Demidova, Olga; Kolyagina, Alena; Pastore, Francesco

Working Paper

Marshallian vs Jacobs effects: which one is stronger? Evidence for Russia unemployment dynamics

GLO Discussion Paper, No. 302

Provided in Cooperation with:
Global Labor Organization (GLO)

Marshallian vs Jacobs effects: which one is stronger? Evidence for Russia unemployment dynamics

Demidova Olga, demidova@hse.ru*

Kolyagina Alena, 4eely94@mail.ru*

*Pastore Francesco**, francesco.pastore@unicampania.it*

* National Research University Higher School of Economics

** University of Campania "Luigi Vanvitelli", Italy

Abstract

This paper is devoted to the study of diversification and specialization influence on one of the main indicators of Russian labour market, the unemployment growth. The purpose of the work is to find out which effects dominate in the Russian regions, Marshallian or Jacobs, and whether this predominance is stable for different time intervals. The following hypotheses were empirically tested: 1) the dependence of the unemployment rate on the degree of concentration or diversification is non-monotonic due to possible overlapping effects of urbanization and localization; 2) the influence of the degree of concentration or diversification on the level of unemployment depends on the time period. To test these hypotheses nonparametric additive models with spatial effects were used. Both hypotheses found empirical confirmation. It was shown that in Russia, depending on the period, various effects dominated: in 2008-2010, and 2013-2016 Marshallian effects predominated, while in 2010-2013, Jacobs effects dominated.

Key words: concentration, diversification, unemployment, spatial effects, nonparametric models

JEL codes: C14, C21, J64

1. Introduction

Better knowledge of the differences between Russian regions allows the state to pursue a more structured national and regional policy in order to avoid negative social and economic consequences from the concentration of regions with high unemployment (Elhorst, 2003). One of the most important factors of inequality is the current concentration of economic activities in regions which have a number of competitive advantages. Possible consequences of this high concentration in a region and its impact on unemployment are interesting because of the existence of two effects of opposite sign. The Jacobs' theory (Jacobs, 1969) posits that due to the higher diversification level urban territories better absorb unemployment shocks: in fact, it's easier to find job in another sector of the economy in case of job loss, which leads to a lower unemployment rate. Marshall's theory, by contrast, suggests that regions with a high level of specialization have better economic indicators and have a lower unemployment rate due to agglomeration economies (Marshall, 1993). In other words, local agglomeration of firms in one industry creates a labour market with a limited set of skills required for this particular industry. Labour resources contribute to the growth of productivity and the reduction of differences in wages during the process of transition from one employer to another. These effects can overlap, especially in heterogeneous regions, and the main objective of the study was to empirically confirm these effects; to find out, which effect dominates; and whether this predominance is constant for different time intervals. Additionally, we want to test the applicability of models with a non-parametric component that work well for European data for modeling labour market indicators (particularly, unemployment growth) also in the case of Russian regions, and justify their advantage over simple linear models.

One should understand the agglomeration effect as the economic benefit deriving from the concentration of firms in a certain territory. Within the borders of agglomeration it becomes possible to save costs for the interacting companies due to close cooperation if certain regions attract manufacturing factors (technologies, labour resources and investments). The agglomeration effect contributes to the emergence of competitive clusters, which, in turn, is an incentive for concentration in a certain territory (Rastvortseva, 2012).

In 1920 Alfred Marshall was the first to notice the existing inclination of industries to the territorial agglomeration, which contributes to the growth of profitability and economies of scale. According to Marshall, workers periodically change their place of work (among those that use this particular kind of labour), which makes it possible to increase productivity and reduce differences in wages. As a result of mobility, workers are able to borrow knowledge and skills from each other in industrial clusters in a short period of time, and enterprises have the opportunity to recruit trained employees with ready to use knowledge and skills, which reduces the cost of training staff inside the company.

On the contrary, Jane Jacobs believed that as the diversification of industries increases, the number of job opportunities for the population increases too, which leads to a reduction in the regional unemployment rate (Jacobs, 1969). The various interrelationships between large diversified cities allow the creation and implementation of innovations, which contributes to increased productivity and economic growth of each of the enterprises in the given territory. These effects were named after the author Jacobs effects.

The disproportions in the spatial development of regions can be explained with the help of the above-mentioned theories of spatial distribution. There may be agglomeration effects from localization (under the Marshallian externalities), contributing to a reduction in production costs due to economies of scale, but the existence of centrifugal processes is also possible due to excessive infrastructure congestion, environmental problems, high population density, increased transportation costs. The total agglomeration effect, which determines the degree of concentration of production in the industry in any limited territory, is of particular interest.

The case of Russia is particularly interesting for a number of reasons. First, the vast territories of the country provide evidence of very different and varied experiences of both agglomeration and diversification. This makes Russia a unique testing ground for the Jacob versus the Marshallian effects. In addition, the historical stratification of industry localization makes several regions of the country traditionally strongly specialized in specific types of industries as a consequence of the past forced industrialization. Partly, agglomeration economies are also linked, at least initially, to the localization of natural resources, especially gas and oil, and the relative mining industry. On the other hand, the “disorganization” of central planning (Blanchard and Kremer, 1997) has changed over the last three decades the past specialization pattern of several regions of the country, breaking down old linkages between industries and, therefore, generating a higher degree of diversification of productions especially in urban areas and new product specializations overlapping with the old ones in other less urbanized areas (for an analysis of the impact of industry diversification on the quality evolution of jobs, see Gimpelson and Kapeliushnikov, 2016). Understanding what is the impact of the two effects on unemployment growth is important for policy makers interested in shaping future decisions regarding investment localization and also interested in forecasting the impact of possible economic crisis of specific sectors on employment outcomes, considering also the fragility over time of the Russian model of labor market, with high wage flexibility and rigid employment rates (Gimpelson and Kapeliushnikov, 2016; Voskoboynikov, 2017). Is this model bound to persist? What is the role of agglomeration factors in shaping it? In case this model becomes not feasible anymore what would be the employment consequence of this change with the occurrence of structural change? This paper aims to address directly or indirectly these types of questions.

This study innovates on previous research under several respects. First, we use Russian regional data over a relatively long period of time (ten years: 2007-2016), which allows us to emphasize the possibility to test for differences in the effects from one period to the next. In particular, thanks to our data, we are able to test whether there is a different dependence relationship in "crisis periods" and in more favorable periods. In fact, we find different effects in periods of ups and downs. Moreover, indices of regional diversification and concentration of production were calculated in two ways using firms’ level data: on the basis of the revenues of companies and on the basis of gross value added by economic activity. Furthermore, we test the robustness of our findings by employing a variety of indices of industrial diversity, including the Vorobyov and the Ellison-Glazer index. Fourth, we used flexible semiparametric dependence for each variable and ANOVA test for the choice between linear and nonlinear functional form.

The structure of the paper is as follows. In the next section, we provided a brief review of the papers highlighting the impact of Jacobs and Marshallians effects on unemployment in different countries. The third section presents our data source, the choice of the explanatory variables and the main hypothesis to be tested. In the fourth section we describe the methodology of econometric modeling and present the results of the estimation and their interpretation. The last section contains some concluding remarks and policy implications.

2. Literature review

Simon and Nardinelli (Simon, 1988; Simon, Nardinelli 1992), Elhorst (Elhorst, 2003), Ferragina and Pastore (Ferragina, Pastore, 2008) empirically came up to a very important conclusion: in more diversified regions there are more job opportunities and, hence, lower unemployment rates, since such regions are able to reduce the negative consequences of labour market shocks through a process of labour reallocation between different sectors. In other words, the more diversified is a region, the less arming are sectoral shocks that affect one or a small number of industries. Quite a large strande of literature provide empirical evidence that confirms the presence of Jacobs effects in a number of countries. However, the authors showed that there were such periods (for example, the beginning of the Great Depression), when unemployment was higher in more diversified regions, which could be explained by the difficulty in distinguishing real shocks from nominal ones for employers (the model of imperfect information using the theory of rational expectations). Real shocks include changes in consumer savings and expenditures, export demand, production functions, terms of trade, etc. A nominal shock is a shock in a supply or a change in the demand for money.

While studying the determinants of unemployment, David Lilien found confirmation of the positive correlation in time between the aggregate unemployment rate and intersectoral variance in the growth of employment, and also created an index for measuring changes in industries. The index proposed by Lilien reflects the degree of labour demand’s dependence on sectoral shifts in production. This index takes a value of 0, if no structural changes occurred during the period. The higher the value of the index, the faster the rate of structural change and more displacements in the labour market between sectors take place (Lilien, 1982). The main criticism of Lilien's index is the fact that it is unable to distinguish sectoral shifts from aggregate shocks in the labour market.

Samson was the first one to confirm Lilian's findings on Canadian data (Samson, 1985). Newell and Pastore also came to the same conclusions for the unemployment rate in Poland: high unemployment is a consequence of a mismatch between the employer's requirements and the worker's capabilities, and the low unemployment rate correlates with greater stability (permanence) in the workplace. This is due to the fact that the main reason for the differences in regional unemployment is industry inconsistency (Newell, Pastore, 2006).

Krajnya`k and Sommer also found confirmation of the Lilien index’s significance, describing a strong correlation between this industry-specific volatility index and the unemployment rate in the Czech Republic in 1998-1999 at the time of restructuring (Krajnya`k, Sommer, 2004).

Robson calculated the Lilien index for the UK macroregions for the time period 1975-2001 and confirmed its positive correlation with unemployment rate (Robson, 2009).

Lehmann and Walsch proposed a possible explanation for the fact that sectoral shifts contribute to a higher level of unemployment: in cases where the human capital can be exchanged, workers do not object to restructuring, which in turn enhances unemployment, but provides a rather rapid recovery and further employment increase (Lehmann, Walsh, 1999).

Simon and Nardinelli confirmed the hypothesis of portfolio theory in the US labour market. They proved that with the growth of sectoral diversification, the influence of sectoral shifts on the production structure is reduced, but the probability of laid-off employees to find work in another industry due to the existence of Jacobs effects is higher. It should be noted that as a measure of diversification, the authors used the Herfindahl-Hirschman index (Simon, Nardinelli, 1992).

Mussida and Pastore found out that sectoral changes lead to the loss of workers' jobs and increase in unemployment level, while the existence of more specialized regions, according to the Marshall’s theory, partially neutralizes the negative consequences of specialization, expressed in greater exposure to external shocks (Mussida, Pastore, 2015a).

The research on the Italian labour market conducted by Mameli et al, as well as Paci and Usai, confirmed the negative impact of specialization externalities and the positive effect of diversification on regional employment growth (Mameli et al. 2008; Paci, Usai, 2008).

Forni and Paba found out that both externalities from specialization and urbanization positively influence on the dynamics of employment (Forni, Paba, 2002).

Beaudry and Schiffauerova (2009) investigated which effects predominate in real life. They investigated 67 studies on this topic and showed that, depending on the methodology and period of analysis, one of the effects predominates. The positive influence of both effects on unemployment level was confirmed in almost the same number of studies.

Maslikhina showed that since the early 1990s in the Russian Federation there has been a gradual process of the region’s divergence (or increasing differences). These regional differences are manifested in the economic and social development of the regions, namely, in their economic growth, unemployment or employment level, migration growth or loss, the standard of citizens’ living (Maslikhina, 2013).

Vorobyov studied the influence of spatial concentration on the productivity of firms over the period 2001-2004. A new methodology for calculating the diversification index was developed, which takes into account both the inequality in the sectoral structure (which classical Herfindahl-Hirschman index and the Gini index reflect) and the diversity of the firms’ industries in a particular region. The authors concluded that positive agglomeration effects dominate up to a certain level of concentration in the region and then these economies are declining. In addition, most organizations are concentrated either at the localization level close to zero, or at a level higher than the optimal value. The state, in turn, can create a positive business environment, develop infrastructure with neighboring regions, promote the development of firms and pursue policies to create organizations in a particular industry (Vorobyov, 2014).

The authors of all the above-mentioned studies on Russia used linear dependencies. A number of foreign authors, including Viladecans-Marsal, use nonlinear dependence and include a quadratic dependence between the unemployment rate and the spatial concentration of enterprises in their models. Due to the empirical analysis of the unemployment rate's dependence on the region's diversification, the author found out that in 1950, 1960 and 1970, diversification effects prevailed. However, at the beginning of the Great Depression in the United States, the unemployment rate in the more diversified areas was higher (Viladecans-Marsal, 2004). The quadratic parametrization is only one of the possible nonlinear models, and nonlinear dependences are better captured in the semiparametric model.

The main goal of this paper is to identify with the help of a nonparametric model which effects predominate (Jacobs or Marshallians ones) and to determine whether their effect on unemployment is constant for different time periods. Our study is based on the article by Roberto Basile et al. (Basile et al., 2012). The authors studied the effects of sectoral shifts and specialization features on regional unemployment in different Italian regions in 2004-2008. The relevance of their paper was due to the fact that Italy is a country with a high level of spatial heterogeneity of local labour markets, and there are significant differences in productivity between the North and South. In addition, the authors wanted to find out in which areas agglomeration effects dominate: in industrial areas with a clear specialization or in diversified areas, and also to study the consequences of sectoral shifts and industry specialization for the regional unemployment rate.

In their study, the authors used a semi-parametric spatial autoregressive model to take into account possible nonlinear effects of explanatory variables and spatial effects. The dependent variable in that study is the average change in the unemployment rate. As a measure of sectoral shifts, the authors use the Lilien index, and as a measure of specialization, they use the logarithm of the Gini index. The authors came to the following conclusions. First, they identify two clusters of regions: with high unemployment in the South and low unemployment in the North of Italy. In addition, local labour market indicators are characterized by significant differences in space (heterogeneity). Moreover, sectoral shifts and the degree of specialization have a negative impact on the dynamics of unemployment (its growth rate is higher). Neighboring regions demonstrate a higher degree of "infection" with unemployment from each other. Some groups of regions are still less efficient than non-specialized areas (unemployment rate is higher). Strongly diversified regions are characterized by a more favorable dynamics of unemployment (its rate of growth is lower). In areas with low specialization, intersectoral mobility helps to absorb shocks in the labour market, which adversely affects unemployment (Jacobs effects are confirmed). However, in areas with a relatively high level of specialization, the importance of Marshallian externalities is increasing, so the overall effect of specialization on unemployment growth is not statistically significant.

3. The main hypotheses and data for their verification

3.1 Main hypothesis

In this paper we analyze the data for 80 Russian regions over 10 years (2007-2016) provided by the Russian statistical agency Rosstat (www.gks.ru). Sevastopol, Kaliningrad Region, the Republic of Crimea and the Chechen Republic were excluded from the sample due to insufficient data, and Moscow and the Moscow Region were merged due to the change in the Moscow border in 2012.

As it was mentioned, Viladecans-Marsal (2004) noticed the nonlinear dependence between degree of concentration and the level of region's unemployment. Simon and Nardinelli (1992) also paid attention to the non-linear influence of diversification: the negative impact on the dynamics of unemployment is mitigated when a certain level of specialization of the region is reached or it completely changes the sign under the existence of the Marshallian externalities. Both Marshallian and Jacobs effects may exist at the same time and influence unemployment level in the opposite directions. These effects may overlap.

Basile et al. (Basile et al., 2012) found evidence of nonmonotonic dependence for Italy. At low specialization values, Jacobs effects dominate due to intersectoral mobility, but in regions with a higher level of specialization, the importance of the Marshallian externalities increases. Thus, in highly concentrated regions, the overall effect of spatial specialization on the unemployment growth is not statistically significant. Russia, as well as Italy, is a very heterogeneous country, so it makes sense to check the validity of the following conclusions for Russia as well. So the first hypothesis was formulated.

Hypothesis 1: The dependence of the unemployment rate on the degree of concentration or diversification is non-monotonic due to the possible overlapping effects of urbanization and localization.

It is assumed that during periods of economic growth regions with a high degree of diversification have more favorable indicators of the labour market (unemployment level) due to the existence of Jacobs effects, as they spread among different industries in one region and labour mobility contributes to a reduction in unemployment.

On the contrary, in the crisis periods localization effects prevail due to the declining demand for products. In addition, the number of firms on the market is decreasing in the crisis due to the closure of small uncompetitive companies, which leads to the process of firms' comprehension of the need for mutual cooperation in order to minimize costs and to use joint innovations. Having studied the transition period in the Russian and Chinese economies (1990's), Galbraith et al. came to the conclusion that the industries with the maximum level of concentration remained in a winning position and were less affected by the crisis, especially, had lower unemployment rate (Galbraith et al., 2004).

Simon and Nardinelli (Simon, 1988, Simon and Nardinelli, 1992), Elhorst (Elhorst, 2003), Ferragina and Pastore (Ferragina, Pastore, 2008) also confirmed the effects of urbanization and portfolio hypothesis. They concluded that with growth of diversification level in the region, employment opportunities increase due to shifts between sectors and lower levels of unemployment are observed. However, the authors showed that there were such crisis periods, when in more diversified regions the unemployment rate was higher.

Based on these previous findings, the study of the unemployment rate's dependence on concentration or diversification level in the Russian regions at various time intervals was of particular interest. Thus, the second hypothesis was formulated.

Hypothesis 2: The direction of influence of the degree of concentration or diversification on the unemployment level depends on the chosen time interval.

The following periods were considered in this paper: 2007-2016 (general period), 2007-2008 (the period before economic crisis), 2008-2010 (crisis period), 2010-2013 (recovery period) and 2013-2016 (slowdown in economic growth).

3.2 An empirical study of indexes of spatial concentration and diversification

Unemployment is the main indicator characterizing the labour market, therefore, the logarithm of unemployment growth, which was used by Roberto Basile et al., was chosen as the dependent variable (Basile et al., 2012). The log difference of unemployment rates is an approximation of the average percentage increase in unemployment over the period [(t-n) – t] in region i and is calculated by the following formula (formula 1):

$$Y_i = \frac{\ln U_{it} - \ln U_{i(t-n)}}{n} \qquad (1)$$

Following the conclusions of Neumann and Topel (Neumann, Topel, 1991), Chiarini and Piselli (Chiarini, Piselli, 2000), and Robson (Robson, 2009) on the need to include an index of diversification (or concentration) in the econometric model as a measure of the diversity of industries, such variables as the normalized and modified diversification index and the modified Ellison-Glaser index reflecting the concentration level were added in the model.

In order to calculate the concentration and diversification indices based on the firms' revenues, data on Russian companies for the period 2007-2016 in various Russian regions were collected. Firms revenues were obtained using a database Ruslana, namely, Bureau Van Dijk. In total there was information about 12116 companies, 24 "Processing industries" (code C) in accordance with the OKVED 2 classification adopted. Based on the work of Mikhailova (Mikhailova, 2017), we decided to consider the classification of manufacturing industries since the extraction of minerals and their primary processing are not of special interest because of the lack of perfect mobility in these industries due to the existence of a territorial reference to the location of mineral deposits. We also collected the data at the firm level because this makes it possible to estimate the agglomeration effects as accurately as possible because of the consideration of individual effects for firms. In addition, these indices were calculated not only on the basis of firms revenues, but also with the help of gross value added (GVA) by types of economic activity, listed on the Rosstat website (www.gks.ru).

Diversification of the region can be measured in two different directions: inequality and diversity (variety). Inequality is understood as the degree of uniformity of the firm's distribution in the region, and variety reflects the number of different industries which exist in the region. The most frequent indices used in the literature (the Herfindahl-Hirschman index and the Gini index) measure inequality, but do not take into account diversity. In order to take into account the variety in measuring inequality, Pavel Vorobyov (Vorobyov, 2014) proposed to measure Jacobs externalities using a normalized and modified diversification index (formula 2):

$$ihh_t^i = \frac{\sum_{j=1}^{S}\left[\frac{pq_t^{ji}}{pq_t^i}\right]^{\frac{1}{s}} - 1}{\left(S^{1-\frac{1}{S}}\right) - 1}, \quad ihh_t^i \in [0;1] \tag{2}$$

where i- number of a region; S – number of industries in the economy; pq_t^{ji} - GVA (revenue) in industry j in region i; pq_t^i - GVA (revenue) in all industries in region i. This index can take values from 0 to 1.

$ihh_t^i = 1$ – equal distribution of firms' turnover between industries (diversification);

$ihh_t^i = 0$ – uneven distribution of firms' turnover in industries (lack of diversification).

The second index was borrowed from the article by Vernon Henderson (Henderson, 2003). Ellison-Glazer index is the sum by regions of the square deviation of the share of each region in the national revenue in the industry i from its share in the national revenue (formula 3):

$$ieg_j(t) = \sum_i \left(\frac{E_{ij}(t)}{E_j(t)} - \frac{E_i(t)}{E_t}\right)^2, \quad ieg_j(t) \in [0,2] \tag{3}$$

where $E_{ij}(t)$ - GVA (revenue) in industry i in region j; $E_j(t) = \sum_i E_{ij}(t)$ - GVA (revenue) in region j; $E_i(t)$ - GVA (revenue) in industry i of the whole country, $E_t = \sum_i E_i(t)$. This index takes values from 0 to 2.

$ieg_j(t) = 2$ – specialization of the region on one industry is observed;

$ieg_j(t) = 0$ – the region does not specialize in one industry.

In this paper, four variables were used that reflect the concentration or diversification: ihhva (diversification index, calculated on the GVA), ihhmn (diversification index, calculated on the revenue), iegva (Ellison-Glaeser index, calculated on the GVA), iegmn (Ellison-Glaeser index, calculated on the revenue). In Table 1 Descriptive statistics are presented: minimum, maximum and average values of each index for the first and last studied period.

Table 1. Minimum, maximum and average values of concentration and diversification indices, 2007-2016

Index	Minimum		Maximum		Average value	
	2007	2016	2007	2016	2007	2016
Diversification (GVA)	0,772	0,797	0,977	0,974	0,88	0,908
Diversification (revenue)	0,084	0,082	0,978	0,973	0,715	0,705
Concentration (GVA)	0,007	0,009	0,506	0,402	0,057	0,052
Concentration (revenue)	0,035	0,02	0,834	0,906	0,218	0,229

Indices calculated on revenue indicate an increase in the concentration of manufacturing industries during 2007-2016, as the average value of the concentration index increased, and the average value of the diversification index, on the contrary, decreased in 2016 compared to 2007. Indices calculated on the GVA by economic activity, on the contrary, indicate an increase in diversification and a decrease in concentration. In addition, in 2016, there is a decrease in the spread between the minimum and maximum values for indices calculated by GVA.

However, a spatial index reflecting the concentration or diversification in the region is only one of the possible variables that can affect the unemployment rate.

3.3 Variables

Based on the previous works, the following variables were used to test Hypotheses 1 and 2: GDP (gross regional product) per capita, calculated in the base prices of 2000, share of urban population, share of population with higher education, coefficient of migration increase per 10000 people, share of people below working age (below 16 years), share of people above working age (55 years for women and 60 years for men), population density (number of persons per square kilometer), Lilien index, initial unemployment level and growth of weighted unemployment in neighboring regions (spatial lag of the dependent variable).

Lilien index – index of variation in the growth of employment in specific industries, which measures sectoral shifts by economic activity. Lilien index is calculated by the following formula (formula 4):

$$lil_i = \left(\sum_{s=1}^{S} \left[\frac{x_{si}}{x_i} \right] \cdot (\Delta \ln x_{si} - \Delta \ln x_i)^2 \right)^{1/2} \quad (4)$$

where x_{si} – regional employment in industry s, x_i – total regional employment, Δ - first difference operator. High values of this index lead to an increase in unemployment growth rates, especially for economically "weak" regions. Lehmann and Walsh proposed a possible explanation: in the case when the human capital can be exchanged, workers do not object to restructuring, which in turn increases unemployment, but provides a fairly quick way out of it (Lehmann, Walsh, 1999). High unemployment arises due to the mismatch of the employer's requirements and the opportunities of the employee, and the low unemployment rate correlates with greater stability in the workplace. The positive impact of the Lilien index on unemployment growth was confirmed by Samson (Samson, 1985), Krajnya`k and Sommer (Krajnya`k, Sommer, 2004), Newell and Pastore (Newell, Pastore, 2006) and Robson (Robson, 2009).

Abraham and Katz (Abraham, Katz, 1986) came to the conclusion that it is necessary to separate the sectoral shifts and general market shocks, and noticed that the Lilien index truly describes sectoral shifts only if a measure of spatial diversity (concentration or diversification index) is included among the regressors (Neuman, Topel 1991).

In a number of empirical works, it was proved that GRP negatively affects the unemployment rate, that is, Okun's law works. However, Elhorst showed that this dependence of unemployment on GRP will not always be observed (Elhorst, 2003). Thus, the relationship between the unemployment rate and the GRP can be nonlinear, it is difficult to predict its exact parametric form, so it is preferable to use a nonparametric form of the dependence.

It is also difficult to predict the parametric form of the relationship between share of urban population and the unemployment rate. On the one hand, unemployment level should increase with the rise in the share of urban population due to higher competition in the labour market, but with the growth of the already high values of the share of urban population, unemployment can decline as there are a lot of jobs in regions with a large number of urban population and job search takes less time due to developed information mechanisms and increased density (Molho, 1995). Due to the ambiguous impact of this variable on the growth of the unemployment rate, we expect non-parametric dependence.

An increase in the share of people with higher education may have a two-way effect on the dynamics of unemployment. On the one hand, in regions with a low share of population with high education, educated people find it difficult to find a job due to a lack of supply, which increases unemployment. But on the other hand, for regions with a high share of the population with higher education, its further growth stimulates a reduction in unemployment since in such regions the equilibrium state in the market is set faster (Aragon et al., 2003). Thus, the nonlinear dependence of the given variable on unemployment growth is expected.

To avoid the problem of endogeneity, the lag of the coefficient of migration increase per 10,000 people is considered. The dependence of this variable on the unemployment rate may be nonlinear. On the one hand, the influx of migrants occurs in favorable regions with low unemployment, where it is easy to find work. But on the other hand, if there are too many such migrants, strong competition for places in the labour market may arise.

The share of people below working age (up to 16 years) should have a positive impact on the unemployment growth, because the change of the age structure of the labour market towards a younger population means an increase in the extra labour force that will appear on the market and will be in active job search process. In other words, young people who are currently studying at school will enter the labour market in a few years and will find it more difficult to find a job due to increased competition, thereby increasing unemployment. In addition, unemployment risk is significantly higher for young people, and a large share of young people increases unemployment. The positive impact of the share of people below working age on unemployment growth was previously proven by Hofler and Murphy (Hofler, Murphy, 1989), as well as Elhorst (Elhorst, 1995).

In the Russian Federation, there is an increase in economic activity among the elderly population, which is the reserve fund for the growth in employment. In recent years, the increase in job search among elderly people is mainly due to "young" pensioners. Also in Russia there is low level of self-employment (about 2%, which is significantly lower than in Europe) which stimulates people above working age to continue working (Sonina, Kolosnitsyna, 2015). In most countries around the globe, there is a trend in increasing the number of years that people work (Sinyavskaya, 2017). However, at very high levels of unemployment, pensioners are likely to no longer be actively seeking work. Thus, there may be a non-linear relationship between the share of people above working age and the increase in unemployment level. The fact that the share of people above working age on average increases the level of unemployment, although not as much as the high proportion of young people, was proved by Partridge and Rickman (Partridge, Rickman, 1995).

Population density is calculated as the number of people per square kilometer. Large and densely populated regions should have greater efficiency in the process of finding work for its residents, hence contribute to a lower unemployment rate (Elhorst, 2003). However, there is an opposite effect: population density reflects the convenience and greater attractiveness of large regions for life, which causes congestion effects, and as a result, a higher level of unemployment (Niebuhr, 2003). In different time periods these effects can overlap, so we suppose to confirm the nonlinear effect on the growth of unemployment. The non-linear effect of population density on the level of unemployment was confirmed by Basile et al. (Basile, 2012).

Based on the work of Overman and Puga, we decided to include the logarithm of the unemployment rate of the region at the beginning of the period to assess whether the processes of beta convergence of regions in terms of unemployment take place (Overman, Puga, 2002). Besides, the significance and nonlinearity of this relationship was confirmed by Basile et al. (Basile, 2012).

One of the explanatory variables is the average increase in unemployment in neighboring regions (wgrunempl), which is calculated by multiplying W (weighting matrix) on the dependent variable. In this paper we used a weighting binary matrix of dimension 80*80, which looks the following (formula 5):

$$W_{ij} = \begin{pmatrix} 0 & w_{12}^{ij} & \dots & w_{1n}^{ij} \\ w_{21}^{ij} & 0 & \dots & w_{2n}^{ij} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}^{ij} & w_{n2}^{ij} & \dots & 0 \end{pmatrix} \qquad (5)$$

The elements of the weighting matrix are defined as follows: w_{ij} = 1, if the regions have common border and $w_{ij} = 0$, if there is no common border between i and j or $i = j$. Then the elements of the weighting matrix were normalized in a row.

The effect of this variable on the unemployment growth can be multidirectional. Basile et al. proved spatial dependence in the Italian regions, since the coefficient for this variable was significant and reflected that neighboring regions showed a greater level of spatial "contamination" than regions located further apart (Basile et al., 2012). However, the impact may be the opposite: it is possible to reduce regional unemployment in response to the rise of unemployment in neighboring regions if the region attracts labour. Due to the possible existence of two opposite effects, a nonparametric dependence of the unemployment growth on weighted unemployment in neighboring regions is used in the model.

4. The model, methodology of its estimation and main results

4.1 Methodology of estimation

As noted earlier in the modeling of unemployment, it is preferable to use a more flexible nonparametric functional form of dependence. So, as a basis we took the methodology and technique of estimation from the article of Basile et al. (2012): an additive semi-parametric model is used, since the additivity property assumes that the effect of each explanatory variable in the model can be interpreted separately from other regressors, just as in linear multiple regression. In addition, this model allows to obtain a graphical representation of the relationship between the dependent variable and the explanatory variables. The classical semiparametric additive model (AM) is as follows (formula 6):

$$Y_i = \alpha_0 + \alpha_1 X_{1i}^* + \alpha_2 X_{2i}^* + \dots + f_1(X_{1i}) + f_2(X_{2i}) + \dots + \varepsilon_i \qquad (6)$$

where $X^*_{1i}, X^*_{2i}, \ldots$ are strictly parametric components, $\alpha_0, \alpha_1, \alpha_2, \ldots$ are the corresponding parameters, $f_1, f_2, \ldots$ – estimated smooth functions, ε – vector of independent identically distributed errors (iid).

The methodology proposed by Wood is used to evaluate additive models with smoothing based on splines (Wood, 2006). The selection of smoothing parameters was carried out using the cross-validation method.

Skipping spatial autocorrelation can lead to omission variable problem, incorrect estimates and conclusions. In order to control the effects of spatial interaction, the spatial lag of the dependent variable $Y_i^o = \sum_{j \neq i} w_{ij} Y_j$ was included in the model (where w_{ij} – elements of the spatial weights matrix, which reflects the interaction between regions i and j).

The final spatial autoregressive additive model used in our paper have the form (formula 7):

$$Y_i = \alpha_0 + \alpha_1 X^*_{1i} + \alpha_2 X^*_{2i} + \ldots + f_1(X_{1i}) + f_2(X_{2i}) + \ldots + f_w(Y_i^o) + \varepsilon_i \tag{7}$$

Since Y and its spatial lag Y^o are interrelated, there is the problem of endogeneity. To avoid this problem, the two-step approach proposed by Blundell and Powell is used (Blundell, Powell, 2003). This is an analog of the Durbin-Wu-Hausman algorithm in the linear case, used in the presence of endogenous regressors.

In the first step, the following auxiliary semiparametric regression is considered (formula 8):

$$\begin{aligned} Y_i^o = {} & f_1^*(X^*_{1i}) + f_2^*(X^*_{2i}) + \ldots + f_1(X_{1i}) + f_2(X_{2i}) + \ldots \\ & + h_1^*(WX^*_{1i}) + h_2^*(WX^*_{2i}) + \ldots + h_1(WX_{1i}) + h_2(WX_{2i}) + \ldots + \nu_i \end{aligned} \tag{8}$$

where explanatory variables $X^*_1, X^*_2, \ldots, X_1, X_2, \ldots$ were used as instruments for Y^o as well as their spatial lags $WX^*_1, WX^*_2, \ldots, WX_1, WX_2, \ldots$, as in the article of Basile et al. (Basile, 2012), ν_i – errors of regression.

The second step is to evaluate the additive model of the following form:

$$Y_i = f_1^*(X^*_{1i}) + f_2^*(X^*_{2i}) + \ldots + f_1(X_{1i}) + f_2(X_{2i}) + \ldots + f_w(Y_i^o) + f_\nu(\hat{\nu}_i) + \varepsilon_i \tag{9}$$

This model includes the same explanatory variables as the original model and additionally a nonparametric function that depends on the model residuals obtained in the first step.

Cubic smoothing splines were used for each function $f_i^*, f_i, \ldots$. For each explanatory we choose between linear and nonparametric dependence: the null hypothesis is that the dependence is linear, and the alternative hypothesis is that the dependence is nonparametric. In the absence of a significant difference, a linear form of the dependence was chosen.

All calculations were performed in a statistical package R and RStudio with the help of the special package MGCV, which includes an estimate of the general additive model (gam). After conducting preliminary tests on the choice of linear or semiparametric dependence and the ANOVA test for each explanatory variable, it was found out that linear dependence took place only for the variables share of people below working age (up to 16 years) and Lilien index (which characterizes the shifts in economic activities).

The results of model (6) estimation for periods 2007-2016, 2007-2008, 2008-2010, 2010-2013, and 2013-2016 are given in Appendixes 1-3.

Graphical representation of unemployment dependence is presented in Appendixes 4-7. The graphs reflect the fitted one-dimensional smooth functions (solid lines), and the confidence intervals (gray areas) at the 95% significance level. On each graph, the vertical axis represents the level of the corresponding unemployment growth rates, and on the horizontal axis - the values of the explanatory variables.

4.2 Testing of the main hypotheses

According to the results obtained, our main hypotheses were empirically confirmed.

The dependence of the dynamics of unemployment on the degree of concentration or diversification in the general case is indeed non-monotonic due to the overlap of the effects of urbanization and localization. In addition, the direction of their influence on the unemployment growth depends on the specific time interval.

Throughout the period under review, from 2007 to 2016, only the coefficient of Ellison-Glaser index, calculated on the GVA, was significant. The dependence in this period is non-linear (see Appendix 4, Fig.1): at low levels of concentration in the region, unemployment decreases with increasing concentration (thus, the localization effect predominates), but when the concentration exceeds a certain threshold value (ca 0.15), its further increase leads to a rise in unemployment (Jacobs externalities dominate).

The period 2007-2016 was quite diverse, due to the fact that either Marshallian or Jacobs effects predominated in different years, so their effects overlapped. That is why special attention was directed to the consideration of periods which reflect different economic situations in the country, and to the identification of the influence on the unemployment growth in each of them.

For the period 2007-2008, the significant impact of the diversification and concentration indexes on the growth of unemployment was not confirmed (see the Appendix 2, results of Models 5 - 8 estimation).

In the crisis period 2008-2010 the significant influence on the dependent variable was proved by both diversification indexes and concentration index calculated on the basis of GVA (see the Appendix 5, Fig.2-4). Along with the diversification growth in the crisis, the unemployment rate increases, indicating the predominance of the Marshallian effects in the crisis period. Therefore, in 2008-2010 specialization effects prevailed.

The time period 2010-2013 is considered as an "exit from the crisis" and an economic upsurge. In these years, the significance of unemployment growth’s dependence on the diversification and concentration indices, calculated on revenue, was confirmed (see Appendix 6, Fig.5-6). So, with the increase in the diversification in the region, the unemployment rate is decreasing, and as concentration increases, unemployment grows, too (Jacobs effects were confirmed).

Finally, in 2013-2016, when the economic situation in the country began to deteriorate again (see Appendix 7, Fig.7), a significant influence was confirmed for the diversification index calculated on revenue: an increase in diversification leads to an upsurge in unemployment (the Marshallian effect predominates). On the level of diversification from 0.7 to 0.9, a small increase in the index leads to a decrease in unemployment (Jacobs effect for fairly diversified regions), but an increase in the index value exceeding 0.9 rapidly increases unemployment. This is true for such regions as St. Petersburg, Yaroslavl Region, Leningrad Region, Moscow and Moscow Region, Krasnodar Territory. The dependence of the unemployment rate on the diversification index calculated by revenue is significant and non-linear.

Thus, during the period of economic recovery (2010-2013), people can find work in various industries and Jacobs effects predominate, and in the difficult crisis periods (such as 2008-2010 and 2013-2016) localization effects predominate, areas of specialization, in which it is easier to find job, survive.

Interpretations of the other results of estimation (characterizing the influence of other variables) are deliberately omitted to avoid obscuring the main research question (but available upon request).

5. Conclusions

For Russia, it is impossible to draw unambiguous conclusions regarding which externalities predominate due to the great heterogeneity of the regions, as well as the imposition of urbanization and localization effects. In addition, their impact on unemployment growth is not constant for different time periods. During the period of economic growth (such as 2010-2013), people move between sectors and can easily find work, so the urbanization effects prevail, and in the difficult periods for the country (for example, 2008-2010 and 2013-2016), the localization effects dominate: local agglomeration of firms from one industry creates a labour market with a limited set of skills that are in demand for a particular industry, and it is easier for people to find a job in industries of specialization.

Understanding the key differences between the regions of the Russian Federation will allow the state to conduct a competent structured socio-economic policy that will help to eliminate the negative social and economic consequences from the high concentration in some regions. So, in the crisis period the state should support enterprises whose specialization does not coincide with the main specialization of the region through tax benefits and special subsidies, and in the period of growth - to develop the most promising sectors in each region. In addition, special attention should be paid to youth policy aimed at lowering unemployment in certain regions.

References

1. Vorobyov P.V., Davidson N.B., Kislyak N.V., Kuznetsov P.D. (2014), Industrial diversity and concentration in the Russian cities as factors of economic efficiency, Vestnik of the UrFU , Series Economics and Management, 6, 4-18.
2. Marshall A. (1920), Principles of economics, London: Macmillan.
3. Maslikhina V.Yu. (2016), The dynamics of spatial inequality in conditions of agglomeration processes, Society: policy, economics, law, 10, 53–56.
4. Maslikhina V.Yu. (2013), Spatial heterogeneity of economic development of regional systems in Russia, Vestnik of the Volga State Technological University, Series Economics and Management, 1 (17), 5–16.
5. Mikhailova T.N. (2017), Joint localization of Russian industries, Moscow, 74.
6. Rastvortseva S.N., Kuga Ya.T. (2012), Regional specialization and geographical concentration of industries in Russia, Scientific bulletins of BelSU, 13 (132), 23/1, 37–47.
7. Sinyavskaya O., Ermolina A., Lyubushina E. (2017), Employment after retirement in Russia: the role of pensions – active longevity or forced necessity? Compilation of the XVIII April International Scientific Conference of the Higher School of Economics.
8. Sonina Yu., Kolosnitsyna M. (2015), Pensioners in the Russian labour market: trends in economic activity of people of retirement age, Demographic Review, 2, 37-53.
9. Federal Service of State Statistics, Russian Regions. Socio-economic indicators, 2006-2017 [Electronic resource]. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156.
10. Abraham K.G., Katz L.F. (1986), Cyclical unemployment: sectoral shifts or aggregate disturbances? The Journal of Political Economy, 94(3), 1, 507–522.
11. Aragon, S.R. (2003), Creating Social Presence in Online Environments, In Steven R. Aragon (eds.), Facilitating Learning in Online Environments, New Directions for Adult and Continuing Education, 100, San Francisco: Jossey-Bass.
12. Basile R., Girardi A., Mantuano M., Pastore F. (2012), Sectoral shift, diversification and regional unemployment: evidence from local labour systems, Empirica, 39, 525-544.

13. Beaudry C., Schiffauerova A. (2009), Who's right, Marshall or Jacobs? The localizationversus urbanization debate, Research Policy, 38, 318–337.

14. Blanchard, O.J. and Kremer, M. (1997) Disorganisation. Quarterly Journal of Economics, 112(4): 1091–1126.

15. Blundell R., Powell J. (2003), Endogeneity in nonparametric and semiparametric regression models. Advances in economics and econometrics. Cambridge University Press, Cambridge, 36, 312-357.

16. Chiarini B., Piselli P. (2000), Unemployment, wage pressure and sectoral shifts: permanent and temporary consequences of intersectoral shifts. Journal of Policy Modeling, 22(7), 777–799.

17. Combes P. P., Thisse T.M. (2008), The integration of regions and nations, Economic Geography, 255–275.

18. Elhorst J.P. (2003), The mystery of regional unemployment differentials: theoretical and empirical explanations, Journal of Economic Surveys, 17, 5, 709—748.

19. Ellison G., Glaeser E. (1997), Geographic concentration in U.S. manufacturing industries: a dartboard approach, Journal of Political Economy, 105 (5), 889–927.

20. Ferragina A.M., Pastore F. (2008), Mind the gap: unemployment in the New EU regions, Journal of Economic Surveys, 22(1), 73–113.

21. Forni M., Paba S. (2002), Spillovers and the growth of local industries, Journal of Industrial Economics, 50, 151-171.

22. Galbraith J.K, Krytynskaia L., Wang Q. (2004), The experience of rising inequality in Russia and China during the transition, The European Journal of Comparative Economics, Castellanza, 1, 1, 87-106.

23. Hofler R., Murphy K. (1989), Using a composed error model to estimate the frictional and excess–supply componets of unemployment, Journal of Regional Science, 29, 213–228.

24. Jacobs J. (1969), The economy of cities, New York 10022, 268.

25. Krajnya`k K., Sommer M. (2004), Czech Republic, IMF country report, August, 4/265.

26. Lehmann H., Walsh P. P. (1999), Gradual restructuring and structural unemployment in Poland: A legacy of central planning, LICOS DP, Centre for Transition Economics, Catholic University of Leuven, 78.

27. Lilien D.M. (1982), Sectoral shifts and cyclical unemployment, Journal of Political Economy, University of Chicago Press, 90(4), 777–793.

28. Mameli F., Faggian A., McCann P. (2008), Employment growth in Italian local labour systems: issues of model specification and sectoral aggregation, Spatial Economic Analysis, 3(3), 343-259.

29. Molho I. (1995), Migrant inertia, accessibility and local unemployment, Economica, 62, 245, 123-132.

30. Mussida C. and Pastore F. (2015a), Geographical labour market imbalances, AIEL Series in Labour Economics, Springer, Berlin and Heidelberg, 370.

31. Neumann G.R., Topel R.H. (1991), Employment risk, diversification, and unemployment, The Quarterly Journal of Economics, 106 (4), 1341-1365.

32. Newell A., Pastore F. (2006), Regional unemployment and industrial restructuring in Poland, Eastern European Economics, 44(3), 5–28.

33. Niebuhr A. (2003), Spatial interaction and regional unemployment in Europe, European Journal of Spatial Development, 5, 1–26.

34. Overman H., Puga D. (2002), Unemployment clusters across European countries and regions, Economic Policy, 34, 117–147.

35. Paci R., Usai S. (2008), Agglomeration economies, spatial dependence and local industry growth, Revue d'économie industrielle, 87-109.

36. Partridge M., Rickman D. (1995), Differences in state unemployment rates. The role of labour and product market structural shifts, Southern Economic Journal, 62, 89–106.

37. Porter M.E. (1990), The competitive advantage of nations, New York, NY: Free Press, 734.

38. Robson M. (2009), Structural change, specialization and regional labour market performance: evidence for the UK, Applied Economics, Taylor & Francis Journals, 41(3), 275–293.

39. Samson, L. (1985), A study of the impact of sectoral shifts on aggregate unemployment in Canada, Canadian Journal of Economics, 18, 518-530.

40. Simon C., Curtis J., Nardinelli C. (1992), Does industrial diversity always reduce unemployment? Evidence from the Great Depression and after, Economic Inquiry, 30, 384-397.

41. Simon C. (1988), Frictional unemployment and the role of industrial diversity, The Quarterly Journal of Economics, 103(4), 715–728.

42. Viladecans-Marsal E. (2004), Agglomeration economies and industrial location: city-level evidence, Journal of Economic Geography, 4, 565–582.

43. Wood S.N. (2006), Generalized additive models. An Introduction with R, Chapman & Hall/CRC, Boca Raton, 397.

Appendix 1. Results of estimations, 2007-2016

Parametric terms (beta and p-values)	Model 1		Model 2		Model 3		Model 4	
time period	2007-2016		2007-2016		2007-2016		2007-2016	
intercept	-0.227***		-0.230***		-0.235***		-0.256***	
	0.000		0.000		0.000		0.000	
lilien	0.047		0.061		0.052		0.072	

	0.322		0.199		0.248		0.117	
young	0.013***		0.013***		0.013***		0.014***	
	0.000		0.000		0.000		0.000	
Nonparametric terms								
F test and p-values		edf						
f(ihhva)	1.225	1.949						
	0.273							
f(ihhmn)			0.360	1.000				
			0.551					
f(egva)					2.665*	1.935		
					0.070			
f(egmn)							0.022	1.000
							0.883	
f(grppercap)	0.009	1.000	0.115	1.000	0.003	1.000	0.004	1.000
	0.926		0.736		0.955		0.948	
f(urbanshare)	1.876	2.574	2.094	1.785	2.008	1.712	3.679*	1.538
	0.149		0.132		0.153		0.053	
f(higheduc)	2.960*	2.043	2.861*	1.995	2.489*	1.827	2.412*	1.750
	0.051		0.053		0.084		0.095	
f(migr)	0.924	2.838	0.008	1.000	0.024	1.000	0.200	1.000
	0.524		0.931		0.877		0.656	
f(old)	12.829***	2.514	14.143***	2.606	13.299***	2.648	15.157***	2.585
	0.000		0.000		0.000		0.000	
f(density)	12.388***	2.027	11.219***	2.019	14.519***	2.035	13.551***	2.008
	0.000		0.000		0.000		0.000	
f(unempl)	28.557***	3.650	28.980***	3.618	34.057***	3.836	32.961***	3.851
	0.000		0.000		0.000		0.000	
f(WY)	2.492*	1.117	2.084*	1.394	3.870**	1.879	3.683**	1.583
	0.094		0.095		0.026		0.024	
R2	0.724		0.726		0.744		0.743	
GVC score	0.000		0.000		0.000		0.000	

Appendix 2. Results of estimations, 2007-2008, 2008-2010

Parametric terms (beta and p-values)	Model 5		Model 6		Model 7		Model 8		Model 9		Model 10		Model 11		Model 12	
time period	2007-2008		2007-2008		2007-2008		2007-2008		2008-2010		2008-2010		2008-2010		2008-2010	
intercept	0.011		0.117		-0.129		-0.024		-0.280		-0.252		-0.184		-0.323	
	0.972		0.684		0.666		0.934		0.090		0.137		0.249		0.068	
lilien	2.162		1.956		1.924		1.851		0.231		0.319		0.261		0.404	
	0.036		0.030		0.035		0.046		0.425		0.305		0.385		0.265	
young	0.000		-0.006		0.009		0.003		0.019		0.017		0.013		0.021	
	0.999		0.746		0.626		0.873		0.056		0.096		0.173		0.048	
Nonparametric terms																
F test and p-values		edf		edf		edf		edf		edf		edf		edf		edf
f(ihhva)	1.437	2.456							4.861**	1.000						
	0.200								0.032							
f(ihhmn)			2.620	1.000							5.797**	1.000				
			0.111								0.019					
f(egva)					1.366	1.378							9.244***	1.000		
					0.183								0.003			
f(egmn)							0.014	1.000							1.727	2.718
							0.905								0.167	
f(grppercap)	2.090	2.296	3.764**	2.999	3.973**	2.593	3.230**	2.794	2.720	1.000	0.009	1.001	0.115	1.000	0.003	1.000
	0.113		0.014		0.012		0.025		0.105		0.926		0.736		0.955	
f(urbanshare)	0.400	1.000	0.173	1.000	0.633	1.000	0.107	1.000	0.719	1.157	1.877	2.386	2.866	1.785	2.009	1.712
	0.530		0.679		0.429		0.745		0.518		0.149		0.132		0.153	
f(higheduc)	2.827**	2.421	3.770**	2.253	2.856*	2.114	3.094**	2.447	3.054**	2.681	2.960*	2.043	2.861*	1.995	2.490*	1.827
	0.045		0.019		0.054		0.035		0.031		0.051		0.053		0.084	
f(migr)	0.739	1.000	1.095	1.000	0.621	1.000	0.832	1.000	3.287**	3.808	0.924	2.838	0.008	1.000	0.024	1.000
	0.394		0.300		0.434		0.366		0.016		0.524		0.931		0.877	
f(old)	0.365	1.000	0.763	1.000	0.895	1.766	0.174	1.000	0.305	1.421	12.830***	2.514	14.144***	2.606	13.300***	2.648

	0.548		0.386		0.387		0.678		0.601		0.000		0.000		0.000	
f(density)	8.381***	2.015	12.152***	2.022	11.538***	2.017	11.148***	2.017	2.829**	3.929	12.389***	2.027	11.220***	2.019	14.520***	2.035
	0.001		0.000		0.000		0.000		0.031		0.000		0.000		0.000	
f(unempl)	12.313***	3.306	16.498***	3.371	14.834***	3.579	17.238***	3.383	13.124***	3.429	28.558***	3.651	28.981***	3.619	34.058***	3.837
	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
f(WY)	4.345**	1.464	5.286***	2.896	3.201**	2.075	3.418**	2.360	2.588*	1.107	2.493*	1.118	2.085*	1.395	3.871**	1.880
	0.025		0.002		0.047		0.026		0.090		0.094		0.095		0.026	
R2	0.570		0.672		0.669		0.659		0.640		0.724		0.726		0.744	
GVC score	0.032		0.026		0. 025999		0.027		0.007		0.000		0.000		0.000	

Appendix 3. Results of estimations, 2010-2013, 2013-2016

Parametric terms (beta and p-values)	Model 13		Model 14		Model 15		Model 16		Model 17		Model 18		Model 19		Model 20	
time period	2010-2013		2010-2013		2010-2013		2010-2013		2013-2016		2013-2016		2013-2016		2013-2016	
intercept	-0.261		-0.279		-0.234		-0.150		-0.318		-0.314		-0.358		-0.306	
	0.012		0.005		0.015		0.139		0.000		0.000		0.000		0.001	
lilien	-0.172		-0.088		-0.124		-0.219		0.107		0.088		0.027		0.081	
	0.354		0.627		0.485		0.225		0.184		0.291		0.727		0.344	
young	0.010		0.011		0.008		0.004		0.017		0.017		0.020		0.017	
	0.080		0.052		0.121		0.496		0.000		0.000		0.000		0.001	
Nonparametric terms																
F test and p-values		edf		edf		edf		edf		edf		edf		edf		edf
f(ihhva)	0.476	1.000							1.778	1.000						
	0.493								0.188							
f(ihhmn)			4.529**	1.000							4.662***	3.928				
			0.037								0.003					
f(egva)					0.955	1.000							1.439	1.647		
					0.333								0.175			
f(egmn)							2.584*	2.870							1.600	1.000
							0.060								0.211	

f(grppercap)	0.004	1.000	2.502	2.708	3.765**	2.100	3.974**	2.594	3.231**	2.646	2.572	1.000	0.009	1.002	0.115	1.000
	0.948		0.113		0.014		0.012		0.025		0.105		0.926		0.736	
f(urbanshare)	3.680*	1.538	0.400	1.001	0.173	1.001	0.633	1.001	0.107	1.001	0.719	1.157	1.878	2.135	2.183	1.785
	0.053		0.530		0.679		0.429		0.745		0.518		0.149		0.132	
f(higheduc)	2.413*	1.750	2.150**	2.391	3.771**	2.253	2.856*	2.114	3.095**	2.448	3.055**	2.682	2.960*	2.043	2.861*	1.995
	0.095		0.045		0.019		0.054		0.035		0.031		0.051		0.053	
f(migr)	0.200	1.000	0.739	1.000	1.095	1.000	0.621	1.000	0.832	1.000	3.133**	3.185	0.924	2.838	0.008	1.000
	0.656		0.394		0.300		0.434		0.366		0.016		0.524		0.931	
f(old)	15.158***	2.585	0.365	1.000	0.763	1.000	0.895	1.766	0.174	1.000	0.305	1.422	12.831***	2.514	14.145***	2.606
	0.000		0.548		0.386		0.387		0.678		0.601		0.000		0.000	
f(density)	13.552***	2.008	8.382***	2.016	12.153***	2.023	11.539***	2.018	11.149***	2.164	2.245**	3.929	12.390***	2.027	11.221***	2.019
	0.000		0.001		0.000		0.000		0.000		0.031		0.000		0.000	
f(unempl)	32.962***	3.852	12.314***	3.307	16.499***	3.372	14.835***	3.580	17.239***	3.384	13.125***	3.429	28.559***	3.652	28.982***	3.620
	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
f(WY)	3.684**	1.584	4.346**	1.465	5.287***	2.897	3.202**	2.076	3.419**	2.816	2.104*	1.108	2.494*	1.119	2.086*	1.396
	0.024		0.025		0.002		0.047		0.026		0.090		0.094		0.095	
R2	0.743		0.570		0.672		0.669		0.659		0.640		0.724		0.726	
GVC score	0.000		0.032		0.026		0. 026		0.027		0.007		0.000		0.000	

Appendix 4. Partial effects of index of concentration, 2007-2016

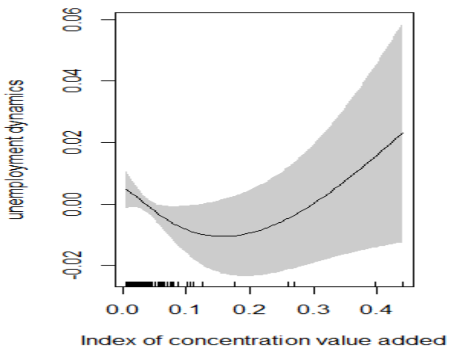

Fig.1. Index of concentration value added

Appendix 5. Partial effects of indexes of concentration and diversification, 2008-2010

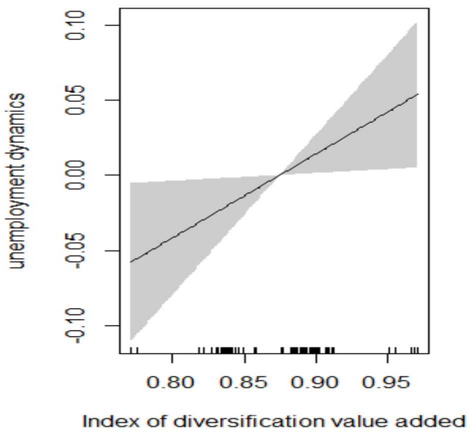

Fig.2. Index of diversification value added

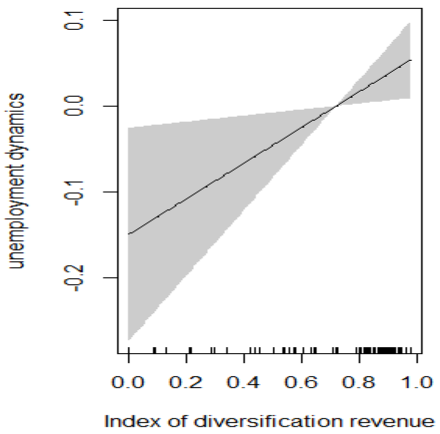

Fig.3. Index of diversification revenue

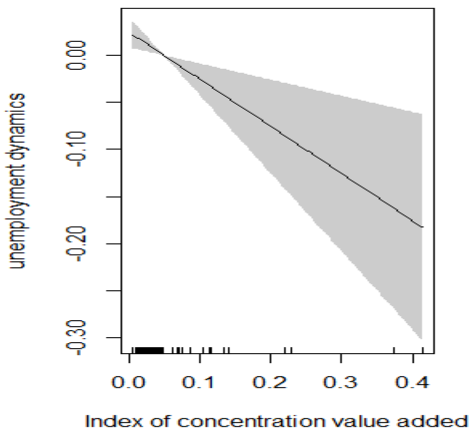

Fig.4. Index of concentration value added

Appendix 6. Partial effects of indexes of concentration and diversification, 2010-2013

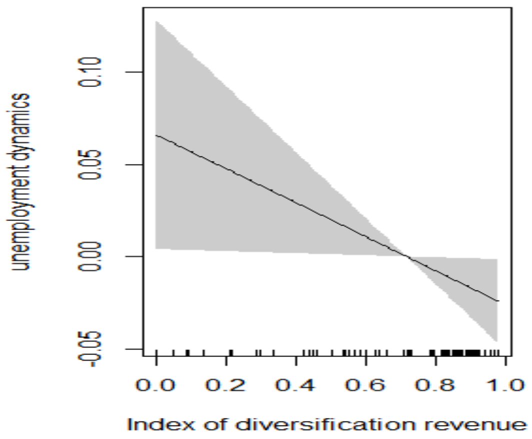

Fig.5. Index of diversification revenue

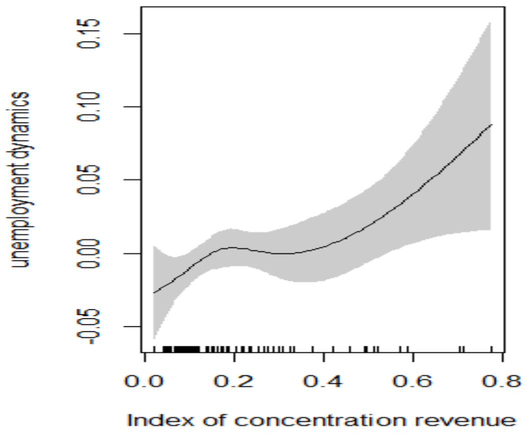

Fig.6. Index of concentration revenue

Appendix 7. Partial effects of index of diversification, 2013-2016

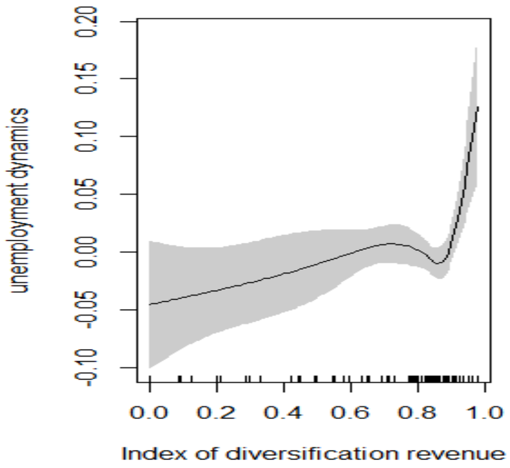

Fig.7. Index of diversification revenue