

Major factors affecting cross-city R&D collaborations in China: evidence from cross-sectional co-patent data between 224 cities

Sheng-qiang Jiang^{1,2} · An-na Shi¹ · Zhi-hang Peng³ ·
Xin Li^{4,5}

Received: 6 January 2016 / Published online: 1 April 2017
© Akadémiai Kiadó, Budapest, Hungary 2017

Abstract Scientific research activities cluster in cities or towns. Modern cities can play a crucial role in the national or regional innovation system. Strengthening R&D collaboration between cities can contribute to perfectly integrating various regional innovation systems. Using the cross-sectional co-patent data of the Chinese Patent Database as a proxy for R&D collaboration, this paper investigates the spatial patterns of R&D collaborations between 224 Chinese cities and the major factors that affect cross-city R&D collaborations in China. A spatial interaction model was used to examine how spatial, economic, technological and political factors affect cross-city R&D collaborations. The degree of centrality shows that cross-city collaborative R&D activities mainly occur in favored regions, advanced municipalities and coastal regions. The mean collaboration intensity for intra-provincial cross-city collaborations is 4.74; however, for inter-provincial collaborations, it is 0.69. The econometric findings reveal that spatial, economic, technological and political bias factors do yield significant influences on the frequency of cross-city R&D

Electronic supplementary material The online version of this article (doi:[10.1007/s11192-017-2358-2](https://doi.org/10.1007/s11192-017-2358-2)) contains supplementary material, which is available to authorized users.

✉ An-na Shi
anshi@hhu.edu.cn

✉ Xin Li
xinli@njmu.edu.cn

¹ Department of Technology Economy and Management, School of Business, Hohai University, Nanjing, China

² Division of Science and Technology, Nanjing Medical University, Nanjing, China

³ Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China

⁴ Department of Health Policy, School of Health Policy and Management, Nanjing Medical University, Nanjing, China

⁵ Department of Clinical Pharmacy, School of Pharmacy, Nanjing Medical University, 101 Longmian Avenue, Nanjing 211166, China

collaboration. Specifically, as evidenced by the model coefficient, it is more likely that R&D collaborations occur among cities that are connected by high-speed railways.

Keywords Cross-city · R&D collaboration · Affecting factors · Spatial interaction model · Co-patent · China

Introduction

The importance of research and development (R&D) collaborations is widely recognized (Chen and Kenney 2007; Gao et al. 2011). To obtain faster access to new knowledge and to learn how to integrate it into research processes or existing products, firms intend to develop various types of collaborative relations with other firms, universities, research institutes that have already this knowledge (Scherngell and Barber 2011). Strategic R&D alliances among firms, universities and research institutes are often described as “engines for growth” that generate skills and research results that are significant sources of innovation for industrial fields (Chen and Kenney 2007; Gao et al. 2011). R&D collaborations are increasingly viewed as vehicles for technology transfer and as a conduit through which knowledge exchange is made more effective in China. In an attempt to foster a national innovation system, the Chinese government has issued a series of policies and measures that point to the crucial importance of inter-regional R&D collaboration (Scherngell and Hu 2011). For instance, the program for enhancing the capacity for innovation of universities (also known as “2011 project”) that was issued by the State Council of China in 2012 emphasizes that a cross-city collaborative innovation system between universities and firms should be established.

Inter-regional R&D collaboration is an essential part of the national or regional innovation system in Western countries; this has been widely touted as a potent means to enhance the flow of new knowledge, modify the spatial diffusion of knowledge and achieve sustained competitive advantage (Jaffe et al. 1993). A vast body of literature has shown that it is more or less easy to establish inter-regional R&D collaboration networks if the scientific, cultural, political, and geographical barriers have been overcome (Katz 1994; Liang and Zhu 2002; Ponds et al. 2007; Hoekman et al. 2009; Fischer et al. 2006; Pan et al. 2012). Recently, a number of empirical studies have been reported regarding scientific collaboration patterns in China by mapping the geographic network (Gao et al. 2011; Scherngell and Barber 2011; Liang and Zhu 2002; Hong 2008; Wang et al. 2005; Henemann et al. 2011; Ma et al. 2014). However, few studies have been conducted that quantitatively analyze the factors that affect inter-regional collaborative R&D activities (Scherngell and Hu 2011; Liang and Zhu 2002; Hong 2008). For instance, using co-authored data as a proxy for collaborative knowledge production, Scherngell and Hu estimated the effects of geographical, technological, and economic factors on the frequencies of cross-provincial collaboration activities in China (Scherngell and Hu 2011). Their findings provide *prima facie* evidence regarding the type of factors that could impede research collaborations at the regional level. Nevertheless, most studies take a provincial perspective. At the city level, we know of only one study, conducted by Ma et al. (2014), which explored the relationship between inter-city geographical distance and scientific collaboration based on co-authored papers between 58 cities. However, the other major factors that affect inter-city scientific cooperation such as culture, politics, traffic, communication, technology and economics have not been investigated.

Usually, R&D activities tend to be spatially concentrated in large cities, and the scientific research clusters in cities or towns (Liefner and Hennemann 2011). As a result of decreased traffic and communication costs, the cities tend to reinforce their important roles as coordination centers of spatially dispersed activities (Sassen 1991; Florida 2002). Thus, to improve the cities' influence within the national and regional innovation system, science policy should consider how to promote R&D collaborations at the city level. Therefore, this paper attempts to answer the following questions: What are the spatial features of the cross-city R&D collaborations networks in China? What are major dynamic and obstructing factors that affect the R&D collaborations between cities in China? To what degree do the major factors impact cross-city R&D collaborations? The answers to these questions could help us provide policy-makers with detailed and valuable guidance when designing China's national and regional innovation systems.

The major differences between this study and previous studies are as follows: First, we use urban regions as units of analysis, in contrast to most previous studies that use provinces as the observational units. Compared with province-level data, we can obtain more accurate estimates of different separation effects on the spatial structure of cross-city R&D collaborations. This study's objective is to examine the effects of spatial distance between the geographical centers of cities rather than the capitals of provinces on the inter-regional R&D collaborations. Given that the area of provinces is much larger than that of cities, it is beneficial to acquire more precise results by measuring the geographical distance between cities. Furthermore, by analyzing the network between the cities in the same or different provinces, we can discover whether the existence of regional protectionism exerts a negative impact on the cross-city R&D collaborations. Second, in contrast to the previous studies (Scherngell and Hu 2011; Liang and Zhu 2002; Ma et al. 2014), we use as a research hypothesis that the patents serve as a proxy for inter-regional R&D collaborations in China. The main research hypotheses are as follows:

First, a number of studies have shown that research collaborations are supposedly made easier by geographical proximity (Scherngell and Hu 2011; Jaffe et al. 1993; Katz 1994; Liang and Zhu 2002; Ponds et al. 2007; Hoekman et al. 2009; Fischer et al. 2006; Pan et al. 2012). Thus, we hypothesize that the frequency of cross-city R&D collaboration significantly depends on the geographical distance between any two cities. Recently, the high-speed railways have had rapid development in China (Zheng and Kahn 2013). By the end of 2012, China's bullet train service length reached 9356 km. By 2020, China's total service length of high-speed railways will reach 12,000 km. The high-speed railways boosted the top speed of some major trains from 100–150 to 200–250 km/h, which could lead to substantial cost savings in travel time and could facilitate cross-city economic integration (Zheng and Kahn 2013). It is reasonably believed that high-speed railways can alleviate the negative impact of spatial distance on R&D collaboration. Consequently, we hypothesize that the possibility of R&D collaboration between any two cities significantly increases if they are connected by the high speed railways.

Second, a considerable number of empirical studies confirmed that the resources of education, know-how, and R&D talents are unequally distributed between the eastern, central, and western regions of mainland China due to the imbalance of regional economic development in China (Gao et al. 2011; Scherngell and Hu 2011; Liang and Zhu 2002; Hong 2008; Wang et al. 2005; Liu and Jia 2008; Xu et al. 2005; Zhang and Kanbur 2005; Tu and Yi 2008). Similarly, there is a great gap of economic development and R&D systems among various cities, including the cities that are located in the same provinces. Jiangsu province, one of the most developed provinces in East China, is such a case. Statistically, the number of patents in the first three cities ranked by gross regional product

(GRP) in Jiangsu represents 56.17% of all cities in 2012, whereas the last three cities ranked by GRP represent 4.21%. Furthermore, R&D expenditures in the highest three cities ranked by GRP are 53.99% of total expenditures in 2012, whereas the last three cities ranked by GRP represent 4.27%. Meanwhile, some empirical studies confirmed that the technological proximity between regions also plays a crucial role in cross-region collaborative R&D activities, indicating that the collaboration probability increases between regions that are located close to each other in technological field (Breschi and Lissoni 2001; Fischer et al. 2006; Leasage et al. 2007; Hoekman et al. 2009; Scherngell and Barber 2011; Scherngell and Hu 2011). Based on previous empirical studies, we hypothesize that the possibility of cross-city R&D collaborations decreases by the difference in economic development and increases by the proximity in technological field.

Third, as described by Scherngell and Hu, the Chinese local authorities at the province level play a crucial role in distributing science and technology resources. In addition, the provincial governments are inward-oriented in their science policies, which protects local firms and research institutes and universities with the objective of maximizing intra-provincial benefits (Chen and Wang 2003; Yoon 2011). Due to provincial protectionism, the R&D collaborations between research organizations located in different provinces encounter more barriers. We hypothesize that the possibility of collaboration between any two cities decreases significantly if they are not located in the same province.

The results of previous empirical research showed that Beijing, the capital of China, is the central hub in the spatial network of inter-regional scientific collaboration (Scherngell and Hu 2011; Hu et al. 2009). Invariably, leading research institutions, universities and firms are always located in Beijing or provincial capital cities. One possible reason is that political priorities are more important than economic factors in selecting technology-related locations, which means the priority for allocating R&D resource is accorded to national or provincial capitals. Thus, it is strongly assumed that Beijing and the provincial capital cities can play vital role in the Chinese innovation network system due to their political centers of nation or provinces. Consequently, we hypothesize that the probability of collaboration between any two cities increases significantly if one of them is a national or provincial capital. Figure 1 presents the proposed hypothesis model based on the supporting theoretical and empirical literature base.

Methods

Study design

To test major external variables on cross-city R&D collaborations, we put forward a spatial interaction model. This model has been applied in some previous studies (Scherngell and Barber 2011; Scherngell and Hu 2011; Ponds et al. 2007; Hoekman et al. 2009; Fischer et al. 2006). The spatial interaction model is described by the following formula:

$$F_{ij} = O_i^{\alpha_1} D_j^{\alpha_2} \exp \left[\sum_{k=1}^k \beta_k S_{ij}^{(k)} \right] + \varepsilon_{ij} \quad i, j = 1, \dots, n \quad (1)$$

Here, F_{ij} denotes the collaboration intensity, which is measured in numbers of collaborations between cities i and j . O_i and D_j are control variables for the R&D scale, which

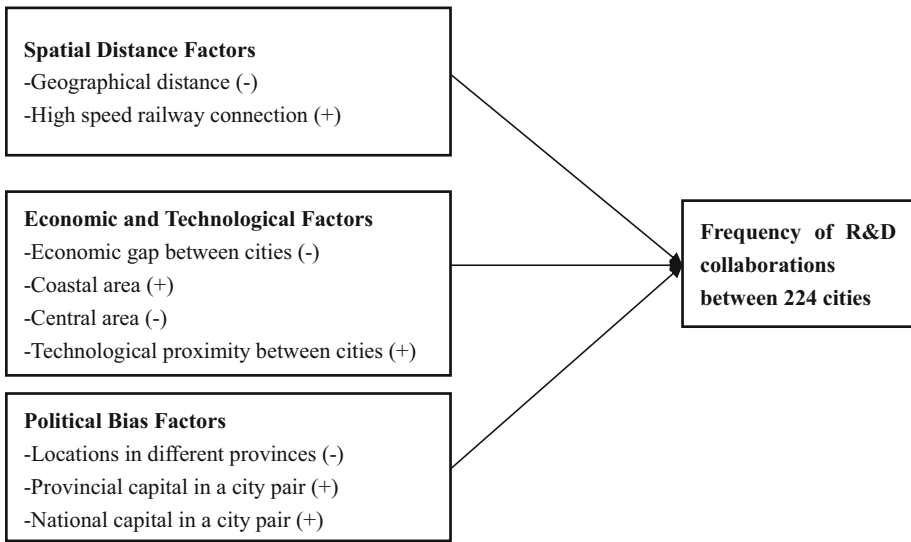


Fig. 1 Major factors affecting cross-city R&D collaborations in China

are measured in total number of firms in city i and city j , respectively. Furthermore, α_1 and α_2 represent the respective parameters to be estimated. $S_{ij}^{(k)}$ denotes K distinct factors, and β_k ($k = 1, \dots, K$) is a parameter that will estimate the impact of all types of factors on R&D collaboration intensity; ε_{ij} is error term. In this thesis, the research work is focused on $K = 9$ distinct factors. Based on the main lines of hypotheses in this study, these variables can be grouped into three categories:

- (i) Variables that represent *spatial factors*: (1) $s_{ij}^{(1)}$ represents the geographical distance as measured by the great circle distance between the geographic centers of cities i and j . (2) $s_{ij}^{(2)}$ is defined as a high-speed railway dummy variable, which takes a value of 1 when the cities i and j are connected by high-speed railways, 0 otherwise.
- (ii) Variables that represent *economic factors*: (3) $s_{ij}^{(3)}$ represents the economic gap of two cities i and j ; it is measured by the absolute difference in GRP for year 2012 between cities i and j . (4) $s_{ij}^{(4)}$ is a dummy variable that is set to a value of 1 if either or both cities i and j are located in the coastal area of China, 0 otherwise. (5) $s_{ij}^{(5)}$ is a dummy variable that is set to a value of 1 if either or both cities i or j are located in the central area of China, 0 otherwise. (6) $s_{ij}^{(6)}$ represents the technological proximity between cities i and j . Data are used from the Chinese Patent Database to measure this variable. It is constructed as a vector $t(i)$ that measures city i 's share of patenting in specific technological classes of the International Patent Classification (IPC). The Pearson correlation coefficient given by $r^2 = \text{corr}[t(i), t(j)]^2$ between the technological vectors of cities i and j is used to define how close they are to each other in technological classes.

- (iii) Variables that represent *political bias factors*: (7) $s_{ij}^{(7)}$ is defined as a different province dummy variable, which takes a value of 1 when cities i and j are located in the different province, 0 otherwise. (8) $s_{ij}^{(8)}$ is defined as a provincial capital city dummy variable, which takes a value of 1 if either or both cities i and j are provincial capitals, 0 otherwise. (9) $s_{ij}^{(9)}$ is defined as a Beijing dummy variable, which takes a value of 1 if either cities i and j is Beijing, 0 otherwise.

Due to the nature of the enumeration data in this context, normal distribution is generally not satisfied, and linear regression models are not appropriate (Long and Freese 2001). A Poisson regression, which is estimated using maximum likelihood techniques, is usually applied to non-normal data. In this study, the observed collaboration intensity between cities i and j has a Poisson distribution with a conditional mean. However, there are many cities with no collaboration between themselves. Consequently, the amount of zeros in the dependent variable is larger than assumed for a Poisson distribution, and the conditional variance is larger than the conditional mean. In an attempt to solve this problem of over-dispersion, a negative binomial model is used in which an extra parameter α is introduced (Hilbe 2011). The key difference between the negative binomial model and the Poisson regression model concerns the conditional variance. The Poisson regression model assumes a Poisson distribution, in which the conditional mean of the dependent variable equals the conditional variance. However, it is common for the conditional variance to exceed the conditional mean, particularly when the count variable has more zeroes than a Poisson-distributed data-generating process would yield. The resulting over-dispersion can be explained in the negative binomial model via an extra parameter denoted as α (Hilbe 2011).

Data source

In mainland China, there are 31 administrative regions, including 22 provinces, 5 autonomous regions and 4 municipalities. According to China Statistical Yearbook (2013), there are 289 prefecture-level cities in 31 administrative regions. We selected 224 prefecture-level cities with populations of more than 500 thousand people in 2012 as the study sample. The 224 cities are located in 31 administrative regions. The detailed list of cities is provided in Supplementary Table S1 online.

Patents are considered to be a particularly appropriate form of R&D activity output; this represents that R&D collaborations will often be reflected in a co-applicant. Patents have long been recognised as an important and fruitful source of data for the study of innovation and technological change. In this study, we took patent data as a proxy for R&D activity. In other words, we collected innovation data (patents) to investigate China's cross-city R&D collaboration. The R&D activities are usually considered as innovation activities. In China, the style of patents is divided into three types: Invention patent, Utility model patent and Appearance design patent. Generally, invention patent will be seen as a high innovation competence, next is utility model patent, last is appearance design patent. Based on China's patent law, only the applications for invention patent request substantive examination. Compared to utility model patent and appearance design patent, invention patent has somewhat higher innovation. To a somewhat larger extent, invention patent could reflect the innovation degree of cross-city R&D collaboration activities. Therefore, invention patents are most closely related to the research idea of this study. On the other hand, the other two types of patent, especially for appearance design patent, are less innovative in

general. Nevertheless, despite the aforementioned shortcoming, the collaborative frequencies of utility model patent and appearance design patent between cities also offer an unobtrusive indicator for the study of cross-city R&D collaboration. As a developing country, China generally lags behind developed countries in scientific and technological innovation level. For instance, utility model patents and appearance design patents make up a significant portion of the total number of patents in China. Based on the patent data from the Patent Database of State Intellectual Property Office of P.R. China (SIPO) in year 2012, the percentage of invention patents, utility model patents and appearance design patents was 34.59, 36.74 and 28.67%, respectively. Therefore, in order to analyze the R&D collaborative relationship between cities completely and systematically, all of the patents were included in this study. We used data from the Patent Database of State Intellectual Property Office of P.R. China (SIPO). In addition, the database includes: the names of inventions, the dates of application, publication, and grant; and the names and addresses of applicants and assignees. Here, the co-occurrence of two or more addresses on a patent is considered as one or more collaborations.

First, data on patents that contain at least two applicants that are research organizations (firms, universities or research institutions) from different cities for 2012 were retrieved from the Patent Database of SIPO. In China, the R&D firms are major players in technological innovation. They have professional R&D department, R&D personnel and R&D teams, all products are designed and manufactured by themselves. Meanwhile, the manufacturing firms and production firms usually purchase the technology or products from the R&D firms, universities or research institutions. They also accept customer's OEM (original equipment manufacturer) production and mainly undertake processing all types of manufacture. Basically, they are irrelevant to the technological innovation. Therefore, considering the research idea of this study, all of the manufacturing firms and production firms should be excluded. If the firms are the applicants for a patent or the patent holders, they are regarded as the R&D firms. By judging whether the firms included in the study have patents or co-patents, we can determine that they are actually carrying out R&D activities. The year 2012 is not a special year with respect to R&D activity or changing patterns in science and technology (S&T) policy; thus, the co-patent data set for 2012 could reflect the current status of collaborative R&D activities in China.

The next step provided the 224×224 city-by-city collaboration matrix. The number of individual collaborative activities was aggregated to the city level, which leads to the observed number of R&D collaborations between two cities, i and j . For example, if there are three participating organizations that are from three different cities (city k , m , and n) in a patent, three co-patent links are counted: from city k to city m , from k to n and from m to n . The data on R&D collaboration was placed into a (224×224) symmetrical matrix that contains all co-patent links between cities. The resulting city collaboration matrix then contains the total collaboration intensities between all (i, j) -city pairs, given the $i = 1, \dots, n = 224$ cities in the rows and the $j = 1, \dots, n = 224$ cities in the columns. The n -by- n matrix is symmetric by construction ($p_{ij} = p_{ji}$). Each cell then includes the number of co-patent links between city i and city j . Because this study only focuses on inter-city collaborations, we exclude intra-city collaboration links in the spatial interaction model. Consequently, there are potentially $(224 \times 224) - 224 = 49,952$ co-patent links in the collaboration matrix. Table 1 describes and lists the data sources of the dependent and independent variables in the spatial interaction models. All analyses were conducted using Stata (version 11.0; StataCorp LP, College Station, TX, USA).

Table 1 Variable description and data sources

Variable type	Variable names	Description	Data source
Dependent variable	F_{ij} : R&D collaborations between city i and j	Numbers of collaborations among 224 cities as captured by co-patent data	Chinese patent database
Independent variable	O_i : origin variables	The natural logarithm of the total number of firms in the city i in year 2012	China City Statistical Yearbook 2013
	D_j : destination variables	The natural logarithm of the total number of firms in the city j in year 2012	China City Statistical Yearbook 2013
	$S_{ij}^{(1)}$: separation variables to measure geographical distance	The natural logarithm of the great circle distance between the centers of the cities i and j	Chinese geographical informational system data
	$S_{ij}^{(2)}$: high-speed railways dummy	1 = cities i and j are connected by high speed railways; 0 = cities i and j are not connected by high speed railways	China Railway Press
	$S_{ij}^{(3)}$: separation variables to measure the economic gap of two cities i and j	The natural logarithm of the absolute value of the difference between the GRP for cities i and j in 2012	China City Statistical Yearbook 2013
	$S_{ij}^{(4)}$: coastal area dummy	1 = either or both cities i and j are located in the coastal area of China; 0 = both cities i and j are not located in the coastal area of China	Chinese geographical informational system data
	$S_{ij}^{(5)}$: central area dummy	1 = either or both cities i and j are located in the central area of China; 0 = both cities i and j are not located in the central area of China	Chinese geographical informational system data
	$s_{ij}^{(6)}$: technological proximity	The Pearson correlation coefficient between the technological vectors of two city i and j	Chinese patent database
	$s_{ij}^{(7)}$: different province dummy	1 = cities i and j in different province; 0 = cities i and j in same province	China City Statistical Yearbook 2013
$s_{ij}^{(8)}$: provincial capital city dummy	1 = either or both cities i and j are provincial capital; 0 = both cities i and j are not provincial capital	China City Statistical Yearbook 2013	
$s_{ij}^{(9)}$: Beijing dummy	1 = either cities i or j is Beijing; 0 = both cities i and j are not Beijing	China City Statistical Yearbook 2013	

Results

Descriptive statistics analysis

Table 2 presents the results of descriptive statistical analysis on R&D collaborations between 224 cities as captured by co-patent data. It can be observed that intra-provincial cross-city R&D collaborations are more frequent than inter-provincial cross-city collaborations; the mean collaboration intensity for intra-provincial cross-city collaborations is

Table 2 Some descriptive statistics on R&D collaborations among 224 cities in China as captured by co-patent data

	ME	Sum	Mean	SD	Min	Max	Skewness	Kurtosis
All R&D collaborations	50,176	65,649	1.31	33.12	0	4731	88.51	10,332.25
Cross-city R&D collaborations	49,552	43,652	0.87	14.98	0	1437	56.90	4522.66
Intra-provincial cross-city R&D collaborations	2242	10,624	4.74	48.12	0	1437	25.22	716.61
Inter-provincial cross-city R&D collaborations	47,710	33,028	0.69	11.20	0	1041	47.81	3545.08

ME matrix elements, SD standard deviations

4.74. However, for inter-provincial, the mean is 0.69. The difference between intra- and inter-provincial collaborations frequency may hint at the existence of different province effects, which will be validated in the spatial interaction model. The results of statistics for kurtosis and skewness indicate extremely left-skewed distributions, which means that there is a relatively small number of city pairs with a high number of R&D collaborations, whereas most city pairs show relatively low or no collaboration frequencies.

Spatial network structure of cross-city R&D collaborations

The spatial city-by-city R&D network is visualized in Figs. 2 and 3. As shown in Fig. 2, the map of mainland China was divided into 289 regions, each region representing one prefecture-level city. The regions in the map that show their names of cities are the 224 sample cities in this study. The intensity of the color for the regions is relative to their

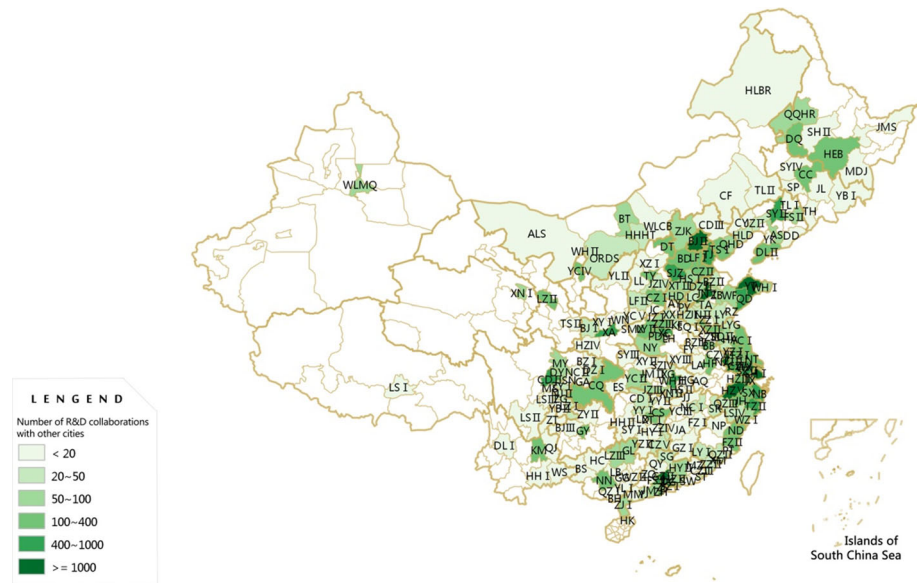


Fig. 2 Degree centrality of cross-city R&D collaboration network in China (SuperMap Deskpro was used to create the map in the Fig. 2)



Fig. 3 Spatial patterns of R&D collaborations as captured by co-patents between 224 cities in China (SuperMap Deskpro was used to create the map in the Fig. 3)

degree centrality, which measures how many times it links with other cities, that is, it gives the number of inter-city collaborations. The deeper the color is, the higher degree centrality the city’s will have. As shown in Fig. 3, the connecting line represents the co-patent link between two cities *i* and *j*. Furthermore, the frequency of R&D collaborations between two cities *i* and *j* is illustrated by means of the intensity of the color of the connecting lines, which corresponds to the number of links. The deeper the color is, the higher number of links the city pair’s will have.

Figure 2 indicates that the cross-city collaborative R&D activities mainly occur in the favored regions, advanced municipalities and coastal regions such as Beijing-Tianjin area, the Yangtze River Delta (Shanghai, Nanjing, Suzhou, Ningbo and Hangzhou), the Pearl River Delta (Guangzhou, Shenzhen, Dongguan and Zhuhai). The top five total number of R&D collaborations of Beijing is 8768, followed by Shanghai (2725), Shenzhen (2723), Guangzhou (2616) and Nanjing (1527). The cities, which are located in these regions, are known for their strong economic development and abundant technological resources.

Initially, Fig. 3 indicates that the central hub in this spatial network is Beijing; a high cross-city interaction density can also be observed for Shanghai, Shenzhen, Guangzhou, Hangzhou, and Nanjing as well as for the provincial capitals in coastal regions. The number of links to interior cities in central and western regions is generally very low. As shown in Table 3, in the top 20 highest cross-city collaboration city pairs, the proportion of the city pairs that include Beijing is 60% (12/20), and the proportion of the city pairs that include provincial capitals is 80% (16/20).

The spatial interaction model

As depicted in Table 4, the spatial interaction model showed the dispersion parameter is significant for the negative binomial model, which indicates that the over-dispersion test

Table 3 Top-20 cross-city collaboration flows in the co-patent data

Rank	City pair	Number of collaborations
1	Guangzhou and Shenzhen	1437
2	Beijing and Yantai	1041
3	Hangzhou and Taizhou 2	562
4	Beijing and Shanghai	501
5	Beijing and Tianjin	442
6	Beijing and Shijiazhuang	345
7	Beijing and Nanjing	309
8	Beijing and Shenyang	307
9	Hangzhou and Ningbo	294
10	Jinan and Xuchang	253
11	Changzhou and Shenzhen	206
12	Beijing and Ningbo	203
13	Beijing and Jinan	202
14	Beijing and Suzhou 1	201
15	Beijing and Chengdu	199
16	Hangzhou and Jinan	195
17	Shenzhen and Suzhou 1	192
18	Beijing and Wuhan	191
19	Dongguan and Ningde	186
20	Beijing and Guangzhou	176

rejected the null hypothesis that the conditional variance of the dependent variable equals its conditional mean. Thus, the spatial interaction model employed the negative binomial regression. Table 4 also shows that the parameter estimates are highly significant in the negative binomial interaction model.

As expected, the parameter estimates for the scale variables ($\alpha_1 = \alpha_2 = 0.798$) indicate that a higher number of firms in a city increase the likelihood of collaboration with other cities. The estimate for β_1 provides evidence that geographical distance between cities exerts a significant negative effect on the likelihood that they collaborate. The parameter estimate of $\beta_1 = -0.685$ indicates that the spatial distance between any two cities adds 100 km each, and the mean collaboration frequency decreases by 49.59%. As evidenced by the estimate for β_2 , it is more likely that collaborations occur among cities that are connected by high-speed railways. The estimate for the coastal area dummy variable ($\beta_4 = 0.280$) indicates that the mean R&D collaboration probability between any two cities increases when at least one city is located in the coastal area. The result suggests that city pairs that involve cities located in the central area have a higher probability for R&D collaboration than other non-coastal area pairs. In contrast, the collaboration probability between any two cities decreases when at least one city is located in the central area ($\beta_5 = -0.300$). However, contrary to expectation, the impact of the economic gap between any two cities was found to have a significantly positive influence on their collaboration probability ($\beta_3 = 0.420$). The estimate for technological proximity variable $\beta_6 = 1.589$ suggests that cross-city R&D activities are more likely to occur between cities that are close to each other in technological field.

Table 4 Estimation results of the negative binomial spatial interaction model for all co-patents between 224 cities

Model variables	Coefficient beta	Standard errors
<i>Scale effects</i>		
Origin variable (α_1)	0.798***	0.032
Destination variable (α_2)	0.797***	0.032
<i>Spatial factors</i>		
Geographical distance (β_1)	-0.685***	0.050
High-speed railways dummy(β_2)	0.485***	0.118
<i>Economic and technological factors</i>		
Economic gap (β_3)	0.420***	0.027
Coastal area dummy (β_4)	0.280***	0.072
Central area dummy (β_5)	-0.300***	0.062
Technological proximity (β_6)	1.589***	0.152
<i>Political bias factors</i>		
Different province dummy(β_7)	-1.415***	0.138
Provincial capital city dummy(β_8)	1.957***	0.066
Beijing dummy(β_9)	2.447***	0.213
Dispersion parameter (α)	2.895***	0.028
Constant	-12.094***	0.441

Observations = 49,952; Log likelihood = -14,663.834; Pseudo $R^2 = 0.1662$

*** Statistically significant at the 0.001 significance level

Table 5 Summary of hypotheses testing

Hypothesis	Regression result
H1: Geographical distance between two cities in a city pair	Supported
H2: Economic gap between two cities in a city pair	Supported (on the opposite side)
H3: One city of a city pair is located in the coastal area	Supported
H4: One city of a city pair is located in the central area	Supported
H5: Technological proximity between two cities in a city pair	Supported
H6: A city pair is connected by high-speed railways	Supported
H7: One city of a city pair is the capital of a province	Supported
H8: A city-pair is located in the different province	Supported
H9: One city of a city pair is Beijing	Supported

The parameter estimate of β_7 indicates that the likelihood of collaboration between any two cities decreases when they are located in different provinces. In addition, the parameter estimates of β_8 and β_9 indicate that the likelihood of collaboration between any two cities increases when either or both are located in a provincial capital city or Beijing, respectively. Table 5 summarizes the final results of the hypothesis in our study.

Discussion

This is the first study that has used more than 220 cities as the unit of spatial analysis to explain major factors that affect inter-regional R&D collaborations in China. The goal of this study was to analyze the variation of cross-city collaborative R&D activities as captured by co-patent data. Due to over-dispersion in the co-patent data, a negative binomial model was used. The overall negative binomial model that incorporates the independent variables was found to be significant.

The probability of cross-city R&D collaboration significantly decreases by geographical distance, which is in accordance with previous empirical studies (Scherngell and Hu 2011; Liang and Zhu 2002; Fischer et al. 2006; Pan et al. 2012; Hong 2008). Moreover, the model coefficient estimate ($\beta_1 = -0.685$) shows that the geographical effect in this study is much higher than those in prior comparable studies, in which the coefficient estimates range from -0.228 to -0.354 (Scherngell and Hu 2011; Liang and Zhu 2002; Fischer et al. 2006; Pan et al. 2012; Hong 2008). In other words, the geographical distance is an important determination of the constitution of cross-city R&D collaborations. Furthermore, this finding is consistent with the previous result of Ma et al.. Their work confirmed that the relationship between inter-city geographical distance and scientific cooperation has been enhanced and strengthened over time (Ma et al. 2014).

Furthermore, there is a significant high-speed railways effect, which implies that the Chinese transport infrastructures could play an important role in overcoming the barriers of geographical distance for inter-regional R&D collaboration; this would accelerate the establishment of a national innovation system and facilitate the integration of regional R&D resources. Predictably, with the rapid development of China's bullet trains, the negative effects of geographical space will become increasingly weaker.

The economic gap between any two cities in the regression model was surprisingly found to have a significantly positive influence on the frequency of cross-city R&D collaborations. This finding suggests that the greater the economic difference between two cities, the higher the number of R&D collaborations. This finding is contrary to our literature review, which suggested that the economic distance between two regions has a significant negative effect on their collaboration probability (Scherngell and Hu 2011; Hu et al. 2009; Acosta et al. 2010). However, according to our finding, this hypothesis does not hold at the city level in China. Several possible factors could be proposed to explain this positive finding. First, from the perspective of descriptive analysis, the degree centrality of mega cities such as Beijing, Shanghai and Guangzhou is very high, which indicates that most cross-city collaborations are related to these mega cities in favored regions. Furthermore, a large number of zero links were observed between the cities in less-favored regions, which means that the absolute differences in GRP between these cities are minimal. In contrast, the absolute differences in GRP between mega cities and other cities are very enormous, which leads to the result that shows that the cities are willing to collaborate with other cities regardless of their economic development level. Second, the center-periphery hypothesis that applies to R&D collaboration provides clues regarding whether differences in economic development between areas may impact the models of R&D collaboration (Schott 1998; Schubert and Sooryamoorthy 2009). It suggests that research organizations in lagging regions tend to collaborate with those in core regions to gain access to resources, whereas research organizations in core regions collaborate by seeking complementarities (Schott 1998; Schubert and Sooryamoorthy 2009). In this study, the economically developed cities usually maintain higher GRP; these are regarded as core

regions. By seeking collaboration with the economically developed cities, the less economically developed cities could garner R&D investment funds, advanced know-how and intellectual resources. Conversely, the economically developed cities collaborate for the purpose of complementary resources such as natural resources or human capital. In addition, the econometric analysis also confirms the Hypothesis 5 for the Chinese case that the cross-city R&D collaborative activities are affected significantly not only by economic gap, but also to a somewhat larger extent by technological proximity.

Hypothesis 3 and 4 have been supported, which confirms that the regional locations of cities are the major positive or negative factors that affect cross-city R&D collaborations. This finding was consistent with the previous work that adopted provincial regions as units of analysis (Scherngell and Hu 2011). Interestingly, taking an urban perspective, the results of our study indicate that the R&D collaboration links between economically developed cities and less economically developed cities usually occurred in coastal and favored regions.

The functions with different provinces, provincial capital cities and Beijing dummies in our spatial interaction model were observed to garner further understanding of the political bias factors such as the capital city effect and the spatial provincial effect. There is a significant different-province effect, which implies that the inter-provincial R&D collaborations between cities encounter significant barriers derived from certain political factors. One of the political factors could be dubbed as a “spatial provincial bias”. It appears that cities in the same province cooperate with each other more than those in a different province. For instance, in this study, Hangzhou in Zhejiang province is more strongly linked to Ningbo (Zhejiang province), but not to Suzhou (Jiangsu province), which is at approximately the same geographical distance. Additionally, Xuzhou in Jiangsu province is more strongly linked to Suzhou (Jiangsu province) than to Bengbu (Anhui province), although the latter is nearer. This finding reflects the presence of a spatial provincial bias as hypothesized, which is very possibly driven by regional protectionism. Indeed, Chinese scholars contend that regional protectionism is a widely documented phenomenon and will be one of the major problems of inter-regional R&D collaborations (Scherngell and Hu 2011; Zhang 2003). Furthermore, a total of 93% of public research institutions in China are mainly funded by local provincial governments (Scherngell and Hu 2011). Therefore, to protect local research organizations with the objective of maximizing intra-provincial benefits, the provincial authorities are usually inward-oriented in their science policies (Chen and Wang 2003). In addition to the fact that a substantial share of Chinese R&D funding is attributable to governmental decision-makers, the Chinese system of R&D is top-down institutionally driven with minimal bottom-up research organization-driven R&D (Jonkers 2010).

In addition, the Beijing dominance in the cross-city R&D network was borne out by the descriptive analysis and econometrics model analysis. This finding is in accordance with the literature (Scherngell and Hu 2011; Hu et al. 2009) and what we expected, which also confirms that the hub of cross-city R&D collaborations is the capital of China. The research organizations in Beijing have more access to research funding, intellectual resources and opportunities for R&D cooperation. Similarly, the research organizations in the 31 provincial capital cities have more access to the opportunities for R&D cooperation than those in other non-provincial capital cities. This finding could epitomize the vital roles of political centers of the country and provinces in R&D networks.

The results of this study must be interpreted given several limitations. First, the cross-sectional study design does not allow the measurement of any annual changes in the dependent and independent variables in the spatial interaction model. A future research

direction would be to establish a time series model based on the panel data for 10 years or more. Thus, the dynamic changes of cross-city R&D collaborations could be observed. Second, the spatial interaction model in this study may have been able to explain more of the variance in factors that affect cross-city R&D collaborations because of the greater complexity of the affecting factors in those other studies. Consequently, this work would be well complemented by adding other new independent variables into the model. Last, this work could not provide insights on the role of the affecting factors across different scientific fields. A possible research direction would be to collect and analyze co-patent data for cross-city R&D collaborations in a sub-scientific field using the spatial interaction model.

In summary, our analysis suggests that geographical distance was negatively associated with the frequency of cross-city R&D collaborations, whereas the high-speed railways ties between the cities were positively associated. The economic gap between the cities was positively associated with the frequency of cross-city R&D collaborations, despite this being a counter-intuitive finding. This gap suggests that the smaller cities in economically backward areas seek collaboration from the core to gain resource access. The regional locations of cities are significant factors that impact cross-city R&D collaborations, which means that an imbalance of the regional economy continues to play a crucial role. Moreover, the existence of so-called technological and political bias effects has been demonstrated by our work. To encourage R&D cooperation between cities with various advantages, the central government needs to take the necessary measures to reduce regional protectionism and promote the rational allocation of R&D resources.

Acknowledgements This work was funded by the research projects of the Philosophy and Social Science Foundation of Nanjing Medical University (Project No. 2013NJZS10) and the Scholarship of Jiangsu Province Government Sponsorship for Overseas Study. The authors would thank anonymous reviewers for their valuable comments.

Author contributions All authors designed and performed the research. Sheng-qiang Jiang and Xin Li collected the data and participated in the writing of the manuscript. Zhi-hang Peng, Xin Li and An-na Shi contributed to data analysis.

Compliance with ethical standards

Conflict of interest Competing financial interests: The authors declared no competing financial interests.

References

- Acosta, M., Coronado, D., Ferrándiz, E., & León, M. D. (2010). Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, 87(1), 63–74.
- Breschi, S., & Lissoni, F. (2001). Knowledge spillovers and local innovation systems: A critical survey. *Industrial and Corporate Change*, 10, 975–1005.
- Chen, K., & Kenney, M. (2007). Universities/research institutes and regional innovation systems: The cases of Beijing and Shenzhen. *World Development*, 35(6), 1056–1074.
- Chen, C. T., & Wang, T. (2003). The influence to the local S&T legislation from the view of CAST operation mode forum on science and technology in China. *Forum on Science Technology in China*, 4, 3–6. (in Chinese).
- Fischer, M. M., Scherngell, T., & Jansenberger, E. (2006). The geography of knowledge spillovers between high-technology firms in Europe: Evidence from a spatial interaction modeling perspective. *Geographical Analysis*, 38(3), 288–309.

- Florida, R. (2002). *The rise of the creative class: And how it's transforming work, leisure, community and every day life*. New York: Basic Books.
- Gao, X., Guan, J., & Rousseau, R. (2011). Mapping collaborative knowledge production in China using patent co-inventorships. *Scientometrics*, 88(2), 343–362.
- Hennemann, S., Wang, T., & Liefner, I. (2011). Measuring regional science networks in China: A comparison of international and domestic bibliographic data sources. *Scientometrics*, 88(2), 535–554.
- Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge: Cambridge University Press.
- Hoekman, J., Frenken, K., & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, 43(3), 721–738.
- Hong, W. (2008). Decline of the center: The decentralizing process of knowledge transfer of Chinese universities from 1985 to 2004. *Research Policy*, 37(4), 580–595.
- Hu, Y. J., Scherngell, T., & Liang, C. Y. (2009). The empirical study on inter-regional R&D collaboration in China. *Science Technology Management Research*, 9, 49–50. (in Chinese).
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3), 577–598.
- Jonkers, K. (2010). *Mobility, migration and the Chinese scientific research system*. London: Routledge Press.
- Katz, J. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31–43.
- Leasage, J., Fischer, M. M., & Scherngell, T. (2007). Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86, 393–421.
- Liang, L., & Zhu, L. (2002). Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics*, 55(2), 287–316.
- Liefner, I., & Hennemann, S. (2011). Structural holes and new dimensions of distance: The spatial configuration of the scientific knowledge network of China's optical technology sector. *Environment and Planning-Part A*, 43(4), 810.
- Liu, N., & Jia, Y. (2008). Regional policy, regional openness and regional disparity in economic development. *Contemporary Economy & Management*, 10, 019. (in Chinese).
- Long, J. S., & Freese, J. (2001). *Regression models for categorical dependent variables using stata*. Texas: Stata Press.
- Ma, H., Fang, C., Pang, B., & Li, G. (2014). The effect of geographical proximity on scientific cooperation among Chinese cities from 1990 to 2010. *PLoS ONE*, 9, e111705.
- Pan, R. K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports*, 2, 902.
- Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3), 423–443.
- Sassen, S. (1991). *The global city: New York, London, Tokyo*. Princeton: Princeton University Press.
- Scherngell, T., & Barber, M. J. (2011). Distinct spatial characteristics of industrial and public research collaborations: Evidence from the fifth EU Framework Programme. *The Annals of Regional Science*, 46(2), 247–266.
- Scherngell, T., & Hu, Y. (2011). Collaborative knowledge production in China: Regional evidence from a gravity model approach. *Regional Studies*, 45(6), 755–772.
- Schott, T. (1998). Ties between center and periphery in the scientific world-system: Accumulation of rewards, dominance and self-reliance in the center. *Journal of World Systems Research*, 4(2), 112–144.
- Schubert, T., & Sooryamoorthy, R. (2009). Can the centre–periphery model explain patterns of international scientific collaboration among threshold and industrialised countries? The case of South Africa and Germany. *Scientometrics*, 83(1), 181–203.
- Tu, C. L., & Yi, W. H. (2008). Factors and regional disparity of self-innovation capabilities in China. *Science Technology Management Research*, 1, 39–42. (in Chinese).
- Wang, Y., Wu, Y., Pan, Y., Ma, Z., & Rousseau, R. (2005). Scientific collaboration in China as reflected in co-authorship. *Scientometrics*, 62(2), 183–198.
- Xu, J. H., Lu, F., Su, F. L., & Lu, Y. (2005). Spatial and temporal scale analysis on the regional economic disparities in China. *Geographical Research*, 24(1), 57–68. (in Chinese).
- Yoon, J. (2011). *Exploring regional innovation capacities of PR China: Toward the study of knowledge divide*. PhD thesis, Georgia Institute of Technology.
- Zhang, G. C. (2003). Study on the environment of science and technology innovation. *Railway Technology Innovation*, 2, 7–10. (in Chinese).
- Zhang, X., & Kanbur, R. (2005). Spatial inequality in education and health care in China. *China Economic Review*, 16(2), 189–204.
- Zheng, S., & Kahn, M. E. (2013). China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences*, 110(14), E1248–E1253.