МРНТИ 06.61.33; 06.61.53
JEL Classification: C23; R10; R11; R12; R15
DOI: https://doi.org/10.52821/2789-4401-2023-1-6-21

SPATIAL CONCENTRATION AND FIRM-LEVEL PRODUCTIVITY IN KAZAKHSTAN

Z. M. Adilkhanova
NAC Analytica, Nazarbayev University, Astana, Republic of Kazakhstan

ABSTRACT

Purpose of the research. This paper studies the effect of spatial agglomeration on firms' total factor productivity in Kazakhstan using panel data from 2009 to 2017.

Methodology. We employ a two-stage estimation strategy and control for endogeneity biases by making use of the GMM approach. The firm-level data is obtained from the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan.

Originality / value of the research. This study contributes to an empirical study of spatial concentration and firm-level productivity in developing countries and provides valuable insights for policymakers to consider before implementing government programs.

Findings. The results suggest that productivity increases with clustering: a 10 % increase in the number of employees of the neighboring firms inside the same industry increases firm-level productivity by 1.36 %, while a 10 % increase in employment in other industries enhances firm performance by 1.95 %. The productivity gains are higher at the 2-digit regional level rather than at the 9-digit sub-regional level of geographical aggregation, implying that the denser geography increases firms' performance.

Keywords: agglomeration economies, total factor productivity, spatial concentration, clusters

INTRODUCTION

The spatial concentration of firms and people is one of the topics of current interest for many researchers in different countries. A denser concentration of economic activity contributes to increased firms' productivity through a number of positive externalities, such as knowledge spillovers, labor pooling, and sharing input markets in agglomerated economies.

The literature has widely emphasized that knowledge spillovers positively affect firms' performance [1; 2; 3]. Industrial clusters promote an exchange of ideas, a transmission of technologies, the diffusion of knowledge through social networks, face-to-face communications between economic agents, and the formation of pools of specialized workers. Moreover, cities facilitate the creation of innovations that are clustered spatially in industries where new knowledge is extensively important [2]. Marshall highlighted that labor market pooling can greatly enhance industry benefits from the localized and constant market of skills [4]. It creates a platform for better matching between workers and employers, as well as, increases the probability of suitable matches due to cluster-specific skills deserved by firms [5]. The advantage of sharing facilities and inputs market externalities lies in the concept of economies of scale: The larger population uses the same facilities and infrastructure for a lesser cost per person. A high concentration of producers motivates suppliers to locate nearby, thereby increasing specialized services for firms and reducing the burden on the infrastructure budget. Nowadays, many countries are actively pursuing policies to increase the concentration of industries and the density of the population. However, congestion and surge in real estate prices are negative externalities of agglomeration, which can adversely affect firms' performance and increase transportation costs. Therefore, it would be useful for policy analyses to understand what are the economic benefits of agglomeration and whether government intervention should be done in favor of industrial or spatial clusters. According to the literature, agglomeration is characterized by two concepts – localization and urbanization. Localization refers to the territorial concentration of firms of one industry at certain geographical points, the so-called industrial clusters of economic activity. Urbanization represents an increase in the density of the population in cities. This work is devoted to the study of the benefits of agglomeration in Kazakhstan, in particular, the effect of localization and urbanization economies on firm-level total factor productivity (TFP).

The analysis is conducted using firm-level data for Kazakhstan covering the 2009 to 2017 period. Our data allows us to control for the endogeneity issues by adopting a two-step estimation approach. First, we estimate firm-level TFP using the GMM method proposed by Wooldridge that resolves simultaneity and selection biases when estimating the Cobb-Douglas production function [6]. In the second stage, the firm-level productivity is regressed on agglomeration economies through a GMM specification. The agglomeration economies are decomposed into intra-industry (localization), inter-industry (urbanization) economies, and competition inside the industry. In this paper, we find that localization and urbanization economies positively affect TFP: our benchmark results after controlling for industrial heterogeneity show that a 10 % increase in the number of employees of the neighboring firms inside the same industry increases firm-level productivity by 1.36 %, while a 10 % increase in the employment in other industries enhance firm performance by 1.95 %. Competition is observed to have a negative impact on TFP by lowering productivity by 0.46 % as a response to the 10 % increase in sectoral employment.

We also explore whether more intense agglomeration is always or not beneficial by employing a nonlinear specification. As cities grow, negative externalities may dominate the positive externalities of clusterization at a certain level of agglomeration. The results show that firms start receiving productivity gains from localization if they choose to locate in a territory with more than 2 860 employees inside their industry. However, the TFP gains decrease if firms choose to locate nearby territory with more than 363 000 workers in the same industry.

The paper is organized as follows. Section 2 presents an overview of the literature, details the empirical strategy, discusses the methodology, and proceeds to the data description. Section 3 presents the results and Section 4 concludes.

MAIN PART

Literature review. The growing firm-level empirical studies on the relationship between TFP and agglomeration economies are mixed. Using Sweden data, Andersson and L¨o¨of find that manufacturing firms operating in larger regions are more productive when taking into account ownership structure, participation in international trade, and industry variations [7]. A similar positive relationship between population density and TFP growth in regions is reported by Combes et al. and Harris and Moffat for the French and British firms [8; 9]. Lopez and Su¨dekum, Martin et al., and Hashiguchi and Tanaka find positive effects of intra-industry side effects but hardly see cross-industry externalities reported in previous papers [10; 11; 12]. This is consistent with Cainelli and Ganau findings of localization economies [13]. They also report that there is a positive effect of localization economies on productivity growth for Italian firms, which increases with the distance after controlling for the characteristics of neighboring firms, and a negative effect of cross-sectoral externalities. However, in the case of Italian manufacturing firms, DiGiacinto et al. show evidence of productivity gains for firms located in both urban and industrial areas, confirming the beneficial effects of both urban and location economies on TFP [14].

Most of the studies use a two-stage estimation procedure and apply the Cobb-Douglas specification at the first stage [11; 8; 14; 9; 13]. The paper of Martin et al., which analyzes the effect of localization and urbanization economies, as well as industry diversity and intensity of competition on the TFP of French manufacturing firms, uses the control functions method proposed by Levinsohn and Petrin for the estimation of the Cobb-Douglas production function [11; 15]. This approach does not solve the endogeneity issue, therefore, as in Cainelli and Ganau, we apply a method proposed by Wooldridge to overcome the arising simultaneity issue [13; 6].

Martin et al. and Carreira and Lopes go a step further and employ a non-linear specification to test the effects of localization economies [11; 3]. The rationale behind this test is in agglomeration diseconomies such as congestion effects that may negatively affect TFP at a certain level of agglomeration, besides the positive externalities. Martin et al. find that there is a bell-shaped relationship with negative effects on TFP for small values of localization [11]. While contradicting the congestion theory, Cainelli et al. report positive effects of specialization and diversity above a certain threshold [16]. Based on these findings, we also analyze whether agglomeration economies are significant above a certain value of industrial concentration and whether the congestion effects dominate positive externalities above a certain threshold for the number of people employees in the industry.

Empirical methodology. Agglomeration is assumed to improve the total factor productivity (TFP) of firms through different channels such as localization and urbanization economies. The empirical analysis is based on the estimation of the Cobb-Douglas production function using firm-level data:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} \tag{1}$$

where subscripts *i* and *t* denote firm and year respectively;

Yit – output for firm *i* at time *t*;

Kit – the capital stock (measured by fixed assets) for firm *i* at time *t*;

Lit – labor (measured by hours worked) for firm *i* at time *t*;

Ait– TFP for firm *i* at time *t*, which is assumed to depend on firm-level component *Uit* for firm *i* at time *t*, and local milieu in terms of localization and urbanization economies:

$$A_{it} = \left(LOC_{it}^{sl}\right)^{\sigma} \left(URB_{it}^{sl}\right)^{\gamma} \left(COMP_{it}^{sl}\right)^{\mu} U_{it} \tag{2}$$

where LOC_{it}^{sl} – localization economies for firm *i*, in sector *s*, location *l*, at time *t*;

URB_{it}^{sl} – urbanization economies for firm *i*, in sector *s*, location *l*, at time *t*;

$COMP_{it}^{sl}$ – competition for firm *i*, in sector *s*, location *l*, at time *t*.

The agglomeration variables (localization and urbanization economies) and competition are constructed using the number of employees at different sectors and locations as in Martin et al. [11]. Localization economies are aimed at dealing with the externalities among the same industries. It is measured as the share of other workers working in the same industry within the same location. Specifically, we calculate the number of workers for firm *i* in sector *s*, location *l*, and year *t*:

$$LOC_{it}^{sl} = workers_{t}^{sl} - workers_{it}^{sl} + 1 \tag{3}$$

where $workers_{t}^{sl}$ – number of workers in sector *s*, location *l*, and year *t*;

$workers_{it}^{sl}$ – number of workers for firm *i* in sector *s*, location *l*, and year *t*.

Urbanization economies capture the externality of cross-fertilization of different industries in the same location and are measured as the number of workers of other sectors where firm *i* operates:

$$URB_{it}^{sl} = workers_{t}^{sl} - workers_{it}^{sl} + 1 \tag{4}$$

$COMP_{it}^{sl}$ is introduced as an additional variable to control for the local competition among firms and industries. This variable aims to test whether more intense competition enhances the productivity of the firms within the sectoral and geographical clusters. We use the inverse of the Herfindahl index of employment concentration to measure competition that a firm face inside sector s on a given location *l* at time *t*:

$$COMP_{it}^{sl} = \frac{1}{Herf_{t}^{sl}} \tag{5}$$

$$Herf_{t}^{sl} = \sum_{j \in S_{t}^{sl}} \left(\frac{workers_{jt}^{sl}}{workers_{t}^{sl}}\right)^{2} \tag{6}$$

where $Herf_{t}^{sl}$ – the Herfindahl index of employment concentration in sector *s* and location *l*, at time *t*.

Estimation issues and strategy. Unobserved heterogeneity and simultaneity bias are the main challenges in assessing agglomeration economies. The unobserved heterogeneity such as transportation infrastructure, climate, proximity to natural resources, or governmental services can increase productivity and therefore correlates with localization and urbanization variables. Whereas the simultaneity bias arises when an economic shock in a specific area or industry affects the performance of other firms. To deal with these endogeneity issues, we implement the two-step estimation approach suggested by Combes and Gobillon [17]. First, we estimate the production function and save the residuals as TFP. Then, the firm-level TFP is regressed against the agglomeration variables and competition [11; 16; 3].

In the first stage, we estimate the Cobb-Douglas production function. Several methodologies have been designed specifically to address the problem of simultaneity and selection bias in estimating capital and labor shares via regression analysis. The simultaneity arises due to the correlation between inputs (capital and labor) with unobservable productivity shocks. Firms choose inputs knowing the level of productivity, and this introduces a bias in OLS parameter estimates. There are numerous approaches to addressing this issue: instrumental variable (IV), fixed-effects approach [18; 15], control functions, and generalized methods of moments (GMM) [18; 15]. Input prices are candidates for the role of instruments in IV estimation. However, finding an appropriate instrument for capital is the main problem of this method. Regarding the fixed-effects model, which controls for unobservable heterogeneity across firms, it requires the productivity shock to be fixed over time, and a strict endogeneity of inputs conditional on firms' heterogeneity, which does not hold in theory [19]. The control functions method is a semiparametric method introduced by Olley and Pakes, Levinsohn and Petrin where investments (intermediate inputs) are introduced through a semiparametric function to control for unobservable productivity shocks [18; 15]. They develop a two-step estimation procedure to resolve the pathologies of simultaneity and selection bias present in OLS. However, Wooldridge proposed a new estimation technique using a GMM framework to modify the control functions method [6]. His approach has several advantages over the two-step approach. First, it addresses an identification problem highlighted by Ackerberg et al. who finds that the assumptions of the previous approach hold if there is some variation in the data. If not, labor and the non-parametric term suffer from collinearity, because firms choose the variable input at some point in time depending on their capital and productivity [20]. Second, it accounts for heteroskedasticity and serial correlation by obtaining robust standard errors. The model equations and descriptions are given below. We specify the Cobb-Douglas production function of the equation (1) in logarithmic form for estimating the model parameters:

$$y_{it} = \alpha_0 + l_{it}\alpha + k_{it}\beta + c_{it}\lambda + e_{it} \tag{7}$$

$$y_{it} = \eta_0 + l_{it}\alpha + k_{it}\beta + \sigma_1\left(c_{i,t-1}\lambda\right) + \cdots + \sigma_G\left(c_{i,t-1}\lambda\right)^G + \epsilon_{it} \tag{8}$$

where y_{it} is logarithm of output for firm i at time t;
l_{it} – logarithm of labor for firm i at time t;
k_{it} – logarithm of capital for firm i at time t;
m_{it} – logarithm of intermediate inputs for firm i at time t;
e_{it} and ϵ_{it} are error terms for firm i at time t;
$h(k_{it}, m_{it})$ is a function of $c_{it}\lambda = c(k_{it}, m_{it})\lambda$ containing polynomials of order three or less:

$$h(k_{it}, m_{it}) = \lambda_0 + c(k_{it}, m_{it})\lambda \tag{9}$$

After obtaining the production function elasticity coefficients, we compute TFP as residuals:

u_{it}

$$\hat{a}_{it} = y_{it} - \hat{\alpha}_0 - l_{it}\hat{\alpha} - k_{it}\hat{\beta} - \mathrm{c_{it}}\,\hat{\lambda} \tag{10}$$

where $\hat{a}_{it}$ estimated TFP for firm i at time t.

In the second stage, TFP is regressed on localization, urbanization and competition as it was indicated in equation (2). The log-linear form of the equation (2) is follows:

$$\hat{a}_{it} = \sigma loc_{it}^{sl} + \gamma urb_{it}^{sl} + \mu comp_{it}^{sl} + u_{it} \tag{11}$$

where loc_{it}^{sl} logarithm of localization economies for firm i, in sector s, location l, at time t;
urb_{it}^{sl} logarithm of urbanization economies for firm i, in sector s, location l, at time t;
:$comp_{it}^{sl}$ – logarithm of competition for firm i, in sector s, location l, at time t;
u_{it} – error term.

Considering that firms do not change location or industry, the fixed-effects approach can be used to take into account firm-level environmental unobserved characteristics. However, at the same time, the agglomeration may also affect these regional characteristics. Therefore, to mitigate problems due to simultaneity and endogeneity, we employ the system of GMM approach with one-year lagged values for all control variables. Finally, to control for possible intertemporal correlation across firms in each industry, each location and each year, we employ robust standard errors clustered by location- industry-year.

Data. The data (1-PF and 1-T forms) (*1-PF and 1-T are annual statistical forms named as "Report on financial and economic activities of the enterprise" and "Report on labor, correspondingly*) are obtained from the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan for all medium and large firms except educational and medical organizations, banks, public associations, and insurance companies [21]. These datasets are collected annually from firms with more than 50 employees from 2009 to 2014 and from firms with more than 100 employees since 2015 after changing the data collection methodology. We limit the entire sample to firms with more than 100 employees from 2009 to 2017. The data covers variables such as location, industries, and various financial indicators. The panel dataset of firms is unbalanced. It consists of about 5873 unique firms observed between 2009 and 2017, for a total of 29 490 observations.

Table 1 shows the location and industrial composition of the data. The geographical composition of the data is presented by a detailed 9-digit classification; in panel A, we use the first 2 digits to identify the main 14 oblasts and 2 republican-level cities, Almaty and Astana. It demonstrates that the largest concentration of firms is located in Almaty (20 %), then about 8.4 % in Astana and 7.8 % in East Kazakhstan region. The industrial variation in our data is provided via a 5-digit classification; we also use the first 2 digits and distinguish 17 sectors. Manufacturing (19.5 %), Construction (14.6 %), Agriculture (11.7 %), and Retail and Wholesale trade (10.5 %) are the main sectors, which account for more than half of observations. Specifically, Table 2 demonstrates within industry-time variation of firms in the dataset. On average, 3277 medium and big firms operate each year, in core industries: on average around 384 in agriculture, 134 in mining, 31 - oil and gas, 637 – manufacturing, 478 – construction, etc. Along with the balance sheet data (output, capital, wage fund, investments, etc.), detailed information about the location, industry, and size of the firm is also provided.

Summary statistics of the variables are provided in Table 3. There is a high variability of most variables due to the large values of standard deviations. The average number of workers is about 339.59 in our data, where the maximum number of workers reaches 40 864 employees. According to the dataset, there are 326 missing values of capital in the whole panel.

Table 1 – Data structure

Panel A: Location composition			Panel B: Sector composition		
Regions	Freq.	Percent.	Sector	Freq.	Percent.
Aktobe region	1332	4.5	Accommodation and Food Services	759	2.6
Almaty region	1649	5.6	Administration and Support Services	1784	6.0
Atyrau region	1450	4.9	Art Entertainment and Recreation	1266	4.3
East Kazakhstan region	2311	7.8	Construction	4305	14.6
Kyzylorda region	1025	3.5	Finance and Insurance	107	0.4
Mangystau region	1316	4.5	Information and Communication	653	2.2
North Kazakhstan region	1529	5.2	Other Services	111	0.4
Pavlodar region	1289	4.4	Professional Sci. and Tech. Services	1930	6.5
South Kazakhstan region	1900	6.4	Real Estate	592	2.0
West Kazakhstan region	1138	3.9	Transportation and Warehouse	2007	6.8
Akmola region	1607	5.4	Agriculture	3455	11.7
Almaty city	5704	19.3	Automobile	138	0.5
Astana city	2480	8.4	Manufacturing	5736	19.5
Karagandy region	2176	7.4	Mining	1206	4.1
Kostanay region	1905	6.5	Oil and Gas	277	0.9
Zhambyl region	679	2.3	Utilities	2070	7.0
			Retail and Wholesale Trade	3094	6.81
Total	29490	100	Total	29 490	100
Note – compiled by the author based on 1-PF and 1-T firm-level data from [11]					

Table 2 – Data structure within industry-time variation

Industry\ Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**Average**	**Total**
Agriculture	460	446	424	405	374	363	335	326	322	384	3455
Mining	121	130	130	142	145	147	133	129	129	134	1206
Oil and Gas	28	26	27	29	32	33	28	37	37	31	277
Manufacturing	599	627	649	658	660	669	637	623	614	637	5736
Utilities	216	228	227	239	235	239	225	232	229	230	2070
Construction	464	499	507	542	528	524	437	417	387	478	4305
Automobile	11	11	12	16	18	21	20	14	15	15	138
Retail and Wholesale Trade	278	300	309	356	377	383	369	358	364	344	3094
Transportation and Warehouse	216	220	227	229	235	234	214	221	211	223	2007
Accommodation and Food Services	67	75	79	80	88	99	87	92	92	84	759
Information and Communication	61	64	67	73	78	83	81	74	72	73	653
Finance and Insurance	11	15	13	13	12	10	12	10	11	12	107
Real Estate	68	67	73	70	71	70	59	61	53	66	592
Professional Sci.and Tech. Services	191	194	234	209	218	246	212	209	217	214	1930
Administration and Support Services	133	154	179	200	225	241	227	215	210	198	1784
Art Entertainment and Recreation	55	65	132	157	167	171	169	181	169	141	1266
Other Services	5	5	6	9	11	17	18	16	24	12	111
Total	2984	3126	3295	3427	3474	3550	3263	3215	3156	3277	29490
Note – compiled by the author based on 1-PF and 1-T firm-level data from [11]											

According to the localization economies, the average number of workers that operate in the same sector and area is about 13 114 employees, and the average number of firms is about 40. The minimum value of localization variables also shows that there are several firms: the only delegates of their sector in their location. As for urbanization variables, the average value of workers in the same location is 7 times higher than in localization economies variables with an average of about 280 firms.

Table 3 – Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Output	3661.88	27262.30	0.64	2053803.64	29490
Capital	3418.46	29228.55	0	1723772.28	29164
Employment	339.59	820.74	100	40864	29490
# of workers, same sector and location	13114.71	12435.09	100	61765	29490
# of workers, same location	91014.29	52802.09	16701	209883	29490
# of firms, same sector and location	40.07	34.12	1	156	29490
# of firms, same location	279.77	183.97	69	705	29490
Note – Compiled by the author based on 1-PF and 1-T firm-level data from [11]. Output and capital are in thousands of real Tenge. Capital is measured as a fixed assets of a firm.					

RESULTS

Measuring agglomeration economies. Table 4 presents the regression results of the effect of the agglomeration variables on the firm-level TFP. All estimations are based on yearly variations of panel data, which address the short-run effects of spatial agglomeration. The signs and the range of coefficients from both OLS and GMM methods coincide and do not change the conclusion, validating the robustness of the results. According to GMM method, the localization coefficient is positive and significant in the short-run, implying that a 10 % increase in the number of employees of the neighboring firms in the same industry increases firm productivity by 2.88 % on average, keeping other things equal. There is a similar positive relationship between urbanization economies and firm-level productivity: a 10 % increase in the number of workers in neighboring firms operating in other industries increases TFP on average by 1.61 %. However, after controlling the industrial heterogeneity, the localization coefficient decreases by half from 0.288 to 0.136, while the coefficient of urbanization economies additionally gains about 0.034 of firms' productivity and is set at 0.195 in the GMM estimation results. Competition is observed to have a negative impact on TFP: a 10 % increase in the number of employees inside the industry in a given territory lowers productivity by 0.46 % on average, taking other things equal. It implies that even if intense competition is assumed to boost innovations and improve productivity, it reduces firm performance in the short run.

Table 4 – Regression results of the effect of agglomeration economies

Dependent variable	Ln TFP			
Method	OLS	**GMM**	OLS	GMM
localization	0.262*** (0.02)	0.288*** (0.02)	0.122*** (0.01)	0.136*** (0.02)
urbanization	0.167*** (0.03)	0.161*** (0.03)	0.203*** (0.02)	0.195*** (0.02)
competition	-0.196*** (0.02)	**-0.231*** (0.03)**	-0.034 (0.02)	-0.046* (0.03)
Time fixed effect	Yes	**Yes**	Yes	Yes
Industry fixed effect	No	**No**	Yes	Yes
Constant	3.948*** (0.33)	3.879*** (0.36)	3.988*** (0.22)	3.982*** (0.24)
N	29106	22963	29106	22963
R^2	0.049	**0.052**	0.217	0.228
Note – Compiled by the author based on 1-PF and 1-T firm-level data from [11]. All regressions are GMM with clustered standard errors at the location-industry-year level, *p < 0.10, ** p < 0.05, ***p < 0.01				

The difference of geographical aggregation. In this subsection 3.2, we examine at what level of aggregation firms benefit most from agglomeration economies. According to Martin et al., the level of geographical aggregation can affect the values of agglomeration [11]. The geographical variation in our data is provided via 2- and 9-digit classifications. In previous subsection, we focused on geographical entities at the 2-digit regional level, constituting 14 oblasts and 2 cities of the republican level. Here, we employ the 9-digit sub-regional level of spatial aggregation that includes 202 subregions (called rayons) and cities.

Table 5 – Regression results with the different level of geographical aggregation

Dependent variable	Ln TFP			
Method	OLS	GMM	OLS	GMM
localization	0.120*** (0.01)	0.147*** (0.01)	0.050*** (0.01)	0.059*** (0.01)
urbanization	0.075*** (0.01)	0.079*** (0.01)	0.076*** (0.01)	0.078*** (0.05)
competition	-0.093*** (0.02)	-0.148*** (0.02)	0.056*** (0.02)	0.044* (0.04)
Time fixed effect	Yes	Yes	Yes	Yes
Industry fixed effect	No	No	Yes	Yes
Constant	6.185*** (0.07)	6.056*** (0.07)	7.541*** (0.12)	7.302*** (0.43)
N	29106	22963	29106	22963
R2	0.072	0.073	0.304	0.318

Note – Compiled by the author based on 1-PF and 1-T firm-level data from [11].
All regressions are GMM with clustered standard errors at the location-industry-year level, *$p < 0.10$, ** $p < 0.05$, ***$p < 0.01$

Table 5 is obtained by using the same estimation strategy. The GMM method results, corrected for the industrial variation, represent a positive effect of localization and urbanization economies but to a lesser extent compared with the previous results of the 2-digit regional level of geographic aggregation. At this lower level of spatial aggregation, competition tends to positively influence firm-level productivity. The reason for this could be that the competition at the rayon's level is lower than the regional level of geographical aggregation. Competing with fewer firms in the rayon or with firms from the whole region makes a huge difference due to the lower market power. Also, when firms compete with each other, they are motivated to innovate and improve their products or services. This can lead to technological advancements and better production methods that can increase TFP. However, the coefficient of competition is observed to be small and weakly significant.

These results suggest that agglomeration economies, represented by localization (intra-sectoral agglomeration) and urbanization (inter-sectoral agglomeration), increase firm-level productivity at both 2-digit and 9-digit levels of geographic agglomeration. However, firms benefit more from higher levels of spatial aggregation (2-digit), deriving positive agglomeration externalities from labor and resource markets.

Who creates externalities: firms or employees? Which is better: supporting and facilitating the growth of existing firms, or encouraging the opening of new small firms? The question of whether the size of firms or the number of firms in an industry in a given location has a greater influence on firm productivity is an important issue for policymakers. To address it, we decompose localization economies into two terms. The first is the number of firms in the industry *s*, location *l* at time *t*, and the second is the mean size of these firms in terms of the number of employees.

Our results are presented in Table 6 at 2- and 9-digit geographical level of aggregation using the GMM approach. The results coincide with Martin et al. who finds that workers generate higher agglomeration externalities than plants, while Henderson claims the opposite [11; 22]. After controlling the number of neighboring firms in their own industry and their average size, it can be seen that the coefficients of the average size of firms increase total factor productivity at both levels of spatial agglomeration, while the coefficient of the number of firms is found to be insignificant. Interestingly, the magnitude of the coefficient of the average size of firms is quite higher than those of the localization economies.

Table 6 – Regression results with localization decomposition

	ln TFP	
The level of spatial aggregation	2-digit level	9-digit level
Ln (Average size of firms), same industry-area	0.935***	0.877***
	(0.06)	(0.15)
Ln (Number of firms), same industry-area	-0.003	0.222
	(0.05)	(0.20)
urbanization	0.148***	0.027***
	(0.03)	(0.01)
competition	0.179***	-0.086
	(0.06)	(0.25)
Time fixed effect	Yes	Yes
Constant	0.146	2.072**
	(0.49)	(0.01)
N	22805	20648
R2	0.091	0.060
Note – Compiled by the author based on 1-PF and 1-T firm-level data from [11]. All regressions are GMM with clustered standard errors at the location-industry-year level, *p < 0.10, ** p < 0.05, ***p < 0.01		

Overall, these findings indicate that the number of employees in the industry in its territory is one of the main channels of external effects of agglomeration. The size of firms in terms of the number of workers is preferable to the concentration of small firms. The results suggest that supporting the development of large firms and encouraging their internal capacity increases the productivity of firms in clusters, rather than a multiplication policy of smaller firms.

Does the size of clusters matter? The results obtained in previous sections suggest that firm-level productivity increases with the growth of localization and urbanization economies, that is, clustering. Carreira and Lopes and Martin et al. investigate a nonlinear relationship between agglomeration economies and TFP [3; 11]. They argue that in addition to positive externalities, there are also negative ones. Congestion can negatively impact market growth and firm productivity, which means that the relationship between clustering and productivity may be nonlinear. To explain whether clustering is always more beneficial, or whether firms always internalize the benefits from locating nearby, we introduce a nonlinear specification of the model via quadratic and cubic terms of localization and urbanization economies.

Table 7 shows nonlinear regression results for different levels of spatial aggregation. All coefficients of localization terms are significant, while the coefficient for urbanization is significant only at the 9- digit subregional level. Figures 1 and 2 show the results for localization economies at 2- and 9-digit location level, respectively. The dark curve represents the fitted value of TFP gains at each level of localization. The gray curve is a distribution of localization economies for Kazakh firms. Figure 1 shows a cubic curve relationship between TFP and localization economies. Quantitative analysis of the model explains that there is a threshold of around 2 180 employees at which firms start receiving positive externalities from industrial clusterization. The thresholds are calculate by taking natural logarithm of localization. The first threshold is equal to Ln (localization), when TFP=0, because it is starting point when firms gain TFP surplus (see Figure 1). The second threshold is a value of Ln (localization), when TFP reaches its peak. The peak value is computed by taking derivative of the non-linear equation subjected to localization. The second threshold implies a peak of around 363 000 employees in an industry that maximizes the benefits from localization economies, but thereafter, TFP gains decrease due to the congestion effects. Two curves move in a similar pattern suggesting that firms do internalize the TFP surplus from clustering while choosing the location. Figure 2 shows similar results with a weaker fit, implying that firms in a lower level of geographical aggregation are less likely to internalize the productivity gains.

Table 7 – Nonlinear regression results

Dependent variable	ln TFP	
The level of spatial aggregation	2-digit level	9-digit level
localization	-1.523***	-0.299***
	(0.16)	(0.08)
localizationˆ2	0.311***	0.083***
	(0.03)	(0.02)
localizationˆ3	-0.015***	-0.003***
	(0.00)	(0.00)
urbanization	2.806	0.117*
	(2.64)	(0.07)
urbanizationˆ2	-0.333	-0.011
	(0.25)	(0.01)
urbanizationˆ3	0.013	0.000
	(0.01)	(0.00)
competition	-0.339***	-0.378***
	(0.04)	(0.04)
Time fixed effect	Yes	Yes
Constant	0.391	6486***
	(49.80)	(0.13)
N	22965	22965
R^2	0.076	0.091
Note – Compiled by the author based on 1-PF and 1-T firm-level data from [11]. All regressions are GMM with clustered standard errors at the location-industry-year level, *p < 0.10, ** p < 0.05, ***p < 0.01		

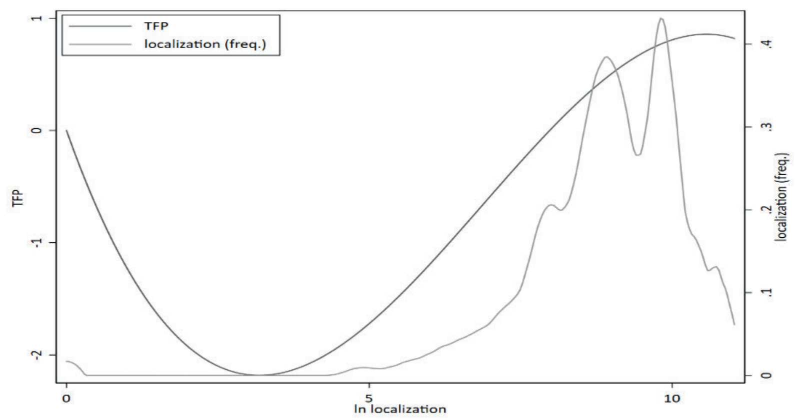

Figure 1 – Localization economies at 2-digit location level
Note – compiled by the author based [11]

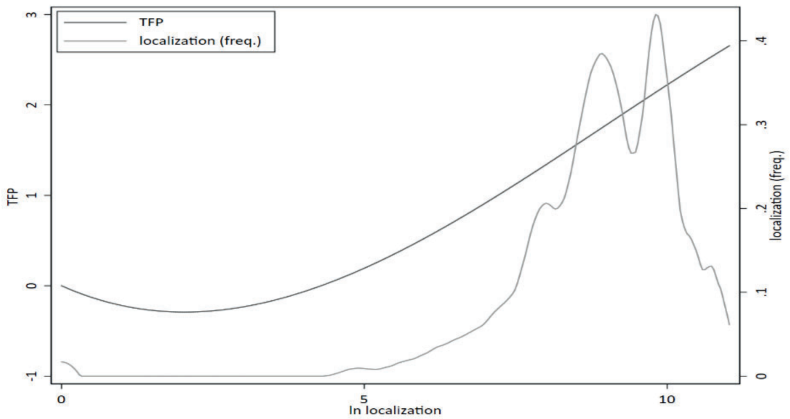

Figure 2 – Localization economies at 9-digit location level
Note – compiled by the author based [11]

Figure 3 shows a linear relationship between urbanization economies and TFP surplus at the 9-digit subregional level of aggregation. It describes that the firm-level productivity gains from urbanization are increasing even if the distribution of urbanization economies passed its peak. This can be explained by the fact that cities and people in Kazakhstan are sparsely located on a large territory, so the level of urbanization is low, and the peak at which negative externalities outweigh positive ones is very high.

In this paper, we considered the location of firms in the regional and subregional levels of aggregation. The subregional level of aggregation is much more accurate in describing the actual level of agglomeration in the country, while the 2-digit regional level of aggregation shows the adjusted level of agglomeration.

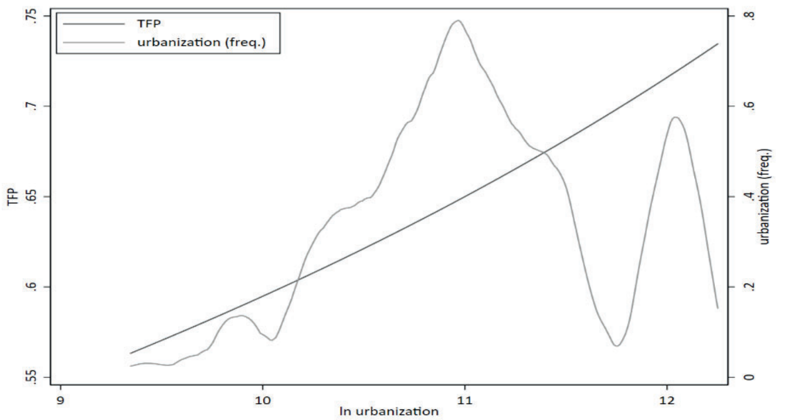

Figure 3 – Urbanization economies at 9-digit location level
Note – compiled by the author based [11]

These findings explain that more clustering is not always better. There are negative externalities of agglomeration, such as congestion, which can outweigh the benefits, holding all other factors constant. Policymakers need to take these factors into account when deciding whether to stimulate cluster growth.

Table 8 – The elasticity coefficients of localization economies at 2-digit regional level

Region	
Astana city	0.193 (11.73)
Almaty city	0.178 (8.72)
Akmola region	0.064 (-4.56)

Aktobe region	0.098 (4.89)
Almaty region	0.110 (5.10)
Atyrau region	0.153 (-1.98)
West Kazakhstan region	0.082 (0.94)
Zhambyl region	0.041 (-3.69)
Karagandy region	0.082 (-0.92)
Kostanay region	0.082 (-1.53)
Kyzylorda region	0.052 (-1.74)
Mangystau region	0.124 (3.75)
South Kazakhstan region	0.082 (-0.92)
Pavlodar region	0.099 (3.94)
North Kazakhstan region	0.054 (-4.74)
East Kazakhstan region	0.082
Note – Compiled by the author on 1-PF and 1-T firm-level data from [11]. Estimated using GMM approach with clustered standard errors at the location-industry-year level. The elasticities are calculated as it is indicated below (e.g., for Astana city = $\hat{\sigma}_0 + \hat{\sigma}_1$). T-statistics of the interaction terms are in parentheses. The reference category is the East Kazakhstan region and takes the value of the localization coefficient ($\hat{\sigma}_0$). All statistically insignificant coefficients for regions' interaction terms takes the value of $\hat{\sigma}_0$.	

Finally, Table 8 presents the elasticities of localization economies to firm productivity at the 2- digit regional level. To obtain regional-level coefficients, we introduce interactive terms of localization variable with regional dummies to Equation (11) and get:

$$\log(A_{it}) = \sigma_0 \log(LOC_{it}^{sl}) + \sum_{i=1}^{15} \sigma_i \mathrm{D_i} \log\left(LOC_{it}^{sl}\right) + \gamma \log(URB_{it}^{sl}) + \\ + \mu log(COMP_{it}^{sl}) + u_{it} \quad (12)$$

The benchmark region is East Kazakhstan oblast (EKO), so that $\mathrm{D_i} = 0$ for $\forall i = 1,2,\ldots,15$.

where $D_1 = \begin{cases} 1\ if\ Astana\ city \\ 0\ if\ otherwise \end{cases}$

where $D_2 = \begin{cases} 1\ if\ Almaty\ city \\ 0\ if\ otherwise \end{cases}$

…

…

where $D_{15} = \begin{cases} 1\ if\ North\ Kazakhstan\ region \\ 0\ if\ otherwise \end{cases}$

The elasticity of TFP with respect to localization would be different. That is, for EKO:

$$\frac{d\ log(A_{it})}{d\ log(LOC_{it}^{sl})} = \sigma_0$$

For Astana city:

$$\frac{d\ log(A_{it})}{d\ log(LOC_{it}^{sl})} = \sigma_0 + \sigma_1$$

The elasticities show how sensitive productivity of firms would be to the localization economies across regions. The higher the elasticity, the more sensitive firms are to an increase in the number of employees in the industry. The results suggest that TFP in Astana city, Almaty city, Atyrau, and Mangystau regions is more sensitive than in EKO, and they will receive higher firm-level productivity gains from localization economies.

The economic rationale for this could be that the level of market development and competition in EKO may potentially prevent TFP gains from the localization economy. At the same time, Zhambyl, Kyzylorda and North-Kazakhstan oblasts are less sensitive for TFP than in East Kazakhstan region, and they will receive modest benefits from higher clustering compared to other regions.

CONCLUSION

This paper analyzes the effect of agglomeration on total factor productivity using firm-level panel data for Kazakhstan. We show that localization and urbanization economies have a strong and positive impact on firms' performance. The results suggest that productivity increases with clustering: a 10 % increase in the number of employees of neighboring firms inside the same industry increases firm- level productivity by 1.36 %, while a 10 % increase in employment in other industries enhances firm performance by 1.95 %.

Our results have several interesting policy implications. First, industrial clusters and urbanization have positive side effects on productivity in the short term. Productivity gains are higher at the 2-digit regional level than at the 9-digit subregional level of geographical aggregation, implying that the denser geography increases firms' performance more than in the observed geography (2-digit regional level of spatial aggregation consists of 14 oblasts and 2 cities of republican significance, 9-digit subregional level consist of 202 rayons and cities). Second, fostering internal growth of existing firms and attracting large firms amplifies the positive externalities for TFP from industrial clustering rather than the proliferation of small firms in the region. Third, besides the positive externalities, other externalities such as congestion effects can reduce productivity and negatively affect spatial growth. We find thresholds at which firms start gaining benefits from locating near industrial clusters and at which negative effects of agglomeration dominate positive externalities. When clusters become overcrowded (more than 363 000 employees in the industry), congestion effects may outweigh localization effects. Consequently, policymakers should also consider the impact of negative externalities before introducing government programs.

REFERENCES

1. Storper M., Venables A. Buzz: face-to-face contact and the urban economy // Journal of Economic Geography. – 2004. – № 4(4). – P. 351-370.

2. Audretsch D., Feldman M. Knowledge spillovers and the geography of innovation // Elsevier: Handbook of Regional and Urban Economics. – 2004. – № 4. – P. 2713-2739.

3. Carreira C., Lopes L. Regional knowledge spillovers: a firm-based analysis of non-linear effects // Regional Studies. – 2018. – № 52(7). – P. 948-958.

4. Marshall A. Principles of economics. – Macmillan, London, 1890. – 320 p.

5. Berliant M., Reed R., Wang P. Knowledge exchange, matching, and agglomeration // Journal of Urban Economics. – 2006. – № 60(1). – P. 69-95.

6. Wooldridge J. On estimating firm-level production functions using proxy variables to control for unobservables // Economics Letters. – 2009. – № 104. – P. 112-114.

7. Andersson M., Loof H. Agglomeration and Productivity – evidence from firm-level data // Working Paper Series in Economics and Institutions of Innovation. No. 170. – CESIS, 2009. – 39 p.

8. Combes P.-P., Duranton G., Gobillon L., Puga D., Roux S. The productivity advantages of large cities: Distinguishing agglomeration from firm selection // Econometrica. – 2012. – № 80(6). – P. 2543-2594.

9. Harris R., Moffat J. Total factor productivity growth in local enterprise partnership regions in Britain // Regional Studies. – 2015. – № 96(6). – P. 1019-1041.

10. Lopez R., Sudekum J. Vertical industry relations, spillovers, and productivity: Evidence from Chilean plants // Journal of Regional Science. – 2009. – № 49(4). – P. 721-747.

11. Martin P., Mayer T., Mayneris F. Spatial concentration and plant-level productivity in France // Journal of Urban Economics. – 2011. – № 69(2). – P. 182-195.

12. Hashiguchi Y., Tanaka K. Agglomeration and firm-level productivity: A Bayesian spatial approach // Papers in Regional Science. – 2015. – № 94. – S95-S114.

13. Cainelli G., Ganau R. Distance-based agglomeration externalities and neighboring firms' characteristics // Regional Studies. – 2018. – № 52(7). – P. 922-933.

14. DiGiacinto V., Gomellini M., Micucci G., Pagnini M. Mapping local productivity advantages in Italy: Industrial districts, cities or both? // Journal of Economic Geography. – 2014. – № 14. – P. 365-394.

15. Levinsohn J., Petrin A. Estimating production functions using inputs to control for unobservables // Review of Economic Studies. – 2003. – № 70(2). – P. 317-341.

16. Cainelli G., Fracasso A., Marzetti G. V. Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach // Papers in Regional Science. – 2015. – № 94. – S39-S67.

17. Combes P.-P., Gobillon L. The empirics of agglomeration economies // Elsevier: Handbook of urban and regional economics. – 2015. – № 5. – P. 247-348.

18. Olley S., Pakes A. Dynamic behavioral responses in longitudinal data sets: Productivity in telecommunications equipment industry // Econometrica. – 1996. – № 64(6). – P. 1263-1297.

19. Wooldridge J. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity // Journal of applied econometrics. – 2005. – № 20(1). – P. 39-54.

20. Ackerberg D., Caves K., Frazer G. Structural identification of production functions? // MPRA Paper, 38349. – 2006. – 42 p.

21. Statistical forms for [Electronic resource] // Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan [website]. – 2021. – URL: https://stat.gov.kz/en/respondents/statistical-forms/ (Accessed: 16.01.2023).

22. Henderson J. V. Marshall's scale economies // Journal of Urban Economics. – 2003. – № 53(1). – P. 1-28.

REFERENCES

1. Storper, M. and Venables, A. (2004). Buzz: face-to-face contact and the urban economy. Journal of Economic Geography, 4(4), 351-370.

2. Audretsch, D. and Feldman, M. (2004). Knowledge spillovers and the geography of innovation. Elsevier: Handbook of Regional and Urban Economics, 4, 2713-2739.

3. Carreira, C. and Lopes, L. (2018). Regional knowledge spillovers: a firm-based analysis of non-linear effects. Regional Studies, 52(7), 948-958.

4. Marshall, A. (1890). Principles of economics. Macmillan, London, 320 p.

5. Berliant, M., Reed, R. and Wang, P. (2006). Knowledge exchange, matching, and agglomeration. Journal of Urban Economics, 60(1), 69-95.

6. Wooldridge, J. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. Economics Letters, 104, 112-114.

7. Andersson, M. and Loof, H. (2009). Agglomeration and Productivity – evidence from firm-level data. Working Paper Series in Economics and Institutions of Innovation. No. 170, CESIS, 39 p.

8. Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica, 80(6), 2543-2594.

9. Harris, R. and Moffat, J. (2015). Total factor productivity growth in local enterprise partnership regions in Britain. Regional Studies, 96(6), 1019-1041.

10. Lopez, R. and Sudekum, J. (2009). Vertical industry relations, spillovers, and productivity: Evidence from Chilean plants. Journal of Regional Science, 49(4), 721-747.

11. Martin, P., Mayer, T. and Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. Journal of Urban Economics, 69(2), 182-195.

12. Hashiguchi, Y. and Tanaka, K. (2015). Agglomeration and firm-level productivity: A Bayesian spatial approach. Papers in Regional Science, 94, S95–S114.

13. Cainelli, G. and Ganau, R. (2018). Distance-based agglomeration externalities and neighboring firms' characteristics. Regional Studies, 52(7), 922-933.

14. DiGiacinto, V., Gomellini, M., Micucci, G. and Pagnini, M. (2014). Mapping local productivity advantages in Italy: Industrial districts, cities or both? Journal of Economic Geography, 14, 365-394.

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

16. Cainelli, G., Fracasso, A. and Marzetti, G. V. (2015). Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach. Papers in Regional Science, 94, S39-S67.

17. Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. Elsevier: Handbook of urban and regional economics, 5, 247-348.

18. Olley, S. and Pakes, A. (1996). Dynamic behavioral responses in longitudinal data sets: Productivity in telecommunications equipment industry. Econometrica, 64(6), 1263-1297.

19. Wooldridge, J. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics, 20(1), 39-54.

20. Ackerberg, D., Caves, K. and Frazer, G. (2006). Structural identification of production functions? MPRA Paper, 38349, 42 p.

21. Statistical forms for. (2021). Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Retrieved, January 16, 2023, from https://stat.gov.kz/en/respondents/statistical-forms/.

22. Henderson, J. V. (2003). Marshall's scale economies. Journal of Urban Economics, 53(1), 1-28.

ҚАЗАҚСТАНДАҒЫ КЕҢІСТІКТІ КОНЦЕНТРАЦИЯ ЖӘНЕ ФИРМА ДЕҢГЕЙІНДЕГІ ӨНІМДІЛІК

З. М. Адилханова
NAC Analytica, Назарбаев Университеті, Астана, Қазақстан Республикасы

АҢДАТПА

Зерттеу мақсаты. Бұл жұмыс 2009-2017 жылдар аралығындағы панельдік деректерді пайдалана отырып, Қазақстандағы фирмалардың жалпы фактор өнімділігіне кеңістік агломерациясының әсерін зерттейді.

Әдіснамасы. Біз моменттердің жалпылама әдісін (GMM) қолдана отырып, екі сатылы бағалау стратегиясын және эндогендік бейімділіктерді бақылауды қолданамыз. Фирма деңгейіндегі деректер Қазақстан Республикасы Стратегиялық жоспарлау және реформалар агенттігінің Ұлттық статистика бюросынан алынған.

Зерттеудің бірегейлігі / құндылығы. Бұл зерттеу дамушы елдердегі кеңістіктік шоғырлану мен фирма деңгейінің өнімділігін эмпирикалық зерттеуге ықпал етеді және саясаткерлерге мемлекеттік бағдарламаларды жүзеге асыру алдында қарастыратын құнды түсініктерді береді.

Зерттеу нәтижелері. Нәтижелер кластерлеу кезінде өнімділіктің жоғарылайтынын көрсетеді: бір саладағы көрші фирмалар қызметкерлерінің санының 10 %-ға артуы фирма деңгейіндегі өнімділікті 1,36 %-ға арттырады, ал басқа салалардағы жұмыспен қамтудың 10 %-ға артуы фирма өнімділігін 1,95 %-ға арттырады. Өнімділіктің артуы географиялық біріктірудің 9-сандық қосалқы аймақтық деңгейіне қарағанда 2-сандық аймақтық деңгейде жоғары болады, бұл тығыз географияның фирмалардың өнімділігін жоғарылататынын білдіреді.

Түйін сөздер: агломерациялық экономикалар, жалпы фактор өнімділігі, кеңістіктік концентрация, кластерлер.

ПРОСТРАНСТВЕННАЯ КОНЦЕНТРАЦИЯ И ПРОИЗВОДИТЕЛЬНОСТЬ ФИРМ В КАЗАХСТАНЕ

З. М. Адилханова
NAC Analytica, Назарбаев Университет, Астана, Республика Казахстан

АННОТАЦИЯ

Цель исследования. В данной статье изучается влияние пространственной агломерации на общую факторную производительность предприятий в Казахстане с использованием панельных данных с 2009 по 2017 год.

Методология. Мы используем двухэтапную стратегию оценки и контролируем отклонения эндогенности, используя подход обобщенного метода моментов (GMM). Данные на уровне предприятий получены от Бюро национальной статистики Агентства стратегического планирования и реформ Республики Казахстан.

Оригинальность / ценность исследования. Это исследование вносит вклад в эмпирическое исследование пространственной концентрации и производительности на уровне компаний в развивающихся странах и дает ценную информацию для политиков, которую следует учитывать перед реализацией государственных программ.

Результаты исследования. Результаты показывают, что производительность увеличивается при объединении в кластеры: 10 % увеличение числа сотрудников соседних фирм в той же отрасли увеличивает производительность на уровне фирмы на 1,36 %, а 10 % увеличение занятости в других отраслях повышает производительность фирмы на 1,95 %. Прирост производительности выше на двузначном региональном уровне, а не на 9-значном субрегиональном уровне географической агрегации, что означает, что более плотная география повышает производительность компаний.

Ключевые слова: агломерационные экономики, общая факторная производительность, пространственная концентрация, кластеры.

ABOUT THE AUTHOR

Adilkhanova Zarina Muratovna – master of arts in economics, senior researcher, Economic Modeling Development Center, NAC Analytica, Nazarbayev University, Astana, Republic of Kazakhstan, e-mail: zarina.adilkhanova@nu.edu.kz, https://orcid.org/0000-0002-7206-6290