PLOS ONE

RESEARCH ARTICLE

Can the establishment of state-level urban agglomeration stimulate enterprise innovation?—Taking Yangtze River Delta and Pearl River Delta as an example

Kai Zhao[1]*, **Huahua Huang**[2], **Wanshu Wu**[3]

1 School of Economics, Qingdao University, Qingdao, China, **2** School of Statistics, Huaqiao University, Xiamen, China, **3** College of Architecture and Urban Planning, Qingdao University of Technology, Qingdao, China

* kzhao_kai@126.com

Citation: Zhao K, Huang H, Wu W (2022) Can the establishment of state-level urban agglomeration stimulate enterprise innovation?—Taking Yangtze River Delta and Pearl River Delta as an example. PLoS ONE 17(8): e0273154. https://doi.org/10.1371/journal.pone.0273154

Editor: Jun Yang, Northeastern University (Shenyang China), CHINA

Received: June 3, 2022

Accepted: August 3, 2022

Published: August 25, 2022

Data Availability Statement: The data underlying the results presented in the study are available from (China Stock Market & Accounting Research (CSMAR), http://www.csmar.com/).

Funding: The research was funded by National Social Science Fund of China [grant number 20FJYB017].

Competing interests: The authors have declared that no competing interests exist.

Abstract

This study uses a quasi-experimental method, Geographic Regression Discontinuity Design (GRDD), to evaluate the actual effect of establishing Yangtze River Delta and Pearl River Delta urban agglomerations on enterprise innovation. GRDD is a design in which a geographic boundary splits the units into treated and control areas in an as-if random fashion, and the shortest distances from each enterprise's location to the boundary of urban agglomeration calculated by ArcGIS are considered as the running variable. The actual effect can be identified by the probability of receiving treatment jumps discontinuously at the known cutoff. It is shown that the establishment of Yangtze River Delta and Pearl River Delta urban agglomerations can significantly improve the enterprise innovation, and this outcome is verified by rigorous robustness tests including the placebo test with pseudo-boundary, the bandwidth sensitivity test, the parametric test with different functional forms and the extreme value test. Further, the influence mechanisms of state-level urban agglomerations promoting enterprise innovation are explored by Staggered DID. It is confirmed that the urban agglomeration construction can promote enterprise innovation through financial support and regional coordination channels.

1 Introduction

Economic globalization needs strengthen the development of urban agglomeration, which has become the most important modern economic development mode. The top 40% of the world's urban agglomerations contribute 66% of the global economy and 85% of scientific and technological innovations [1]. "First experimenting and then spreading" is the consistent thinking of China's central and local governments to promote various reforms. As a creative achievement to promote China's economic development and accelerate the urbanization process, the exploration and practice of urban agglomeration construction is surpassing the national and administrative boundaries and becoming a "node" connecting the global economy, which is related to the realization of the 14th "Five-Year Plan" and the "Long-term Goal of 2035" [2].

Under the background of economic globalization and regional integration, the importance of innovation is self-evident. No matter a country, a region or even an enterprise, the economic competitive advantage comes from the constantly improving innovation ability. Although some new global problems (*i.e*. epidemic of COVID-19, Sino-US trade friction) have changed the external environment in recent years, innovation is still the key to gain competitive advantage and promote the high-quality development. Enterprises, as the main body of innovation, play an important role in promoting the high-quality development of China, while state-level urban agglomerations, as the key carrier of innovation-driven, provide a novel perspective for understanding local government behavior, enterprise innovation strategies and China's economic growth. Building a scientific and rigorous analysis framework to clarify the relationship between China's urban agglomeration construction and enterprise innovation can expand the evaluation perspective of China's urban agglomeration policy, and provide useful ideas for realizing high-quality economic development by encouraging enterprise innovation [3].

Yangtze River Delta (henceforth, YRD) is the most comprehensive state-level urban agglomeration in China, located in the alluvial plain at the mouth of the Yangtze River. The YRD urban agglomeration covers 26 cities, including Shanghai, Nanjing and Hangzhou, with a land area of 211,700 square kilometers and a total population of 150 million (in 2020), accounting for one-fifth of China's GDP. Pearl River Delta (henceforth, PRD) is the state-level urban agglomeration with the strongest economic vitality in China. It is composed of 9 cities including Guangzhou, Shenzhen and Foshan, with a total area of 56,000 square kilometers and a total population of 70 million people (in 2020). Both YRD and PRD urban agglomerations are bases for scientific and technological innovation, important platforms for China to participate in economic globalization, strategic areas for China's modernization. They are the two most advantageous urban agglomerations in China in terms of location and policy. On the whole, the YRD and PRD have become China's urban agglomerations with the largest population inflow and the highest innovation output. Therefore, clarifying the actual effect of establishment of YRD and PRD urban agglomerations on enterprise innovation has important values for optimizing government system and encouraging enterprise innovation. Specifically, as far as the government is concerned, the accurate evaluation of the impact of urban agglomeration establishment on enterprise innovation and the identification of the preconditions and conditions for the policy implementation can provide important reference for the central and local governments to formulate effective development strategies, and provide a reliable basis for the introduction of future policies. As far as enterprises are concerned, discussing the relationship between the construction of YRD and PRD urban agglomerations and the innovation of enterprises is helpful to make innovation investment decisions at the micro-level and provide certain reference and theoretical guidance for enterprises to optimize the investment structure.

Selection and endogeneity are often key threats to inference in the social science. Recently, analysts have turned to natural experiments and quasi-experimental methods as one way to overcome these obstacles in observational studies [4]. Although the Difference-In-Difference (henceforth, DID) method can effectively identify the impact of the establishment of YRD and PRD urban agglomerations, there is still a lack of analysis on the spatial role of urban agglomeration. To this end, we use a quasi-experimental method, Geographic Regression Discontinuity Design (henceforth, GRDD), to evaluate the actual effect of establishing YRD and PRD urban agglomerations on enterprise innovation from the spatial viewpoint.

The contributions of this study are mainly reflected in the following two aspects. First, the establishment of urban agglomeration is a strategic plan put forward by the state according to the current economic development situation, and it is the comprehensive product of regional economic situation and administrative planning. Investigating the impact of state-level urban agglomeration establishment across administrative divisions on micro enterprise innovation, provides a beneficial supplement for China's location-oriented policy evaluation. Second, examining the actual effect of urban agglomeration establishment by GRDD, can effectively take the impact of "spatial layout" and "policy events" into account. While overcoming the possible interference of endogenous problems on the empirical results, this study provides a solid micro-foundation for how to practice and promote "innovation-driven" and "high-quality development" in the new era.

The paper is organized as follows. In the next section, we provide a brief literature review. In section 3, we present the research design, identification strategy and estimation method. In section 4, the data source and the preliminary exploration are discussed. In section 5, we discuss the estimated results, and verify the reliability of the results. In section 6, we draw the main conclusions and policy recommendations.

2 Literature review

2.1 Urban agglomeration

There has been a long debate about the role of urban agglomeration in the literature on why some regions successfully achieve growth, while other regions stagnate or decline [5]. Since Gottmann puts forward the concept of urban agglomeration (Megalopolis) [6], a large number of scholars begin to explore urban agglomeration from the perspectives of geography, regional economics, urban economics, environmental science and other disciplines [7]. These studies not only reflect the characteristics and the evolution of urban agglomerations, but also include the prediction and optimization suggestions for the future development, providing theoretical basis for the strategic planning of world urban agglomerations. The literature on urban agglomeration can be roughly divided into three stages: the initial stage (2000–2007), the developing stage (2008–2014), and the mature stage (2015–2022).

At the initial stage, scholars focus on the agglomeration form of "city-industry", and try to explore the geographical scope of external economic operation [8]. The related research in this stage is mainly to investigate the formation process, population flow patterns and industrial cluster types of urban agglomerations in the northeastern United States and Europe based on qualitative methods and case studies [9, 10]. At the developing stage, scholars pay attention to the development of urban agglomeration caused by urban network and urban cooperation. It is found that urban expansions not only lead to regional fragmentation, but also generate new regional associations, and the huge spatial scale will help to form new regional networks and spatial associations across metropolitan areas [11]. The study of urban agglomeration in this stage changes from qualitative to quantitative models, and the cellular automata simulation, the multiple linear regression and the gravity model become analytical tools to study the evolution of urban agglomeration [12–14]. In addition, the night-light data and the traffic data are more and more widely used in urban agglomeration study [15, 16]. With the development of urbanization and urban agglomeration, the evaluation and the mechanism analysis of sustainable development of economy and environment become the center of attention, and the research on urban agglomeration enters a mature stage. Many empirical studies evaluate the development benefits of urban agglomerations from the perspective of economic efficiency. These studies mainly include: building the empirical framework of spatial structure affecting the economic efficiency of urban agglomeration [17]; discussing the spatial scale conditions of urban agglomeration scale benefit [18]; evaluating the policy performance of establishing urban agglomeration [7]; exploring the relationship between urban agglomeration spatial integration and industrial coordinated development [19]. Besides, the development of urban agglomeration inevitably involves resource consumption [20] and environmental pollution [21].

The research on China's urban agglomeration mainly involves four aspects: first, the horizontal comparison of the development of urban agglomerations and longitudinal spatiotemporal evolution analysis [22]; second, the scheme optimization and adjustment of urban agglomeration construction [23]; third, the impact of urban agglomeration on economic development, culture, ecological environment and ecological efficiency [24, 25]; fourth, the spatial role of urban agglomerations [26, 27]. Scholars mainly investigate the effects of China's urban agglomeration from macro and meso perspectives. Firstly, from the macro perspective, the actual impact of urban agglomeration on high-quality economic development [28] and the promoting effect of urban agglomeration construction on China's regional development [29] are investigated by taking the Yangtze River Delta, Beijing-Tianjin-Hebei and other urban agglomerations in China as examples. Secondly, from the meso perspective, the promoting effects of Yangtze River Delta urban agglomeration on regional labor productivity [30], total factor productivity, efficiency change and technology change [31] are discussed in-depth.

2.2 Enterprise innovation

Generally, enterprise innovation is affected by external environmental factors and internal control factors. The external environmental factors mainly involve the horizontal competition environment [32], market operation environment [33], thermal comfort environment [34], and policy implementation environment [35] faced by enterprises in the process of innovation. The horizontal competition environment can significantly adjust the performance and innovation spirit of enterprises. Under the precondition that the performance is fixed, the fiercer the competition in the same industry, the lower the willingness of enterprises to innovate [36]. The marketization degree is one of the important indexes to evaluate the market operation environment, and the difference of marketization degree in different regions is often an important reason for the heterogeneity of innovation behaviors [37].

The internal control factors that affect enterprise innovation mainly involve ownership and management [38, 39], and the impact of internal control factors on enterprise innovation is still inconclusive. First, from the perspective of ownership, Chen et al. confirms that the stronger the control over enterprises, the more beneficial it is to encourage enterprises to increase innovation input and improve innovation output [40]. However, some studies find that there is an "inverted U-shaped" relationship between the level of ownership and the enterprise innovation, and too high or too low ownership will reduce the willingness of enterprises to invest in R&D [41], and may hinder the innovation output [42]. Second, from the perspective of management, moderate concentration of management can ensure the efficient execution of enterprises and provide a strong guarantee for enterprises to carry out innovation activities [43]; excessively centralized management may lead to inefficient decision-making quality and further generate a negative impact on enterprise innovation [44]; excessively decentralized management may lead to decision-making conflicts, which in turn inhibits the innovation of enterprises [45].

2.3 Policy evaluation and GRDD

Knowledge resource related to innovation activities play an important role in high-quality economic development. As the government's knowledge resource investment has been rising, the importance of policy evaluation has become pronounced [46, 47]. Regression Discontinuity Design (henceforth, RDD), as a mainstream quasi-experimental approach, has been gradually applied to economics since the late 1990s [48]. Lee and Lemieux summarize the promise that surrounds this design, attributing the recent wave of RDD studies to "the belief that the RDD is not 'just another' evaluation strategy and that causal inferences from RDD are potentially more credible than those from typical 'natural experiment' strategies" [49]. Wang et al. use the RDD to empirically test the influence of the establishment of the Shanghai pilot free-trade zone on the green total factor productivity [50]. Moreover, analysts use RDD to recover experimental benchmarks, which have only bolstered their credibility [51]. By taking the boundaries of regions or regional policies as running variables, RDD can judge individuals in the treatment group or control group by their relative geographical boundary positions, thus constituting the Geographic RDD (GRDD).

One of the earliest and most famous examples of exploiting geographic variation to estimate causal effects is the study by Card and Krueger, who estimate the effect of increasing the minimum wage on employment by comparing fast-food restaurants in New Jersey (where the minimum wage is increased) to restaurants in adjacent eastern Pennsylvania [52]. In political science, political boundaries are often associated with variation in key treatments such as national or state institutions. For example, Posner uses the colonial border between Zambia and Malawi, which is drawn by the British South African Company and split two different ethnic groups, to study the political salience of cultural cleavages [53]. GRDD are an increasingly popular type of natural experiment in political science, and have been recently used to study a variety of topics [54]. Dell is the first to take geographical distance as a running variable, and uses GRDD to explore the impact of the labor service system implemented by the Spanish colonial government in Peru on the local economic development [55]. Subsequently, scholars begin to use GRDD to study Chinese problems. Chen et al use the "Qinling Mountains and Huaihe River" central heating boundary as the geographical cutoff line to study the impact of central heating on air quality, and it is found that coal burning for heating in the north of "Qinling Mountains and Huaihe River" line reduces average life expectancy [56]. Xin and Xu start with a border of the puppet Manchukuo, and deeply discusses the long-term impact of the puppet Manchukuo colonial rule on the regional economy of their jurisdiction [57]. Yu and Wang take the lead in applying GRDD to urban agglomeration performance evaluation [58]. Due to the large expansion scope and strong observability of the YRD urban agglomeration, Deng and Li take the capacity expansion of the YRD urban agglomeration as a quasi-natural experiment by means of GRDD [2]. The actual impact of urban agglomeration expansion in the YRD is identified by comparing the performance changes of 60 cities near the expansion boundary in 2010.

To sum up, there are still the following problems to be improved. First, there are still differences on the delineation of urban agglomeration, and few literatures discuss the micro-enterprise innovation within the scope of urban agglomeration. Second, existing studies mostly use OLS or DID methods to investigate the impact of location-oriented policies such as urban agglomeration, but there is still a lack of analysis on the spatial role of urban agglomeration.

3 Research design

3.1 Setup and notation

GRDD is a design in which a geographic or administrative boundary splits the units into treated and control areas in an as-if random fashion. It is a special case of the RDD with two arbitrary scores (coordinate systems like latitude and longitude). We compare the enterprises in a treated area (YRD and PRD urban agglomerations) to enterprises in a control area, which we denote by A^T and A^C, respectively. We adopt the potential outcomes framework and assume that enterprise i has two potential outcomes, Y_{i1} and Y_{i0}, which correspond to levels of treatment $T_i = 1$ and $T_i = 0$, respectively. In this context, $T_i = 1$ denotes that enterprise i is within YRD and PRD urban agglomerations (A^T) and $T_i = 0$ denotes that enterprise i is out of urban agglomeration (A^C). The observed outcome is $Y_i = T_i Y_{i1}+(1-T_i) Y_{i0}$, and the fundamental problem of causal inference is that we cannot observe both Y_{i1} and Y_{i0} simultaneously for any given enterprise, which implies that we cannot recover the individual effect $\tau_i = Y_{i1}-Y_{i0}$.

We define the variable that uniquely represents enterprise i's geographic location, and allows us to compute enterprise i's distance to any point on the border. We use vectors, in bold, to simplify the notation. The geographic location of enterprise i is given by latitude and longitude, $(S_{i1},S_{i2}) = \mathbf{S}_i$. We consider the set that collects the locations of boundary points $\mathbf{B}$ (cutoff), and denote a single point on the boundary by $\mathbf{b} = (S_1,S_2)\in B$. Thus, A^T and A^C are the sets that collect, respectively, the locations that receive treatment and control. The treatment assignment is a deterministic function of $\mathbf{S}_i$, and can be written as $T_i = T(\mathbf{S}_i)$, with $T(\mathbf{s}) = 1$ for $\mathbf{s}\in A^T$ and with $T(\mathbf{s}) = 0$ for $\mathbf{s}\in A^C$. This assignment has a discontinuity at the known boundary B, and the actual effect of the establishment of YRD and PRD urban agglomerations on the innovation activities of enterprises can be identified by comparing whether the innovation activities of enterprises on both sides of the border jump at the known boundary B.

3.2 Identification

GRDD is a particular case of the two-dimensional RDD, and geography creates a number of complications that are not necessarily common in nongeographic designs. Therefore, it is better to define the running variable (score) S as the shortest distance to the boundary. Enterprises that are close to the boundary in terms of this distance but on opposite sides of it are taken as valid counterfactuals for each other [59]. Specifically, enterprise i has distance $S_i = dist$ if the distance from enterprise i's location to the point on the boundary that is closest to i is equal to $dist$. If $\Pr(T_i = 1) = 1$ for all i such that $\mathbf{s}_i\in A^T$ and $\Pr(T_i = 0) = 1$ for all i such that $\mathbf{s}_i\in A^C$ (the discontinuity is sharp), then

$$\begin{aligned}\tau(\mathbf{b}) &\equiv E\{Y_{i1} - Y_{i0}|\mathbf{S}_i = \mathbf{b}\} \\ &= \lim_{\mathbf{s}^{\mathrm{T}}\to\mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^{\mathrm{T}}\} - \lim_{\mathbf{s}^{\mathrm{C}}\to\mathbf{b}} E\{Y_i|\mathbf{S}_i = \mathbf{s}^{\mathrm{C}}\} \quad \textit{for all } \mathbf{b}\in\mathbf{B}\end{aligned} \tag{1}$$

In other words, we can identify one treatment effect $\tau(\mathbf{b})$ for every point $\mathbf{b}$ on the boundary, defining a treatment effect curve. The average parameter $\tau = E\{Y_{i1}-Y_{i0}|\mathbf{S}_i\in\mathbf{B}\}$ could be obtained by integrating the $\tau(\mathbf{b})$ effects over the entire boundary. Following previous studies [60], we can write the average treatment effect τ as

$$\tau = \int_{\mathbf{s}\in\mathbf{B}} \tau(\mathbf{s}) f(\mathbf{s}|\mathbf{S}\in\mathbf{B}) d\mathbf{s} = \frac{\int_{\mathbf{s}\in\mathbf{B}} \tau(\mathbf{s}) f(\mathbf{s}) d\mathbf{s}}{\int_{\mathbf{s}\in\mathbf{B}} f(\mathbf{s}) d\mathbf{s}} \tag{2}$$

which can be easily recovered once the local effects $\tau(\cdot)$ and the density $f(\cdot)$ are estimated at multiple boundary points.

3.3 Estimation

To estimate a conditional expectation of the outcomes as a function of the distance to the boundary, we follow the existing studies [4] and define $\mu(x) = E(Y|X = x)$ as the regression function of the observed outcome Y on some univariate X. Assume that the first p+1 derivatives of $\mu(x)$ at the point $X = x_0$ exist, we can approximate $\mu(x)$ in a neighborhood of x_0 by a

Taylor expansion:

$$\mu(x) \approx \mu(x_0) + \mu^1(x_0)(x - x_0) + \frac{\mu^2(x_0)}{2}(x - x_0)^2 + \ldots + \frac{\mu^p(x_0)}{p!}(x - x_0)^p \tag{3}$$

where $\mu(x_0)$,$\mu^1(x_0)$,. . .,$\mu^p(x_0)$ denote the first $(p+1)^{th}$ derivatives of $\mu(x_0)$. In local regression estimation, this polynomial is fitted locally, minimizing a weighted sum of squared residuals. The estimated coefficients $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_p)'$ are defined as

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \left[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right]^2 w_i \tag{4}$$

with weight $w_i = \frac{1}{h} K\left(\frac{X_i - x_0}{h}\right)$ for a given kernel function $K(\cdot)$ (for example, uniform kernel: $K(u) = \frac{1}{2}\mathbf{1}\{|u| < 1\}$; triangular kernel: $K(u) = (1 - |u|)\mathbf{1}\{|u| < 1\}$) and bandwidth h. Using this local polynomial estimator, we borrow the basic estimation approach from RDD, which involves estimating the left and right limits of $\mu(c)$, denoted $\mu^l(c)$ and $\mu^r(c)$ respectively, with a local polynomial of degree one, where c is a known cutoff in the score. The estimation of $\mu^l(c)$ uses only observations to the left of c, similarly, estimation of $\mu^r(c)$ uses only observations to the right of the cutoff. For given weights w_i and a scalar score S_i, this involves computing the weighted regression of the observed outcome Y_i on a constant and S_i–c. The estimated effect is then

$$\widehat{\tau} = \widehat{\mu^r(c)} - \widehat{\mu^l(c)} \tag{5}$$

For a given point $\mathbf{b}$ on the boundary, we calculate the Euclidean distance between the location $\mathbf{S}_i$ of enterprise i and $\mathbf{b}$. For every enterprise i in the sample, this distance is defined as $f_{\mathbf{b}}(\mathbf{S}_i)$. Consider

$$\begin{aligned} \mu(\mathbf{b})^{\mathrm{C}} &\equiv \lim_{\mathbf{s}^{\mathrm{C}} \to \mathbf{b}} E\{Y_{i0} | f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s}^{\mathrm{C}})\} \\ \mu(\mathbf{b})^{\mathrm{T}} &\equiv \lim_{\mathbf{s}^{\mathrm{T}} \to \mathbf{b}} E\{Y_{i1} | f_{\mathbf{b}}(\mathbf{S}_i) = f_{\mathbf{b}}(\mathbf{s}^{\mathrm{T}})\} \end{aligned} \tag{6}$$

Local linear regression is used to estimate the above functions and reduce estimation bias [61], namely

$$\begin{aligned} (\widehat{\alpha}_{\mathbf{b}}^{\mathrm{C}}, \widehat{\beta}_{\mathbf{b}}^{\mathrm{C}}) &= \arg\min_{\alpha_{\mathbf{b}}^{\mathrm{C}}, \beta_{\mathbf{b}}^{\mathrm{C}}} \sum\nolimits_{i \in A^{\mathrm{C}}} \{Y_i - \alpha_{\mathbf{b}}^{\mathrm{C}} - \beta_{\mathbf{b}}^{\mathrm{C}} [f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})]^2\} w_{i\mathbf{b}} \\ (\widehat{\alpha}_{\mathbf{b}}^{\mathrm{T}}, \widehat{\beta}_{\mathbf{b}}^{\mathrm{T}}) &= \arg\min_{\alpha_{\mathbf{b}}^{\mathrm{T}}, \beta_{\mathbf{b}}^{\mathrm{T}}} \sum\nolimits_{i \in A^{\mathrm{T}}} \{Y_i - \alpha_{\mathbf{b}}^{\mathrm{T}} - \beta_{\mathbf{b}}^{\mathrm{T}} [f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})]^2\} w_{i\mathbf{b}} \end{aligned} \tag{7}$$

where $w_{i\mathbf{b}} = \frac{1}{h_{\mathbf{b}}} K\left(\frac{f_{\mathbf{b}}(\mathbf{S}_i) - f_{\mathbf{b}}(\mathbf{b})}{h_{\mathbf{b}}}\right)$ are a set of spatial weights, $h_{\mathbf{b}}$ is the bandwidth at the boundary point $\mathbf{b}$. Kernel function $K(\cdot)$ mainly includes triangular kernel, uniform kernel and epanechnikov kernel. Since triangular kernel is more suitable for boundary estimation in local linear regression [49], we use triangular kernel function to minimize the above equations, and test the robustness with other kernel functions.

4 Data and preliminary exploration

This study takes listed enterprises within 100 kilometers on both sides of the boundary of YRD and PRD urban agglomerations in China from 2007 to 2019 as research samples. There are two reasons for selecting enterprises within 100km inside and outside the boundary of urban agglomeration as research samples: first, 100km inside the boundary of the urban agglomeration can cover all listed enterprises within the urban agglomeration. If the distance is too small (for example, 50km), some enterprises within urban agglomeration will not be included in the analysis; second, when designing the geographic regression discontinuity, the distance between the two sides of the boundary should be symmetrical far as possible.

4.1 Data source and variable description

The research data in this study consist of micro-enterprise data and urban-agglomeration-map data. Micro-enterprise data stem from the China Stock Market & Accounting Research (CSMAR) database (http://www.csmar.com/), mainly involving the China's A-share listed enterprises from 2007 to 2019. Specifically, the patent data (such as patent applications, invention patent applications and patent grants) of enterprises comes from the "Patent Database of Listed Companies and Subsidiaries" in CSMAR database. Considering that the index of research and development investment in CSMAR database changed its statistical caliber in 2007, sample data after 2007 are selected for analysis to ensure the consistency of data statistical caliber. In addition, this study also processes the original data as follows: First, we eliminate ST, PT and samples with variable deletion. Secondly, in order to ensure the comparability of samples before and after the policy of YRD and PRD urban agglomerations, we eliminate enterprises that are listed after 2007 and delisted before 2019. Urban-agglomeration-map data stem from the "*Yangtze River Delta Urban Agglomeration Development Plan*" approved by the Chinese government in 2016 and the "*Pearl River Delta Urban Agglomeration Coordinated Development Plan (2004–2020)*" approved by the Chinese government in 2004. Fig 1 shows the location information of China's listed enterprises and the boundary information about YRD and PRD urban agglomerations. The symbol and definition of the variables involved in this study are summarized in Table 1.

In this study, the quantity of patent applications (*RDoutput*) is chosen to measure the innovation of enterprises for two reasons: first, the quantity of patent applications is not easily interfered by external factors, such as bureaucratic factors, patent maintenance fees; second, the patent application data is easy to obtain, and can be used as a stable and objective standard to effectively measure the innovation [62].

Running variable is set as the shortest distance (*dist*) from the longitude and latitude coordinates of the geographical location of the enterprise to the boundary of urban agglomerations. To obtain the shortest distance for each enterprise, we adopt the following steps. First, the shape of YRD and PRD urban agglomerations defined by government is projected to the ArcGIS software (version 10.5), to obtain the boundary information about urban agglomeration. Second, the geocoding technique, which is the process of converting addresses of enterprises into a coordinate system (typically latitude and longitude), is used to help us to know the precise location of enterprises. Third, we use a buffer to identify which enterprises are within 100km from the border on either side. Fourth, we calculate the shortest distance from each enterprise location to the boundary of urban agglomerations by ArcGIS software, and develop a running variable (score) that reflects the two-dimensional geographic space.

Control variables mainly cover enterprise operating status, enterprise profitability, and owner's equity of enterprises. Operating status usually affect the innovation behavior and willingness of enterprises. Considering that operating income can reflect the change of short-term demand, the operating income processed by logarithm (*sale*) is used in this study to measure the operating status of enterprises. Enterprise innovation needs a large amount of capital investment. Generally, the higher the profit rate, the higher the level of innovation [63]. Thus, we use the logarithm of net profit (*profit*) to measure the profitability of enterprises. Owners' equity is the portion of an enterprise's assets that belongs to the owner after the deduction of creditors' equity, which can reflect the case of owner investment capital preservation and

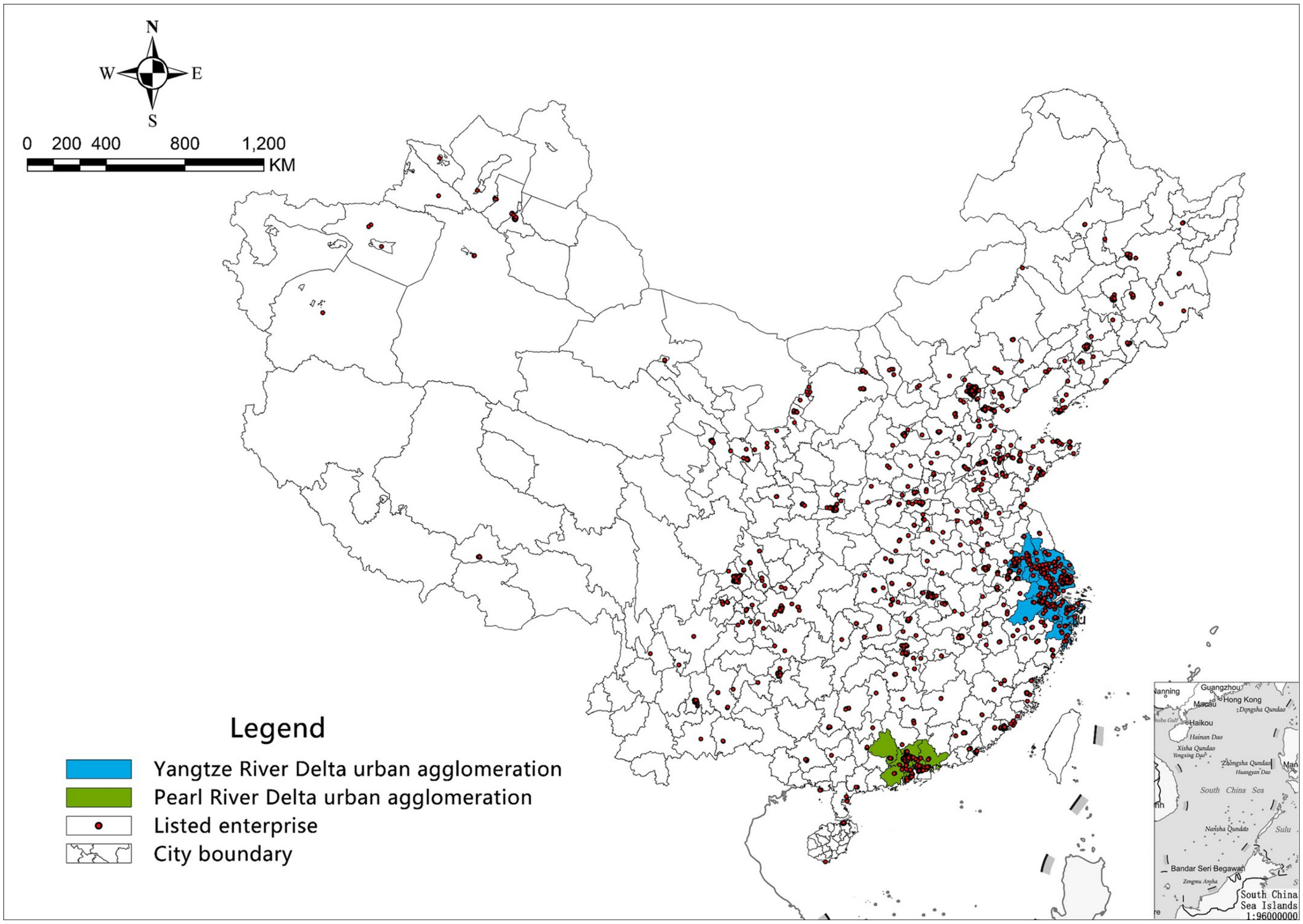

Fig 1. Urban agglomerations and listed enterprises.

https://doi.org/10.1371/journal.pone.0273154.g001

appreciation. We use the owners' capital input, other comprehensive income and retained income to measure the owners' equity (*equity*) of enterprises.

It is worth noting that there are differences in the distribution of enterprise samples on both sides of the boundary of urban agglomerations, specifically more samples within urban agglomeration and less samples outside urban agglomeration. Therefore, we randomly select 288 listed enterprises distributed on both sides of the boundary of urban agglomeration and less than 100 kilometers away from the boundary. After random sampling, there are 149 listed enterprises (109 enterprises in YRD urban agglomeration, 40 enterprises in PRD urban agglomeration) in the treatment group and 139 listed enterprises (121 enterprises in YRD

Table 1. Variable definitions.

Type	Name	Symbol	Definition
Outcome variable	Enterprise innovation	*RDoutput*	Ln (number of patent applications +1)
Running variable	Score	*dist*	The shortest distance from the location of enterprises to the boundary of urban agglomerations
Control variable	Operating status	*sale*	Ln (operation income)
	Profitability	*profit*	Ln (net profits)
	Owner's equity	*equity*	Ln (total owners' equity)

https://doi.org/10.1371/journal.pone.0273154.t001

urban agglomeration, 18 enterprises in PRD urban agglomeration) in the control group. Details are shown in Fig 2.

Table 2 lists descriptive statistics for the main variables. It is worthwhile to note that according to $dist_{-}$, there are 139 samples of listed enterprises outside the demarcated scope of YRD and PRD urban agglomerations, and the number of enterprises within YRD and PRD urban agglomerations is 149 according to $dist_{+}$.

4.2 Preliminary exploration

GRDD can be tested as a randomized experiment. Therefore, before in-depth analysis, we should first observe the breakpoint caused by the policy of the establishment of YRD and PRD urban agglomerations, to verify whether the outcome variable will show systematic changes due to the urban agglomeration setup. Fig 3 shows the relationship between the outcome variable (*RDoutput*) and the running variable (*dist*). It can be seen that enterprise innovation has an obvious jump at the cutoff. This indicates that the innovation level of enterprises in treated group near the boundary of urban agglomerations is significantly higher than that in control group.

Besides, predetermined covariates (control variables) can be treated as outcomes using the similar approach outlined above to verify whether the change of enterprise innovation is related to the characteristics of enterprises. We would hope to find that there is no obvious jump for operating status (*sale*) and other covariates (*profit*,*equity*) at cutoff (boundary line). Fig 4 shows the changes of covariates. The scatter points are the average value of covariates, and the straight line is the regression fitting value of all the scatter points on both sides of the cutoff. It is shown that all control variables have continuity at the cutoff, and the difference in enterprise innovation at cutoff is not due to the influence of control variables.

5 Empirical discussion

5.1 Impact of urban agglomeration construction on enterprise innovation

Considering the operating status, profitability, and owners' equity of enterprises shown in Fig 4 meet the continuity assumption, we introduce the abovementioned control variables into the model estimation. By this way, the actual impact of urban agglomeration establishment on enterprise innovation can be more accurately obtained, and the estimated values is closer to the real values. The specific estimated results are shown in Table 3.

Column (1) of Table 3 shows the estimated results without control variables, columns (2) to (4) are the estimates for introducing control variables and using different kernel functions. It is shown that at the statistical level of 5% or 10%, the average treatment effect (ATE) of urban agglomeration construction on enterprise innovation is significantly positive, and the estimated values of this jump remains between 0.753 and 0.921. No matter whether to introduce control variable or change the form of kernel function, the estimated results are basically consistent, which indirectly confirms the robustness of the study. On the whole, this finding indicates that the urban agglomeration construction can significantly improve the innovation of enterprises within the urban agglomeration. The reason behind this may be that the construction of urban agglomeration is conducive to the formation of agglomeration effect among enterprises. Agglomeration effect can be divided into specialized economic effect and diversified economic effect. The former refers to the gain of enterprises in the same industry due to spatial agglomeration, while the latter refers to the benefit from different industries due to spatial agglomeration. Enterprises in the same industry gather in urban agglomeration, which can bring many benefits to enterprises [64]. First, the agglomeration of enterprises in the same industry can generate knowledge spillover effect [65], which will reduce the learning cost

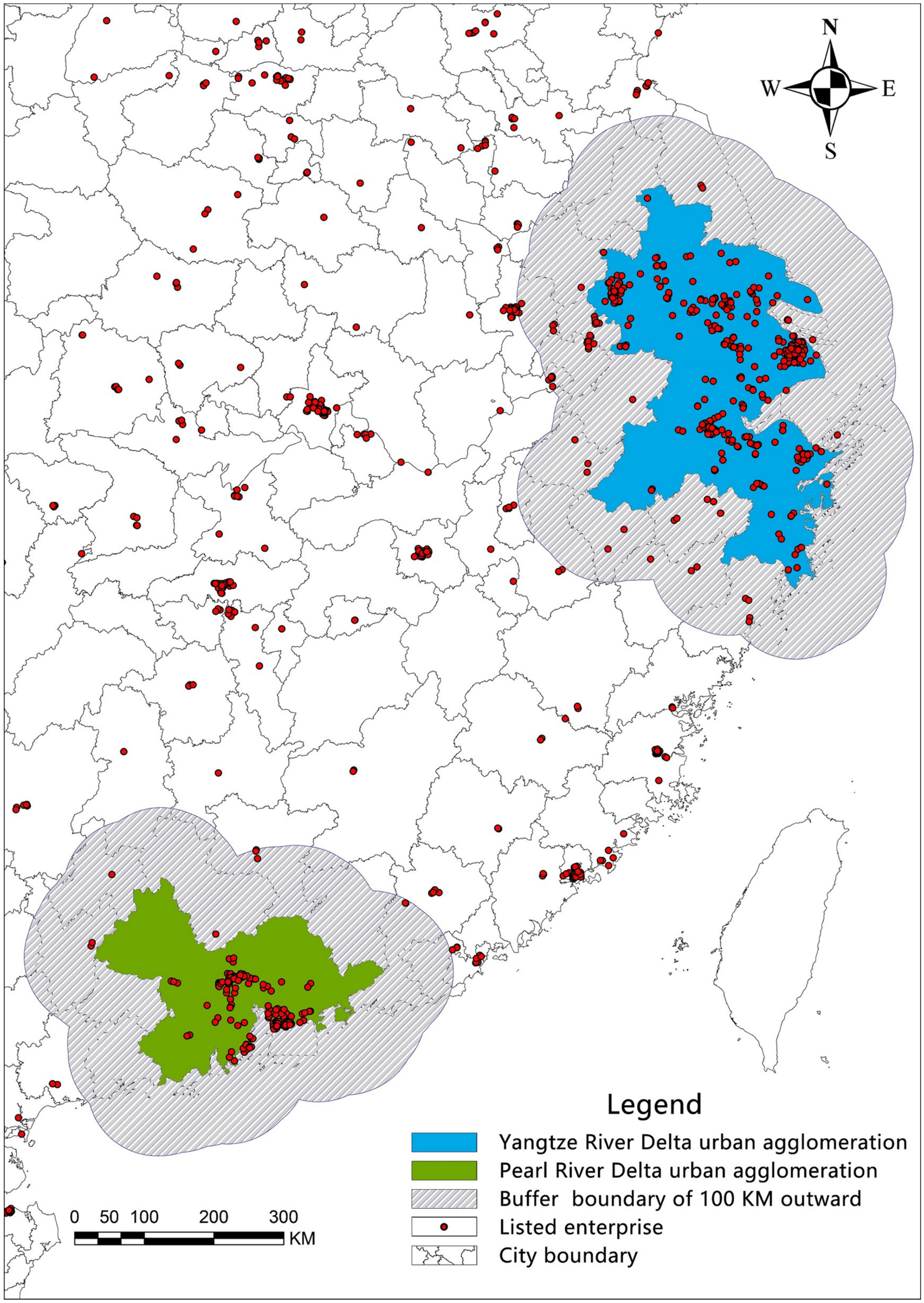

Fig 2. Boundary and buffer of urban agglomerations.

https://doi.org/10.1371/journal.pone.0273154.g002

Table 2. Descriptive statistics.

Name	Mean	Median	Variance	Minimum	Maximum	Observation
RDoutput	3.56	3.95	2.21	0	10.68	288
dist	-13.64	0.90	35.18	-96.07	35.63	288
$dist_-$	-43.88	-52.20	25.86	-96.07	-1.90	139
$dist_+$	14.58	13.15	10.77	0.09	35.63	149
sale	22.13	22.06	1.279	19.09	26.47	287
profit	21.76	21.71	1.115	19.41	25.49	287
equity	19.34	19.20	1.37	15.45	23.85	269

https://doi.org/10.1371/journal.pone.0273154.t002

among enterprises and accelerate the speed of knowledge updating [66]. Second, the agglomeration of enterprises in the same industry is also conducive to realizing economies of scale. The convenience of raw material supply and customer group search will be greatly increased, thus reducing the production cost and sales cost of unit product [67]. Third, enterprises in the same industry have a high similarity in the demand for professional personnel. The agglomeration of enterprises expands the development space of professionals, which is conducive to attracting a large number of talents, thus making it easier for enterprises to gather all kinds of technical personnel needed [68]. Fourth, the similarity in the use of factories and equipment among enterprises in the same industry makes it easier to realize the special assets invested by enterprises, which will greatly increase the financial flexibility of enterprises and better meet the capital needs of enterprise innovation [69]. It is thus clear that the specialized economic effects produced by urban agglomeration provide superior conditions for enterprise innovation [70]. The agglomeration of enterprises in different industries in the same urban agglomeration will also generate increasing benefits, even more than that in the same industry [71], namely Jacobs external economy. First, infrastructure can be shared among diversified industries, for

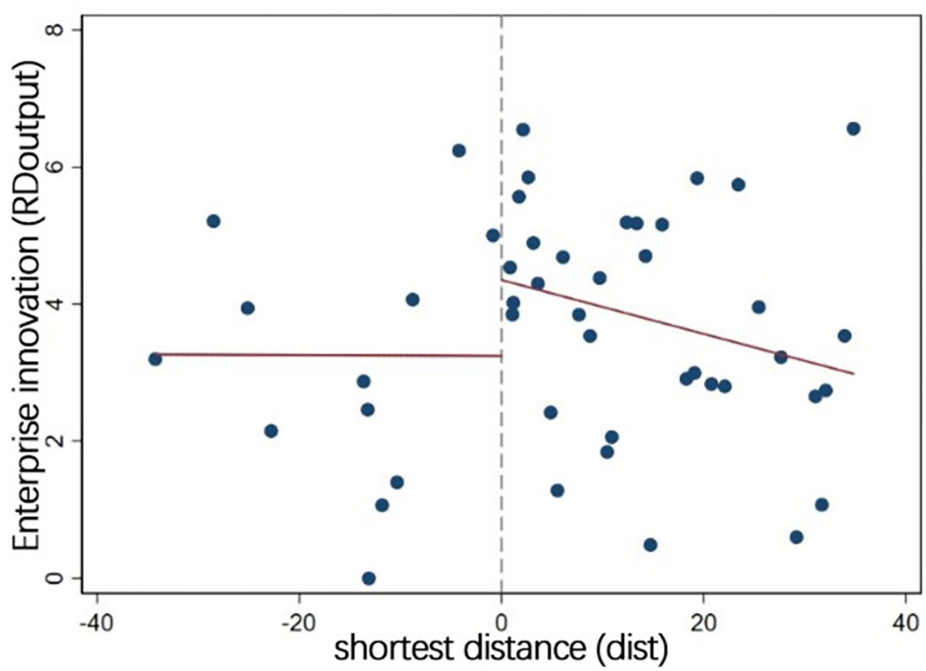

Fig 3. Changes of enterprise innovation at the cutoff.

https://doi.org/10.1371/journal.pone.0273154.g003

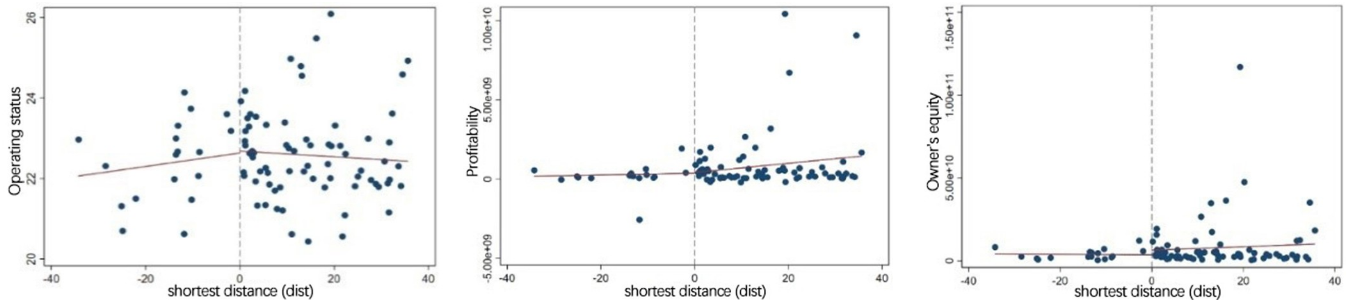

Fig 4. Changes of covariates at the cutoff. a (Operating status). b (Profitability). c (Owner's equity).

https://doi.org/10.1371/journal.pone.0273154.g004

example, more banking institutions make corporate financing more convenient and cheaper. Second, the agglomeration of enterprises from different industries in the same urban agglomeration will produce spillover effects of knowledge and technology, and the knowledge spillover or networks could improve the innovation performance of enterprises [72], even lead to the breakthrough innovation [73].

5.2 Robustness check

To verify the validity of the previous estimates and prove that the GRDD results do not depend on the special settings of the model, we will conduct the following robustness tests: (1) test for precise control of running variable by enterprises; (2) test for the continuity of control variables; (3) test for bandwidth sensitivity; (4) placebo test with pseudo-boundary; (5) extreme value test.

(1) Test for precise control of running variable by enterprises. The premise of GRDD is that the groups on both sides of cutoff (line) have randomness. If sample enterprises can manipulate and even select groups, the estimated results of GRDD will be invalid. To ensure the accuracy and validity of GRDD, it is necessary to test whether the sample enterprise can control the running variable. If the GRDD is valid (there is no manipulation around the cutoff), then there should be no discontinuity observed in the number of observations just above or below the cutoff. We use the nonparametric statistical test provided by McCrary to check whether the distribution of running variable is continuous [74]. Fig 5 depicts the running variable does not jump significantly at the cutoff, and the shortest distance from the enterprise location to the urban agglomeration boundary is continuously distributed. This indicates that

Table 3. Estimated results of GRDD.

Outcome	Enterprise innovation			
	(1)	(2)	(3)	(4)
ATE	0.831*	0.753*	0.913**	0.921*
	(0.4716)	(0.4294)	(0.4559)	(0.4794)
Kernel function	Tri	Tri	Epa	Uni
Bandwidth (km)	50	50	50	50
Control variable	no	Yes	yes	yes
Observation	211	197	197	197

Note

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses; Kernel function "Tri" stands for Triangular; "Epa" stands for Epanechnikov; "Uni" stands for Uniform.

https://doi.org/10.1371/journal.pone.0273154.t003

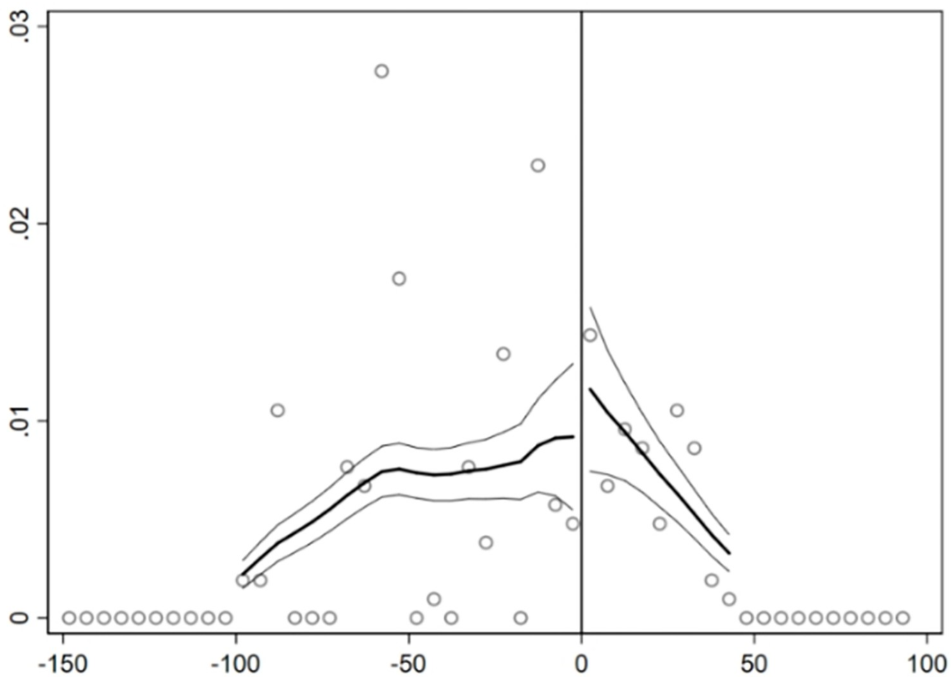

Fig 5. Distribution of running variable.

https://doi.org/10.1371/journal.pone.0273154.g005

the individual enterprise cannot accurately control the running variable, the research design meets the assumptions of GRDD, and the estimated results are valid.

(2) Test for the continuity of control variables. To ensure that the urban agglomeration establishment is the cause of the significant change of enterprise innovation, the continuity of the conditional density of control variables near the cutoff should be guaranteed. If the control variable jumps significantly at the cutoff, it means that the jump of enterprise innovation at the cutoff cannot be attributed to the establishment of urban agglomeration. On the basis of Fig 4, we use RDD to further empirically check whether enterprise characteristics change significantly at the cutoff, and to verify the continuity of covariate. Table 4 reports the estimated results of the jump at the cutoff for control variables (covariate). By using three different kernel functions (Triangular, Epanechnikov, and Uniform), we find that the estimated results for all

Table 4. Continuity test of control variables.

	Control variables		
Kernel function	***sale***	***equity***	***profit***
Triangular	0.179	0.225	0.535
	(0.3362)	(0.2376)	(0.3512)
Epanechnikov	0.236	0.331	0.0720
	(0.3249)	(0.2405)	(0.6224)
Uniform	0.315	0.443	-0.468
	(0.3218)	(0.2472)	(0.7253)

Note

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses.

https://doi.org/10.1371/journal.pone.0273154.t004

Table 5. Sensitivity test to bandwidth choice.

	(1) BW-4	(2) BW-2	(3) BW	(4) BW+2	(5) BW+4
ATE	0.720*	0.738*	0.753*	0.765*	0.818*
	(0.4235)	(0.4266)	(0.4294)	(0.4319)	(0.4202)
Bandwidth selection	46	48	50	52	54
Control variable	Yes	yes	yes	yes	yes
Observation	204	206	211	212	220

Note

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses.

https://doi.org/10.1371/journal.pone.0273154.t005

control variables at the cutoff are insignificant. Therefore, there is no empirical evidence that these control variables are discontinuous at the cutoff, which satisfies the validity assumption of GRDD.

(3) Test for bandwidth sensitivity. Since there is a "bias-variance tradeoff", the choice of bandwidth is fundamental for the analysis and interpretation of GRDD. Table 5 reports the estimated results by using different bandwidths. "BW" represents the initial bandwidth; "BW +1" means adding 1 to the initial bandwidth and "BW+2" means adding 2 to the initial bandwidth, and so on. The test results show that despite the change of bandwidth, the estimated ATEs are significant at the statistical level of 10%, and the difference of estimated values is small (between 0.720 and 0.818). It can be considered that the bandwidth change has no significant impact on the estimated results, and the findings by GRDD shown in Table 3 are robust.

(4) Placebo test. This falsification test replaces the true boundary by another pseudo-boundary at which the treatment status does not really change, and performs estimation and inference using this fake or placebo cutoff line. The expectation is that no significant treatment effect will occur at placebo cutoff line (pseudo-boundary). If the outcome variable at the pseudo-boundary has a significant jump, it can be considered that some unobserved factors may affect the outcome variable. To this end, we define some pseudo-boundaries, such as "real boundary -1km", "real boundary +1km", "real boundary -2km" and "real boundary +2km". The estimated results based on these four pseudo-boundaries are sorted out in Table 6. It is shown that the ATEs are not statistically significant, and the outcome of interest does not jump discontinuously at the artificial cutoffs considered.

(5) Extreme value test. The sample can be limited to a value of running variable between the upper and lower quartiles, and the extreme value test is carried out by eliminating potential extreme enterprises. GRDD is used to estimate the policy effect with three different kernel functions, and the estimated results are shown in Table 7.

Table 6. Placebo test with pseudo-boundary.

	(1)	(2)	(3)	(4)
ATE	-0.552	0.326	-1.378	0.065
	(0.5735)	(0.3618)	(0.9121)	(0.3559)
Pseudo-boundary	real boundary -1km	real boundary +1km	real boundary -2km	real boundary +2km
Optimal bandwidth (km)	8.8	6.3	9.3	5.3
Kernel function	Triangular	Triangular	Triangular	Triangular
Observation	62	53	63	52

Note

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses.

https://doi.org/10.1371/journal.pone.0273154.t006

Table 7. Extreme value test.

	(1)	(2)	(3)	(4)
ATE	1.160**	1.050**	1.201**	1.211**
	(0.5028)	(0.4622)	(0.4901)	(0.5111)
Kernel function	Triangular	Triangular	Epanechnikov	Uniform
Bandwidth (km)	50	50	50	50
Control variable	no	yes	yes	yes
Observation	139	127	127	127

Note

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses.

https://doi.org/10.1371/journal.pone.0273154.t007

According to results shown in Table 7, the ATEs of the establishment of urban agglomeration on enterprise innovation have some changes compared with the full sample, but the differences are not obvious. The results are not affected by the extreme values, which confirms the reliability of the GRDD estimation in this study.

5.3 Influence mechanism of state-level urban agglomeration

Considering that the establishment time of state-level urban agglomerations is different, we follow the practice of Hoynes et al. [75], and adopt Staggered DID to explore the influence mechanisms of state-level urban agglomerations promoting enterprise innovation, specific as follows.

$$RDoutput_{it} = \alpha_0 + \alpha_1 \cdot treat_{it} + \alpha_2 \cdot treat_{it} \times \mathbf{M} + \mathbf{X_{it} g} + u_i + \lambda_t + \varepsilon_{it} \tag{8}$$

where i indexes the enterprise, t represents the year, $RDoutput$ is enterprise innovation. $\mathbf{X}$ is a set of control variables that change with time (shown in Table 1). u_i is the individual effect, λ_t is the time fixed effect, and ε_{it} is the random disturbance term. $treat_{it}$ represents the policy treatment variable. If the enterprise i is within the scope of the state-level urban agglomeration in t year, this enterprise will be in the treatment period, and the value *treat* of the current year (t) and the subsequent period (t+1,t+2,. . .) will be 1, otherwise 0. $\mathbf{M}$ is the mediator variable which stands for financial support channel (*sub*) and regional coordination channel (*coor*), respectively. *sub* is calculated by natural logarithm of government subsidies received by each enterprise plus 1, and the data of government subsidies stem from CSMAR database and annual reports of listed enterprises. As far as the measure of regional coordination effect is concerned, for enterprises out of urban agglomeration, *coor* is measured by the attractiveness score of the city where enterprises are located; for enterprises within urban agglomeration, *coor* is measured by the average score of all node cities in this urban agglomeration. The data of attractiveness score are compiled from the annual "Ranking of Cites' Business Attractiveness in China" published by Yicai (https://www.yicai.com/).

Column (1) of Table 8 reports the impact of state-level urban agglomeration construction on enterprise innovation using Staggered DID method. The estimation results show that urban agglomeration can significantly promote enterprise innovation, which further verifies the previous findings. According to the results in column (2), the coefficient of cross-term *treat*×*sub* is significantly positive, which proves that urban agglomeration has significant financial support effect on enterprise innovation. The financial support (for example, subsidy) enjoyed by enterprises in urban agglomeration is an important driving force to encourage enterprise innovation. Enterprises in state-level urban agglomerations often have more opportunities to get financial support from the government, which stimulates them to carry out

Table 8. Influence mechanism.

	(1)	(2)	(3)
treat	1.268***	1.032***	1.041***
	(0.0773)	(0.2909)	(0.2706)
treat×sub		0.079*** (0.0171)	
treat×coor			0.763*** (0.1511)
Control variable	yes	Yes	yes
Adj R2	0.129	0.131	0.132
Observation[a]	3288	3288	3288

Note

[a] no sampling process and no control of the distance between sample enterprises and the boundary of urban agglomeration

***, **, and * represent the significance levels at 1%, 5%, and 10%, respectively; robust standard error is presented in parentheses.

https://doi.org/10.1371/journal.pone.0273154.t008

innovation activities [76]. In addition, financial support will indirectly guide the flow of financial resources into urban agglomerations (for example, banks will be more willing to provide low-interest loans to enterprises within urban agglomerations), thus improving innovation by reducing the financing costs of enterprises. Column (3) reports the effect of cross-term *treat×coor* on enterprise innovation. The results show that urban agglomeration promotes enterprise innovation through regional coordination channel. The construction of urban agglomeration not only expands the market scale, but also broadens the scope of innovation resources flow, greatly promoting the flow of talents, capital and information in the region [77]. On this basis, the agglomeration of innovation resources and the acceleration of their flow force enterprises in urban agglomerations to improve their R&D modes, thus enhancing their innovation ability.

6 Conclusion

At present, China's high-quality development needs to be driven by point to area policy, and the establishment of urban agglomeration is undoubtedly in line with this. The YRD is an important intersection of the Belt and Road (B&R) and the Yangtze River Economic Belt; the PRD is one of the most dynamic economic zones in the Asia-Pacific region. Both of them are bases for scientific and technological innovation, important platforms for China to participate in economic globalization, strategic areas for China's modernization. This study evaluates the impact of YRD and PRD urban agglomeration establishment on enterprise innovation measured by the number of patents using the quasi-natural experimental method GRDD. Unlike the existing literature focusing on regional outcomes, this study explores the impact on individual enterprises based on the data of Chinese listed enterprises. The shortest distances from the enterprise location to the boundary of urban agglomerations are calculated by ArcGIS, and we consider the shortest distance as the running variable. Compared with other methods, the causal inference of GRDD is clear, and the assumptions are easy to test. It is shown that the establishment of YRD and PRD urban agglomerations can significantly improve the enterprise innovation, and this outcome is verified by some robustness tests including bandwidth sensitivity test, placebo test, extreme value test, *etc*. In addition, the influence mechanisms of state-level urban agglomerations promoting enterprise innovation are explored by Staggered DID. It is confirmed that the urban agglomeration construction can promote enterprise innovation through financial support and regional coordination channels.

Overall, the establishment of YRD and PRD urban agglomerations achieves good results in stimulating the innovation of enterprises. It is an important way for China to implement and promote this innovation-driven and place-based strategy to weaken the restrictions of administrative divisions. Although some cities in YRD and PRD are still in an environment with relatively backward resources, the establishment of urban agglomeration can integrate and share the resources within the region, which is conducive to the enterprise innovation and promote the high-quality development. To further strengthen the positive effect of YRD and PRD urban agglomeration on enterprise innovation, relevant government departments should weaken the restrictions of traditional administrative divisions, take urban agglomerations as regional units [78], gather innovative talents in science and technology, optimize the allocation of innovation resources in urban agglomerations and improve infrastructure construction. In addition, the policy implementation environment [79], the thermal comfort environment [80, 81], the infrastructure environment [82, 83] and the innovation environment [84, 85] have important influence on enterprise innovation. Therefore, relevant government departments should build the relatively fair and transparent policy implementation environment, improve the quality of infrastructure, and strengthen Industry-University-Research cooperation, to further promote enterprise innovation. In particular, the relevant departments should pay attention to promoting the steady development of enterprise innovation by building an innovation-driven policy system with urban agglomeration as the carrier. Besides, relevant departments can ameliorate the innovative mode of YRD and PRD urban agglomeration, and take advantage of the "joint development" and "relying" mode of urban agglomeration integration to achieve "introduction, absorption and re-innovation".

It is unavoidable that some shortcomings remain in this study. First, the measurement of enterprise innovation is rather rough, without considering the differences in patents (for example patent for Invention, patent for Utility Model and patent for Industrial Design). The follow-up research can make a more detailed division of enterprise innovation according to Chinese patent classification standards, so as to reflect the differences in quantity and quality of enterprise innovation. Second, the industries to which the enterprises belong are not refined. In the future, we can further study the influence of YRD and PRD urban agglomerations on innovation activities of enterprises in different industries, and explore the impact of urban agglomeration construction for different industries. Third, the establishment of YRD and PRD urban agglomerations may generate heterogeneous effects on enterprise innovation, it will be better that the different regional policies and the differences of GRDD results can be highlighted in the future study. Fourth, there is a lack of investigation on the dynamic effect of YRD and PRD urban agglomerations. In the future, a variety of causal identification methods can be used to evaluate the actual effect of the "development" and "capacity expansion" of urban agglomeration, and explore the radiation effect of YRD and PRD urban agglomeration on the innovation activities of enterprises outside the boundary of urban agglomerations.

Author Contributions

Conceptualization: Kai Zhao, Huahua Huang.

Data curation: Huahua Huang.

Formal analysis: Kai Zhao, Huahua Huang.

Funding acquisition: Kai Zhao.

Methodology: Kai Zhao, Wanshu Wu.

Software: Huahua Huang.

Supervision: Wanshu Wu.

Visualization: Wanshu Wu.

Writing – original draft: Kai Zhao, Huahua Huang.

Writing – review & editing: Kai Zhao, Wanshu Wu.

References

1. Fang C L, Yu D L. Urban agglomeration: An evolving concept of an emerging phenomenon. *Landscape and Urban Planning*. 2017; 162, 126–136. https://doi.org/10.1016/j.landurbplan.2017.02.014
2. Deng H H, Li H R. Regional integration and enterprise growth: A micro perspective of domestic circulation. *Economic Review*. 2021; 3, 3–17.
3. Zhao Y, Mao J, Li Y. Local governments' environmental emphasis and corporate green innovation: evidence from China. *Economic Change and Restructuring*. 2022; https://doi.org/10.1007/s10644-022-09406-3
4. Keele L J, Titiunik R. Geographic boundaries as regression discontinuities. *Political Analysis*. 2015; 23 (1), 127–155. https://doi.org/10.1093/pan/ mpu014
5. Ejdemo T, Örtqvist D. Exploring a leading and lagging regions dichotomy: Does entrepreneurship and diversity explain it? *Journal of Innovation and Entrepreneurship*. 2021; 10, 6. https://doi.org/10.1186/s13731-021-00146-8
6. Gottmann J. Megalopolis or the urbanization of the northeastern seaboard. *Economic Geography*. 1957; 33(3), 189–200. https://doi.org/10.2307/142307
7. Zhang X C, Guo Q E, Cheung D M W, Zhang T Y. Evaluating the institutional performance of the Pearl River Delta integration policy through intercity cooperation network analysis. *Cities*. 2018; 81, 131–144. https://doi.org/10.1016/j.cities.2018.04.002
8. Phelps N A, Ozawa T. Contrasts in agglomeration: proto-industrial, industrial and post-industrial forms compared. *Progress in Human Geography*. 2003; 27(5), 583–604. https://doi.org/10.1191/0309132503ph449oa
9. Alonso-Villar O. Urban agglomeration: knowledge spillovers and product diversity. *Annales of Regional Science*.2002; 36(4), 551–573. https://doi.org/10.1007/s001680200090
10. Musterd S, Bontje M, Ostendorf W. The changing role of old and new urban centers: the case of Amsterdam region. *Urban Geography*. 2006; 27(4), 360–387. https://doi.org/10.2747/0272-3638.27.4.360
11. Lang R, Knox P K. The new metropolis: rethinking megalopolis. *Regional Studies*. 2009; 43(6), 789–802. https://doi.org/10.1080/00343400701654251
12. Lagarias A. Urban sprawl simulation linking macro-scale processes to micro-dynamics through cellular automata, an application in Thessaloniki, Greece. *Applied Geography*. 2012; 34, 146–160. https://doi.org/10.1016/j.apgeog.2011.10.018
13. Marull J, Galletto V, Domene E, Trullen J. Emerging megaregions: a new spatial scale to explore urban sustainability. *Land Use Policy*. 2013; 34, 353–366. https://doi.org/10.1016/j.landusepol.2013.04.008
14. Wang K Y, Deng Y, Sun D W, Song T. Evolution and spatial patterns of spheres of urban influence in China. *Chinese Geographical Science*. 2014; 24(1), 126–136.
15. Florida R, Gulden T, Mellander C. The rise of the mega-region. *Cambridge Journal of Regions Economy and Society*. 2008; 1, 469–476. https://doi.org/10.1093/cjres/rsn018
16. De Goei B, Burger M J, Van Oort F G, Kitson M. Functional polycentrism and urban network development in the Greater South East, United Kingdom: Evidence from commuting patterns, 1981–2001. *Regional Studies*. 2010; 44(9), 1149–1170. https://doi.org/10.1080/00343400903365102
17. Huang X L, Li G P, Liu J G. The impact of spatial structure on economic efficiency of Beijing-Tianjin-Hebei megalopolis in China. *Complexity*. 2020; 18, 6038270. https://doi.org/10.1155/2020/6038270
18. Meijers E J, Burger M J. Stretching the concept of "Borrowed Size". *Urban Studies*. 2017; 54(1), 269–291. https://doi.org/10.1177/0042098015597642
19. Liu Y L, Zhang X H, Pan X Y, Ma X X, Tang M Y. The spatial integration and coordinated industrial development of urban agglomerations in Yangtze River Economic Belt, China. *Cities*. 2020; 104, 102801. https://doi.org/10.1016/j.cities.2020.102801
20. Tian Z Z, Fang D L, Chen B. Three-scale input-output analysis for energy and water consumption in urban agglomeration. *Journal of Cleaner Production*. 2020; 268, 122148. https://doi.org/10.1016/j.jclepro.2020.122148

21. Chen M X, Guo S S, Hu M G, Zhang X P. The spatiotemporal evolution of population exposure to PM2.5 within the Beijing-Tianjin-Hebei urban agglomeration, China. *Journal of Cleaner Production*. 2020; 265, 121708. https://doi.org/10.1016/j.jclepro.2020.121708

22. Lyu Y M, Jiang F Y. Spatial and temporal distribution of population in urban agglomerations changes in China. *Scientific Reports*. 2022; 12, 8315. https://doi.org/10.1038/s41598-022-12274-6 PMID: 35585107

23. Gao X L, Xu Z N, Niu F Q, Long Y. An evaluation of China's urban agglomeration development from the spatial perspective. *Spatial Statistics*. 2017; 21, 475–491. https://doi.org/10.1016/j.spasta.2017.02.008

24. Yang Z D. Comparison and empirical analysis of the urban economic development level in the Yangtze River urban agglomeration based on an analogical ecosystem perspective. *Ecological Informatics*. 2021; 64, 101321. https://doi.org/10.1016/j.ecoinf.2021.101321

25. Zeng P, Wei X, Duan Z C. Coupling and coordination analysis in urban agglomerations of China: Urbanization and ecological security perspectives. *Journal of Cleaner Production*. 2022; 365, 132730. https://doi.org/10.1016/j.jclepro.2022.132730

26. Yu H S, Yang J, Sun D Q, Li T, Liu Y J. Spatial responses of ecosystem service value during the development of urban agglomerations. *Land*. 2022; 11, 165. https://doi.org/10.3390/land11020165

27. Zhu X H, Wang Q L, Zhang P F, Yu Y J, Xie L L. Optimizing the spatial structure of urban agglomeration: based on social network analysis. *Quality & Quantity*. 2021; 55, 683–705. https://doi.org/10.1007/s11135-020-01016-3

28. Zhang Y, Liu L, Huang S J. Does regional integration promote the high-quality development of urban agglomeration economy: A quasi-natural experiment based on the Yangtze river delta urban economic coordination commission. *Studies in Science of Science*. 2021; 39(1), 63–72.

29. Yu W J, Zhao D L. The rise of urban cluster, economic performance and regional income disparity based on the metropolitan area of Jing-Jin-Ji, Yangtze River Delta and Pearl River Delta. *Journal of Zhongnan University of Economics and Law*. 2010; 4, 15–20.

30. Zhang X L, Li P X, Li L X. Government cooperation, market integration and economic performance of city cluster: Evidence from the Yangtze River Delta urban economic coordination committee. *China Economic Quarterly*. 2017; 16(4), 1563–1582.

31. Li Z, Jiang W G, Wang W J, Chen Z, Ling Z Y, Lv J X. Ecological risk assessment of the wetlands in Beijing-Tianjin-Hebei urban agglomeration. *Ecological Indicators*. 2020; 117, 106677. https://doi.org/10.1016/j.ecolind.2020.106677

32. Seddighi H, Mathew S. Innovation and regional development via the firm's core competence: Some recent evidence from North East England. *Journal of Innovation & Knowledge*. 2020; 5(4), 219–227. https://doi.org/10.1016/j.jik.2019.12

33. Chen L, Wu B D. Marketization, education and family business's R&D investment. *Science Research Management*. 2014; 35(7), 44–50. https://doi.org/10.19571/j.cnki.1000-2995.2014.07.006

34. Yang J, Xin J X, Zhang Y Q, Xiao X M, Xia J H. Contributions of sea-land breeze and local climate zones to daytime and nighttime heat island intensity. *NPJ Urban Sustainability*. 2022; 2, 12. https://doi.org/10.1038/s42949-022-00055-z

35. Zhao K, Wu W S, Ye J M. The impact of "Mass Entrepreneurship and Innovation" policy on SMEs' innovation: using quasi-natural experiments. *Bulletin of Economic Research*.2022; 1–18. https://doi.org/10.1111/boer.12359

36. Li J, He X G, Lian Y L, Lv F F. Performance driven, market transition and innovative spirit of family business. *Management Review*. 2016; 28(1), 96–108. https://doi.org/10.14120/j.cnki.cn11-5057/f.2016.01.009

37. Chen W T, Li X C. Ownership concentration, risk propensity and market value of listed family firms—An empirical study based on market development degree grouping. *China Industrial Economics*. 2008; 10, 139–149. https://doi.org/10.19581/j.cnki.ciejournal 2008.10.014

38. Calabro A, Vecchiarini M, Gast J, Campopiano G, De Massis A, Kraus S. Innovation in family firms: A systematic literature review and guidance for future research. *International Journal of Management Reviews*. 2019; 21(3), 317–355. https://doi.org/10.1111/ijmr.12192

39. Xu Y B, Wan Y X, Xie H L. Impact of family control intensity on technological innovation behavior of firms. *Science Research Management*. 2020; 41(6), 29–36. https://doi.org/10.19571/j.cnki.1000-2995.2020.06.004

40. Chen V Y S, Tsao S M, Chen G Z. Founding family ownership and innovation. *Asia-Pacific Journal of Accounting & Economics*. 2013; 20(4), 429–456. https://doi.org/10.1080/16081625.2012.762971

41. Roessl D, Fink M, Kraus S. Are family firms fit for innovation? Towards an agenda for empirical research. *International Journal of Entrepreneurial Venturing*. 2010; 2(3–4), 366–380. https://doi.org/10.1504/IJEV.2010.037118

42. Min B S, Smyth R. Corporate governance, globalization and firm productivity. *Journal of World Business*. 2014; 49(3), 372–385. https://doi.org/10.1016/ j.jwb.2013.07.004
43. Ode E, Ayavoo R. The mediating role of knowledge application in the relationship between knowledge management practices and firm innovation. *Journal of Innovation & Knowledge*. 2020; 5(3), 210–218. https://doi.org/10.1016/j.jik.2019
44. Wang Y Z, Lo F Y, Weng S M. Family businesses successors knowledge and willingness on sustainable innovation: The moderating role of leader's approval. *Journal of Innovation & Knowledge*. 2019; 4(3), 188–195. https://doi.org/10.1016/j.jik. 2019.05.001
45. Davis P S, Harveston P D. The phenomenon of substantive conflict in the family firm: A cross-generational study. *Journal of Small Business Management*. 2001; 39(1), 14–30. https://doi.org/10.1111/0447-2778.00003
46. Kim K, Choi S O, Lee S. The effect of a financial support on firm innovation collaboration and output: does policy work on the diverse nature of firm innovation? *Journal of the Knowledge Economy*. 2021; 12, 645–675. https://doi.org/10.1007/s13132-020-00667-9
47. Nikolaidis Y, Fouskas K, Carayannis E G. Assisting regional policy by rapidly comparing enterprise innovation between regions. *Journal of Innovation and Entrepreneurship*. 2013; 2, 16, https://doi.org/10.1186/2192-5372-2-16.
48. Calonico S M, Cattaneo D, Titiunik R. Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica*. 2014; 82(6), 2295–2326. https://doi.org/10.3982/ECTA11757
49. Lee D S, Lemieux T. Regression discontinuity designs in economics. *Journal of Economic Literature*. 2010; 48(2), 281–355. https://www.jstor.org/stable/20778728
50. Wang H Y, Zhang Y H, Liu Z K, Liu R K, Li K Y. The impact and mechanisms of the Shanghai pilot free-trade zone on the green total factor productivity of Yangtze River Delta urban agglomeration. *Environmental Science and Pollution Research*. 2022; https://doi.org/10.1007/s11356-021-17758-y PMID: 35083685
51. Caughey D, Sekhon J S. Elections and the regression discontinuity design: Lessons from close U.S. House races, 1942–2008. *Political Analysis*. 2011; 19(4), 385–408. https://www.jstor.org/stable/41403727
52. Card D, Krueger A B. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*. 1994; 84(4), 772–793. https://www.jstor.org/stable/2677856
53. Posner D N. The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review*. 2004; 98(4), 529–545. https://doi.org/10.1017/S0003055404041334
54. Gerber A S, Kessler D P, Meredith M. The Persuasive effects of direct mail: A regression discontinuity based approach. *Journal of Politics*. 2011; 73(1), 140–155. https://doi.org/10.1017/S0022381610000927
55. Dell M. The persistent effects of Peru's mining mita. *Econometrica*. 2010; 78(6), 1863–1903. https://doi.org/10.3982/ECTA8121
56. Chen Y Y, Ebenstein A, Greenstone M, Li H B. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(32), 12936–12941. https://doi.org/10.1073/pnas.1300018110 PMID: 23836630
57. Xin X, Xu Y C. The Long-term effects of the Colonial rule of the Puppet Manchukuo on the local economy: Regression discontinuity design based on township and village levels. *Economic Science*. 2019; 3, 106–118.
58. Yu W J, Wang C C. Analysis on endogenous mechanism of metropolitan area in stimulating the regional economic growth: Based on the metropolitan area of Jing-Jin-Ji,Yangtze River Delta and Pearl River Delta. *Economic Review*. 2011; 1, 69–78.
59. Black S E. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*. 1999; 114(2), 577–599. https://www.jstor.org/stable/2587017
60. Reardon S F., & Robinson J P. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*. 2012; 5(1), 83–104. https://doi.org/10.1080/19345747.2011.609583
61. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*. 2001; 69(1), 201–209. https://doi.org/10.1111/1468-0262.00183
62. Yoon W, Kwon S. The impact of technological and non-technological innovative activities on technological competitiveness. *Journal of the Knowledge Economy*. 2022; https://doi.org/10.1007/s13132-021-00868-w

63. Fung K C, Aminian N, Tung C Y. Some characteristics of innovation activities: Silicon Valley, California, China and Taiwan. *Economic Change and Restructuring*. 2016; 49, 221–240. https://doi.org/10.1007/s10644-015-9162-x

64. Stuart S R, William C S. Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*. 2003; 85(2), 377–393. https://doi.org/10.1162/003465303765299882

65. Panne G V D. Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics*. 2004; 14(5), 593–604. https://doi.org/10.1007/s00191-004-0232-x

66. Yuan Y, Gao K. The synergetic agglomeration of industries, spatial knowledge spillovers and regional innovation efficiency. *Studies in Science of Science*. 2020; 38(11), 1966–1975.

67. Dong X F, Yuan Y. Firm innovation, life cycle and agglomeration economies. *China Economic Quarterly*. 2014; 13(2), 767–792.

68. Cao X F, Huo P, Yu L L. The research of relationship between the high-tech talent agglomeration and the high-tech industrial agglomeration. *Studies in Science of Science*. 2017; 35(11), 1631–1638.

69. Mao R. Firm innovation, productivity growth and convergence: Effects of industry clustering. *Journal of Financial Research*. 2017; 8, 83–99. https://doi.org/10.12094/1002-7246(2017)08-0083-17

70. Henderson J V. Marshall's scale economies. *Journal of Urban Economics*. 2003; 53(1), 1–28. https://doi.org/10.1016/S0094-1190(02)00505-3

71. Feldman M P, Audretsch D B. Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*. 2004; 43(2), 409–429. https://doi.org/10.1016/S0014-2921(98)00047-6

72. Nunes S, Lopes R, Fuller-Love N. Networking, innovation, and firms' performance: Portugal as illustration. *Journal of the Knowledge Economy*. 2019; 10(3), 899–920. https://doi.org/10.1007/s13132-017-0508-7

73. Lan F, Jiao C C, Deng G Y, Da H L. Urban agglomeration, housing price, and space–time spillover effect—Empirical evidences based on data from hundreds of cities in China. *Managerial and Decision Economics*. 2021; 42(4), 898–919. https://doi.org/10.1002/mde.3281

74. McCrary J. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*. 2008; 142(2), 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005

75. Hoynes H, Schanzenbach D W, Almond D. Long-run impacts of childhood access to the safety net. *American Economic Review*. 2016; 106(4), 903–934. https://doi.org/10.1257/aer.20130375

76. Cuerva M C, Triguero-Cano A, Corcoles D. Drivers of green and non-green innovation: Empirical evidence in Low-Tech SMEs. *Journal of Cleaner Production*. 2014; 68, 104–113. https://doi.org/10.1016/j.jclepro.2013.10.049

77. Liu N Q, Wu Y. Can the enlargement in Yangtze River Delta boost regional economics common growth. *China Industrial Economics*. 2017; 6, 79–97. https://doi.org/10.19581/j.cnki.ciejournal.2017.06.019

78. Günther J, Meissner D. Clusters as innovative melting pots?—The meaning of cluster management for knowledge diffusion in clusters. *Journal of the Knowledge Economy*. 2017; 8, 499–512. https://doi.org/10.1007/s13132-017-0467-z

79. Xu X X, Gan W Y. Impact of institutional environment and technological capability on governance transformation of family firms. *Science Research Management*. 2018; 39(12), 131–140. https://doi.org/10.19571/j.cnki.1000-2995.2018.12.016

80. Yang J, Wang Y C, Xiao X M, Jin C, Xia J H, Li X M. Spatial differentiation of urban wind and thermal environment in different grid sizes. *Urban Climate*. 2019; 28, 100458. https://doi.org/10.1016/j.uclim.2019.100458

81. Yang J, Yang Y X, Sun D Q, Jin C, Xiao X M. Influence of urban morphological characteristics on thermal environment. *Sustainable Cities and Society*. 2021; 72, 103045. https://doi.org/10.1016/j.scs.2021.103045

82. Zhang D H, Zhang G J, Zhou C S. Differences in accessibility of public health facilities in hierarchical municipalities and the spatial pattern characteristics of their services in Doumen district, China. *Land*. 2021; 10(11), 1249. https://doi.org/10.3390/ land10111249

83. Zhang D H, Zhou C S, Sun D Q, Qian Y. The influence of the spatial pattern of urban road networks on the quality of business environments: the case of Dalian city. *Environment, Development and Sustainability*. 2022; 24, 9429–9446. https://doi.org/10.1007/s10668-021-01832-z

84. Zhang D H, Zhou C S, Xu W W. Spatial-temporal characteristics of primary and secondary educational resources for relocated children of migrant workers: the case of Liaoning province. *Complexity*. 2020; 7457109. https://doi.org/10.1155/ 2020/7457109

85. Zhou C S, Zhang D H, He X. Transportation accessibility evaluation of educational institutions conducting field environmental education activities in ecological protection areas: a case study of Zhuhai city. *Sustainability*. 2021; 13(16), 9392.https://doi.org/10.3390/su13169392