824.

Hannah, L., & Kay, J. A. (1977). *Concentration in modern industry*. Theory, Measurement and the UK Experience.

Hartog, M., Boschma, R., & Sotarauta, M. (2012). The impact of related variety on regional employment growth in Finland 1993–2006: High-tech versus medium/low-tech. *Industry and Innovation*, *19*(6), 459–476.

Hausmann, R., & Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, *16*, 309–342. https://doi.org/10.1007/s10887-011-9071-4

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

Hausmann, R., & Klinger, B. (2006). *Structural transformation and patterns of comparative advantage in the product space* CID Working paper 128. Harvard University Press.

He, C., Zhu, S., & Yang, X. (2017). What matters for regional industrial dynamics in a transitional economy? *Area Development and Policy*, *2*(1), 71–90. https://doi.org/10.1080/23792949.2016.1264867

Hidalgo, C. A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., & Zhu, S. (2018). The principle of relatedness. In *International conference on complex systems* (pp. 451–457). Springer.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487. https://doi.org/10.1126/science.1144581

HSE University. (2021). Atlas of economic specialisation of Russian regions. https://publications.hse.ru/books/473034767, https://doi.org/10.17323/978-5-7598-2379-7

Isaksen, A. (2015). Industrial development in thin regions: Trapped in path extension? *Journal of Economic Geography*, *15*(3), 585–600. https://doi.org/10.1093/jeg/lbu026

Jacobs, J. (1969). *The economy of the cities*. Random House.

Kadochnikov, S., & Fedyunina, A. (2013). Export diversification in the product space and regional growth: Evidence from Russia, Papers in Evolutionary Economic Geography# 13.27.

Kerr, W. R. (2008). Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics*, *90*(3), 518–537. https://doi.org/10.1162/rest.90.3.518

Ketels, C., & Protsiv, S. (2016). *European cluster panorama*. Center for Strategy and Competitiveness, Stockholm School of Economics.

Ketels, C., & Protsiv, S. (2021). Cluster presence and economic performance: A new look based on European data. *Regional Studies*, *55*(2), 208–220. https://doi.org/10.1080/00343404.2020.1792435

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping knowledge space and technological relatedness in U.S. cities. *European Planning Studies*, *21*(9), 1374–1391. https://doi.org/10.1080/09654313.2012.755832

Kolomak, E. (2020). Spatial development of the post-Soviet Russia: Tendencies and factors. *Regional Science Policy & Practice*, *12*(4), 579–594. https://doi.org/10.1111/rsp3.12299

Kolomak, E. (2021). The urban system of Russia from 1991–2020: Gradual development instead of radical transformation. *Area Development and Policy*, *7*, 427–441. https://doi.org/10.1080/23792949.2021.2002168

Krugman, P. R. (1993). On the relationship between trade theory and location theory. *Review of International Economics*, *1*(2), 110–122. https://doi.org/10.1111/j.1467-9396.1993.tb00009.x

Lyubimov, I. (2019). Russia's diversification prospects. *Russian Journal of Economics*, *5*(2), 177–198. https://doi.org/10.32609/j.ruje.5.34753

Lyubimov, I., Gvozdeva, M., & Lysyuk, M. (2018). Towards increased complexity in Russian regions: Networks, diversification and growth, BOFIT discussion papers, no. 8/2018, ISBN 978-952-323-221-1, Bank of Finland, Institute for Economies in Transition (BOFIT), Helsinki, https://nbn-resolving.de/urn:NBN:fi:bof-201804171448

Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge University Press.

Mao, X., & He, C. (2019). Product relatedness and export specialization in China's regions: A perspective of global–local interactions. *Cambridge Journal of Regions, Economy and Society*, *12*(1), 105–126. https://doi.org/10.1093/cjres/rsy031

Maskell, P., & Malmberg, A. (1999). Localized learning and industrial competitiveness. *Cambridge Journal of Economics*, *23*(2), 167–185. https://doi.org/10.1093/cje/23.2.167

Mikheeva, N. (2020). Qualitative aspect of the regional growth in Russia: Inclusive development index. *Regional Science Policy & Practice*, *12*(4), 611–626. https://doi.org/10.1111/rsp3.12289

Mironko, A. (2020). To make or to serve–industrial specialization drivers of foreign subsidiaries in service and manufacturing sectors across agglomerations in a transition economy. Available at SSRN 3574531.

Motała, D., Bystryakov, A. Y., & Pizengolts, V. M. (2018). Level of specialization and management methods in small and medium enterprises of the gas industry. *Management and Production Engineering Review*, *9*(2), 80–87.

Nefedova, T., & Treivish, A. (2020). Russia's early developed regions within shrinking social and economic space. *Regional Science Policy & Practice*, *12*(4), 641–655. https://doi.org/10.1111/rsp3.12278

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, *87*(3), 237–265. https://doi.org/10.1111/j.1944-8287.2011.01121.x

Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management and Governance*, *4*, 69–92. https://doi.org/10.1023/A:1009941416749

Panne, G. V. D. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics*, *14*, 593–604. https://doi.org/10.1007/s00191-004-0232-x

Pasinetti, L. L. (1993). *Structural economic dynamics*. Cambridge University Press.

Petralia, S., Balland, P.-A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5), 956–969. https://doi.org/10.1016/j.respol.2017.03.012

Pilyasov, A., & Molodtsova, V. (2022). Resilience capacity of contemporary Russian Arctic cities: Methodological approaches and quantitative assessments. *Regional Science Policy & Practice*, *14*(1), 99–126. https://doi.org/10.1111/rsp3.12409

Polyachenko, A. (2022). The resilience of Russian Arctic cities 1989–2017. *Regional Science Policy & Practice*, *14*(4), 796–825. https://doi.org/10.1111/rsp3.12527

Porter, M. (2003). The economic performance of regions. *Regional Studies*, *37*(6-7), 549–578.

Quatraro, F. (2010). Knowledge coherence, variety and economic growth: Manufacturing evidence from Italian regions. *Research Policy*, *39*(10), 1289–1302. https://doi.org/10.1016/j.respol.2010.09.005

Rastvortseva, N. S., & Chentsova, S. A. (2015). Regional specialization and geographical concentration of industry in Russia. *Regional Science Inquiry*, *7*(2), 97–106.

Rigby, D. L. (2015). Technological relatedness and knowledge space: Entry and exit of US cities from patent classes. *Regional Studies*, *49*(11), 1922–1937. https://doi.org/10.1080/00343404.2013.854878

Saviotti, P. P., & Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, *18*(2), 201–218. https://doi.org/10.1007/s00191-007-0081-5

Schmookler, J. (1966). *Invention and economic growth*. Harvard University Press.

Schumpeter, J. A. (1912). *The theory of economic development, tenth printing 2004*. Transaction Publishers.

Stein, E., & Daude, C. (2007). Longitude matters: Time zones and the location of foreign direct investment. *Journal of International Economics*, *71*(1), 96–112. https://doi.org/10.1016/j.jinteco.2006.01.003

United Nations. (2018). *Industrial development report 2018. Demand for manufacturing: Driving inclusive and sustainable industrial development*. United Nations Industrial Development Organization.

Wang, Y., Ning, L., Li, J., & Prevezer, M. (2016). Foreign direct investment spillovers and the geography of innovation in Chinese regions: The role of regional industrial specialization and diversity. *Regional Studies*, *50*(5), 805–822. https://doi.org/10.1080/00343404.2014.933800

Xiao, J., Boschma, R., & Andersson, M. (2018). Industrial diversification in Europe: The differentiated role of relatedness. *Economic Geography*, *94*(5), 514–549. https://doi.org/10.1080/00130095.2018.1444989

Young, D. S. (2017). *Handbook of regression methods* (pp. 109–136). CRC Press.

Yu, B. (2022). The impact of the internet on industrial green productivity: Evidence from China. *Technological Forecasting and Social Change*, *177*, 121527. https://doi.org/10.1016/j.techfore.2022.121527

Zemtsov, S., Barinova, V., & Semenova, R. (2019). The risks of digitalization and the adaptation of regional labor markets in Russia. *Foresight and STI Governance*, *13*(2), 84–96. https://doi.org/10.17323/2500-2597.2019.2.84.96

Zemtsov, S., & Kotsemir, M. (2019). An assessment of regional innovation system efficiency in Russia: The application of the DEA approach. *Scientometrics*, *120*(2), 375–404. https://doi.org/10.1007/s11192-019-03130-y

Zemtsov, S. P., Komarov, V. M., & Barinova, V. A. (2022). Uncovering new economy potential of Russian regions on the basis of the last 20 years dynamics' analysis. *Ekonomika Regiona [Economy of Regions]*, *18*(1), 92–104. https://doi.org/10.17059/ekon.reg.2022-1-7

Zhang, H., Sun, Z., Elahi, E., & Zhang, Y. (2021). Internet development, level of industrial synergy, and urban innovation. *Sustainability*, *13*(22), 12410. https://doi.org/10.3390/su132212410

Zhao, Q., & Guan, J. (2013). Love dynamics between science and technology: Some evidences in nanoscience and nanotechnology. *Scientometrics*, *94*, 113–132. https://doi.org/10.1007/s11192-012-0785-7

How to cite this article: Kutsenko, E., Tripathi, S., & Tyurchev, K. (2023). Does complementarity matter for the emergence of new specialization industries in the regions of Russia? *Regional Science Policy & Practice*, *15*(9), 2126–2155. https://doi.org/10.1111/rsp3.12668

APPENDIX A

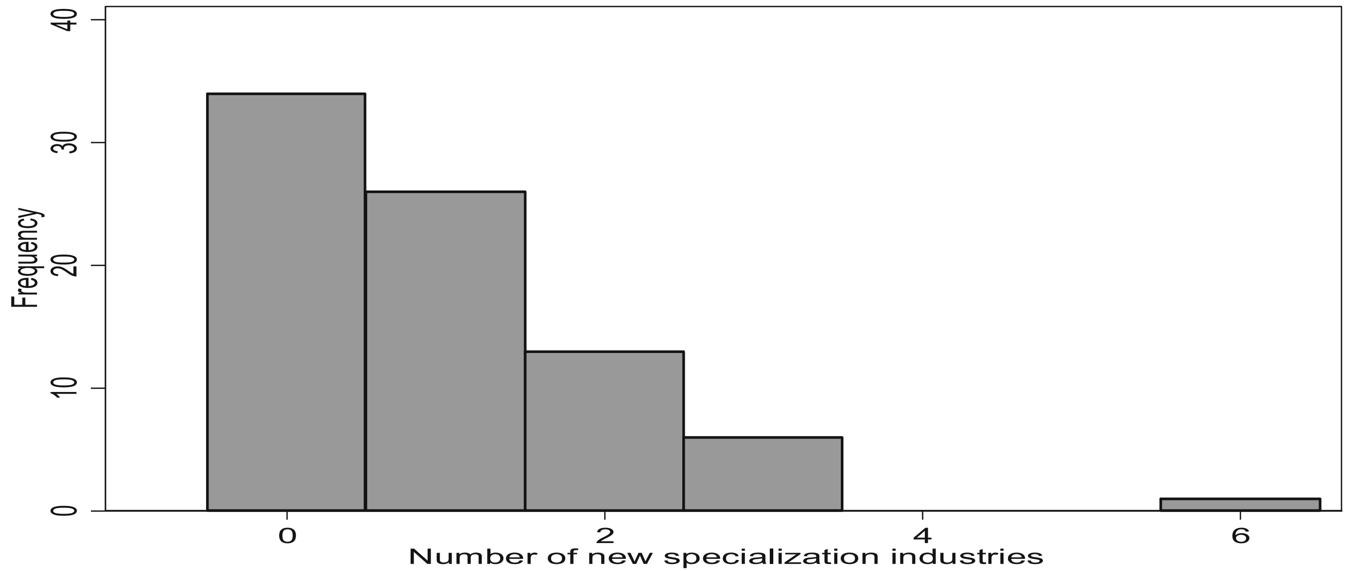

FIGURE A1 Histogram of the dependent variable.

TABLE A1 Data sources for modeling.

Variable	Indicator	Source
Number of new specialization industries from 2005 to 2015		Calculated by authors
Dummy for new specialization industries (1/0) from 2005 to 2015		Calculated by authors
Number of specialization industries (2005–2007 average)	Number of employees	Rosstat
	Annual payroll	Rosstat
Dummy for the share of new specialization industries related to existing specialization industries in the region (1 indicates if more than 50%, otherwise 0)		Calculated by authors on the basis of US cluster mapping[a]
Dummy for the share of new specialization industries related to existing specialization industries in neighboring region (1 indicates if more than 50%, otherwise 0)		Calculated by authors on the basis of US cluster mapping[b]
Log of the total number of employees (2005–2007)	Number of employees	Rosstat
Log of GRP per worker employed in the region (2005–2007 average)	Number of employees	Rosstat
	Regional GDP	Rosstat
Dummy for special economic zones (industrial and logistics)		Russian special economic zones[c]
Log of the share of companies using the internet (2005–2007 average)	Number of companies using the internet in their activities	Rosstat
Log of the share of small companies (2005–2007 average)	Number of small companies	Rosstat
Log of volume of foreign direct investment (2005–2007 average)	Volume of foreign direct investment	Rosstat
Distance from the administrative center of the region to the nearest million-plus city (by road)	City population	Rosstat
	Distance	Google maps
Square distance from the administrative center of the region to the nearest million-plus city (by road)		
Log of the share of the working population with higher education (2002)	Population with higher education aged 25–64 years	Rosstat (population census)
Log of number of domestic patent applications per 1 million workers aged 15–72 years (2005–2007 average)	Domestic patent applications	Rospatent
	Labor force aged 15–72 years	Rosstat
Log of gross domestic expenditure on R&D (GERD) as a percentage of GRP (2005–2007 average)	Gross domestic expenditure on R&D	Rosstat
	Regional GDP	Rosstat
Log of market potential (2005–2007 average)	Regional GDP	Rosstat
	Distance	Google maps

Abbreviation: GRP, gross regional product.
[a]US Claster Mapping. http://www.clustermapping.us/ (22.06.2021).
[b]US Claster Mapping. http://www.clustermapping.us/ (22.06.2021)).
[c]Russian Special Economic Zones. http://eng.russez.ru/ (22.06.2021).
Source: Author creation.

TABLE A2 Simple correlation coefficients.

	v1	v2	v3	v4	v5	v6	v7	v8
Number of new specialization industries (v1)	1.00							
Dummy of new specialization industries (v2)	0.75*	1.00						
Dummy of the share of new specialization industries related to existing specialization industries (v3)	0.53*	0.65*	1.00					
Dummy of the share of new specialization industries related to existing specialization industries in the neighboring region (v4)	0.69*	0.88*	0.63*	1.00				
Number of specialization industries (v5)	0.35*	0.36*	0.42*	0.33*	1.00			
Total number of employees (v6)	0.24*	0.34*	0.40*	0.28*	0.74*	1.00		
Market potential (v7)	0.29*	0.15	0.26*	0.15	0.45*	0.48*	1.00	
Share of working population with higher education (v8)	0.01	0.00	0.05	−0.01	0.33*	0.37*	0.32*	1.00
Gross domestic expenditure on R&D (v9)	0.19	0.18	0.23*	0.14	0.61*	0.39*	0.24*	0.52*
Number of domestic patent applications (v10)	0.29*	0.35*	0.39*	0.32*	0.64*	0.69*	0.44*	0.50*
GRP per worker employed in the region (v11)	−0.05	−0.03	−0.08	−0.05	0.25*	0.29*	−0.05	0.29*
Special economic zones (v12)	0.01	0.09	−0.05	0.04	0.05	0.06	−0.05	0.06
Share of companies using the internet (v13)	0.05	0.01	0.01	0.00	0.27*	0.29*	0.06	0.48*
Share of small companies (v14)	0.21	0.11	0.20	0.08	0.34*	0.29*	0.20	0.35*
Volume of foreign direct investment (v15)	0.16	0.06	0.00	0.10	0.33*	0.39*	0.29*	0.29*
Distance from the administrative center of the region to the nearest million-plus city (v16)	−0.21	−0.18	−0.19	−0.22	−0.39*	−0.49*	−0.62*	−0.07
Square distance from the administrative center of the region to the nearest million-plus city (v17)	−0.16	−0.14	−0.15	−0.17	−0.33*	−0.45*	−0.55*	−0.01

Note: See Table 2 for full variable definitions. The correlation coefficients are based on 78 observations; GRP, gross regional product; * indicates the significance level at 5% and below.

TABLE A2 (Continued)

	v9	v10	v11	v12	v13	v14	v15	v16	v17
Number of new specialization industries (v1)									
Dummy of new specialization industries (v2)									
Dummy of the share of new specialization industries related to existing specialization industries (v3)									
Dummy of the share of new specialization industries related to existing specialization industries in the neighboring region (v4)									
Number of specialization industries (v5)									
Total number of employees (v6)									
Market potential (v7)									
Share of working population with higher education (v8)									
Gross domestic expenditure on R&D (v9)	1.00								
Number of domestic patent applications (v10)	0.47*	1.00							
GRP per worker employed in the region (v11)	0.15	0.15	1.00						
Special economic zones (v12)	0.12	0.10	0.14	1.00					
Share of companies using the internet (v13)	0.24*	0.37*	0.54*	0.07	1.00				
Share of small companies (v14)	0.26*	0.57*	0.25*	0.07	0.55*	1.00			
Volume of foreign direct investment (v15)	0.21	0.23	0.47*	0.04	0.39*	0.12	1.00		
Distance from the administrative center of the region to the nearest million-plus city (v16)	−0.07	−0.42*	0.28*	0.04	0.11	−0.10	−0.11	1.00	
Square distance from the administrative center of the region to the nearest million-plus city (v17)	0.03	−0.38*	0.31*	0.09	0.17	−0.05	−0.05	0.97*	1.00

Note: See Table 2 for full variable definitions. The correlation coefficients are based on 78 observations; GRP, gross regional product; * indicates the significance level at 5% and below.
Source: Author calculations.

DOI: 10.1111/rsp3.12668

Resumen. El objetivo de este artículo es evaluar cómo los vínculos de complementariedad interna y externa son responsables de la aparición de nuevas industrias especializadas en el contexto de las regiones rusas en el período comprendido entre 2005 y 2015, teniendo en cuenta otros factores significativos para el surgimiento de dichas industrias. Los vínculos de complementariedad interna se miden por la proporción de nuevas industrias de especialización relacionadas con industrias de especialización existentes en la misma región. Los vínculos de complementariedad externa se miden por la proporción de nuevas industrias especializadas relacionadas con industrias especializadas existentes en regiones vecinas. Los datos muestran que los vínculos de complementariedad interna y externa tienen un fuerte efecto positivo en la aparición de nuevas industrias especializadas, siendo mayor el efecto de la externa. El número de industrias especializadas, el número de empleados, el potencial de mercado, el gasto interior bruto en investigación y desarrollo (I+D), las solicitudes de patentes nacionales, la población activa con estudios superiores y la distancia entre el centro administrativo de una región y la ciudad con más de un millón de habitantes más cercana también influyen en la aparición de nuevas industrias especializadas, mientras que el producto regional bruto (PRB) por trabajador, la proporción de empresas que utilizan Internet y las zonas económicas especiales no influyen. Además, se evaluaron los modelos que predicen el inicio del proceso de diversificación en una región. Por último, se sugieren algunas medidas políticas para la creación de nuevas industrias especializadas en Rusia.

抄録: 本稿では、2005~2015年の期間におけるロシアの地域という文脈において、新しい専門産業の出現に重要な他の要因を考慮に入れて、内部および外部の相補的なリンケージがどのようにそのような産業の出現の原因となるかを検証する。内部の相補的なリンケージは、同じ地域の既存の専門産業に関連する新しい専門産業のシェアによって測定される。外部の相補的なリンケージは、近隣地域における既存の専門産業に関連する新規専門産業のシェアによって測定される。我々のデータは、内部と外部の相補的なリンケージが新しい専門産業の出現に大きなプラスの影響を与え、外部のリンケージがより大きな影響を与えることを示している。専門産業の数、従業員数、市場の潜在性、国内の研究開発(R&D)への総支出額、国内の特許出願件数、高等教育を受けた労働人口、地域の行政中心地から最も近い100万人以上の都市までの距離は、新しい専門産業の出現に影響するが、労働者1人当たりの地域総生産、インターネットを利用する企業の割合、経済特区の影響はない。さらに、地域における多様化プロセスの開始を予測するモデルを評価した。最後に、ロシアのための新しい専門産業の創造のためのいくつかの政策措置を提案する。