Research in Economics and Management
ISSN 2470-4407 (Print) ISSN 2470-4393 (Online)
Vol. 9, No. 1, 2024
www.scholink.org/ojs/index.php/rem

Original Paper

Spatial Agglomeration, Industrial Land Price and Enterprise Innovation: Evidence from Micro-data of Land Transactions in Industrial Enterprises

Zichen Lei[1]
[1] Zhongnan University of Economics and Law, Wuhan City, Hubei Province, China

Received: October 31, 2023 Accepted: November 22, 2023 Online Published: December 13, 2023
doi:10.22158/rem.v9n1p15 URL: http://dx.doi.org/10.22158/rem.v9n1p15

Abstract

It has become a consensus in academic circles that enterprise spatial agglomeration promotes technological innovation through comparative advantage of factor cost and knowledge spillover effect. With the increase of agglomeration degree, the advantage of low cost is gradually lost, and even under the constraint of financing, it will inhibit the innovation and development of enterprises. By constructing a new micro-level spatial agglomeration index, this paper investigates the dynamic impact of corporate spatial agglomeration on corporate innovation from the perspective of agglomeration cost. On the basis of using the longitude and latitude of industrial enterprises to construct a new index of agglomeration, this paper further carries out a comprehensive matching on the land market transaction data, patent application data and industrial enterprise data, obtains the micro-data at the enterprise level from 2007 to 2014, and constructs a non-linear intermediary effect model with industrial land price as the breakthrough point. The empirical results show that corporate spatial agglomeration has an "inverted U" effect on corporate innovation. The reason is that agglomeration costs such as increased competition caused by excessive spatial agglomeration and the rise in industrial land prices tighten corporate capital constraints, affect corporate resource allocation, and cause enterprises to reduce research and development investment. Research on heterogeneity shows that industrial enterprises have obvious preference characteristics for the direction of innovation investment in different stages of spatial agglomeration. According to this, the local government should scientifically control the agglomeration layout, on the one hand, give full play to the knowledge spillover effect of agglomeration, and on the other hand, mitigate the adverse impact of the rapid increase in land prices on innovation.

Keywords

spatial agglomeration, industrial land price, enterprise innovation, agglomeration cost

1. Introduction

For a long time, corporate spatial agglomeration has been expected to enhance the innovation capability by enhancing Marshall's externalities, but the actual situation is that only companies "cluster" without agglomeration effect (Zheng et al., 2008). The national "14th Five-Year Plan" and the long-term target for 2035 point out that the organization and management of industrial clusters and the promotion mechanism of specialization should be improved, an innovation and public service complex should be constructed, and a number of strategic emerging industry growth engines with unique features, complementary advantages and reasonable structure should be constructed. China's entry into the new development stage puts forward higher requirements for the spatial agglomeration layout of enterprises, which needs to change from simple scale agglomeration to innovation agglomeration.

Krugman (1991) and Krugman and Venables (1995) bring spatial factors into the traditional economic analysis framework, trying to explain the causes of economic growth and the distribution of economic activities from a new perspective. This positive externality brought by agglomeration, which promotes the economy, is called "agglomeration economy", and the existence of this effect has also been confirmed by many scholars (Sun et al., 2013; Wang X. & Wang Q., 2023). For innovation, corporate spatial agglomeration can effectively promote the enhancement of regional innovation capability through sharing, matching and learning mechanisms (Peng & Jiang, 2011). In the existing relevant research, most mainstream viewpoints agree that spatial agglomeration promotes the transmission path of innovation through externalities such as technological knowledge spillovers, which does exist in the agglomeration environment in reality. However, most of the researches pay attention to the agglomeration effect and unconsciously ignore the existence of agglomeration cost, and underestimate the impact of agglomeration cost on other economic objects. After the spatial agglomeration reaches a certain scale, the agglomeration cost will be generated with the entry of more and more enterprises, including environmental pollution, space and resource restrictions, congested transportation and storage facilities, etc. (Lin & Tan, 2019). Therefore, no matter which aspect of economic phenomenon is studied, the cost brought by agglomeration has developed to a point that cannot be ignored. Relevant research should include it in the scope of investigation. A few literatures have realized the importance of agglomeration cost, and have carried out research on the impact of spatial agglomeration formed by specialization and diversification on enterprise innovation. However, judging from the conclusion, this part of research has different effects on spatial agglomeration, and has drawn two opposite conclusions, positive and negative. From the perspective of process, this paper finds that it is precisely because of the different measurement indicators of spatial agglomeration used by these studies and the neglect of agglomeration cost in the transmission mechanism of spatial agglomeration impact that the conclusions of the impact of spatial agglomeration on innovation activities are different. Then, can we obtain more accurate effect results from two aspects: spatial agglomeration index and influence transmission mechanism? This paper will give new ideas and methods.

In terms of spatial agglomeration indicators, most of the agglomeration indicators commonly used in the existing literature are constructed from the macro level, that is, to measure the agglomeration within the whole region on a larger scale. Common agglomeration indicators include spatial Gini coefficient, Herfindal index, EG index, entropy index, etc. Most of them stay at the provincial and municipal levels in the scope of research, and a few more microscopic literatures only touch the agglomeration at the county level. Although there are many factors, such as large workload of data statistics and limited means of information technology, the long-term use of macro-indicators is still difficult to reflect the true agglomeration in China, which is a vast country with more complicated economic forms. Another big problem with the use of these indicators is that most of the basic data that make up these indicators involve output value and employment, so for some regions where there are high-tech and labor-intensive enterprises, their measurement values will appear “more concentrated”, and it is also possible to strengthen the endogenous links with other economic indicators. From the perspective of economic intuition, it is obvious that measuring the degree of spatial agglomeration by the regional output value or the proportion of employment does not grasp the essence of “agglomeration”, i.e. the proximity of spatial location. In recent years, relevant scholars at home and abroad have gradually started to construct new spatial agglomeration indicators from the micro level (Duranton & Overman, 2005; Shao & Li, 2017). Inspired by these studies, this paper argues that latitude and longitude coordinates can be achieved not only through the conversion of text addresses in terms of data availability, but also as a more direct and objective information in an economic sense. Therefore, it attempts to construct a new spatial agglomeration index from the micro-individual level of enterprises based on latitude and longitude coordinates. The advantage of this approach is that, on the one hand, the research level can be analyzed in depth at the individual level of the enterprise, and the research is no longer limited to the overall perspective; On the other hand, it can separate the data source from labor force quantity, output value level and other indicators, reduce the influence of endogeneity, and help to obtain more objective and credible research conclusions.

In conclusion, based on the comprehensive data set obtained by matching China’s land transaction data, patent application data and industrial enterprise database from 2007 to 2014, taking industrial land price as the breakthrough point, this paper analyzes the impact of system spatial agglomeration degree on enterprise innovation at the individual level, and analyzes its impact mechanism and heterogeneity. The study found that spatial agglomeration has an “inverted U” effect on enterprise innovation. When the level of spatial agglomeration is low, agglomeration will promote technological innovation. When agglomeration reaches a certain level, the increase of agglomeration level will only expand the “crowding effect” and hinder technological innovation. In addition, this paper also found that the land transfer price plays an intermediary role in the nonlinear model, spatial agglomeration has a significant impact on the industrial land price, and squeeze out the research and development investment of the enterprise through the land cost, and consume the innovation environmental benefits of the enterprise. Research on heterogeneity shows that spatial agglomeration has obvious preference for different types of innovation.

It prefers substantive innovation when the degree of agglomeration is low and the competitive pressure is low, and it prefers product differentiation through design innovation when the degree of agglomeration is high and the competitive pressure is high.

2. Theoretical Analysis and Research Assumptions

2.1 Theoretical Analysis

Although spatial agglomeration can bring economic benefits such as scale expansion and market sharing, with the increase of agglomeration degree, it will also bring negative costs such as overcrowding and intensified competition to economic entities in the region. Fujita and Thisse (2013) calls these two positive and negative effects “agglomeration economy” and “agglomeration cost”, and calls the trade-off between them “fundamental trade off of spatial economy”. Similarly, there is a basic trade-off problem for innovation activities. On the one hand, the regional production network formed by enterprises in the process of agglomeration will increase the frequency of forward and backward contact and peer learning, and improve the product innovation and process innovation capabilities while enhancing cooperation among enterprises (Mitra, 2000); On the other hand, spatial agglomeration will also lead to higher prices of production factors, shortage of materials, deterioration of ecological environment, crowded public facilities, etc., which will have a negative impact on innovation activities (Fan & Shao, 2011). Therefore, the substantial impact of spatial agglomeration on enterprise innovation is a complex and comprehensive role, which needs to be judged by combining the positive and negative effects.

2.2 Research Assumptions

On the whole, the overall impact of spatial agglomeration on enterprise innovation depends on the comprehensive effect of the two, that is, in the early stage of the dynamic evolution process of spatial agglomeration, it can still play a positive role of agglomeration and help the development of innovation, and as the scale of agglomeration expands, it will create a burden on innovation at a certain point in time. Therefore, hypothesis 1 is proposed in this paper.

Hypothesis 1: there is an “inverted U” relationship between the impact of corporate spatial agglomeration on corporate innovation.

In recent years, the rise in land price has become an important phenomenon in China’s current economy (Yan & Sun, 2020). To a certain extent, the benefits generated by spatial agglomeration are realized through the comparative advantages of basic factor costs. Once land, as a limited resource, cannot be supplied at a sustainable low price, the cost advantages that used to be relied on among regions will gradually disappear, and may even become the disadvantages of the dynamic evolution of spatial resources in turn. In some agglomeration areas with good policies and market environment, the entry cost brought by land price competition will cause great pressure on enterprises, especially for small and medium-sized enterprises with weaker ability to cope with the increase in factor cost. In the internal decision-making environment of the enterprise, the consequence of the increase in land price is to increase the capital constraint, which in turn affects the enterprise to redistribute the expenditure on research and development. At the same time, land prices are closely related to real estate prices, and the rise in land prices is bound to affect housing prices. Under the temptation of high real estate returns, industrial enterprises will also have more incentives to allocate resources to non-main businesses than to new product development with higher risk of failure and uncertain return on investment, thus adversely affecting enterprise innovation. Therefore, this paper further proposes hypothesis 2.

Hypothesis 2: Spatial agglomeration of enterprises can inhibit enterprise innovation by increasing land cost.

3. Estimation Methods and Data

3.1 Estimation Methodology

The empirical model is designed as follows:

$$patent_{it} = \alpha_0 + \beta_1 agglo_{it} + \beta_2 [agglo_{it}]^2 + X'_{it} \cdot B + \theta_i + \eta_t + \varepsilon_{it} \tag{1}$$

In formula (1), the explained variable is the logarithm of the number of patents applied for by an enterprise; And represent the first enterprise and year respectively; The core explanatory variable is the degree of agglomeration of industrial enterprises; For other control variables, see Table 1 for specific variables and their calculation methods; It is firm fixed effect, year fixed effect and random disturbance term. $patent_{it}$ is logarithm of the number of patents applied for for an enterprise ； i and t respectively represent the i -th year t ; $agglo_{it}$ is X_{it} and is other control variables. For specific variables and their calculation methods, see Table θ_i for enterprise fixed effects, η_t for year fixed effects, and ε_{it} for random disturbance terms.

3.2 Variable Selection

3.2.1 Interpreted Variables

The measurement of corporate innovation has always been a controversial topic. The existing literature mainly measures corporate innovation from two dimensions: research and development input and research and development output. From the perspective of research and development investment, some literatures use research and development expenditure and the number of research and development personnel to measure innovation ability. As the R&D expenditure only represents the enterprise’s emphasis on innovation and the degree of capital surplus, and does not directly reflect the conversion effect of R&D, this approach has certain defects. From the perspective of research and development output, some literatures use the output value of new products to overcome the shortage of research and development investment to a certain extent, but this paper still does not use it. The specific reasons are as follows: (1) The output value of new products cannot reflect the innovative outputs such as trademark right, copyright, improvement of production process or technology. (2) The National Bureau of Statistics has increased the “activity space” for enterprises to report the output value of new products as defined in the First Economic Census Plan, which may expand the innovative achievements. (3) In terms of data availability, the China Industrial Enterprise Database has been missing the relevant data on the output value of new products since 2009, which overlaps too little with the scope selected in this paper, thus affecting the reliability of the results.

Based on the above considerations, this paper refers to Hall and Harhoff (2012), Li Wenjing and Zheng Manni (2016), and chooses to measure the enterprise’s innovation capability by the number of patent applications from the output perspective. There is a certain threshold for an enterprise to apply for a patent, which can partially reflect the results of the enterprise’s early research and development investment and the conversion success rate. Therefore, this paper refers to the practice of most relevant literature, using the number of corporate patent applications as the core explained variable.

Core explanatory variables

In the choice of core explanatory variables of spatial agglomeration, the most common indicators used by domestic economists to measure the degree of spatial agglomeration are:

① Gini coefficient or spatial Gini coefficient, such as Chen Jianjun (2009);

② Herfindal index, such as Ji Shuhan (2016);

③ EG index, e.g. Tan Hongbo (2013).

Although these indicators have long been recognized by relevant academic research, this does not mean that it is entirely appropriate to use these indicators to measure the degree of agglomeration.

Specifically:

④ From the point of data sources, whether the number of workers or the number of output value is used, it naturally determines that the agglomeration index cannot be independent of the degree of economic development, thus losing the exogeneity that it should have in the research process;

⑤ From the results, the final calculation value of such indexes is still the degree of dispersion of economic indicators within a certain range, which is relatively low in relation to specific geographical locations and does not reflect the essence of “agglomeration”;

⑥ Other relevant factors, such as the production characteristics of labor-intensive industries at the industry level and the productivity differences brought about by the technological level at the enterprise level, will cause such indexes to have large deviation in comparison. Based on the above considerations, none of these indicators can perfectly reflect the meaning of the word “spatial agglomeration”. It is necessary to jump out of the traditional thinking of simply considering the accumulation of economic factors and look for other ideas.

In the research, Shao Yihang and Li Zeyang (2017) used the longitude and latitude of industrial enterprises to construct the spatial agglomeration index, which has a better improvement over the problems existing in the above indexes, reducing the possibility of endogeneity, and more reflecting the essence of agglomeration. The specific idea is to use the precise longitude and latitude information of industrial enterprises to calculate the coefficient of variation of longitude and latitude of enterprises in each city, and the product of the two coefficients of variation is the spatial agglomeration index.

Although this method is quite pioneering in the long-term exploration of the construction of agglomeration indicators, and avoids the inherent problems of traditional spatial agglomeration indicators, it still has certain deficiencies from the perspective of rigour and sufficiency. First, the aggregation index artificially separates the longitude and latitude of the same enterprise and calculates the coefficient of variation, which makes the element of geographical location lose its original integrity. For example, if the enterprises in the region have one-way concentration in longitude and are relatively dispersed in latitude, the product of the coefficient of variation of the two based on the index is still large in value, but obviously this does not reflect the concentration of spatial economic activities from the original intention of spatial concentration. Second, although this method uses micro-level industrial enterprise data, it still chooses to set agglomeration variables from the urban level when constructing the measurement model, which is based on its research direction and model setting, but this paper believes that this agglomeration index cannot expand the research scope of spatial agglomeration to the micro level, and inevitably has certain limitations.

Therefore, this paper uses the experience of the above-mentioned documents for reference, on the one hand, it continues the idea of calculating spatial agglomeration using latitude and longitude information, on the other hand, it optimizes the design method, makes full use of the data characteristics, and tries to construct a new micro-level spatial agglomeration index. The specific method is as follows: after obtaining the longitude and latitude data of each industrial enterprise, calculate the average longitude ($longitude_{mean}$) and average latitude ($latitude_{mean}$) of each industrial enterprise respectively in the same city and the same year, and then from each enterprise individual in the data set required for empirical research Starting from this point, calculate the spatial distance ($distance$) between the virtual coordinate point ($longitude_{mean}, latitude_{mean}$) composed of the average longitude and the average latitude respectively. In addition, due to the vast land area of our country, if spatial distance is treated equally, there will be problems in the measurement dimension. Therefore, this article calculates the average spatial distance of each enterprise ($distance_{mean}$), and finally uses the above distance integration to construct the core index of the full text - The degree of spatial agglomeration, that is:

$$agglo_{it} = -ln\left(distance / distance_{mean} + 1\right) \tag{2}$$

Mediation variables

As mentioned above, spatial agglomeration has an impact on economic benefits through the cost of basic elements of land. For industrial enterprises, the agglomeration of other surrounding enterprises will bring about a re-decision assessment of production land. The assessment results will also create a new trade-off on the input proportion of different production elements. Therefore, this paper focuses on the research on the relationship between the input cost of industrial land in the face of changes in the surrounding environment and the enterprise’s own research and innovation, and takes the unit price (ten thousand yuan/hectare) of each industrial land purchased by the enterprise as an intermediate variable, i.e. the land transaction price/land area. In terms of land use, this paper only examines enterprises that have purchased industrial land. Commercial and residential land and land with other uses of public utility nature are excluded from the scope of this paper because they do not conform to the research logic.

3.3 Data Sources and Processing

3.3.1 Data Sources

The sample data used in the research process of this paper are mainly from the following four sources:

The first is the “China Industrial Enterprise Database” of the National Bureau of Statistics, which covers all state-owned industrial enterprises and non-state-owned enterprises above designated size. Specific data include, for example, address, basic operating information, assets, liabilities, equity, inputs, outputs, etc.

The second is the data on patents, which are derived from the patent database of China National Intellectual Property Administration, including all the micro patent data since the implementation of the Patent Law in 1985, specifically including the number of patent applications in three categories of invention patents, utility model patents and appearance patents, and relevant data such as applicants.

The third is the land parcel transaction data collected from China Land Market Network (http://www.landchina.com/) for each year. The “Standard for the Transfer of State-owned Land Use Rights by Bidding, Auction and Listing (for Trial Implementation)” requires local land authorities to publish the transfer information and transaction results of each state-owned land parcel on the website, including the area, land use, location and specific address of the parcel, transaction amount, name of the purchaser and other information of each parcel.

The fourth is the level of prefecture-level city level data. The macro-indicators at the city level used in this paper are mainly from China City Statistical Yearbook and China Statistical Yearbook. The two yearbooks contain data on city-level control variables such as population, GDP and industrial proportion of each city and the price index required for price revision.

Considering the availability of data, the time span selected in this paper is 2007-2014. It is worth mentioning that there are many data quality problems such as extreme values in the financial data of 2010 in the industrial enterprise database. Although the median method can be technically used for certain repairs (Wang Guidong, 2017), this paper still chooses to eliminate the data of that year for research due to the consideration of stringency.

3.3.2 Data Processing

A. industrial enterprise database matching and cleaning

This paper refers to the sequential matching method of Brandt et al. (2012) to match the industrial enterprise data, i.e. identify the same enterprise one by one based on the same organization code, enterprise name, legal representative name, telephone number, administrative division code, main business, etc. The database contains information such as missing data and abnormal data, so the original data is cleaned up according to the methods of Nie Huihua et al. (2012) and Brandt et al. (2012).

B. calculation of longitude and latitude of the enterprise

The background database of Baidu Map API (Geocoding API) is connected with the information such as administrative divisions, streets and names of each enterprise in the previous step, and each enterprise is geocoded to obtain its latitude and longitude. Therefore, the core explanatory variables of this paper are obtained by using the aggregation index construction method mentioned above.

C. land market network data cleansing

Due to the large amount of information that needs to be filled in manually and the inevitable errors in web page grabbing, the data of the land market network needs to be cleaned and screened on some indicators. The specific process is as follows: only the samples of corporate enterprises with land use rights are retained, the missing data of the users are deleted, and the samples purchased by individuals (based on the length of characters), collectives (such as committees, task forces, cooperatives, etc.), government agencies (such as offices, acquisition reserve centers, seed management stations, etc.) and other institutions (such as reservoirs, farms, etc.) are deleted; Deleted samples with a transaction area of 100 square meters (0.01 hectare) or less or with a price less than RMB50,000; Only samples of land use for purchase of industrial land are retained.

Before 2007, the land sold in our country did not require the issuance of bidding, auction or listing announcement in the land tangible market or designated places and media. In September 2007, the Ministry of Land and Resources required all industrial and commercial land to be sold by way of “bidding, auction and hanging”, which means that the data of land sales after 2007 will be gradually improved and standardized. At the same time, this paper finds that there were only 1,733 land purchase enterprises from 2000 to 2006, far less than the number after 2007, by sorting out the data of China land market network. Therefore, this paper comprehensively determines the research interval based on the database of industrial enterprises.

Table 1. Land Purchase Enterprises, 2000-2014

year	2000—2006	2007	2008	2009	2010	2011	2012	2013	2014	all
Number of enterprises purchasing land	1,733	4,619	3,833	4,747	10,082	9,137	9,341	9,711	6,033	59,236
Percentage by year	2.93%	7.80%	6.47%	8.01%	17.02%	15.42%	15.77%	16.39%	10.18%	100%
Percentage of companies purchasing land	0.11%	1.37%	0.93%	1.42%	2.28%	3.04%	2.89%	2.83%	1.95%	2.11%

Sources: Land Market Network, China Urban Statistical Yearbook and Industrial Enterprise Database.

D. Matching industrial enterprise data with land transfer data

One of the major challenges in the data processing work in this paper is to match the land transfer with the corresponding industrial enterprises in many-to-one way to realize cross-database connection. In the existing researches on how to achieve this work, most of them simply give matching ideas without analyzing specific methods (Zhang et al., 2019; Yan & Sun, 2020; Li, 2020). The matching idea of this paper is that firstly, the two databases are matched exactly according to the enterprise name and year, and the result of the exact match is exported. Because the same enterprise may change its name in different years of operation, and at the same time, due to the inherent expression and writing habits in Chinese, errors such as typos, homophones and similar characters may occur in the registration, the same enterprise may have different names in the two databases, which means that an enterprise that does not have an exact match may have purchased land. Therefore, only using samples on an exact match will result in a certain degree of omission.

In order to make the sample selection more accurate, this paper performs fuzzy matching on the remaining data, and the processing process is roughly as follows:

(1) firstly, preprocessing the enterprise name, and sequentially removing various abnormal symbols in the original name of the enterprise; The English letters shall be in uppercase and half-angle format, and the uppercase numerals shall be replaced by Arabic numerals; In order to unify the Chinese and English symbols of enterprise names, all Chinese punctuation marks and numbers are converted into English symbols using ASCII coding; Unify different kinds of symbols caused by different writing habits, such as quotation marks, dashes, dots, etc.; Remove redundant symbols at the end of the name, such as comma, brackets, etc. After completing all the pretreatment process, the cleaned enterprise name will be obtained.

(2) Considering that some enterprises often change the relevant words of “company”, the following words at the end of the enterprise name are removed in turn: limited, liability, shares, company, group, head office, branch and factory, from which “key enterprise name” is obtained.

(3) Through “key enterprise name” and year, all industrial enterprises are fuzzy-matched with the remaining plot data after accurate matching year by year, the reclink2 command is used and the accurate value of 0.95 is set, and finally the matching results with problems are screened out through manual inspection, and the remaining samples are combined with the precisely matched data.

(4) combining the statistical yearbook data with the matched database. From the “China City Statistical Yearbook” to sort out the relevant macro-data of the city, and adjust the names of the cities of each caliber to keep them consistent. By combining the city level data with the data set by city and year, the matching of enterprise information, land parcel information, patent information and city information is completed, and the comprehensive data set is taken as the basic data for this study.

4. Basic Regression Results and Robustness Test

4.1 Basic Regression

Table 2. Regression To Basics

	Total patent applications		
	（1）	（2）	（3）
$agglo_{it}$	0.425***	0.426***	0.400***
	（3.65）	（3.59）	（3.38）
$[agglo_{it}]^2$	-0.130***	-0.130***	-0.131***
	（-3.53）	（-3.56）	（-3.60）
debt			0.025
			（0.49）
profit			-0.101
			（-1.45）
indval			0.091***
			（6.87）
scale			-0.060***
			（-4.73）
landrank			-0.002
			（-1.61）
pgdp			-1.068***
			（-7.69）
sndratio			0.021***
			（5.56）
open			0.030*
			（1.88）
med			0.372***
			（4.26）
tran			0.035
			（0.75）
edu			0.116
			（1.58）
inf			0.070***
			（2.64）

time-fixed effect	no	yes	yes
individual fixed	yes	yes	yes
constant term	0.240***	-0.110	3.988***
	(3.45)	(-1.18)	(2.87)
observations	59631	59631	59631
R-Square	0.942	0.945	0.946
Adj.R-Square	0.75	0.76	0.76

Note. Robust standard errors in parentheses; *** is p<0.01, ** is p<0.05, * is p<0.1, same table below.

Source: Calculated by the author using Stata software based on the Land Market Network, the State Intellectual Property Office, the China Urban Statistical Yearbook and the database of industrial enterprises, the following table is the same.

Based on the above basic regression model, this paper first estimates the impact of industrial enterprise spatial agglomeration on enterprise innovation. Table 2 reports the basic estimation results of the fixed effect model. Among them, column (1) only considers the individual fixed effect of the enterprise, column (2) adds the time fixed effect, and column (3) adds the effect on the set of control variables. The results of column (1)-(3) show that, regardless of whether the time-fixed effect and the set of control variables are considered, the degree of agglomeration and the coefficient of its quadratic term are significantly not 0 at the significance level of 1%, with the former coefficient being positive and the latter coefficient being negative, i.e. the degree of agglomeration has an "inverted U" effect on the number of patents. When the degree of agglomeration is low, the strengthening of spatial agglomeration will increase the number of patents filed by enterprises, while when the degree of agglomeration is high, the strengthening of spatial agglomeration will reduce the number of patents filed by enterprises, which is in line with the assumption 1 made above.

4.2 Robustness Test

The following robustness tests are used to test the reliability of firm conclusions.

4.2.1 Reverse the Original Time Point

As enterprises consider the factors of location selection, the innovation environment in the region will also become an important condition for enterprises to enter. This means that the number of corporate patent applications is large, which often indicates that they have stronger innovation ability or are located in regions with strong innovation atmosphere, which will attract more new enterprises to enter, which may promote the degree of spatial agglomeration in the region, i.e. the core explanatory variable in this paper may have the endogenous problem of reverse causality. In order to deal with the degree of spatial agglomeration of enterprises and the endogeneity of patent applications, this paper tests the reverse causal relationship with reference to Aghion et al. (2016) through the variable timing of variables: the independent variables are current values, the geographical location of enterprises has been determined, the current timing of spatial agglomeration index is between the dependent variable current period and the lag phase, and the dependent variable and the control variable are lag phase, so as to test whether the future spatial agglomeration changes can predict the current innovation changes of enterprises. And the test results are shown in Table 3. It can be seen that the spatial agglomeration index and the coefficient of its square term are not significant. Therefore, it can be concluded that there is no reverse causal relationship between the variables.

Table 3. Results of Reverse Causality Test

	Total number of patent applications (lagged by one period)		
	（1）	（2）	（3）
$agglo_{it}$	0.781	0.809	1.252
	(1.06)	(1.28)	(1.60)
$[agglo_{it}]^2$	-0.279	-0.122	-0.320
	(-0.97)	(-0.47)	(-0.99)
debt			-0.758***
（one stage behind）			(-4.02)
profit			-0.139
（one stage behind）			(-0.87)
indval			0.039
（one stage behind）			(0.76)
scale			-0.006
（one stage behind）			(-1.05)
landrank			-1.374**
（one stage behind）			(-2.10)
pgdp			0.070***
（one stage behind）			(3.89)
sndratio			0.040
（one stage behind）			(0.51)
open			-0.719**
（one stage behind）			(-2.21)
med			-0.235
（one stage behind）			(-1.13)
tran			0.006
（one stage behind）			(0.02)
edu			0.297***

（one stage behind）			(2.81)
inf			-0.245***
（one stage behind）			(-4.62)
observations	4123	4123	3954
R-Square	0.969	0.970	0.969
Adj.R-Square	0.91	0.92	0.91

4.2.2 Replace the Core Explanatory Variable

First of all, as most of the current researches still choose to study the innovation capability from the city level, this paper decides to test the above assumptions again on the urban agglomeration and urban innovation, and obtains the empirical results by changing the explanatory variables and the explained variables. The degree of spatial agglomeration at the city level This paper first selects the index construction method used by Shao Yihang and Li Zeyang (2017) mentioned above, which uses the product of longitude and latitude variation coefficients of each enterprise in the city as one of the choices of urban spatial agglomeration index. The specific formula is: $agglo\_city1$，As one of the choices for urban spatial agglomeration indicators, the specific formula is:

$$agglo\_city1 = -\ln(CV_{longitude} \cdot CV_{latitude}) \tag{3}$$

In equation (3), and are the coefficients of variation of longitude and latitude of all industrial enterprises in the city respectively; $CV_{longitude}$ and $CV_{latitude}$ are the coefficients of variation of the longitude and latitude of all industrial enterprises in the city respectively;

Secondly, this paper uses location entropy to measure the degree of agglomeration based on the practice of most mainstream literature in the agglomeration field (Yang, 2013). Considering the special household registration system in China, the statistical data of the regional employment population cannot truly reflect the local employment population, so this paper uses the total industrial output value to calculate the entropy index, and the calculation formula of the location entropy of the regional industry is:The calculation formula of location $agglo\_city2$ for industry r in region i is:

$$agglo\_city2 = (e_{ir} / \sum_i e_{ir}) / (\sum_i e_{ir} / \sum_i \sum_r e_{ir}) \tag{4}$$

In formula (4), which represents the gross industrial output value of the regional industry, this paper mainly studies the spatial agglomeration of manufacturing industry based on the idea of benchmark regression. In the alternative index selection of the explained variables, this paper still chooses to use the output innovation data to represent the regional innovation capability. Continue the basic regression indicator selection thinking, using the sum of the number of urban inventions, utility models and design patents granted. As shown in Table 4, after regression using the replaced core explanatory variables and the explained variables respectively, all results are similar to the structure of the basic results, in which the estimation coefficients of the primary term and the secondary term are significantly negative at the statistical level of 1%. In equation (4), e represents the total industrial output value of industry r in region i . Based on the idea of benchmark regression, this article mainly studies the spatial agglomeration of manufacturing industry.

Table 4. Replacement Indicator Test Results

	Number of city patents	
	(1)	(2)
$agglo\_city1$	1.086***	
	(5.271)	
$[agglo\_city1]^2$	-0.037***	
	(-4.236)	
$agglo\_city2$		0.939***
		(8.778)
$[agglo\_city2]^2$		-0.100***
		(-5.468)
$pgdp$	0.505***	0.538***
	(8.776)	(9.238)
$sndratio$	-0.005***	-0.009***
	(-2.708)	(-3.935)
$open$	0.139***	0.144***
	(8.915)	(8.975)
med	0.563***	0.429***
	(11.567)	(9.049)
$tran$	0.210***	0.116***
	(5.364)	(2.920)
edu	0.021	-0.026
	(0.421)	(-0.510)
inf	0.712***	0.799***
	(16.969)	(18.971)
observations	1832	1830
R-Square	0.829	0.820
Adj.R-Square	0.83	0.82

This indicates that the robustness of the basic regression is not affected whether from the individual level or the city level, regardless of whether to change the calculation of the explanatory variables and the explained variables. Hypothesis 1 is thus established: the degree of spatial agglomeration of industrial enterprises has an "inverted U" effect on enterprise innovation, that is, when the level of spatial agglomeration is low, agglomeration will promote technological innovation, and when the agglomeration reaches a certain level, the increase of agglomeration level will only expand the "crowding effect" and hinder technological innovation.

4.3 Heterogeneity Analysis

Considering that there is a clear division of patent types in China, which is mainly divided into invention patents, utility model patents and appearance patents, this paper will analyze the heterogeneity from the explanatory variables. Among them, the invention patents and utility model patents protect the technical solutions, the number of which can reflect the technological innovation ability of the enterprise, while the appearance patents are innovations in the product appearance, and do not reflect the innovation level. In addition, the invention patents have great difficulty and high technical content, while utility model patents pay more attention to practicality and have low technical difficulty, and are easier to be converted into practical value (Li & Zheng, 2016).

The estimation results show that spatial agglomeration plays a greater role in promoting enterprises to apply for utility model patents. Columns (1) - (3) of Table 5 report the estimation results of the number of applications for invention patents, utility model patents and design patents of enterprises for explanatory variables which are spatial agglomeration. First of all, the first term of invention patents and utility model patents are significantly positive and the second term is significantly negative, which is consistent with the basic regression results, indicating that with the increase of spatial agglomeration, the innovation level of enterprises will show a trend of "first increase and then decrease". Secondly, the significance level and coefficient of invention patents are lower than those of utility model patents, which can show that enterprises choose to allocate resources to utility model patents with less investment and easier transformation, instead of invention patents with higher difficulty and higher risk when facing the intensified competition and land cost squeeze. Finally, the significance level of design patents is lower than the first two, and the first term is significantly negative and the second term is significantly positive, which indicates that the change of spatial agglomeration level has little impact on design patents, and reflects that in the initial stage when enterprises did not gather in large quantities, each enterprise did not invest too much resources in the external design of products, and with more and more enterprises pouring in and intensifying competition, enterprises gradually attach importance to the refined differentiation of products and increase the proportion of research and development of design.

Table 5. Heterogeneity Analysis

	Patents for inventions	Utility model patents	Design Patents
	（1）	（2）	（3）
$agglo_{it}$	0.206**	0.457***	-0.119*
	（2.42）	（4.54）	（-1.74）
$[agglo_{it}]^2$	-0.085***	-0.136***	0.043**
	（-3.22）	（-4.38）	（2.05）
debt	0.135***	0.093***	0.017
	（4.65）	（2.71）	（0.71）
profit	-0.072**	-0.053	-0.069***
	（-2.39）	（-1.47）	（-2.84）
indval	0.036***	0.058***	0.039***
	（3.72）	（5.12）	（5.10）
scale	-0.055***	-0.057***	-0.003
	（-6.00）	（-5.28）	（-0.41）
landrank	-0.000	-0.000	-0.001
	（-0.37）	（-0.13）	（-1.31）
pgdp	-0.789***	-1.185***	-0.236***
	（-7.88）	（-10.03）	（-2.95）
sndratio	0.020***	0.025***	0.002
	（7.29）	（7.70）	（0.92）
open	0.020*	0.011	0.025***
	（1.72）	（0.84）	（2.74）
med	0.022	0.312***	0.137***
	（0.34）	（4.18）	（2.71）
tran	0.060*	-0.006	-0.040
	（1.79）	（-0.14）	（-1.50）
edu	0.076	0.083	0.017
	（1.44）	（1.33）	（0.40）
inf	0.058***	0.033	0.052***
	（3.05）	（1.45）	（3.41）
observations	59631	59631	59631
R-Square	0.937	0.939	0.919
Adj.R-Square	0.72	0.73	0.65

5. Conclusion

This paper attempts to construct a new micro-level spatial agglomeration index, and tests the impact of corporate spatial agglomeration on innovation activities from micro-individual level by matching with land transaction data and patent application data. The analysis shows that the following main empirical conclusions of this paper have certain robustness: (1) there is a nonlinear relationship between spatial agglomeration of enterprises and patent applications, which increases first and then decreases; (2) Land price plays an intermediary role in the impact of corporate spatial agglomeration on patent applications; (3) Spatial agglomeration has an "inverted U" effect on invention patents and utility model patents, and a "positive U" effect on design patents. The research shows that the spatial agglomeration of manufacturing industry in China has an "inverted U" effect on enterprise innovation by increasing the land cost. At the same time, the preference of enterprises for innovation is heterogeneous in different stages of spatial agglomeration. It pays attention to substantive technological innovation in growth period and non-functional attribute innovation in maturity period.

Based on the above analysis, as China's manufacturing industry plays an important role in the economic level, efforts to reduce the above-mentioned negative impact that is not conducive to spatial agglomeration and to play the incentive effect of spatial agglomeration on innovation and growth of manufacturing industry should be an important direction for local governments to focus on development. The analysis results of this paper may contain the following policy recommendations: (1) Local government should take the lead to build a platform for technology exchange and cooperation among enterprises in the region, encourage intra-industry and inter-industry exchanges in the manufacturing industry, and enhance the spillover of technological knowledge to amplify the positive effect of spatial agglomeration; (2) Local governments should mitigate the rapid increase in land prices and reduce the impact of land costs on innovation as much as possible, ensure that enterprises have access to sufficient capital flows for research and development expenditures, at the same time avoid the crowding-out effect of high land prices and high house prices on normal investment in the manufacturing industry, and increase policies such as infrastructure and public service supply to support the use of the manufacturing industry to effectively control "agglomeration costs"; (3) The local government can guide the scientific layout of the industry as a whole. On the one hand, it should not only prevent some regions from losing the potential positive impact on innovation because they cannot reach the corresponding concentration level, but also avoid the negative impact on enterprise innovation caused by excessive concentration in some regions. On the other hand, it should reasonably encourage enterprises to make flexible choices in the allocation of innovation resources, which is of vital importance to the long-term development of cities and enterprises.

References

Aghion, P., Akcigit, U., Cage, J., & Kerr, W. R. (2016). Taxation, Corruption, and Growth. *European Economic Review*, *86*, 24-51. https://doi.org/10.1016/j.euroecorev.2016.01.012

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, *97*(2), 339-351. https://doi.org/10.1016/j.jdeveco.2011.02.002

Bruilhart, M., & Mathys, N. A. (2008). Sectoral Agglomeration Economies in a Panel of European Regions. *Regional Science and Urban Economics*, *38*(4), 348-362. https://doi.org/10.1016/j.regsciurbeco.2008.03.003

Chen, J. J., Chen, G. L., & Huang, J. (2009). Research on Producer Services Agglomeration and Its Influencing Factors from the Perspective of New Economic Geography: Empirical Evidence from 222 Cities in China. *Management World*, *2009*(4), 83-95.

Combes, P. P., Mayer, T., & Thisse, J. F. (2008). *Economic Geography: The Integration of Regions and Nations*. New Jersey: Princeton University Press. https://doi.org/10.2307/j.ctvcm4h9k

Combes, P.-P., Duranton, G., & Gobillon, L. (2019). The Costs of Agglomeration: House and Land Prices in French Cities. *The Review of Economic Studies*, *86*(4). https://doi.org/10.1093/restud/rdy063

Duranton, G., & Overman, H. G. (2002). *Testing for Localization Using Micro-Geographic Data*. CEPR Discussion Papers.

Fan, J. Y., & Shao, T. (2011). Housing price level, location distribution of differentiated products and urban system. *Economic Research*, *46*(2), 87-99.

Fujita, M., & Thisse, J. F. (2013). *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139051552

Fujita, M., Krugman, P. R., &. Venables, A. (2001). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge: MIT press.

Han, F., & Zhuang, Z. W. (2021). Li kai hang. how does the distortion of land market affect the upgrading of the quality of manufacturing export products-an empirical analysis based on land market transaction data and manufacturing enterprise data. *Economic Theory and Economic Management*, *41*(3), 68-83.

Ji, S. H., Zhu, Y. M., & Zhang, X. (2016). Research on the Improvement Effect of Industrial Agglomeration on Resource Mismatch. *China Industrial Economy*, *2016*(6), 73-90.

Kou, Z. L., & Liu, X. Y. (2020). Patent Behavior of Enterprises in China: Characteristic Facts and Impact from Innovation Policy. *Economic Research*, *55*(3), 83-99.

Krugman, P. (1991). *Geography and Trade*. MIT Press.

Krugman, P., & Venables, A. J. (1995). *The Seamless World: A Spatial Model of International Specialization*. NBER working paper. https://doi.org/10.3386/w5220

Li, B., Yang, X. M., & Yang, M. Y. (2020). Does Land Purchase Affect Enterprise Innovation? —Evidence from industrial enterprises in China. *Economic Management*, *42*(5), 37-53.

Li, W. J., & Zheng, M. N. (2016). Substantive innovation or strategic innovation? —The impact of macro-industrial policies on micro-enterprise innovation. *Economic Research*, *51*(4), 60-73.

Mackinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, *58*, 593-614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mitra, J. (2000). Making connections: Innovation and collective learning in small businesses. *Education + Training*, *42*(4/5), 228-237. https://doi.org/10.1108/00400910010373679