Identifying Knowledge Spillovers from Universities: Quasi-experimental Evidence from Urban China

Jing Li[a], Shimeng Liu[b], Yifan Wu[a]

[a] School of Economics, Singapore Management University, Singapore
[b] Institute for Economic and Social Research, Jinan University, China

2022 AREUEA/ASSA Conference

Background

Economists and policy makers have long been interested in the relationship between higher education and economic growth (Acemoglu 1995; Redding 1996; Aghion et al. 2009).

An important channel is by inducing ground-breaking innovations through:

- A direct expansion of the research capacity;
- Externalities that indirectly foster technological breakthroughs outside the institutions.

The latter channel on externalities is important and is the focus of this paper.

Question

This paper examines the causal impact of university activity on the creation of local patents and new products by taking advantage of a unique quasi-experiment in China that has exogenously expanded higher education institutions since 1999.

We focus on identifying the role of knowledge spillovers.

A few key highlights:

- We uncover the localized nature and striking geographic attenuation of university spillovers at a refined geographic level that is new in this literature.
- We utilize novel datasets that contain comprehensive information on patents and new products from firms geocoded at the address level.
- We merge our core datasets with patent citation information and reveal consistent spatial decay patterns of citation links.

To Better Understand Economic Growth

R&D and innovation have played a central role in advancing the world technology frontier (Acemoglu 2008).

However, in innovation-based growth models, the R&D production function has been taken as a reduced-form representation and the specific steps leading to practical innovation is not yet clear (Akcigit 2017).

Understanding how the fundamental knowledge transforms into patentable findings and practical products forms the cornerstone of the R&D production function that is at the center of innovation-based economic growth theory.

To Provide Policy Implications

It justifies public investments on higher education that has witnessed enormous growth in recent decades.

It provides guidance for creating technology hubs near education institutions.

- Policy makers have formed a general consensus that proximity to universities is a key condition for a vibrant high-tech community.
- A careful policy design requires a good understanding on how quickly the positive externalities decay with geographic distance.

Identification

The unique national university expansion policy introduced a structural break in the time series of the city-specific university capacity that was exogenous and independent of unobserved local economic conditions.

- We make use of the kinked relationship created by the shock to identify the causal impact of the university expansion in a difference-in-differences framework.

We focus on within-city variations to characterize the geographic nature of university spillovers and to identify the role of knowledge spillovers.

- This within-city focus allows us to adopt a triple-differences approach in which we control for a rich set of interacting fixed effects to tighten our identification.

Findings

- University innovation activities increase nearby patents, and the impact decays sharply within 2 kilometers of the universities.
- The spatial attenuation of university spillovers is ubiquitously present in different regions and industries in China but is more pronounced in the Eastern region and for industries more reliant on high-skilled labour.
- University expansion increases nearby patents that cite university patents, the magnitude of which also decays quickly across space and stabilizes beyond the 2 km radius. The spatial attenuation pattern, however, is not present for nearby patents that cite patents far away from universities.
- University expansion boosts new products from firms and the impact follows a similar spatial decay.

Literature on Knowledge Spillovers and Agglomeration Economies

- Since Marshall (1890), researchers have rationalized the observed geographic concentration of economic activities on account of positive externalities associated with the clustering of firms.
  - Duranton and Puga 2004; Holmes 1999; Glaeser and Mare 2001; Moretti 2004; Ellison et al. 2010
- It is also recognized that different microfoundations are associated with different spatial attenuation of agglomeration externalities.
  - Rosenthal and Strange 2004; Combes and Gobillon 2015; Li et al. 2020; Rosenthal and Strange 2020
- The fast attenuation speed documented in this paper directly points to the role of knowledge spillovers in explaining university spillover benefits.

Literature on the Impact of Education and Research Institutions

- Previous studies investigated the impact of universities by taking advantage of various quasi-experiments and examined a range of local economic outcomes.
  - Andersson et al. 2004, 2009; Kantor and Whalley 2014; Liu 2015; Aghion et al. 2009; Kantor and Whalley 2019; Andrews 2019; Hausman 2020
- The broad scope of the analysis prohibits researchers from drawing detailed conclusions on the channel through which the localized spillovers take place.
- By documenting a sharp spatial attenuation of the impact of university expansion on intermediate innovation outcomes (patents and citation links) and final output measures (new products), we highlight the role of knowledge spillovers on innovation as an important contributing mechanism of the university impact.

Conceptual Framework

The impact of universities on local innovation can be mediated through a collection of channels (Valero and Van Reenen 2019).

We highlight the specific mechanisms through which the impact is especially pronounced at narrowly defined geographic distances (within 2-3 km in general).

The number of new ideas, NI, is assumed to be a function of the existing knowledge stock, A, and the number of researchers, R, who spent time producing them:

$$NI = f(A, R). \tag{1}$$

The number of new products, NP, is assumed to be a function of new ideas, NI, and the necessary facility, equipment, and personnel, X, to convert the new ideas into new products:

$$NP = g(NI, X). \tag{2}$$

Conceptual Framework

We also assume that the knowledge stock, A, is affected by a nearby university's scale or its innovation capacity, U, through the channel of knowledge spillovers as well as the distance to the university, D, which captures the sharp spatial attenuation of knowledge spillovers:

$$A = a(U, D), \tag{3}$$

where $\frac{\partial a}{\partial U} > 0$, $\frac{\partial a}{\partial D} < 0$, and $\frac{\partial^2 a}{\partial U \partial D} < 0$.

However, this is not the only channel through which universities affect local innovation. For instance, the number of researchers, R, could be a function of the local university's own innovation capacity, U:

$$R = r(U). \tag{4}$$

Conceptual Framework

A further difference of the university impact along the spatial dimension helps tease out the labour market mechanism and highlight the role of knowledge spillovers.

We assume linearity of the functional form and write the determinants of new ideas at "close" and "far" distances as follows:

$$NI_{D(\text{close})} = A_{D(\text{close})} + R_{D(\text{close})} = \alpha_{D(\text{close})} U + \beta U, \tag{5}$$

$$NI_{D(far)} = A_{D(far)} + R_{D(far)} = \alpha_{D(far)} U + \beta U. \tag{6}$$

Any differences in the impact of universities across various spatial distances can be attributed to the differences in A.

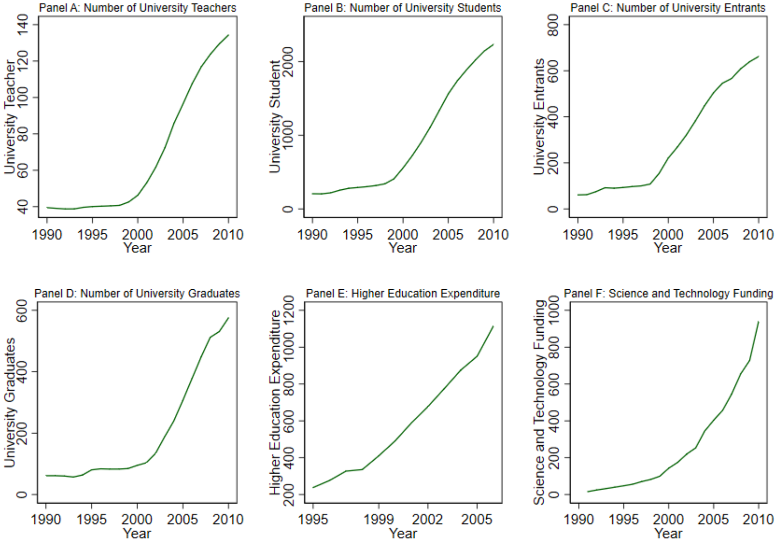
$$NI_{D(\text{close})} - NI_{D(\text{far})} = A_{D(\text{close})} - A_{D(\text{far})} = \left[\alpha_{D(\text{close})} - \alpha_{D(\text{far})}\right] U. \tag{7}$$

Institutional Background

- Concerns about a recession in the late 1990s triggered a need to expand the higher education sector to stimulate the domestic demand for educational services and other related consumption.
- In June 1999, the Ministry of Education and the National Development and Planning Commission jointly announced a new higher education recruitment plan, with expected enrollment of 1.53million students in 1999, which is a 42 percent year-on-year increase.
- Around the same time, the ministry announced an increase in college tuition by 15-20 percent across different regions.

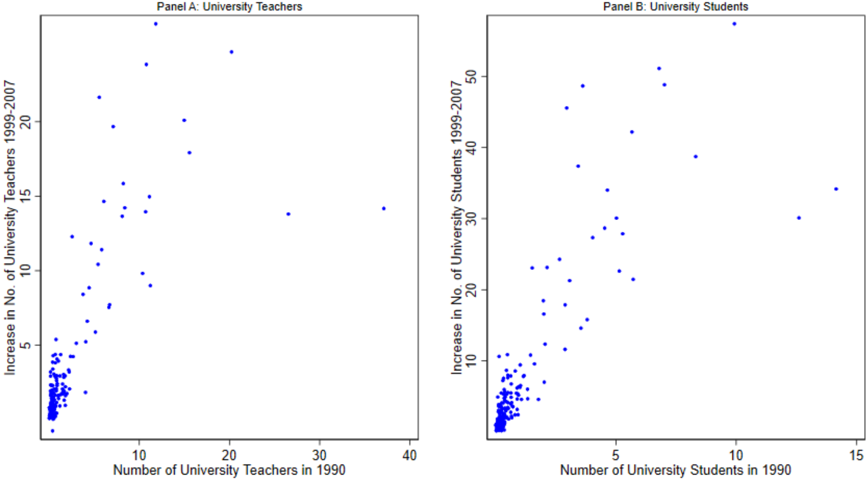
Institutional Background

Various Aspects of University Expansion



Institutional Background

Growth in the University Scale between 1999 and 2007



Empirical Framework

Empirical Challenges:

- Unobserved factors that are correlated with both the university activity and local economic outcomes.
- The challenge on disentangling various mechanisms and to identify the role of knowledge spillovers.
- Measurement for innovation.

Strategies:

- We utilize the higher education expansion policy in China as a quasi-experiment.
- We focus on the extremely localized effects of universities to identify the role of knowledge spillovers in boosting local innovation.
- We supplement our patent analysis with two additional measures: patent citation links and new commercial products.

The First-Stage Regression

$$UniversityInno_{c,t} = \beta \times (Treatment_c \times Post_t) + \alpha_c + \gamma_t + \varepsilon_{c,t}$$

$UniversityInno_{c,t}$ is proxied by the number of university teachers, university students, or university patents in city c and year t after removing city-specific pre-expansion time trend. $Treatment_c$ is a continuous variable of the number of university teachers (or students) in city c in year 1990 (a proxy for treatment intensity). $Post_t$ is a dummy variable that equals 1 if the year t is 2000 or after α_c and γ_t are city and year fixed effects.

The Reduced-Form Regression

$$OtherInno_{c,t} = \beta \times (Treatment_c \times Post_t) + \alpha_c + \gamma_t + \varepsilon_{c,t}$$

$OtherInno_{c,t}$ represents the number of collaboration patents or non-university-related patents in city c and year t after removing the city-specific pre-expansion time trend.

- The core element of our empirical analysis is the focus on within-city variations to characterize the geographic nature of university spillovers and to identify the role of knowledge spillovers.
- This within-city focus allows us to adopt a triple-differences approach in which we control for a rich set of interacting fixed effects to tighten our identification.
- We achieve this by specifying a rich set of concentric ring variables that capture the innovation activities at various distances from the research universities.
    - Each concentric ring spans 500 meters.
    - We include 10 or 20 rings to cover places up to 5 km or 10 km away in our estimation, depending on the specific model.

Empirical Framework

Within-City Variation

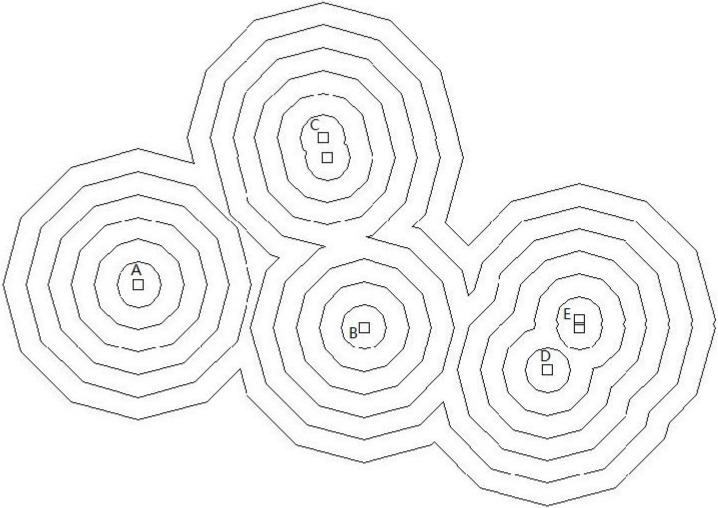


Figure A1: Illustrative Graph for the Construction of Concentric Rings

Empirical Framework

Within-City Variation

A Triple-differences Framework at Ring Level

$$
\begin{aligned}
Inno_{c,r,t} = & \sum_{r=1}^{9} \beta_r \times (Treatment_c \times Post_t \times Ring_r) + d_{c,r} + d_{c,t} \\
& + d_{r,t} + \varepsilon_{c,r,t}
\end{aligned}
$$

$Inno_{c,r,t}$ represents the number of patents or the number of cases when industrial patents cite university patents in city c, ring r, and year t, after removing ring- and city-specific time trends; $Treatment_c$ and $Post_t$ are defined in the same way as before; $Ring_r$ is a dummy variable that equals 1 if the patent is within the concentric ring r and 0 otherwise (ring 10 is set as the reference group and is omitted); $d_{c,r}$, $d_{c,t}$, and $d_{r,t}$ are city by ring, city by year, and ring by year fixed effects, respectively.

Empirical Framework

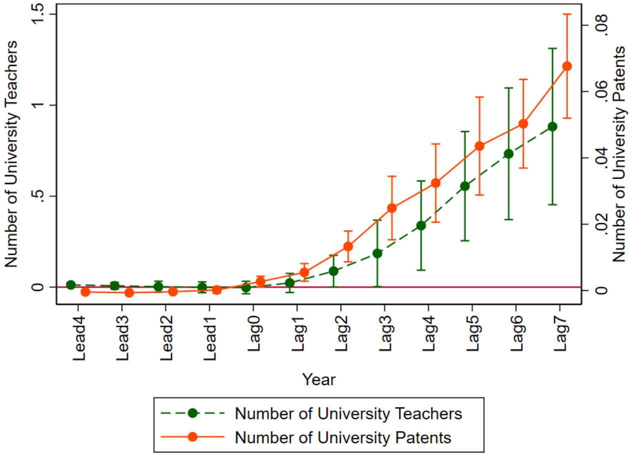
Within-City Variation

A Triple-differences Framework at Firm Level

$$Inno_{i,c,r,t} = \sum_{r=1}^{9} \beta_r \times (Treatment_c \times Post_t \times Ring_r) + d_{c,r} + d_{c,t} + d_{r,t} + \boldsymbol{X_{i,c,r,t}} \boldsymbol{\rho} + \varepsilon_{i,c,r,t}$$

$Inno_{i,c,r,t}$ represents firm i's new commercial product ratio in city c, ring r, and year t, after removing ring- and city-specific time trends; $\boldsymbol{X_{i,c,r,t}}$ is a set of firm-specific controls, including the age of a firm, fixed assets, a dummy for whether a firm is an SOE, and the employment size; and $\varepsilon_{i,c,r,t}$ is a firm-specific error term.

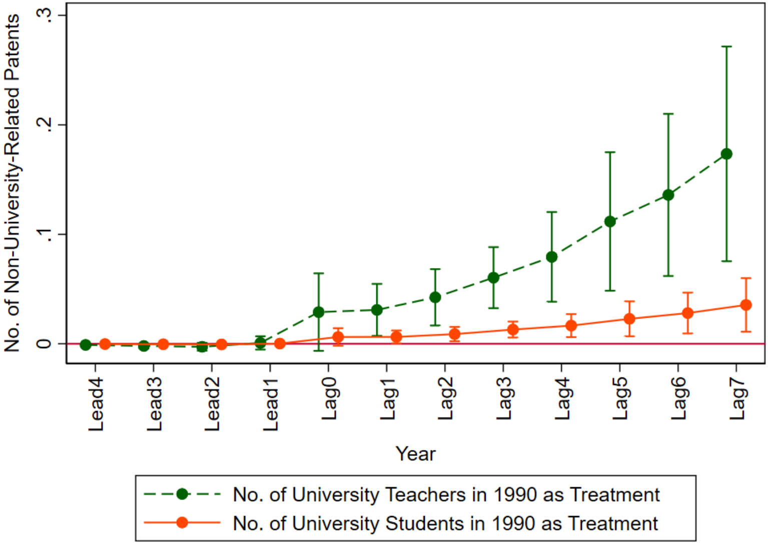
The Dynamic First-Stage Regressions



No. of University Teachers in 1990 as Treatment

Figure: The Dynamic Effects of University Expansion on the Number of Teachers,

The Dynamic Reduced-Form Regressions

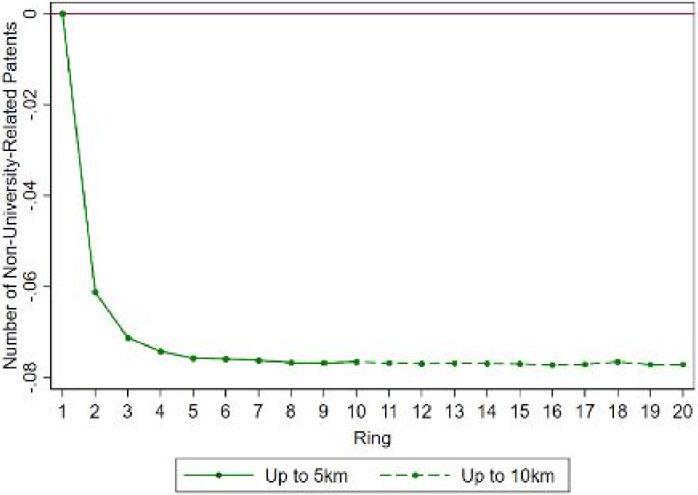


Ring Regresions

Table 4: Effects of University Expansion on Innovation — Ring Regressions

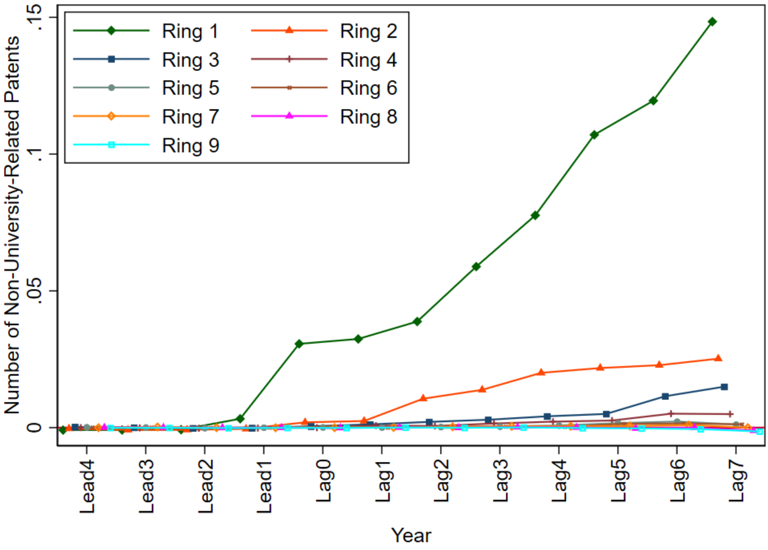
	No. of Teachers in 1990 as Treatment		No. of Students in 1990 as Treatment	
Dependent Variable	Number of Patents			
	(1)	(2)	(3)	(4)
Treatment × After × 0.5km	0.0769***	0.0765***	0.0157***	0.0156***
	(3.80)	(3.79)	(3.08)	(3.08)
Treatment × After × 1km	0.0157***	0.0153***	0.00333***	0.00324***
	(3.54)	(3.57)	(3.24)	(3.26)
Treatment × After × 1.5km	0.00570**	0.00529**	0.00125**	0.00117**
	(2.08)	(2.04)	(2.21)	(2.18)
Treatment × After × 2km	0.00271**	0.00230*	0.000607**	0.000524**
	(2.03)	(1.93)	(2.28)	(2.21)
Treatment × After × 2.5km	0.00121	0.000806	0.000275	0.000192
	(1.36)	(1.07)	(1.53)	(1.26)
Treatment × After × 3km	0.00105*	0.000646	0.000227**	0.000144*
	(1.91)	(1.55)	(2.00)	(1.67)
Treatment × After × 3.5km	0.000749	0.000342	0.000167	0.0000843
	(1.39)	(0.80)	(1.57)	(0.99)
Treatment × After × 4km	0.000243	-0.000164	0.0000520	-0.0000311
	(1.12)	(-1.01)	(1.16)	(-0.83)
Treatment × After × 4.5km	0.000180	-0.000227	0.0000357	-0.0000475
	(0.79)	(-1.38)	(0.72)	(-1.25)
Treatment × After × 5km	0.000407*	-	0.0000831	-
	(1.81)	-	(1.64)	-
Observations	23920	23920	23920	23920
Treatment × Ring Dummies	Yes	No	Yes	No
City FE	Yes	No	Yes	No
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	No	Yes	No	Yes
City × Ring FE	No	Yes	No	Yes
Dependent Variable Mean	10.91	10.91	10.59	10.59
Adj. R^2	0.255	0.482	0.228	0.479

Ring Regresions



(a) No. of University Teachers in 1990 as Treatment

The Dynamic Ring Regressions



Effects on Patent Citations

Table 8: Effects of University Expansion on Patent Citations — Ring Regressions

	No. of Teachers in 1990 as Treatment		No. of Students in 1990 as Treatment	
Dependent Variable: No. of	(1) Citations to University Patents	(2) Citations to Patents beyond 5km	(3) Citations to University Patents	(4) Citations to Patents beyond 5km
Treatment ×	5.59E-04***	7.29E-05	1.14E-04**	1.52E-05
After × 0.5km	(1.15)	(1.15)	(0.98)	(0.98)
Treatment ×	8.23E-05**	4.81E-06	1.85E-05**	1.03E-06
After × 1km	(0.34)	(0.34)	(0.30)	(0.30)
Treatment ×	1.86E-05*	2.72E-05*	4.49E-06**	5.38E-06
After × 1.5km	(1.91)	(1.91)	(1.55)	(1.55)
Treatment ×	3.04E-05***	-1.41E-06	6.65E-06***	2.52E-07
After × 2km	(-0.10)	(-0.10)	(0.08)	(0.08)
Treatment ×	7.07E-06	8.19E-06	1.87E-06	2.13E-06
After × 2.5km	(0.96)	(0.96)	(1.07)	(1.07)
Treatment ×	7.71E-06	1.64E-05	1.82E-06	3.88E-06*
After × 3km	(1.65)	(1.65)	(1.71)	(1.71)
Treatment ×	8.03E-06*	2.43E-05***	1.86E-06*	5.19E-06**
After × 3.5km	(3.14)	(3.14)	(2.57)	(2.57)
Treatment ×	-1.58E-06	7.77E-06	-2.62E-07	1.81E-06
After × 4km	(1.16)	(1.16)	(1.19)	(1.19)
Treatment ×	3.86E-06	8.22E-06	9.78E-07	1.70E-06
After × 4.5km	(1.28)	(1.28)	(1.08)	(1.08)
Treatment ×	-	-	-	-
After × 5km	-	-	-	-
Observations	4500	4500	4500	4500
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.32	0.46	0.32	0.46
Adj. R^2	0.668	0.522	0.653	0.522

Effects on New Products

Table 9: Effects of University Expansion on New Product Ratio — Ring Regressions

Dependent Variable	New Product Ratio			
	(1)	(2)	(3)	(4)
After Dummy	2000	2002	2004	2006
Treatment ×	4.16E-06**	4.84E-06**	6.35E-06**	6.68E-06**
After × 0.5km	(2.17)	(2.15)	(2.19)	(2.17)
Treatment ×	2.74E-06**	3.32E-06**	4.26E-06***	4.58E-06**
After × 1km	(2.57)	(2.59)	(2.68)	(2.46)
Treatment ×	1.76E-06***	2.07E-06***	2.74E-06***	2.89E-06***
After × 1.5km	(4.98)	(5.17)	(4.87)	(4.60)
Treatment ×	1.28E-06***	1.47E-06***	1.75E-06***	1.77E-06***
After × 2km	(4.91)	(4.80)	(5.14)	(4.78)
Treatment ×	1.66E-06***	2.03E-06***	1.73E-06***	1.63E-06***
After × 2.5km	(3.65)	(3.90)	(3.62)	(3.20)
Treatment ×	1.07E-06***	1.31E-06***	1.63E-06***	1.68E-06***
After × 3km	(4.51)	(5.10)	(4.69)	(4.59)
Treatment ×	6.20E-07***	7.28E-07**	1.04E-06***	1.21E-06***
After × 3.5km	(2.74)	(2.54)	(3.12)	(3.39)
Treatment ×	6.66E-07***	7.85E-07***	1.24E-06***	1.56E-06***
After × 4km	(2.78)	(2.99)	(3.97)	(4.31)
Treatment ×	6.06E-07	6.96E-07	9.91E-07	1.14E-06**
After × 4.5km	(1.19)	(1.24)	(1.61)	(2.01)
Treatment ×	5.54E-07**	6.84E-07**	1.11E-06***	1.38E-06***
After × 5km	(2.17)	(2.13)	(2.84)	(3.38)
Observations	1196263	996185	759980	589233
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.034	0.035	0.035	0.037
Adj. R^2	0.091	0.097	0.107	0.106

Robustness Checks

Within-City Variation

Ring Level Regression – Linear Pre-trend Specification

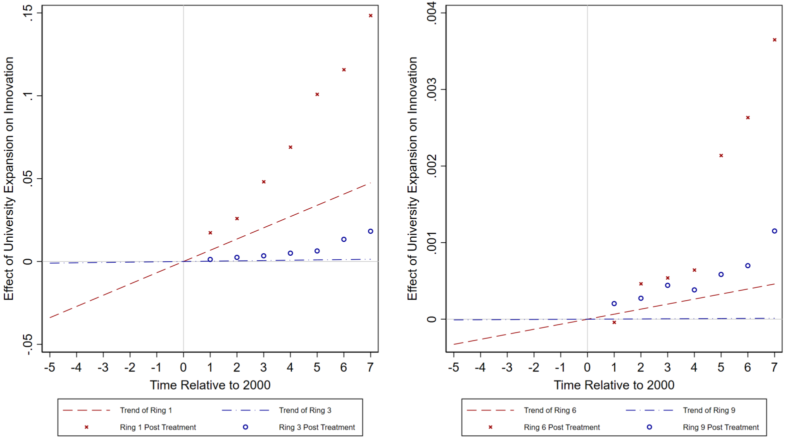
$$Inno_{c,t} = \delta \times \ell + \sum_{\ell=1}^{7} \mu_\ell \times Treatment_c \times \mathbf{1}\{t = 2000 + \ell\} + \alpha_c + \gamma_t + \varepsilon_{c,t}$$

ℓ is the time relative to 2000. δ captures the slope of the trend, and it is the slope of the dash line in the figures (next page). The crosses and dots in the graphs capture the term $\delta \times \ell + \sum_{\ell=1}^{7} \mu_\ell \times Treatment_c \times \mathbf{1}\{t = 2000 + \ell\}$.

- Borrow the idea from Dobkin et al. (2018)

Robustness Checks

Linear Pre-trend Specification



No. of University Teachers in 1990 as Treatment

A Triple-differences Framework at Ring Level – Trend-break Specification

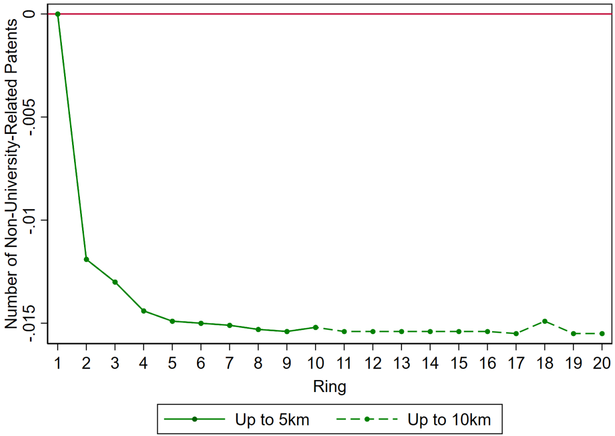
$$\begin{aligned} Inno_{c,r,t} = & \sum_{r=2}^{10} \alpha_r \times (Treatment_c \times \ell \times Ring_r) \\ & + \sum_{r=2}^{10} \beta_r \times (Treatment_c \times \ell \times Post_t \times Ring_r) \\ & + d_{c,r} + d_{c,t} + d_{r,t} + \varepsilon_{c,r,t} \end{aligned}$$

ℓ is the time relative to 2000; $Ring_r$ is a dummy variable that equals 1 if the patent is within the concentric ring r and 0 otherwise.

- Borrow the idea from Almond et al. (2019)

Robustness Checks

Trend-break Specification



No. of University Teachers in 1990 as Treatment

Conclusion

- This paper exploits a unique quasi-experiment of university expansion in China to study the impact of university activities on local innovation. In particular, we utilize rich geocoded data on patent generations, patent citation links, and new products from firms to examine the geographic nature of the university impact and to identify the role of knowledge spillovers.
- We find that the university expansion significantly increases universities own innovation capacity, which results in a dramatic boom in nearby firms patenting activities. The impact attenuates sharply with spatial distance. The evidence justifies the continually increasing support for research universities as a viable policy instrument for the government to promote long-term economic growth.