



Higher Education Agglomeration Promoting Innovation and Entrepreneurship Based on Spatial Dubin Model

Xingyuan Shao^{1*}, Xiaoping Bao², Ahlam Al-Ethawi³

1 Headquarter of NAU new campus construction, Nanjing Agriculture University, Nanjing, 210095, China

2 Business College, Nanjing University of Science and Technology ZiJin College, Nanjing, 210000, China

3. College of Administrative Sciences, Applied Science University, Bahrain

Submission Info

Communicated by Juan Luis García Guirao

Received January 09th 2022

Accepted March 26th 2022

Available online July 15th 2022

Abstract

We use the panel data of innovation and entrepreneurship for global spatial autocorrelation and model selection testing. At the same time, we applied the Spatial Dubin Model (SDM) to empirically study the knowledge spillover effects of higher education campuses on regional innovation capabilities. Experimental research found that there are significant spatial differences in the innovation efficiency of universities. Specialized agglomeration can effectively improve the innovation efficiency and scale efficiency of colleges and universities. At the same time, the diversified agglomeration has no obvious effect on the innovation efficiency of colleges and universities. The accumulation of high-tech industries can effectively promote the innovation efficiency of universities. Human capital and opening to the outside world can effectively improve the innovation efficiency of universities.

Keywords: Higher education; Knowledge spillover; Innovation and entrepreneurship; Spatial Dubin model

AMS 2010 codes: 62H20

†Corresponding author.

Email address: sxy@niau.edu.cn

 sciendo

This work is licensed under the Creative Commons Attribution alone 4.0 License.

1. Introduction

Universities are the main body of knowledge creation and transmission, and they are also important gathering places for scientific research talents. Since China implemented the strategy of "rejuvenating the country through science and education," the R&D (Research and Experimental Development) activities of universities have played an important role in the implementation of social development, economic construction, and independent innovation strategies [1]. Moreover, the products invested by R&D have typical characteristics of public goods. It acts on the micro-level to show strong externalities to the regional industrial structure and total output. The final manifestation is the overall improvement of the industry and the improvement of the overall technical level.

From a realistic perspective, the output of university R&D activities is the intermediate knowledge output and many market-based outputs closely related to enterprises' business behavior [2]. Because although university R&D focuses on basic research, basic research as a public product will stimulate technological innovation. This positive externality benefits various enterprises and manufacturers in the form of knowledge spillovers. Therefore, the impact of university R&D on regional innovation capabilities should also include the two aspects mentioned above. But judging from the existing literature, it mainly revolves around the knowledge spillover of the intermediate output of university R&D, even if there is little literature involving two aspects of regional innovation capabilities. The main reason is that it has not adopted a reasonable empirical method and has a relatively large limitation. Therefore, the use of practical analytical tools and standardized empirical analysis is of great significance to clarify the R&D knowledge spillovers in universities.

2. Model setting

2.1 Theoretical measurement model

The theoretical analysis foundation of the spillover effect of university R&D on the regional economy is based on the conceptual framework of the knowledge production function [3]. In essence, the knowledge production function is still a typical Cobb-Douglas production function [4]. Since university R&D investment involves two aspects: expenditure and personnel input, we give a revised knowledge production function based on related theories:

$$K = A \cdot UE^\alpha UN^\beta e^\mu \quad (1)$$

K is the knowledge output, which is used to measure the regional innovation capability. UE is the expenditure of university R&D. UN provides personnel input for university R&D. α and β are the output elasticity of expenditure and personnel input, respectively. A stands for other influencing factors. e^μ is the error term. Take the logarithm of both sides of equation (1) and expand it to get:

$$\ln K = \ln A + \alpha \ln UE + \beta \ln UN + \mu \quad (2)$$

We add the model's R&D expenditure (FE) and personnel input (FN) of large and medium-sized enterprises as control variables. Modify the model based on the 2018-2020 inter-provincial panel data selected in this article:

$$\ln Patent_{it} = \beta_0 + \beta_1 \ln UE_{it} + \beta_1 \ln UN_{it} + \beta_2 \ln FE_{it} + \beta_3 \ln EN_{it} + \xi_{it} \quad (3)$$

$$\ln NewRevenue_{it} = \gamma_0 + \gamma_1 \ln UE_{it} + \gamma_1 \ln UN_{it} + \gamma_2 \ln FE_{it} + \gamma_3 \ln FN_{it} + \eta_{it} \quad (4)$$

Models (3) and (4) are the basic measurement models used later in this article.

2.2 Spatial Autoregressive (SAR) and Spatial Error Model (SEM)

The SAR model is mainly used to measure the spatial spillover effects of dependent variables, and its setting form is as follows:

$$y_{it} = \rho \cdot \sum_{j=1}^N W_{ij} y_{jt} + X_{it} \phi_i + \theta + \varphi_i + \eta_i + \varepsilon_{it} \quad (5)$$

y_{it} is the observed value ($i = 1, 2, \dots, N; t = 1, 2, \dots, T$) of the dependent variable in the area i during the t period. W_{ij} is the spatial weight matrix of order $N \times N$. θ is the constant term, and ε_{it} is the error term. SEM is used to reflect the spatial dependence of the error term. In this article, SEM measures the error impact of R&D expenditures and personnel inputs of universities in a region on the region's innovation capabilities. And its impact on the changes in innovation capabilities in other regions [5]. The specific form of the model is:

$$y_{it} = X_{it} \phi_i + \theta + \varphi_i + \eta_i + \tau_{it} \quad (6)$$

$$\tau_{it} = \lambda \cdot \sum_{j=1}^N W_{ij} \tau_{jt} + \varepsilon_{it}$$

τ_{it} is the error term, $\sum W_{ij} \tau_{jt}$ is the error as mentioned above impact, and λ is the coefficient; other parameters have the same meaning as a model (5).

2.3 Spatial Doberman Model (SDM)

SEM does not consider the spatial spillover effects of dependent variables. The previous research literature on the knowledge spillover effects of R&D in Chinese universities is mainly based on the estimation results of the SEM model. The SAR model ignores the spatial dependence in the error term, which reduces the effectiveness of the estimation. Because of the characteristics and limitations of the SAR and SEM models, some scholars have proposed an SDM model that can simultaneously measure the spatial spillover effect of the dependent variable and the spatial dependence of the error term [6]. The characteristic of this model is that it contains both the spatial lag term of the dependent variable and the spatial lag term of the explanatory variable. Whether the real data generation process is SEM or SAR, the SDM model has relatively general and robust estimation results. Its form is as follows:

$$y_{it} = \rho \cdot \sum_{j=1}^N W_{ij} y_{jt} + X_{it} \phi_i + \delta \cdot \sum_{j=1}^N W_{ij} X_{ijt} + \theta + \varphi_i + \eta_i + \varepsilon_{it} \quad (7)$$

$\sum W_{ij} X_{ijt}$ measures the impact of R&D expenditures and personnel inputs of universities in one region on the changes in innovation capabilities in other regions. δ is the coefficient. The meaning of other parameters is the same as a model (5).

3. Selection of data sources and spatial measurement models

3.1 Data sources and related processing

Only in 2008 did university R&D expenditures. Since 2011, the statistical caliber of industrial enterprises has been changed from large and medium-sized to large-scale [7]. Relative quantities cannot examine the overall scale of university R&D expenditures and personnel input knowledge spillovers in various regions, so all variables in models (3) and (4) are absolute numbers. In 1994, China implemented the tax-sharing system reform, so we set the 1994 CPI index as the base price. Table 1 shows the running results of STATA12.0, and the rest are obtained by Matlab R2012b.

Table 1 LM sarsem test

InPatent	π	0.1	0.2	0.3	0.4	0.5
	LM-LAG	1.126	3.002	5.865	9.739	14.61
	P	0.289	0.083	0.015	0.002	0

	LM-ERROR	56.01	56.15	55.76	54.46	51.75
	P	0	0	0	0	0
lnNewRevenue	π	0.1	0.2	0.3	0.4	0.5
	LM-LAG	11.94	10.20	8.519	6.92	5.433
	P	0.001	0.001	0.004	0.009	0.02
	LM-ERROR	13.39	11.90	10.11	8.012	5.647
	P	0	0.001	0.001	0.005	0.017

3.2 Selection of spatial weight matrix form

First, select the spatial weight matrix. In the past, the research literature of university R&D knowledge spillover only considered the geographical distance between regions and did not consider the gap in economic growth between regions [8]. This paper adds the regional economic distance weight matrix to the geographical distance weight matrix:

$$\begin{aligned}
 wd_{ij} &= 1/d_{ij} \\
 we_{ij} &= 1/|Income_i - Income_j| \\
 W_{ij} &= \pi \cdot WD_{ij} + (1 - \pi)WE_{ij}
 \end{aligned} \tag{8}$$

d_{ij} is the geographic distance, and wd_{ij} is the geographic distance weight matrix. $Income_i$ and $Income_j$ represent the per capita disposable income of urban residents in regions i and j . $|Income_i - Income_j|$ is the economic development gap between regions [9]. We choose the average value of urban per capita disposable income from 2018 to 2020 as the economic distance between regions. WD_{ij} and WE_{ij} are the standardized forms of the geographical distance weight matrix and the economic distance weight matrix, respectively. It is used to control the share of geographic distance and economic distance. The estimation results of SAR, SEM, and SDM models are largely determined by the value of π . W_{ij} is the weighted spatial weight matrix.

3.3 LM spatial autocorrelation-spatial error test (LMsarsemtest) and determination of the value of π

Some scholars have proposed that the LM statistic obeys the asymptotic $\chi^2(1)$ distribution. In this way, it is judged whether to choose a space measurement model and what form of the space measurement model to choose. This model is more concise and convenient than Moran's I statistics.

When the explained variable is lnPatent, regardless of the value of π , the LM-ERROR statistic is very significant. Explain that the data generation process is in SEM format. But with the increase of the geographic distance share, the significance of the LM-ERROR statistics diminishes. At this time, the model completely uses geographic distance as the weight matrix. When $\pi=0.1$, the SAR model is not suitable. When $\pi=0.2$, the LM-LAG statistic is significant at the 1% level. SAR models can be selected starting from $\pi=0.3$. When the explained variable is lnNewRevenue, when π takes a value of 0 to 0.6, both SAR and SEM models are suitable. It shows that the data generation process conforms to both the SAR format and the SEM format. At this time, it is very necessary to choose the SDM model. Otherwise, the validity and consistency of the parameter estimation will be biased. When the explained variable is lnPatent, the π values are 0.7, 0.7, and 0.6, respectively. When the explained variable is lnNewRevenue, π is 0.3, 0.6, and 0.3, respectively.

4. Empirical analysis

4.1 Non-spatial econometric model estimation results

The results of the hybrid OLS estimation show that regardless of whether the explained variable is $\ln\text{Patent}$ or $\ln\text{New Revenue}$, the expenditure and personnel input of university R&D and enterprise R&D are significant. Hybrid OLS is the benchmark and reference model for subsequent spatial measurement model estimation. The LM space autoregressive-error test is based on the OLS model. The table also shows the estimated results of the ordinary panel's fixed effects and random-effects models. $\ln\text{Patent}$ and $\ln\text{New Revenue}$ reject the assumption that there is no difference between fixed effects and random effects at the significance levels of 1% and 5%, respectively [10]. The estimation method of the fixed effect model is better, so the fixed effect model is also used to estimate the spatial measurement model. When the explained variable is $\ln\text{Patent}$, the effect of university R&D personnel input is uncertain. When the explained variable is $\ln\text{New Revenue}$, the explanatory variables are significant, and the degree of fit is relatively higher. The DW value under the mixed OLS model is also close to 2, which shows that heteroscedasticity does not affect the model.

4.2 Estimated results of SAR and SEM models

The SEM model measures the error impact of the R&D expenditure and personnel input of universities in a region on the region's innovation capability. At the same time, it also reflects the extent of its impact on the changes in innovation capabilities in other regions. Tables 2 and 3 show the parameter estimation results of the SAR and SEM models of the knowledge spillover effects of university R&D on regional innovation capabilities as the explained variables $\ln\text{Patent}$ and $\ln\text{New Revenue}$ respectively.

Table 2 The knowledge spillover effect of university R&D when the explained variable is $\ln\text{Patent}$

	$\ln\text{Patent}$					
	SAR			SEM		
	Regional fixed	Fixed time	Double fixation	Regional fixed	Fixed time	Double fixation
$\ln\text{UE}$	-0.1061	0.1495	-0.2030	-0.1549	0.1984	-0.2071
	(-2.8704)	-3.3179	(-4.3726)	(-3.2648)	-4.2377	(-4.1984)
$\ln\text{UN}$	-0.1433	0.1595	-0.1322	-0.1493	0.1182	-0.1278
	(-1.8661)	-2.1488	(-1.5826)	(-1.7383)	-1.525	(-1.4396)
$\ln\text{FE}$	0.2418	0.6328	0.1634	0.1591	0.7266	0.1691
	-5.7823	-11.0594	-3.2596	-3.0969	-12.0784	-3.1735
$\ln\text{FN}$	0.1205	-0.0379	0.1545	0.0984	-0.1156	0.1653
	-2.7329	(-0.634)	-3.3443	-2.0245	(-1.8273)	-3.3719
ρ	0.7510	0.5060	0.2080			
	-19.0213	-8.8495	-1.6646			
λ	0.9220	0.2700	0.2604			
	-62.9079	-2.2429	-2.1436			
R^2	0.9726	0.9047	0.9732	0.8871	0.8944	0.973
$\text{Adjust } R^2$	0.8782	0.8714	0.8696	0.8672	0.8687	0.8623
LogL	-2.0509	-242.1127	10.6244	-2.0134	-254.54	9.8759

Table 3 The knowledge spillover effect of university R&D when the explained variable is In New Revenue

	In New Revenue					
	SAR			SEM		
	Regional fixed	Fixed time	Double fixation	Regional fixed	Fixed time	Double fixation
lnUE	0.1895	0.2529	0.1893	0.2565	0.3005	0.1886
	-3.4387	-4.8188	-3.0238	-5.1687	-5.6397	-2.8485
lnUN	0.2960	0.1432	0.2253	0.2025	0.1120	0.2281
	-2.8203	-1.6604	-1.9983	-1.7871	-2.2665	-1.9115
lnFE	0.3786	0.8486	0.3258	0.3679	0.9332	0.3269
	-6.1334	-12.3513	-4.8205	-5.5543	-13.0891	-4.5652
lnFN	- 0.1013	- 0.2566	- 0.1162	-0.0662	- 0.3292	- 0.1181
	(- 1.6818)	(- 3.6053)	(- 1.8667)	(- 1.0265)	(- 4.4152)	(- 1.7947)
ρ	0.4820	0.4390	0.1139			
	-9.4609	-8.3856	-1.6582			
λ				0.8660	0.2540	0.1068
				-33.3533	-5.2703	-1.6751
R²	0.9665	0.9158	0.9681	0.906	0.9013	0.9681
Adjust R²	0.8917	0.8724	0.8741	0.8733	0.8614	0.8542
LogL	-121.1745	-300.0728	-105.3039	-149.3385	-316.9979	-101.2397

When the explained variable is lnPatent, the π value of the SAR and SEM spatial weight matrices are both 0.7. This shows that the influence of geographic distance is greater than that of economic distance. From the LM inspection in Table 1, it can be seen that the data generation process at this time is both SAR and SEM forms. In the SEM model, when they are 0.1691% and 0.1653%, respectively, the effect of capital investment in enterprise R&D is only slightly higher than the effect of personnel input. For every increase of 1% of the knowledge spillover effect of the innovation capability level of the region on the innovation capability of the surrounding area, the innovation capability of the surrounding area will increase by 0.208%.

When the explained variable is In New Revenue, the π value in the SAR model is 0.3. The impact of economic distance is 70%, and the value of π in the SEM model is 0.6. The impact of geographic distance is greater than that of economic distance. Although the value of π is different, the data generation process is still in both SAR and SEM formats. Both university R&D expenditures and personnel input are significantly conducive to the increase in sales revenue of new products in the region. In SAR, every 1% increase in university R&D funding will increase sales revenue by 0.1893%. For every 1% increase in staff full-time equivalent, new product sales revenue will increase by 0.2253%. The relevant data in SEM were 0.1886% and 0.2281%. This is not much different from SAR. Corporate R&D expenditures and personnel inputs have different effects on new product sales revenue. Expenditures are significantly conducive to the increase in sales revenue of new products. The SAR and SEM models are 0.3258% and 0.3269%, respectively, which are significantly greater than the contribution of college R&D expenditures. However, the personnel investment of enterprise R&D is not conducive to the increase of sales revenue of new products.

4.3 Spatial Doberman Model (SDM)

When the explained variable is $\ln \text{Patent}$ or $\ln \text{New Revenue}$, the data generation process is both in the form of SAR and SEM. Therefore, it is necessary to use the SDM model. It can be seen from Table 4 that when the explained variable is $\ln \text{Patent}$ or $\ln \text{New Revenue}$, the fitting degree of the SDM model under the double fixed effects is higher than that of the SAR and SEM models. When the explained variable is $\ln \text{Patent}$, university R&D expenditure and personnel input are still not conducive to improving regional basic innovation capabilities. This point is consistent with the estimated results of SAR and SEM, and the spatial spillover effects of the two to other regions are not significant. Enterprise R&D expenditure and personnel input have a significant positive effect on improving regional basic innovation capabilities. The effect of capital investment is greater than the effect of personnel investment. The elasticity is 0.1739% and 0.1257%, respectively. The SAR and SEM models respectively underestimate the former and significantly overestimate the role of enterprise R&D personnel input in promoting regional basic innovation capabilities. The spatial spillover effect of the dependent variable is -0.3860%.

When the explanatory variable is $\ln \text{New Revenue}$, university R&D expenditures and personnel input are conducive to improving regional secondary innovation capabilities. The improvement ratio was 0.1762% and 0.2768%, respectively. The SAR and SEM models overestimate the former and significantly underestimate the role of university R&D personnel input. Personnel investment is also not conducive to improving regional secondary innovation capabilities, which are close to the estimated results of SAR and SEM models. The spatial spillover effect of the dependent variable is positive. This shows that the increase in sales revenue of new products in this region is conducive to improving secondary innovation capabilities in surrounding areas. This effect is 0.0249%.

Table 4 Spatial Dubin model estimation results

Explanatory variables	SAR			SEM		
	Regional fixed	Fixed time	Double fixation	Regional fixed	Fixed time	Double fixation
lnUE	-0.1607	0.0998	-0.1636	0.1695	0.1811	0.1762
	(-3.6828)	-2.0982	(-3.6731)	-2.7025	-3.3192	-2.7827
lnUN	-0.1291	0.2047	-0.0949	0.2978	0.2179	0.2768
	(-1.6639)	-2.6456	(-1.1611)	-2.6684	-2.455	-2.3919
lnFE	0.1707	0.6013	0.1739	0.3742	0.8327	0.3454
	-3.6306	-9.8824	-3.6084	-5.5388	-11.775	-5.0811
lnFN	0.1274	-0.0121	0.1257	-0.1263	-0.2361	-0.1170
	-2.8754	(-0.1927)	-2.7945	(-1.9710)	(-3.1813)	(-1.8350)
WlnUE	0.0618	-0.1944	-0.1554	-0.2507	-0.2236	-0.0996
	-0.5387	(-0.6875)	(-0.7059)	(-1.9035)	(-1.1171)	(-0.5241)
WlnUN	-0.0451	-0.408	0.5391	0.6837	0.2259	0.5859
	(-0.2282)	(-1.0500)	-1.2123	-2.7406	-0.7605	-1.4445
WlnFE	0.5128	2.0471	0.5202	0.5288	1.2255	0.7136
	-3.2092	-6.7424	-1.8907	-2.9566	-5.0426	-2.839
WlnFN	0.7663	-1.3748	0.9643	-0.3629	-1.3439	-0.3925
	-5.0543	(-4.5872)	-4.4216	(-2.3054)	(-5.8152)	(-2.0111)
p	-0.3040	0.104	-0.3860	0.2680	0.3450	0.0249
	(-2.2383)	-0.9781	(-2.5783)	-3.0599	-4.2713	-1.6621
R2	0.9755	0.915	0.9762	0.967	0.9218	0.9686
Adjust R2	0.8917	0.8943	0.8752	0.8973	0.8831	0.8667
LogL	27.5879	-214.061	32.514	-114.031	-283.615	-102.539

5. Conclusion

This paper uses China's provincial data from 2018 to 2020 and uses the Spatial Durbin Model (SDM) to study the knowledge spillover effect of R&D in Chinese universities on regional innovation capabilities. The results show that (1) University R&D expenditures are not conducive to improving basic innovation capabilities in the region, and the impact of personnel input is not significant. (2) Expenditure and personnel input have a significant role in promoting the secondary innovation capabilities of the region, and the promotion effect of personnel input is greater than the effect of expenditures. (3) The expenditure and personnel input of university R&D did not produce direct knowledge spillover effects on surrounding areas. This spillover effect is more indirectly realized through changes in the region's innovation capabilities.

References

- [1] Zhu, Y. How International Trade Affects the Spatial Pattern of China's Economic Activities?. *American Journal of Industrial and Business Management.*, 2019 9(3): 627-646
- [2] Gong, Z., Kong, Q., & Tan, J. Analysis on multidimensional poverty reduction of industrial structure upgrading based on provincial panel data and spatial Durbin model. *International Journal of Sustainable Development and Planning.*, 2020 15(8): 1197-1204
- [3] Zhu, Y. How International Trade Affects the Spatial Pattern of China's Economic Activities?. *American Journal of Industrial and Business Management.*, 2019 9(3): 627-646
- [4] Stephens, M. A. Collisions in the dark: Invisible intersectionality and the Black female psychoanalyst. *Contemporary Psychoanalysis.*, 2021 57(2): 165-197
- [5] Skillman, N. Lyric Reading Revisited: Passion, Address, and Form in Citizen. *American Literary History.*, 2019 31(3): 419-457
- [6] Scott, G. J., Petsakos, A., & Suarez, V. Not by bread alone: Estimating potato demand in India in 2030. *Potato Research.*, 2019 62(3): 281-304
- [7] Krivorotov, V. V., Kalina, A. V., Erypalov, S. E., & Kobekina, D. I. Dynamic Assessment of the Company's Competitiveness, Taking into Account the Implementation of Its Innovative Development Strategy. *of Applied Economic Research.*, 2020 19(4): 512-542
- [8] Aghili, A. Complete Solution For The Time Fractional Diffusion Problem With Mixed Boundary Conditions by Operational Method. *Applied Mathematics and Nonlinear Sciences.*, 2021 6(1): 9-20
- [9] Zhao, W., Shi, T. & Wang, L. Fault Diagnosis and Prognosis of Bearing Based on Hidden Markov Model with Multi-Features. *Applied Mathematics and Nonlinear Sciences.*, 2020 5(1): 71-84
- [10] Chenard, M. G., Martin, Y. R., Echarte, G. V., Pérez, N. N., & Pérez, A. L. V. Retrospective study of the frequency of anger presentation in the period 2000-2017 in Havana, Cuba. *Brazilian Journal of Health Review.*, 2019 2(2): 836-849