Cross-City Travel Speed Enhances Researcher Productivity: Evidence from China's Investment in Bullet Trains

Xiaofang Dong (Xiamen University)
Matthew E. Kahn (University of South California)
Siqi Zheng (MIT)

April 2017

Abstract

China's best universities are located in its mega cities such as Beijing and Shanghai. Consistent with the localized human capital spillovers theory, high tech industrial clusters have co-agglomerated close to these universities. The "localization" of such spillovers hinges on travel speeds. If people can move at higher speeds, then such knowledge diffusion and idea flows are likely to reach greater distances. This paper uses the construction of China's bullet train network as a natural experiment to this claim. Since the bullet trains reduce cross-city commute times, they reduce the cost of face-to-face interactions between cities. We study a city's researcher productivity, and also cross-city university co-authoring patterns before and after the construction of the bullet train. We document a complementarity effect between knowledge production and the transportation network such that research pair productivity rises and more new co-author pairs emerge from connected second tier cities.

Introduction

The urban human capital externality literature posits that workers and firms in cities featuring a large percentage of well-educated people benefit from knowledge spillovers (Glaeser et. al. 1995 1999). One of the most important features of knowledge spillovers is that they are localized. A number of studies document that spillovers are constrained by geography (Rosenthal and Strange, 2003, 2004). This may explain some of the variation in productivity across regions, and that worker's wages are higher in cities with greater human capital shares (Moretti 2004, Rauch 1992).

Lowering the cost of transportation facilitates matching and learning. Imagine a case in which people can move at infinite speed within cities but cannot travel across cities. In this case, a great scholar can only work with researchers who work in the same city. Younger researchers have less opportunity to directly learn from this scholar. This example highlights how travel speed both increases the superstar researcher's private benefits (through building a research pyramid that directly benefits him) and through providing social benefits (his public lectures can be attended by many).

Our starting point is that cross-city travel speeds are a key determinant of the geographic spillover effects of superstar scientists. Helsley and Strange (2007) model the optimal transportation subsidy in an economy featuring learning but costly transportation. They document that the optimal transportation subsidy for encouraging learning increases for those who live at the periphery.

We use the opening of China's bullet train network as a natural experiment to study the relationship between higher cross-city commute speeds and the productivity of pairs of scholars who work in different cities that now feature lower commute times.   We find that increases in speed between cities raises productivity of scholars working in different cities that are now bullet train connected.

China's best universities are located in a small number of its major cities. In China's 287 prefecture-level cities, more than 60%of the top universities (those in "985-Program") are located in the six mega cities – Beijing, Shanghai, Nanjing, Guangzhou, Wuhan and Xi'an. Congestion forces in those mega cities are high – expensive housing prices, polluted air and traffic congestion, which prevent scholars from clustering in those mega cities. In a world where transportation costs are high within cities and cross cities, relatively few people can access the brilliant scholars and the key pieces of the advanced laboratories that are housed in those world-renowned universities. Such an "intellectual monopoly power" will create and reinforce human capital inequalities cross cities and is "socially cost" because scholar agglomeration generates great human capital externality (Rauch, 1991).

Over the last thirty years, the Chinese central government has made enormous cross-city investments in transport infrastructure. Bullet trains move at high speed and effectively convert second tier cities that are too far to drive to the mega cities but too close to fly (in the 50 mile to 300 mile range) into being suburbs of those mega cities, which allows people to have comfortable and easy transport to the superstar cities. Such declining transportation costs transform China's best universities from being "local public goods" to "regional public goods". Transportation innovations increase the geographic reach of scholar's idea flows. The researchers in the 2nd tier universities can collaborate with the stars, and learn from them. If remote interactions through Skype and telephones was a perfect substitute for face to face interaction, then we would not expect to observe an economically meaningful effect from bullet train integration (Gaspar and Glaeser 1998).

The provision of transportation infrastructure is one of the major policies that China's central and local governments implement to spur regional economic growth through its effect on productivity, employment and investment. But its impact on knowledge creation and diffusion has been overlooked by the literature and the cost-benefit evaluation of such transportation policies. Our paper provides insight for the social benefit of such huge high speed rail investments in enhancing knowledge diffusion and spillovers, and thus alleviating human capital inequalities across Chinese cities.

The rest of the paper is organized as follows. In Sections 2 we present a simple model of research production, to guide our research design. Section 3 introduces the geography of Chinese universities and bullet trains. Section 4 and Section 5 describe data, empirical models and identification strategies. Section 6 and 7 present our main results on city academic productivity and regional idea flows, respectively. We conclude in Section 8.

A Simple Model of Research Production

Let there be two cities called "a" and "b" and define t as the commuting time between them. Let researcher 1 be a superstar researcher in city a.

This researcher co-authors papers. His production function (measured in quantity and quality of papers) is a function of his own inputs $ability_1$ and his co-author's inputs $ability_2$ and the number of face to face interactions they face $meetings_{12}$

This production can be written as:

$$output_{1a} = f(ability_{1a}, ability_2, count\ of\ meetings_{12}, length\ of\ meeting_{12})$$

Each researcher has a time budget constraint. Researcher 2's time budget constraint features a fixed time budget that can be allocated to other activities or to meeting with co-authors on different projects. For each meeting there is a count of annual meetings and the duration of each meeting.

$$Time_2 = \sum_{l=1}^{h} count_{l2} * length_{l2} + count_{l2} * commute_{l2}$$

As shown in the time budget constraint the commute time for researcher 2 to visit co-author l (and he works with h total co-authors) acts to reduce the count of meetings. Commuting is a fixed cost technology encouraging researchers to have fewer face to face meetings (a lower count) and to have longer meetings. If there are diminishing returns in the production of knowledge from the length of meetings then this is inefficient.

The bullet train reduces the commute time and encourages more face to face meetings. This increases the superstar researcher's output because he know is better matched with the right co-authors (the extensive margin) and he optimizes the quantity of meetings versus their length.

Our empirical approach really focuses on the intensive margin of scientist interactions without exploring the increase in match permutations as the local labor market grows in size thanks to the decline in commuting costs.

Given that the superstar scientist's utility is an increasing function of his own publication quality, there are no externality spillovers here. Where there are externality knowledge spillovers is in an overlapping generations model in which there are young scientists who can now more easily travel to city a to watch the scientist teach and to learn from his students.

The Geography of Chinese Universities and Bullet Trains

China started to build its modern universities in late Qing Dynasty (around 1890s). Peking University and Tsinghua University were founded in 1898 and 1911, respectively. The major universities were founded in big cities, and have continuously received investment during the Republic of China era, and later from the Chinese Communist Party after 1949. There is a hierarchy system of universities in China – on the top there are 39 superstar universities defined in the "985 Program"[1], and they receive favorable research and teaching funds from the State; the second tier includes73universities defined in the "211 Program"[2](but not included in the "985 Program"), they can also get some special funds from the State and the provincial governments, but they are not as prestigious as the superstar universities in the "985 Program". All the other universities can be regarded as the third tier – they receive funds from provincial or lower-level governments. Private university is very rare in China. The qualities of professors and students vary significantly along this hierarchy – the superstar cities have excellent laboratories, libraries and databases, professors there have a much higher probability to obtain research funds from

[1]The "985 Program" is a constructive project for founding world-class universities in the 21st century conducted by the government of the People's Republic of China on May 4, 1998. In the initial phase, 9 universities were included in the project. The second phase, launched in 2004, expanded the program until it has now reached 39 universities.

[2] The "211 Program" is the Chinese government's new endeavor aimed at strengthening about 100 institutions of higher education and key disciplinary areas as a national for the 21st century, and to facilitate the development of higher education in the context of the country's advancement in social and economic fields.

national research foundations (such as National Natural Science Foundation of China) and also have smart students, and students there can meet with great professors and other brilliant students. Therefore, the best professors and students are attracted to those best universities and this process has reinforced itself in the last 100 years.

The best universities are disproportionally concentrated in a small number of mega cities (Figure 1-A). For Beijing, Shanghai, Nanjing, Guangzhou, Wuhan and Xi'an, each has more than 2 superstar universities, and altogether they account for 61% of all the first-tier universities in the "985-Program". If you cannot win the fierce competition in such best universities and you cannot get the high salary (and also compensation from research projects), it will be hard for you to find a job there and even live in those mega cities because the living costs there have been bid up to very high levels. A 100-square meter apartment near Tsinghua University costs 10 million RMB (1.4 million USD). The housing prices in the six mega cities have an average annual appreciation rate of 15% in the last ten years (Hang Lung Center for Real Estate at Tsinghua University).

Improving transport infrastructure between nearby cities offers one strategy for mitigating the mega city quality of life challenge. China's recent investment in "bullet trains" allows individuals to move at speeds of roughly 175 miles per hour and this increases the menu of locations that have access to mega cities. If individuals can swiftly move from nearby cities to mega cities then they can enjoy the benefits of mega city access, including the intellectual power there, without suffering the social costs associated with mega city growth.

The concept of the "bullet train" (or High Speed Railways, HSR) was born in 1964 with the formal opening of Japan's "Shinkansen". The bullet train is regarded as one of the most significant technological breakthroughs in passenger transportation developed in the second half of the 20th century. In 1990s, the average speed of Chinese conventional trains was below 60 kilometers per hour. The speed had been raised several times in late 1990s and early 2000s, but the highest speed did not exceed 150 kilometers per hour. The Ministry of Railway (MOR) announced its ambitious bullet train plan in 2006.[3] The first set of bullet train lines opened in April 2007, boosting the speed of some major trains to 200 to 250 km/h. In August 2008, to coincide with the 2008 Beijing Olympic Games, new bullet trains opened between Beijing and Tianjin. They reached a higher top speed of about 350 km/h. By the end of 2010, China's bullet train service length reached 8,358 kilometers. By 2020 China's total service HSR length will reach twelve thousand kilometers.[4] Figure 2 shows the geographical overview where the high-speed rails are distributed across space. Here is an example. Tianjin is 120 kilometers from Beijing and there is no flight between them. The Beijing-Tianjin bullet train ships 400 thousand passengers (one-way) per week.[5] Before the bullet train was introduced in 2008, the conventional train shipped 150 thousand passengers per week but this number has declined to about 45 thousand and highway traffic has also declined.[6]

Data

As we discuss below, we seek to study the association between city pair bullet train connections and cross-city research productivity. To study this cross-city trade in ideas, we need

[3]The typical financing arrangement for constructing bullet train lines is that the MOR pays 50 to 60% of the total cost and the destination cities pay the remainder. See: http://finance.people.com.cn/GB/1037/8743758.html

[4]The average construction cost is about 100 million RMB per kilometer for newly-constructed bullet train lines, and this cost is much lower for upgraded railway lines. The average operating cost is about 0.3 RMB per person per kilometer.

[5]The one-way price per ride of the Beijing-Tianjin bullet train is 55 RMB Yuan (8.9 US dollars), roughly 2.5 times the conventional train fare. The fare price of the Wuhan-Guangzhou bullet train is roughly 60% of the flight fare. China's middle class can afford bullet train travel. Poor rural migrants would not choose bullet trains, but they do not travel much. Most of them only travel once a year when they return to their hometowns for the Chinese New Year.

[6]http://www.cnbuses.com/news/201105/36656.htm

to collect geocoded data on academic partnerships and information on the construction and opening of specific bullet train stations.

We collect all the international journal publications from Chinese universities during the years 2001 to 2016 (featuring at least one author from Mainland China) from the website of "Web of Science" (run by Thomson Reuters). It covers all the papers published in the academic or science world. For each paper, we obtain the paper title, the names and affiliations of the authors, publication date, journal field, journal impact factor, and the number of citations (as of December 2016), etc. After dropping possible duplicates papers, we have around 1.5 million journal papers. Figure 1-B shows that those papers are also disproportionally concentrated in a small number of mega cities where the best universities are located. For an author who appears more than one time in our database, we check if his/her affiliation has changed in our study period – if yes, he/she is defined as a “mover”, otherwise a “non-mover”.

*** Insert Table 1 about here ***

We construct both quantity and quality measures for academic papers. For the quantity dimension, we count the number of papers. For the quality dimension, we have two variables – the citation weighted number of papers (as a measure of how important and well-cited the papers are), and the journal impact factor weighted of papers (as a measure of how many papers are published in important journals).We count the number of published papers (both quantity and quality) by city by year as a measure of a city’s academic productivity, which is expected to increase when the city is connected by the bullet train.

97% of the papers in our data set are co-authored papers. Co-authoring represents team production of knowledge. Each research partner must volunteer to join the team and thus must receive some benefits from the participation. Face-to-face interactions are crucial for the knowledge flows the diffusion between coauthors. That is why in this modern era with advanced online communication technologies (Email, Skype, mobile phone, etc.), scholars still need to go to conferences and seminars to present their research work, and interact with other scholars.

Below, we will argue that the bullet train causes a large decline in transportation costs and thus facilitates face-to-face communications between coauthors. We take advantage of this co-authorship structure to examine the idea flows before and after the introduction of bullet train. For the papers with multiple authors, we keep the first three authors at the most, and construct one-to-one author pairs between the coauthors. We then match the university affiliation and city name to those coauthors so that we can count the number of paper publications (both quantity and quality) for each university-pair and city-pair by year. Figure 2 shows which cities are connected together by those co-publications. The figure displays the spatial patterns of co-authorship– the mega cities with superstar universities are the cores in the co-authorship rays.

*** Insert Figure 2 about here ***

Figure 3 shows that over time, the total international journal publications from Chinese universities have increased dramatically, and the secondary cities have a clear lead comparing with those mega cities since 2008. The bullet train program also start at the year 2008, which suggests the possibility of knowledge diffusion from mega cities to those secondary cities. The increase is relatively slow in mega cities, perhaps because the scholars in superstar universities already have intensive connections with other scholars all over the world, while the researchers in the secondary universities have enjoyed an increase in access to the best professors in the major cities who have excellent labs and seminars after their cities are connected to those mega cities by bullet trains.

In Figure 4 we only include secondary cities. The benchmark (the red line) is the annual number of publications for those city-year observations without bullet train connection, and the blue line is this indicator for the city-year observations with bullet train connection. It shows that after the secondary cities are connected by bullet trains, their academic publications do have a jump. The 2008 huge jump is possibly due to selection effect (the secondary cities with relatively more universities are connected), but latter we do see the blue line has a stable gap over the non-connected city-year observations and also has a faster growth trend.

*** Insert Figure 3 about here ***

*** Insert Figure 4 about here ***

The bullet train network information is collected from the official website (www.12306.cn) of National Railway Administration of the People Republic of China (see Figure 2). On this website, we can identify whether and when a city is connected by bullet train, and also calculate the travel time between any two cities (by bullet train or regular train).The "*CONNECT*" dummy will turn on once a city is connected by bullet train, or once the two cities (or two universities) in a city-pair (or a university-pair) are connected by bullet train. We then construct city-city matrix including all cities with at least one co-publication where in each cell we have the travel time between these two cities by train, and this travel time will shrink after the two cities are connected by bullet trains.

We recognize that the state is unlikely to randomly choose which cities will be connected by bullet trains. To address this concern, we implement an instrumental variables (IV) regression approach and compare these results to the results based on ordinary least squares regressions. We will introduce how we construct the instrument variables in the next section.

Following the market potential literature (Harris, 1954; Hanson, 2005), we use the city-city travel distance matrix to construct an "intellectual potential" variable for each city each year. The introduction of bullet trains reduces the cross-city passenger travel time and thus enlarges the opportunities for idea flows and knowledge spillovers. If more scholars outside a city now become "closer" to this city due to the bullet train connection, the scholars in this city will have more opportunities to interact with outside scholars by traveling to those cities that are physically close but not very close to their own city to attend seminars, use labs, and talk and work with scholars there, therefore this city will enjoy an increase in academic production. We define "intellectual potential" (*INT_POTENTIAL*) as the distance weighted size of scholars of neighboring cities. For city *i*in year *t*, we have

$$INT\_POTENTIAL_{i,t} = \sum_{\mathrm{j \neq i}} SCHOLAR_{j,t}.\mathrm{f}(T_{ijt}) = \sum_{\mathrm{j \neq i}} SCHOLAR_{j,t}/T_{ijt} \quad [1]$$

Where $SCHOLAR_{j,t}$ is the number of scholars in all the colleges of city *j* in year *t*; The function $\mathrm{f}(T_{ijt})$ measures the spatial decay rate of a neighbor city *j*'s influence on city *i* as city *j* moves further away from city *i*. $T_{ij,t}$is the distance between city *i* and *j* in year *t* measured in train travel time (in minutes) and it will change when city *i* and *j* are connected by bullet train. In this paper, we set the function $\mathrm{f}(T_{ijt})$to be a simple form as the inverse of the time varying travel time between city *i* and city *j,* as $1/T_{ijt}$.

We collect additional time varying city variables as controls, mostly from the China City Statistical Yearbook from 2004 to 2016. Specifically, we collect the passenger volumes in both airport and highway for each city to control for other transportation modes. If the local government expects that the bullet train will bring in new opportunities for its universities, it may also invest in them. Therefore, we include the total government expenditure in high-tech area and the number of scholars in universities for each city as controls. Furthermore, the total population and GDP per capita are also included.

Table 1 provides the summary statistics of the key variables, for both city level, city-pair level, and university-pair level.

*** Insert Table 1 about here ***

The Empirical Framework and Instrument Variables

We have a two-fold empirical strategy: First, we examine the relationship between measures of academic productivity *Y* in city *i* in year *t* (Y_{it}, representing both research quantity and quality measures) and the city's status in the bullet train network – whether it is connected by bullet train (*connect*), or the time-variant intellectual potential calculated using train network (*int_potential*).

$$Y_{it} = \alpha_0 + \alpha_1 * X_{it} + \alpha_2 * int\_potential\ (connect)_{it} + \mu_{it} \quad (2)$$

Where X_{it} are time-variant controls. Y can be the number of papers (*papers*), the number of citation-weighted papers (*c_w_papers*), the number of impact factor-weighted papers (*i_w_papers*). We also control for regional fixed effects. We expect that, when second- and third-tier cities are connected to mega cities with superstar universities, they will enjoy a significant increase in intellectual potential, and thus an increase in academic productivity *Y*.

The second set of models focuses on the "idea flows" across cities. Such idea flows represents cross-city trade. We aim to provide evidence of the mechanism through which bullet trains increase scholars' productivity – their ability to ease the flow of ideas and local knowledge, which may serve as an important input to journal paper production. The co-authorship structure in our journal paper data set provides us the opportunity to observe such idea flows. We count the number of co-author pairs for city *i* and city *j* (in a pair, one author in city i, and the other in city j), or for university *m* and university *n*, and estimate:

$$CO_{ij(mn)t} = \beta_1 + \beta_2 * X_{ij(mn)t} + \beta_3 * connect_{ij(mn)t} + \varepsilon_{ij(mn)t} \quad (3)$$

Where CO_{ijt} (or CO_{mnt}) is the quantity or quality measure of the co-publications between city *i* and city *j*(or university *m* and *n*) in year *t*, which can also be *papers*, *c_w_papers, i_w_papers*. *CONNECT* will turn on after the two cities are connected by bullet train. We understand that it takes time for scholars to publish a paper, so we use the lagged measures of *connect* and *int_potential* in some of our regressions.

Our Web of Science data base allows us to disaggregate the data along five major categories including (life science and biomedicine; natural science; applied science, humanity, social science) and 151 minor categories. This variation allows to test for heterogeneity. For example, in the humanities labs are less important than in the physical sciences.

The main empirical challenge in estimating the above equations using OLS is the possible correlation between unobservables (μ_{it}, ε_{ijt}, and δ_{mnt}) and the city's status of bullet train connection.   Given that a city's bullet train treatment status is not randomly assigned, we must address concerns of omitted variable bias.

Here we seek to explain two different cases featuring interesting economic content. In the first case, if the Chinese central and local governments anticipate that there are beneficial synergies between second- or third-tier universities and easier access to the mega cities, then the leaders will invest more in the local universities in those newly bullet train connected second and third tier cities. Such a complementarity between public infrastructure (the bullet train) and university investment (private capital) would lead a researcher to over-estimate the role of the bullet train alone. Conversely, if we seek to estimate the "total effect" of the bullet train and if these new university investments would not have taken place in the absence of the bullet train then OLS estimates would yield the total effect.

To explore these issues, we obtain the researcher count and the value of investment in each university by year using data from the China Education Statistical Yearbook and other sources, so that we can partially measure the changes in such investments. We study this below.

In the second case, investments in universities and cities are correlated with bullet train connection but not caused by it. For instance, booming cities have a rising demand for transportation so that the State places bullet train stations there, and at the same time those cities have a larger fiscal capacity to invest in universities and other infrastructures. Or the State chooses to connect the weak cities into the high speed rail network to help them to grow, but the investment on universities there still lags behind. We employ an instrument variables approach to address the omitted variables problem in this case. We seek city level instrumental variables that are correlated with the likelihood that a city is connected by bullet train but that are unlikely to be correlated with the unobserved determinants of a city (or a university)'s academic output.

Following the transportation economics literature (Duranton and Turner 2012; Duranton el al.,2014), the first instrument variable we use comes from the nation's historical railway network. Baum-Snow et al. (2012) rely on the Chinese railroad networks from 1962 as sources of quasi-random variation in their regressions predicting roads' effect on regional economic growth. Zheng and Kahn (2013) also use China's 1961 railway road map as the IV to investigate bullet train's effect on housing prices. In this paper, we use the same instrument variable as that

Zheng and Kahn (2013) from a Chinese railroad map in – whether the city was connected by train in 1961 ($rail1961_i$). The validity of this historical instrument lies on the assumption that those factors which determined the railroad network in China have a weak effect on China's later economic development in this new twenty-first century. However, this old railway network IV is time-invariant. In order to fit with our panel data structure, we multiply our static historical IV times the annual construction rate of the high speed rail network in China, which generates a dynamic IV:

$$IV\_hist_{it} = rail1961_i * constructionrate_t \quad (4)$$

Where IV_hist_{it} represent the dynamic history IV for city *i* in year *t*. The $constructionrate_t$ is calculated as the ratio of the miles finished for the whole China in year *t* over the planned total miles in the State's "four vertical and four horizontal" high speed rail construction plan. We believe that this construction rate is a country level measure which is uncorrelated with a single city *i*'s possibility of being connected by bullet train or the city *i*'s paper publications. Therefore, IV_hist_{it} provide a credible quasi-random variation both across cities and over years.

For the second IV, we follow Faber (2015) to construct a hypothetical least-cost network connecting major Chinese cities. Faber's novel IV idea relies on identifying the routes which are more likely to be built if the sole policy objective is to connect all these target cities subject to construction cost minimization. In our paper, we first choose 20 targeted cities[7] based on their GDP level and population level in 2006. Secondly, we calculate the geographical distance between all city pairs in China, and use them as the proxy measure for construction cost for each

[7] We follow Faber (2015) to choose the number of target cities, and this number cannot be too small, otherwise, it can't guarantee to pass all other cities when we try to connect these target cities.

city pair. We then feed the distance data into the Kruskal's (1956) minimum spanning tree algorithm to identify the subset of routes that can connect all these 20 targeted cities to minimize the total construction costs. This algorithm can help us to indentify the minimum number and the possible connected cities of least cost connection to connect all the targeted cities on a single high speed rail network. We then have a planned IV for each city$i$ named as " $mstree_i$". For the same reason as forthe historical IV, we also need a time-variant IV for our panel data set, so we also multiple $mstree_i$ with the national high speed rail construction speed and get a time-variant planned instrument variable IV_plan_{it}.

The two instrument variables ($IV_hist_{it}$and IV_plan_{it}) will be used for the IV regressions of equation (2) and (3), where in the first stage we use these two IVs to predict the probability of a city's connection to bullet rain (or the intellectual potential increase due to bullet train connection), and in the second stage we use the predicted measures to estimate bullet train's impacts on city academic productivity and also regional idea flows.

Testing for Growth in City Academic Productivity

For city-level analysis, our main hypothesis is that after a city is connected by bullet train, the scholars in that city can enjoy a much easier and faster access to other scholars in those cities thatare too far to drive but too close to fly from their own city. The enhancement of such face-to-face interactions will increase this city's academic productivity, measured in both publication quantity and quality.

Table 2 presents our city-level regression results. The study period is 2006-2016. Panel A and B are OLS regressions, and Panel C and D are IV regressions. We first discuss about columns (1) to (3), in which we control for a long list of city attributes (GDP per capita, population, airport ridership, highway ridership, sizes of researchers and fund, tertiary industry share, and the city's latitude and longitude), and also province fixed effects and year fixed effects. Standard errors are clustered by city. Panel A shows that when a city is connected by bullet train, both the quantity (the number of papers) and the quality (citation-weighted and impact factor-weighted) of this city's academic publications see a significant jump of about 25%. However, the intellectual potential is insignificant in Panel B.

We now switch to the IV results in Panel C and D. Clearly the OLS estimates are biased to zero. Our IV results show that the introduction of bullet train, measured either by the connection dummy or the intellectual potential the city gains, has significantly positive effects on the city's academic productivity, and the coefficients become larger and more significant than those in OLS regressions. In columns (4) to (6) we replace province fixed effects with city fixed effects, and still control for the time-variant city attributes. All the coefficients shrink in size and their significance level also reduces. Nevertheless, in IV regressions in Panel C and D, we still see a significant effect of bullet train connection on academic productivity, especially for paper quantity. If we use Panel C of columns (4)-(6) as the benchmark, we observe a 13.5% increase in paper quantity and a 25.6% increase in citation-weighted papers after a city is connected to the bullet train network.

In later regressions, we will mainly discuss the results with province fixed effects (standard errors clustered by city) and time-variant city attributes, so that the spatial variation across cities within a province can also help us with the identification. This is also because our two IVs mainly have the cross-sectional variation, so we should use the cross-city variation for identification. The regression results with city fixed effects are presented in the Appendix.

It takes time for scholars to write a paper together and get it published. This time length varies in different disciplines – in some fields such as bio science and some science fields, this time is relatively short (the median waiting time from submission to acceptance is around 100 days)[8]; while in some other fields such as economics, it can be quite long. Therefore, it is reasonable to put the lagged measure of bullet train connection into our regressions (columns (7)-(9), with province fixed effects). The IV results show that the coefficients are similar than those in the previous columns.

*** Insert Table 2 about here ***

Since there is a huge disparity in the concentration of researchers in China's mega cities versus other cities, we expect that when those secondary cities are connected to the bullet train network, they can benefit more because they are now exposed to a much larger pool of scholars compared to the small stock in their own city. Table 3 confirms our prediction. No matter whether we use the connection dummy or the intellectual potential measure, and whether these variables are based on the current year or the year before, the effects on mega-cities are quite weak, while researchers in smaller bullet train connected cities enjoy a productivity boost.

*** Insert Table 3 about here ***

Regional Idea Flows

We take advantage of the pair structure of co-publications to examine the idea flows between cities. In this section we have two unit of analysis: city-pair and university-pair. For such pair-level analysis, we will always control for pair fixed effects and other time-variant controls. We will mainly discuss the results with one-year lagged bullet train connection

[8] See: , http://www.nature.com/news/does-it-take-too-long-to-publish-research-1.19320

indicator (*CONNECT_1*), and focus on citation-weighted paper quality measure. Other results (paper quantity, impact factor-weighted quality measure, and non-lagged) can be found in the Appendix.

Table 4 presents the city-pair results. We focus on impact factor-weighted papers, and the results for the paper quantity and the citation-weighted papers are similar (in Appendix). Panel A contains OLS results. Since the co-publications between some city-pairs are not too many, in Panel B we use negative binomial regression approach. Panel C shows the IV regression results. In column (1) for all city-pairs, the bullet train connection has significantly positive effect on their co-publications (citation-weighted papers). Panel A's estimate shows that, when two cities are connected by bullet train, they will see a 14.5% increase in their impact factor-weighted co-publications in a year after the connection. The negative binomial regression and IV regression also show the significantly positive effect of bullet train connection on impact factor-weighted co-publications.

In principle, there are a number of mechanisms through which bullet trains may affect the creation and diffusion of knowledge. It is possible that when cities are connected by bullet trains, scholars themselves will move to other cities. Research in urban economics has emphasized that transportation infrastructure generates regional growth through agglomeration economies, typically modeled as an inflow of new workers (Duranton and Turner, 2012). An important channel we want to highlight is that it does not require an influx of new innovators (Agrawal, Galasso and Oettl, 2017). Our results are robust when focusing on a sample of non-mover scholars, whose locations do not change during our study period. In our sample, 80% of the coauthor-pairs are non-movers. In column (2) we constrain our sample to those non-mover scholars, and the bullet train connection still has a significant effect on weighted publications in all three panels (OLS, negative binomial, and IV), and the size of its effect is just slightly smaller than that for all scholars. This reinforces our view that bullet trains facilitate the circulation of ideas even in the absence of the move of scholars.

We are also interested in the heterogeneous treatment effect that a city-pair receives if the two cities are both mega cities, or both are non-mega cities, or one is mega city and the other one is not (column (3)-(5) in Table 4). Our analysis in Section 2 indicates that when a non-mega city is connected with a mega city with abundant intellectual resources (brilliant scholars and advanced labs in the superstar universities there), this non-mega city will benefit a lot from the latter's spillovers. The regression results in columns (3) – (5) in Table 4 are consistent with our hypothesis. The subgroup of mega city and secondary city pairs receive the largest benefit from bullet train connection, while the connection between two mega cities have little effect (perhaps because they are already well connected by other transportation means and the academic exchanges between them are already very intensive).

*** Insert Table 4 about here ***

We repeat this exercise for university pairs (Table 5).We control for the total number of scholars, total research funds of the two universities within a pair as well as the two cities' attributes. We also include university-pair fixed effects. Column (1) shows that for all the university-pair observations, after the two cities where the two universities locate are connected by bullet train, the co-publications between the two universities significantly increase. The coefficient in Panel A (OLS) indicates that this jump is about 20.3% for impact factor-weighted papers. If we look at the non-mover scholars, the effect is also significant and is only slightly smaller than that for all scholars. Interestingly, we also find that, the bullet train connection between a weak university and a superstar one has a very strong effect on co-publications.

*** Insert Table 5 about here ***

Conclusion

In the modern university setting, most active researchers are aware of other researchers in their general research field. Such scholars may differ by age, quality and geographic location. While any scholar can email a leading scholar working at an elite Chinese university, the research synergies between the two are unlikely to be maximized without face to face interaction. If information technology was a perfect substitute for face to face interaction, then reductions in transportation costs should have no effect on research productivity. In this paper, we have rejected this claim based on our natural experiment focused on the impact of China's bullet train on researcher productivity.

The empirical results in this paper shows that once a city is connected into the bullet train network, the scholars in that city have a much easier access to nearby scholars and this leads to a significant increase in their academic productivity. We also show that after two cities or two universities are connected by bullet trains, co-publications between them will rise, indicating more intensive idea flows between them. We also find that weak cities and universities benefit more from this travel cost reduction. Our main finding that faster cross-city commuting speeds enhances productivity takes the original Gaspar and Glaeser (1998) work in a new direction. They argue that cities and information technology are complements and not substitutes. The benefits of face to face interaction increase if strangers recognize that once they have met that they can subsequently connect again by phone, Skype and email. Cities exist because they economize on transportation costs. If new technologies such as bullet trains effectively make nearby cities "closer" to superstar cities (through moving at a faster speed), then agglomeration benefits spread out further. The boundary of the agglomeration area is endogenous and hinges on speed.

References

References

Baum-Snow, Nathaniel, et al. "Roads, railways and decentralization of Chinese cities." Unpublished mimeo, Brown University. 2012.

Duranton G, Puga D. Micro-foundations of urban agglomeration economies. Handbook of regional and urban economics. 2004 Dec 31;4:2063-117.

Duranton G, Turner,MA. Urban Growth and Transportation, Review of economic studies, 2012 OCT; 79(4): 1407-1440.

Duranton G, Morrow,PM, Turner,MA. Roads and Trade: Evidence from the US, Review of economic studies, 2014 APR; 81(2): 681-724.

Faber, B. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System, Review of economic studies, 2014 JUL;81(3):1046-1070

Gaspar J, Glaeser EL. Information technology and the future of cities. Journal of urban economics. 1998 Jan 31;43(1):136-56.

Glaeser EL. Learning in cities. Journal of urban Economics. 1999 Sep 1;46(2):254-77.

Glaeser EL, Scheinkman J, Shleifer A. Economic growth in a cross-section of cities. Journal of monetary economics. 1995 Aug 31;36(1):117-43.

Harris CD. The market as a factor in the localization of industry in the United States. Annals of the Association of American Geographers. 1954 44:315-348.

Hanson GH. Market potential, increasing returns and geographic concentration. Journal of International Economics. 2005 67:1-24.

Helsley RW, Strange WC. Urban interactions and spatial structure. Journal of Economic Geography. 2007 Mar 1;7(2):119-38.

Jaffe AB, Trajtenberg M, Henderson R. Geographic localization of knowledge spillovers as evidenced by patent citations. the Quarterly journal of Economics. 1993 Aug 1:577-98.

Kruskal J B. On the Shortest Spanning Tree of a Graph and the Travelling Salesman Problem, Proceedings of the American Mathematical Society, 1956 7(1):48-50.

Moretti E. Human capital externalities in cities. Handbook of regional and urban economics. 2004 Dec 31;4:2243-91.

Rauch JE. Productivity gains from geographic concentration of human capital: evidence from the cities. National Bureau of Economic Research; 1991 Nov 1.

Rosenthal SS, Strange WC. Evidence on the nature and sources of agglomeration economies. Handbook of regional and urban economics. 2004 Dec 31;4:2119-71.

Rosenthal SS, Strange WC. Geography, industrial organization, and agglomeration. review of Economics and Statistics. 2003 May;85(2):377-93.

Rosenthal SS, Strange WC. The attenuation of human capital spillovers. Journal of Urban Economics. 2008 Sep 30;64(2):373-89.

Zheng S, Kahn ME. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceedings of the National Academy of Sciences. 2013 Apr 2;110(14):E1248-53.

Figure 1 Superstar Universities in China's Mega Cities

Figure 1-A Superstar Universities (those in the "985" Program)

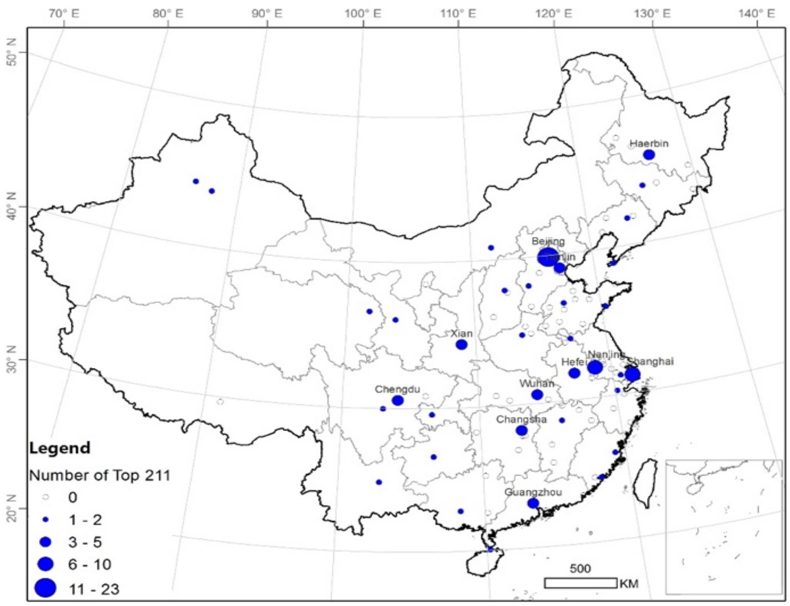


Figure 1-B International Journal Papers (science citation index & social science citation index, 2001-2015)

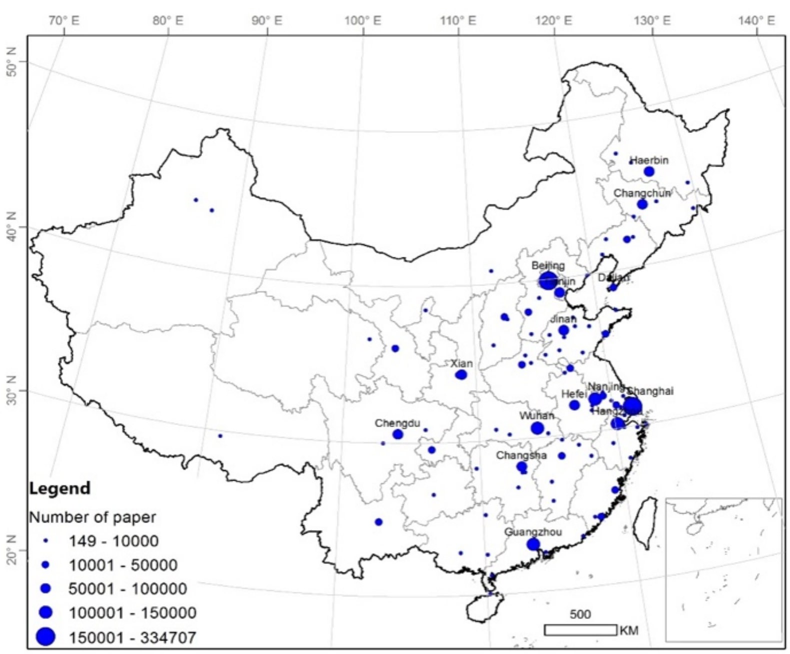
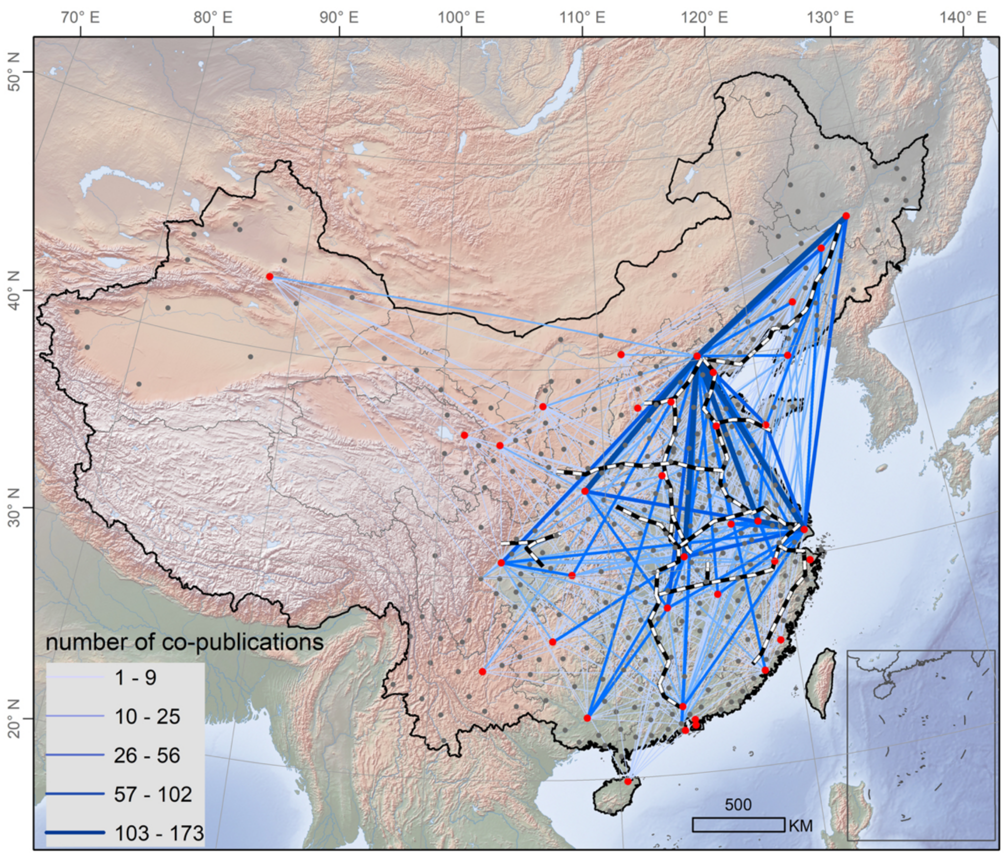


Figure 2 Co-publications between Cities and Bullet Trains



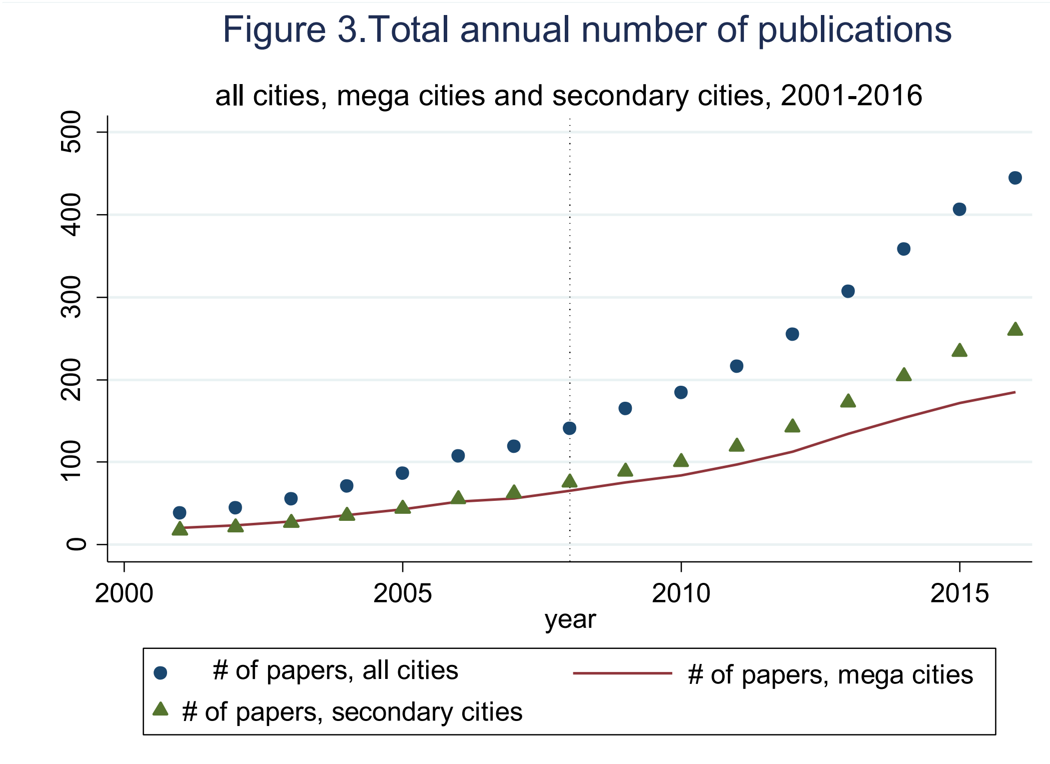


Figure 3.Total annual number of publications
all cities, mega cities and secondary cities, 2001-2016
0
100
200
300
400
500
2000
2005
2010
2015
year
# of papers, all cities
# of papers, mega cities
# of papers, secondary cities

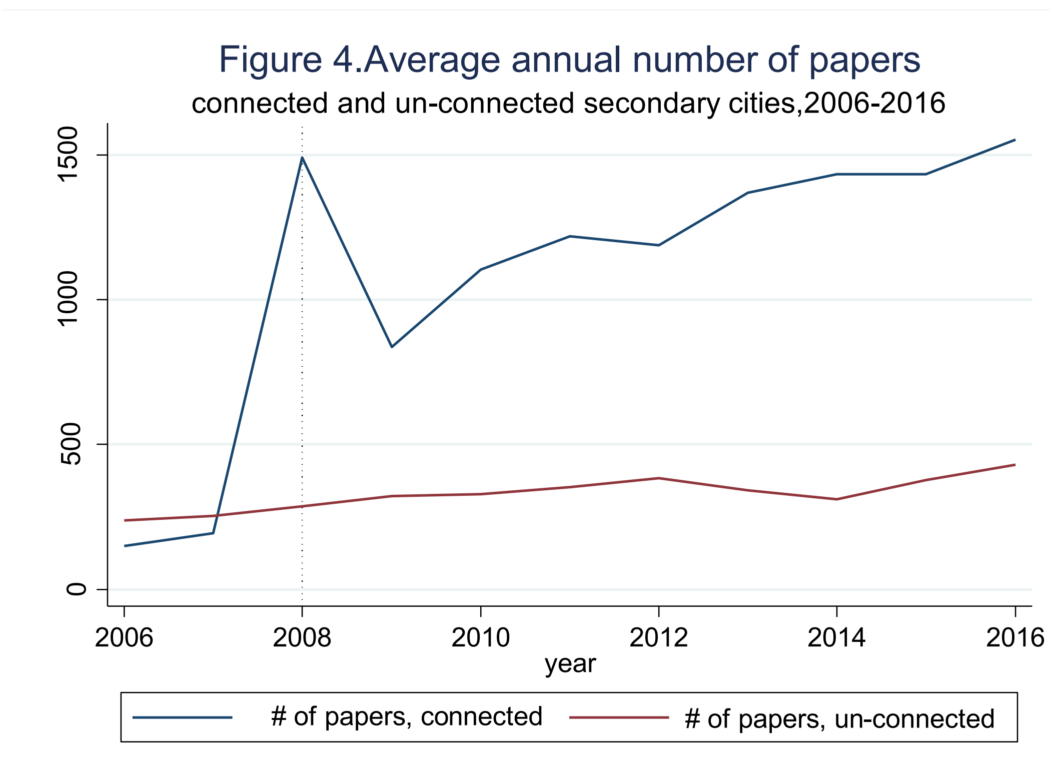


Figure 4.Average annual number of papers
connected and un-connected secondary cities,2006-2016
0
500
1000
1500
2006
2008
2010
2012
2014
2016
year
# of papers, connected
# of papers, un-connected

Table 1: Summary Statistics

Variable	Explaination	Obs	Mean	Std. Dev.	Min	Max
Panel A: City level (by city by year)						
year	Year	3,141	2010.982	3.157	2006	2016
papers	# number of papers in SCI&SSCI journals	2,924	922.498	3968.377	1	77331
c_w_papers	citation weighted papers	2,924	8773.845	41579.43	0	673600
i_w_papers	journal impact factor weighted papers	2,924	2561.652	12207.87	0	254488.5
connect	whether connected by bullet train	3,141	0.189	0.392	0	1
connect_1	connected by bullet train in last year	2,875	0.179	0.384	0	1
int_potential	distance weighted scholars outside the city	2,218	83558.33	230338	45.04111	2414220
iv_hist	IV: whether connected in 1961 railroad network	2,548	0.568	0.496	0	1
iv_plan	IV: hypothetical least-cost network	2,548	0.197	0.398	0	1
highway	ridership of highway	2,541	9197.205	15236.64	0	286557
airport	ridership of airport	2,390	643696.4	3715477	0	6.20E+07
scholar_size	# of scholars in college	2,496	4738.151	9210.13	26	67549
Fund_size	# of education funds by local government	2,560	47560.52	176556.3	0.13	2827117
Gdp_pc	city GDP per capita	2,498	37285.8	40940.88	2757.462	481692.3
population	city population	2,542	431.97	308.738	0.1	3375.2
tertiary	the ratio of tertiary over total employment	2,515	0.527	0.132	0.099	0.948
latitude	latitude of the city	2,548	32.934	6.685	16.8318	50.2451

longitude	longitude of the city	2,548	113.959	7.188	84.8892	131.159
Panel B: City-pair level						
year	year	41,902	2011.745	3.201	2006	2016
papers	# number of papers in SCI&SSCI journals	41,902	22.775	92.577	1	3248
c_w_papers	citation weighted papers	41,902	190.543	1051.502	0	49742
i_w_papers	journal impact factor weighted papers	41,902	52.808	261.793	0	11583.57
connect	whether the two cities are connected by bullet train	13,165	0.559	0.496	0	1
connect_1	whether the two cities are connected by bullet train in last year	11,572	0.525	0.499	0	1
iv_hist	city pair IV using 1961 railroad network	26,364	0.107	0.202	0	0.66
iv_plan	city pair IV using hypothetical least-cost network	26,429	0.057	0.157	0	0.66
scholar_size	# of scholars in the two universities	19,780	34290.07	20298.38	1059	124862
fund_size	Size of fund in the two universities	19,832	438094.3	646031.6	1040	5450030
Panel C: University-pair level						
year	sample period	39,460	2012.549	2.599	2006	2016
papers	# number of papers in SCI&SSCI journals	39,460	8.236	12.908	1	534
c_w_papers	citation weighted papers	39,460	62.778	171.834	0	7166
i_w_papers	journal impact factor weighted papers	39,460	20.181	45.475	0	2201.396
connect	whether the two cities (where the two universities locates) are connected by bullet train	19,004	0.549	0.498	0	1

connect_1	whether the two cities are connected by bullet train in last year	12,869	0.439	0.496	0	1
iv_hist	city pair IV using 1961 railroad network	34,448	0.278	0.251	0	0.66
iv_plan	city pair IV using hypothetical least-cost network	34,448	0.218	0.250	0	0.66
scholar_size	# of scholars in the two universities	25,775	2493.445	2127.892	60	22256
fund_size	Size of fund in the two universities	25,775	1246532	1131378	1777	8066039

Notes: The data is collected from Web of Science, 12306 website, the Chinese City Yearbook, and the author's manual collecting of the news of China High Speed Railway program during 2006 to 2016. Panel A presents the summary statistics of all variables in city level. There are 283 prefectures are included and cover the period from 2006 to 2016. Panel B reports the city pair level data summary, and Panel C shows the university pair level summary of data.

Table 2: Bullet train connection and city productivity increase

	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)
	papers	*c_w_papers*	*i_w_papers*	*papers*	*c_w_papers*	*i_w_papers*		*papers*	*c_w_papers*	*i_w_papers*
Panel A: OLS Regression with connection dummy										
connect	0.269**	0.266**	0.266**	-0.024	-0.047	-0.047	*connect* _1	0.173*	0.184	0.180*
	(0.112)	(0.121)	(0.121)	(0.031)	(0.062)	(0.062)		(0.094)	(0.116)	(0.101)
# of observations	2,067	2,067	2,067	2,067	2,067	2,067	# of observations	1,859	1,859	1,859
R-squared	0.840	0.786	0.786	0.982	0.948	0.948	R-squared	0.838	0.784	0.825
Panel B: OLS regression with intellectual potential measure										
int_potential	0.093	0.062	0.097	0.051	-0.027	0.028	*int_potential*_1	0.160**	0.128	0.148*
	(0.095)	(0.116)	(0.107)	(0.053)	(0.104)	(0.073)		(0.072)	(0.089)	(0.080)
# of observations	1,892	1,892	1,892	1,892	1,892	1,892	# of observations	1,670	1,670	1,670
R-squared	0.850	0.796	0.835	0.983	0.952	0.974	R-squared	0.851	0.795	0.837
Panel C: IV Regression with connection dummy										
connect	0.524***	0.535**	0.540***	0.136*	0.256*	0.165	*connect* _1	0.242**	0.234	0.286**
	(0.155)	(0.208)	(0.176)	(0.080)	(0.147)	(0.106)		(0.122)	(0.165)	(0.139)
# of observations	2,061	2,061	2,061	2,061	2,061	2,061	# of observations	1,854	1,854	1,854
R-squared	0.841	0.787	0.828				R-squared	0.840	0.787	0.827
Panel D: IV regression with intellectual potential measure										
int_potential	0.309***	0.320**	0.326**	0.216*	0.318	0.237	*int_potential*_1	0.371***	0.365	0.436
	(0.111)	(0.150)	(0.127)	(0.123)	(0.211)	(0.159)		(0.137)	(0.330)	(0.279)

# of observations	1,889	1,889	1,889	1,889	1,889	1,889	# of observations	1,668	1,505	1,505
R-squared	0.849	0.795	0.835				R-squared	0.848	0.791	0.832
city controls	YES	YES	YES	YES	YES	YES		YES	YES	YES
year dummies	YES	YES	YES	YES	YES	YES		YES	YES	YES
province FEs	YES	YES	YES	No	No	NO		YES	YES	YES
city FEs	NO	NO	NO	YES	YES	YES		NO	NO	NO

Notes: All regressions include a constant. Standard errors in column (1) (2) and (3) are clustered at the city level.*,**,***:significant at 10%,5%1%. All regressions control for the city GDP per capita, population, number of researchers, funds, airport and highway ridership, and tertiary industry ratio.

Table 3: Academic productivity increase in mega-cities and other cities

	Mega-cities	Secondary cities
Dependent variable (in log)	*i_w_papers*	*i_w_papers*
Panel A: IV regression with connection dummy		
connect	0.185	0.558***
	(0.255)	(0.187)
# of observations	44	2,017
R-squared	0.996	0.802
Panel B: IV regression with intellectual potential measure		
int_potential	0.158	0.321**
	(0.316)	(0.127)
# of observations	44	2,017
R-squared	0.996	0.802
Panel C: IV regression with connection lag dummy		
connect_1	-0.013	2.043**
	(0.047)	(0.826)
# of observations	40	1,816
R-squared	0.998	0.733
Panel D: IV regression with intellectual potential lag measure		
int_potential_1	-0.241	0.381**
	(0.290)	(0.163)
# of observations	39	1,620
R-squared	0.994	0.809
city controls	YES	YES
year dummies	YES	YES
province FEs	YES	YES

Notes: All regressions include a constant. *,**,***: significant at 10%,5%1%. All regressions include province fixed effects and year fixed effects. All regressions control for the city GDP per capita, population, number of researchers, funds, airport and highway ridership, and tertiary industry ratio. "Mega cities" are defined as Beijing, Shanghai, Nanjing, Guangzhou, Wuhan and Xi'an.

Table 4: Examining Regional Idea Flows Using City-Pair Co-publications

	(1)	(2)	(3)	(4)	(5)
	All	Non-movers	Mega-secondary	Secondary-secondary	Mega-mega
	i_w_papers	*i_w_papers*	*i_w_papers*	*i_w_papers*	*i_w_papers*
Panel A: OLS					
*connect*_1	0.145***	0.126**	0.238***	0.099	-0.169**
	(0.053)	(0.054)	(0.066)	(0.086)	(0.072)
# of observations	4,683	3,697	1,883	2,725	75
R-squared	0.859	0.878	0.887	0.766	0.973
Panel B: NBREG					
*connect*_1	0.134***	0.127***	0.140***	0.106***	-0.106**
	(0.025)	(0.026)	(0.035)	(0.034)	(0.053)
# of observations	7,448	6,162	2,757	4,592	99
# of city pairs	1,100	1,035	348	742	10
Panel C: IV					
*connect*_1	1.591***	0.881***	1.854***	1.375***	--
	(0.265)	(0.266)	(0.436)	(0.380)	--
# of observations	4,683	3,697	1,883	2,725	75
R-squared	0.307	0.346	0.411	0.166	0.351
city-pair controls	YES	YES	YES	YES	YES
year dummies	YES	YES	YES	YES	YES
city-pair FEs	YES	YES	YES	YES	YES

Notes: All regressions include a constant. *,**,***: significant at 10%,5%1%. All regressions include city-pair fixed effects, and year fixed effects. Panel A reports the results of OLS regression with city -pair controls. Panel B presents the Negative binomial regression results without city-pair control, and Panel C shows the IV regression results with city-pair controls. City-pair controls here refer as the sum of two cities GDP per capita, population, number of researchers, funds, airport and highway ridership. Column (1) is for all the pairs, and column(2) is for the non-movers group. "Mega cities" are defined as Beijing, Shanghai, Nanjing, Guangzhou, Wuhan and Xi'an.

Table 5: Examining Regional Idea Flows Using University-Pair Co-publications

	(1)	(2)	(3)	(4)	(5)
	All	Non-movers	985-non985	non985-non985	985-985
	i_w_papers	*i_w_papers*	*i_w_papers*	*i_w_papers*	*i_w_papers*
Panel A: OLS					
connect_1	0.203***	0.174***	0.137***	0.111*	0.023
	(0.035)	(0.038)	(0.050)	(0.067)	(0.121)
# of observations	11,172	8,278	4,599	2,492	1,187
R-squared	0.239	0.251	0.264	0.316	0.423
city-pair FEs	YES	YES	YES	YES	YES
Panel B: NBREG					
connect_1	0.036*	0.027	0.039	0.023	-0.003
	(0.020)	(0.022)	(0.026)	(0.039)	(0.047)
# of observations	15,371	11,435	8,762	4,103	2,506
Number of univid	3,342	2,819	1,870	1,115	357
Panel C: IV					
connect_1	0.585**	0.707***	0.588***	0.409***	0.700**
	(0.276)	(0.095)	(0.104)	(0.135)	(0.308)
# of observations	16,534	8,274	6,245	3,420	1,502
R-squared	0.067	0.134	0.132	0.119	0.363
province-pair FEs	YES	YES	YES	YES	YES
city-pair and university-pair controls	YES	YES	YES	YES	YES
year dummies	YES	YES	YES	YES	YES

Notes: All regressions include a constant. *,**,***: significant at 10%,5%1%. All regressions include year fixed effects and city-pair and university pair controls. Panel A reports the results of OLS regression with city -pair fixed effects. Panel B presents the Negative binomial regression results, and Panel C shows the IV regression results with province-pair fixed effects. University-pair controls are the sum of two universities researchers and funds. City-pair controls here refer as the sum of two cities GDP per capita, population, airport and highway ridership. Column (1) is for all the pairs, and column (2) is for the non-movers group."985" here represents these 39 universities which selected by the Chinese government in the Program 985.