NBER WORKING PAPER SERIES

THE ROLE OF TRANSPORTATION SPEED IN FACILITATING HIGH SKILLED TEAMWORK

Xiaofang Dong
Siqi Zheng
Matthew E. Kahn

Working Paper 24539
http://www.nber.org/papers/w24539

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2018

We thank Jeff Zabel, Andrea Craig, Nolan Jones and the participants at the 2017 Urban Economics and Public Finance Conference at the Lincoln Institute of Land Policy for their helpful comments. Siqi Zheng acknowledges the research support from MIT China Future City Lab. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Role of Transportation Speed in Facilitating High Skilled Teamwork
Xiaofang Dong, Siqi Zheng, and Matthew E. Kahn
NBER Working Paper No. 24539
April 2018
JEL No. O15,O31,R4

ABSTRACT

High skilled workers gain from face to face interactions. If the skilled can move at higher speeds, then knowledge diffusion and idea spillovers are likely to reach greater distances. This paper uses the construction of China's high speed rail (HSR) network as a natural experiment to test this claim. HSR connects major cities, that feature the nation's best universities, to secondary cities. Since bullet trains reduce cross-city commute times, they reduce the cost of face-to-face interactions between skilled workers who work in different cities. Using a data base listing research paper publication and citations, we document a complementarity effect between knowledge production and the transportation network. Co-authors' productivity rises and more new co-author pairs emerge when secondary cities are connected by bullet train to China's major cities.

Xiaofang Dong
The Wang Yanan Institute for
Studies in Economics
Xiamen University
Xiamen, Fujian 361005
China
xfdong@xmu.edu.cn

Siqi Zheng
MIT China Future City Lab
Center for Real Estate
Department of Urban Studies and Planning
Massachusetts Institute of Technology
Cambridge, MA 02139
sqzheng@mit.edu

Matthew E. Kahn
Department of Economics
University of Southern California
KAP
Los Angeles, CA 90089
and NBER
kahnme@usc.edu

1. Introduction

Worker productivity is higher in high human capital cities (Moretti 2004, Rauch 1991). These cities offer greater learning opportunities but the knowledge spillovers are localized (Glaeser and Mare 2001, Rosenthal and Strange 2003, 2004). The localization of spillovers emerges in an economy where low transportation speeds limits who can interact with whom.

The most productive and ambitious workers often migrate to major cities and then become even more productive through interactions and learning. This dynamic is playing out now in China. In China today, basic research takes place at the major universities that are disproportionately concentrated in a small number of major cities (Figure 1-A).[1] If cross-city transportation costs are high, researchers working in secondary cities will have fewer opportunities to visit major cities and to interact with leading researchers there. Transportation costs limit the ability of a talented professor at a top university to work with researchers with complementary skills who works in another city.

This paper uses China's recent investment in the creation of the high speed rail (HSR) network as a natural experiment to test for the role of cross-city transportation speed as a key determinant of urban productivity. We study whether reductions in cross-city transportation costs facilitate matching and interactions between scientists. This process results in more research ideas and higher quality research ideas. The academic production process creates published papers and yields research citations. This output offers us quantitative metrics for measuring the economic geography of productivity dynamics and for identifying cross-city flows

[1]These nine cities are Beijing, Shanghai, Nanjing, Guangzhou, Wuhan, Tianjin, Chengdu, Changsha and Xi'an. Each has more than two top universities.

of ideas and teamwork (Jaffe, Trajtenberg and Henderson 1993, Henderson, Jaffe and Trajtenberg 1998). While our empirics focus on university researcher knowledge creation, we posit that cross-city transportation speed also facilitates private sector R&D as firms that site their labs in different locations can more frequently meet.[2]

China's HSR allows individuals to move across cities at speeds of roughly 175 miles per hour. Such increased speed (a doubling of past train speeds) increases the menu of locations that have access to mega cities. By the end of 2020, the HSR network will connect 113 Chinese cities with population greater than one million. These trains offer a high quality comfortable ride and greatly reduce the travel time of commuting across cities. There are many pairs of cities such as Beijing and Shijiazhuang, Shanghai and Hefei, that are too far apart to drive and too close to fly. Below, we will explicitly test for whether HSR is associated with a larger increase in trade in ideas between these cities.

Using a data set of academic publications in the Web of Science (WoS, run by Thomson Reuters), we document that researchers working in second tier cities enjoy a productivity boost when their city is HSR-connected with the major cities. With increased travel speeds, individual researchers in HSR connected secondary cities can visit top universities in mega cities more frequently, listening to lectures and attending conferences. Such spillover effects may directly improve their productivity by facilitating the learning process. Furthermore, in this age of specialized knowledge production, complex research tasks need to be completed by a team composed of individuals with complementary skills. Of all the papers in our data set, 97% are co-authored.

[2] Chinese patent citation data is publicly unavailable, which makes it difficult to test similar mechanisms in the private sector.

We explore the matching and interaction of research teams. We test three possible channels through which the HSR could cause a growth in researcher productivity. First, existing pairs of co-authors located in two bullet train connected cities might interact more ("the intensive margin"). Second, new matches might now take place because the bullet train increases the interactions between scientists in these two cities ("the extensive margin"). This mechanism builds on the claim that information technology is a complement for urbanization (Gaspar and Glaeser 1998). They document that information technology accelerates the speed that ideas flow across cities. In our case, the bullet train increases the speed that people move across cities. If face-to-face interaction did not matter in the scientific idea creation process, then the bullet train would have little marginal value added in an economy featuring widespread information technology access.

The third channel focuses on the locational choice of scholars. Once cities are connected to the HSR network, these secondary cities become more attractive locations for young scholars because of their cheaper living cost and greater access to the mega city scholars. Such rising scholars can now more easily visit the mega city to attend conferences, and network with the big city star scientists.

We recognize that the Chinese central government is unlikely to randomly choose which cities to connect by high speed rail. To address this concern, we implement an instrumental variables (IV) regression approach and compare these results to the results based on ordinary least squares (OLS) regressions. This endogeneity issue is less of a concern for our city-pair analysis because cities are not connected by the bullet train to achieve the goal of promoting more cross-city research collaboration.

We find that after a city is connected by high speed rail, the city's academics production of academic papers increases on average by 30.1% in quantity (# of papers) and by 30.9 % in quality (impact-factor weighted # of papers) in the following years. At the city-pair level, higher travel speed facilitates matching and idea flows between two HSR-connected cities. We estimate that there are larger productivity gains for the subset of secondary cities most likely to gain the most from HSR. These are cities too close to the major cities for there to be active air travel but too far from the big cities to facilitate short vehicle rides. It is this subset of cities that would be expected to gain the most from the HSR.

Our finding that HSR connections increase cross-city trade in ideas and research partnerships has implications for the urban growth and human capital literature. Rauch (1991) and Glaeser et. al. (1995) have implicitly assumed that human capital spillovers benefit the geographic area in question (such as the metro area or county) but do not spillover across geographic units. In contrast, our study highlights that the extent of the interaction across geographic boundaries is related to travel speeds. As travel speeds increase (and the cost of cross-city interactions decline), the stock of human capital in one area may increase economic growth in HSR connected cities that now have greater "labor market access". In this sense, our paper contributes to the market potential literature (Harris 1954, Hanson 2005). The past market potential literature has emphasized the earnings possibilities for a geographic area increases as a function of its proximity to populations with high incomes (consumers). In our setting, the productivity of scientists in smaller cities increases as their "proximity" to major cities in China increases.

Our findings also have implications for cross-city inequality across China. Since more people living in HSR connected secondary cities directly benefit from greater access to the superstar researchers in the small number of mega cities, high speed rail effectively reduces the sharp human capital inequality across cities and contributes to urban economic growth in the 2nd tier cities.

Our study builds on previous research investigating the consequences of HSR for different aspects of the Chinese economy. Zheng and Kahn (2013) document that this transport innovation is associated with rising real estate prices in the nearby secondary cities. Lin (2017) reports that an HSR connection increases the city's passenger flows by 10% and employment by 7%. Industries with a greater reliance on non-routine cognitive skills benefit more from HSR-induced market access to other cities.

The rest of the paper is organized as follows. We present our conceptual framework in Section 2. Section 3 discusses the economic geography of China cities, the spatial distribution of research productivity and the high speed rail system's development. Section 4 presents our empirical models and results.

2. Travel Speed and Knowledge Creation

Given the specialized nature of knowledge creation, many research teams engage in complex tasks completed by individuals with complementary skills. If knowledge is produced by scientists employed by for-profit firms, then the firm will locate its scientists efficiently because all of the spillovers are internalized within the firm (Rossi-Hansberg and Sarte 2009). In cases where there are cross firm spillovers, then firms will co-agglomerate in close physical proximity (Arzaghi and Henderson, 2008).

Within a university, the spatial proximity of various researchers facilitates learning and spillover effects (Claudel, et. al., 2017). A major city will be the home to many universities so the local matching pool is larger. However, in this age of specialized research, even for the major cities with thousands of researchers, the gains to trade knowledge may not be exhausted. In this sense, transportation costs between cities limit the ability of research teams to effectively work together, and the marginal benefits of connections to more scholars in other cities could be large. Although researchers can discuss ideas via email and Skype, such information technologies are not a perfect substitute for face to face interactions for these high-skilled workers (Catalini, Fons-Rosen and Gaule, 2016).

High speed rail increases the travel speed across cities, and thus creates a larger local market for highly productive workers. Starting with the work of Mortenson and Pissarides (1999), labor economists have modeled the search process such that workers and firms are matched through a matching function. The extent of the creation of new hires is a function of the count of people seeking a job and the count of posted jobs and a stochastic process determining how quickly they are matched. The matching possibilities in a city are a local public good that all entrants enjoy (Helsley and Strange,1990). Low transportation speeds limit the entry of individuals from other cities to enjoy these local public goods.

In the formation of new research teams, scientists who seek to co-author with other scientists face a search process for meeting and interacting. Major academic conferences in large cities can solve this co-ordination issue. In this case, new research teams may form due to better matching of researchers with complementary skills in a larger academic labor market. Such matching is facilitated by the increased cross-city travel speed. We call this channel the "extensive margin".

There is also an "intensive margin". For those research teams that have already formed, high speed rail reduces the price of face to face interaction. This encourages such teams to work together more and perhaps to invest in pair specific human capital because they expect to interact more often in the future. If scientists scattered across universities in different cities can easily meet and interact with each other, then aggregate productivity increases.

Scholars can move across cities and universities. New Ph.D. graduates will choose which city to start their academic career, and senior researchers may change their job and move to other cities. The increase in cross-city transportation speed expands the "menu" of cities that researchers can consider when searching for a job. This will influence human capital investment strategies given the uncertainty in researchers' future locations. Suppose a superstar researcher is training a new set of PhD students. She will make more matched specific investments in these young scholars' skills if she knows her students will continue to work with her in the future. Young researchers will be more likely to invest in relationship specific human capital if they anticipate that they will have easy access to the superstar researcher in the future (Azoulay, Graff-Zivin and Wang, 2010). This channel highlights the complementarity between specific skill development and the extent of the market.

2.1 Testing for the Productivity Effects of Increased Cross-City Travel Speeds

Our empirical work features three different units of analysis: the city, the city-pair, and the researcher level. At the city level, we examine whether the HSR connection increases a city's academic productivity. We construct two instrumental variables to address the possible endogeneity in HSR placement (see Section 4.2). The rise in a city's academic productivity may come from better matching and interactions between scientists in connected cities (knowledge trade), or from individual scientists' learning process facilitated by easier and more frequent travels to great conferences (knowledge spillover). We are not able to disentangle these two sources. Nevertheless, in this age of specialized knowledge production and the prevalence of teamwork in the academia (97% of the papers in our data set are co-authored papers), better matching and interactions may be the dominant mechanisms here.

After establishing that HSR-connected cities do enjoy an increased research productivity, we then test several plausible mechanisms at the city-pair level and at the researcher level. City-pair level models are estimated to study idea flows across cities. For example, one city-pair is Beijing and Tianjin. Such idea flows represent a cross-city trade of knowledge. The HSR facilitates this knowledge trade. The co-authorship structure in our journal paper data set provides us with the opportunity to track such idea flows.

Throughout this paper, we define treatment as representing whether a city is connected to the HSR network, or whether two specific cities are HSR connected. We test whether treatment is associated with an increase in the quantity and quality of co-authored papers written by scientists located in HSR-connected cities. We then restrict our sample to "non-movers" (scientists who continue to live in the same city after its connection to the HSR network) to highlight that high speed rail facilitates inter-city matching and interactions without requiring an influx of new scholars. We also document evidence of the role of both the extensive and intensive margins by splitting our researcher sample into "first-time co-authors" and "incumbent co-authors". Travel time should matter here –those city pairs with the distance within HSR's comparative advantage (i.e., within a three-hour travel time) should experience larger gains.

At the researcher level, we focus on a subset of "movers" (scientists who move to another city during our study period). For each researcher in our sample who has moved across cities, we know her origination city and destination city. We use a conditional logit model of locational choice to explore whether she is more likely to move to cities with an HSR connection. Each scholar faces a tradeoff. The mega cities with top universities have excellent academic resources and a large pool of star scholars, but living costs (e.g. home prices) there are extremely high (Zheng et. al. 2016). After being connected by HSR, those secondary cities with a reasonable commute time to the major cities may gain by attracting more researchers to live there.

3. The Geography of Research Productivity and the High Speed Railway

3.1 Knowledge Production in Chinese Universities

China started to build its modern universities in the late Qing Dynasty (around 1890). Peking University and Tsinghua University were founded in 1898 and 1911, respectively. The elite universities were founded in major cities and have received state investment during the Republic of China era, and later from the Chinese Communist Party starting in 1949. As of 2016, there are 803 universities in China and 98% of them are public universities. This system is quite hierarchical. At the top, there are 39 first-tier universities that receive favorable research and teaching funds from the State (called the "top universities" thereafter).[3] They have excellent laboratories, libraries and databases and attract the best scholars and talented students. This human capital agglomeration process reinforces itself.

China has 287 prefecture-level cities, but 60% of the top universities in China are concentrated in nine major cities.[4] In China, the best universities are disproportionately concentrated in a small number of mega cities (Figure 1-A). Beijing, Shanghai, Changsha, Nanjing, Guangzhou, Wuhan, Tianjin, Chengdu and Xi'an, all have more than two top universities, and together they account for 64% of all top universities. Beijing has eight top universities, and Shanghai has four . On the other end of the spectrum, almost 100 cities (most are small cities) have no universities. The medium sized cities feature the largest count of average quality universities. Such a spatial concentration of top universities generates a clear disparity in academic productivity across Chinese cities. To develop world-class universities, the Chinese government allocates most funding to a few elite universities which house the most productive researchers and the most advanced laboratories (Freeman and Huang 2015). During the years 2006-2016, more than half of peer-reviewed international journal papers were produced by scholars at the universities in the aforementioned nine mega cities (Figure 1-B).

To measure academic productivity, we collect all of the international journal publications from Chinese universities during the years 2001 to 2016 (featuring at least one author from Mainland China) from the database of "Web of Science". To increase the international visibility of Chinese research, the number and quality of WoS papers have been widely used by the

[3]These universities are included in the "985 Program". This 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.

Chinese government to evaluate the research performance since late 1990s. Chinese scholars are required to publish WoS papers to be promoted, while their affiliated universities need WoS papers to rise in the rankings and to increase their funding from the government. Universities offer preferential policies and monetary rewards to encourage their scholars to publish in the WoS-indexed journals with high impact factors (Quan, et. al., 2017). Such policies provide a further incentive for researchers to work with stars at elite universities.

For each paper, we obtain the paper's title, the names and affiliations of the authors, the publication date, journal field, journal impact factor, and the citation count (as of December 2016).[5] Our final data set includes roughly 1.5 million journal papers. Figure 1-B shows that these papers are disproportionally authored by researchers who are concentrated in a few mega cities where the top universities are located.

We construct both quantity and quality measures for academic papers. We count the number of papers as the quantity measure. To measure quality, we create two indicators–the journal impact factor weighted number of papers (as a measure of how many papers are published in impactful journals) and the citation weighted number of papers (as a measure of how impactful the papers are). Both indicators yield very consistent and similar results. We will mainly report the results using the first indicator.

The importance of face-to-face interaction and the publication cycle vary across fields. Using each journal's academic field information in the "Web of Science" database, we classify all papers into five fields:(1) Arts& Humanities; (2) Social Science; (3) Life Sciences and

[5]For the author records with the same first name, last name and affiliation, we assume they are written by the same person. Miscoding may occur when two scholars at one affiliation have exactly the same first name and last name, but such cases should be very rare.

Biomedicine; (4) Physical Sciences; and (5) Technology. In our regressions, we control for city fixed effects and field fixed effects. When testing for the heterogeneous effect of high speed rail connection on academic publication, we combine (1) and (2) into a broad category of social science, and (3), (4) and (5) into a broad category of science and technology. We expect that researchers in the social sciences may benefit more from increased travel speeds across cities because social scientists spend less time in labs but more time in face-to-face discussions.

3.2 The Economic Geography of the High Speed Rail Routes

The HSR network information is collected from the official website (www.12306.cn) of the National Railway Administration of the People Republic of China (see Figure 2). On this website, we identify whether and when a city is connected by HSR, and also calculate the travel time between any two cities (by HSR or regular train). The "*CONNECT*" dummy equals one once a city is connected by HSR, or once the two cities in a city-pair are connected by HSR. We then construct a 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. This cross-city travel time shrinks after the two cities are connected by HSR.

Given that a city's HSR treatment status is not randomly assigned, we must address the site selection issue. If the local governments in second-tier Chinese cities anticipate that there are beneficial synergies between their ordinary universities and easier access to the mega cities, then the leaders will invest more in the local universities in those newly HSR connected secondary cities. Such a complementarity between public infrastructure (high speed rail) and university investment would lead the econometrician to over-estimate the role of HSR alone. We note that if we seek to estimate HSR's "total effect" on research productivity then OLS estimates yield the right answer. In this case, we would be assuming that all of the new investment by universities was caused by the construction of the HSR. To test this complementary investment hypothesis, we collect the amount of investment in each university by year using data from the China Education Statistical Yearbook and other sources. We do not observe significant increases in the research funding after HSR connection (See Appendix 3). But our investment variable might not capture all the resources city leaders put on the universities in their jurisdiction.

In the second case, investments in universities and cities are correlated with HSR connection but not caused by it. For instance, booming cities have a rising demand for transportation so that the central government places HSR stations there, and at the same time those cities have a greater fiscal capacity to invest in universities and other infrastructure. Or the State chooses to connect the weak cities into the HSR network to help them to grow, but the investment on universities there still lags behind.

We employ an instrumental variables approach to address this omitted variables concern. We seek city level instrumental variables that are correlated with the likelihood that a city is connected by HSR but that are unlikely to be correlated with the unobserved determinants of a city's academic output. Following the transportation economics literature, we construct two instrumental variables for the city level analysis. The first one (IV_hist_{i}) is based on the nation's historical railway network (Baum-Snow et. al. 2017; Zheng and Kahn 2013), and the second one (IV_military_i) is based on the spatial distribution of major military troop deployments in 2005

(Zheng and Kahn 2013). One purpose for China's central government to build high speed rail network is to ship troops in case of emergency.[6]

The endogeneity issue is a lesser concern for the city-pair analysis. It is not plausible that cities are connected by the bullet train because of the goal of promoting more cross-city research collaboration. Therefore, we employ the IV estimator in the city-level analysis but not in the city-pair level analysis.

3.3 Descriptive Evidence on Team Production and Cross-City Travel Speed

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 and diffusion between coauthors. That is why in this modern era with advanced online communication technologies (Email, Skype, mobile phone, etc.), scholars are still willing to attend conferences and seminars to present their research work, and interact with other scholars. High speed rail 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 HSR. For papers with multiple authors, we work with the first three authors, and their corresponding affiliations and cities[7]. We then construct one-to-one author

[6]See: Hai J (2010) The military significance of and our reflection on the high-speed railways of our country. *Traffic Engineering and Technology for National Defense* 5:5-7. (in Chinese)

[7]In Chinese universities, the order of authors is very important for judging a specific author's contribution to the paper. This information is used for promotion and performance evaluation purposes.

pairs between the coauthors. For instance, if a paper has two coauthors: A and B, then we will construct one coauthor pair A-B. If another paper has three coauthors: C, D and E, then we will construct three co-author pairs: C-D, C-E, and D-E. Since we know the city name each author was working in when publishing this paper, we count the number of paper publications (or weighted by impact factor, citation count) for each city-pair by year. Figure 2 shows the spatial distribution of the co-publications between cities. Mega cities with top universities represent the cores in the co-authorship rays.

*** Insert Figure 2 about here ***

We seek to identify whether an author in our data set is a "mover" (who moves to another city in our study period) or a "non-mover" (who continues to live in the city). CV information is publicly unavailable for most of the researchers in our data set, so we have to rely on the affiliation information in their papers. For an author who appears more than once 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".[8] For those authors who only published one paper (accounting for 36% of all authors in our data set during our whole study period), we are unable to tell whether they are mover or non-mover. Our labels of movers and non-movers are "narrowly-defined" and they only apply to a subset of authors with multiple publications. For co-author pairs, we define this pair as a non-mover pair only if both authors in a pair did not move in our study period. When we use the author-level data to explore moved researchers' location

[8]We do not observe the migration of a new researcher before she publishes her first paper.

choices, we only track the affiliation change of each paper's first author to construct the subsample of moved researchers.

We collect additional time varying city variables as controls. The main data source is the China City Statistical Yearbook covering the years 2004 to 2016. We collect data on the passenger volumes in both airport and highway for each city to control for other transportation modes. The total population and GDP per capita are also included as controls.

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

*** Insert Table 1 about here ***

4. Empirical Models and Results

At the city-level and city-pair level, we employ fixed effects and instrumental variable estimation techniques to measure the association between HSR connection and the quantity and quality of cross-city research scholarship. Then we employ a conditional logit model to explore a moved researcher' city choice.

4.1 Growth in city research productivity

At the city level, we examine the relationship between the academic productivity measure Y in city i in field s in year t (Y_{ist}, representing either research quantity or quality measure) and the city's status in the HSR network –the year which the city was firstly connected by the bullet train network (with at least one high speed railway station) (*connect*$_{it}$). Our benchmark fixed effects specification is:

$$Y_{ist} = \alpha_0 + \alpha_1 connect_{i,t-t0} + \beta' X_{it} + \delta_i + \sigma_s + \tau_t + \varepsilon_{ist} \qquad (1)$$

Where X_{it} are time-variant controls for city i.$Y_{ist}$is either the number of papers (*papers*), or the number of impact factor-weighted papers (*weighted_papers*)in city *i* in field *s* in year *t* for the quantitative and qualitative analyses, respectively. We include city fixed effects(δ_i), year fixed effects(τ_t) , and field fixed effects (σ_s), which account for many omitted variables and some sources of endogeneity. For example, cities with a strong industrial base may be more likely to have a HSR connection, and at the same time they may also have good universities in the engineering field and thus produce more papers in that field. To account for serial correlation, standard errors are clustered at province level.

We expect that when a city is connected by bullet train (the *connect* dummy turns on), it will enjoy a significant increase in academic productivity *Y*. Since the academic publication review process takes time, we introduce a lagged term for *connect* (the lag period is t_0).

Table 2 reports the effects of HSR connection on the academic productivity at city level, based on a panel data set. Panel A of Table 2 presents OLS regressions of equation (1). The dependent variable is log of the (1+count of papers) or the log of (1+quality weighted papers).[9] Column (1) and column (2) present the OLS regressions based on the full sample. The number of papers and impact factor weighted papers for cities after the HSR connection increases by 7.8% and 6.3%, respectively, although the latter estimate is statistically insignificant.

[9]The regression controls include the GDP per capita, population, airport ridership, and highway ridership. Standard errors are clustered by province level, and we also include city fixed effects, field fixed effects and year dummies.

*** Insert Table 2 about here ***

There are two possible reasons for this insignificant effect. First, 38% of the observations have the value of zero publication. Although we add one to the number of publications and then take logarithm, the distribution of this dependent variable is still far from being normally distributed. Second, as discussed above, there may be endogeneity issue between a city's academic productivity and its probability of HSR connection.

We employ zero-inflated negative binomial regression technique to address the first issue (see panel B).The coefficients of *connect_1* is statistically significant with a value of 0.339. This indicates that the HSR connection is associated with a 40% increase in the quantity of publications in a city ($e^{0.339} - 1$). Column (2) shows that this increase is 36% for quality-weighted papers. We now restrict our sample to the "narrowly-defined" non-movers, and the effect of bullet train connection is larger for the paper quality improvement for this subgroup in Panel B. Note that such effects include both the benefit from knowledge trade (generated by interactions between researchers) and knowledge spillovers (received by individual researchers) brought by the HSR connection. In Appendix 2 we report the regressions of equation (1) with the current year or 2-year lag HSR connection indicators. The results are similar with those in Table 2.

The effect of HSR connection differ by discipline. Social science is expected to rely more on face-to-face communication. We report the results for publication quality by discipline in column (5) and (6). As expected, the HSR effect is statistically significant for both disciplines but it is slightly larger for social science papers.

We now employ the IV estimator. Since our two instruments are time-invariant, we are unable to run the IV regression using the panel data set. Instead, we estimate the following long-difference equation over the years 2006 to 2015:

$$\log(Y_{isp2015}) - \log(Y_{isp2006})$$
$$= \alpha + \beta connect_{ip2006-2015} + \theta_1(X^1_{i2015} - X^1_{i2006}) + \theta_2 X^2_{i,pre-existing}$$
$$+ \gamma \log(Y_{isp2006}) + \sigma_s + \mu_p + \varepsilon_{isp} \qquad (2)$$

Where $log(Y_{isp2015}) - log(Y_{isp2006})$ measures the growth rate in the number of papers (*papers*), the number of impact factor-weighted papers (*weighted_papers*)in field *s* in city *i* of province *p* during 2006 to 2015. $connect_{ip2006-2015}$ equals one if this city is connected by HSR during this study period. Time-varying city attributes (X_i) are also first-differenced. Field fixed effects (σ_s) and province fixed effects(μ_p)are included. Following Agrawal et. al.(2017) and Faber(2014), we also control for the research productivity in the initial year ($Y_{isp2006}$), and a set of variables measuring pre-existing demographic, economic and political conditions of city *i* ($X'_{i,pre-existing}$). The identifying assumption is that our two instrument variables affect changes in academic productivity only through HSR connections, conditional on all the controls we include in Equation (2).

Table 3 presents both OLS and IV regression results based on the long difference equation (2). Column (1) and (2) are for all cities. The OLS coefficients in Panel A are still insignificant, but the IV regressions in Panel B yield significantly positive effects of HSR connection on the city's academic productivity. HSR connection is associated with around a 30% increase in both publication quantity and quality. This magnitude is quite similar to the zero-inflated negative binomial regressions reported in Table 2. Such effects are even larger for non-movers. Columns

(5) to (8) illustrate that scholars in social science benefit more from this new transportation technology than those in science and technology.

*** Insert Table 3 about here ***

4.2 HSR Boosts Productivity at the Intensive Margin

We take advantage of the pair structure of the coauthored papers to examine the idea flows between cities. In this section, our unit of analysis is the city-pair by year. We count the number (or the quality-adjusted number) of publications for the co-author pairs in city *i* and city *j* (for each pair, one author in city *i*, and the other in city *j*)in year *t* ($CO_{ij,t}$), and estimate:

$$CO_{ijt} = \gamma_1 + \gamma_2 connect_{ijt-t0} + \rho' X_{ijt} + \tau_{ij} + \sigma_t + \mu_{ijt} \quad (3)$$

In equation (3) X_{ijt} are time-varying controls for city *i* and *j in year t*. We include year fixed effects (σ_t), and city pair fixed effects (τ_{ij}) which account for time-invariant omitted variables for these two cities, and standard errors are clustered at the province pair level. $connect_{ijt-t0}$ measures whether the city pair was first connected by the bullet train (with at least one bullet train connecting two cities directly) in this year.

Table 4 presents the city-pair results. The dependent variable is impact factor-weighted papers. Panel A and B show the results of OLS and Zero-inflated binomial regression (ZINB) results, respectively. In column (1) for all city-pairs, for the ZINB specifications, HSR connection (with one year lag) has significantly positive effect on the coauthored papers between HSR-connected city pairs.

As discussed in Section 2, it is possible that when cities are connected by HSR, scholars will permanently move to other cities (we will explicitly test this in section 4.3). An important channel we want to highlight here is that the HSR effect also exists for those who do not choose to move (Agrawal, Galasso and Oettl, 2017). Our results are robust when focusing on a sub-sample of "narrowly-defined" non-movers, whose affiliations do not change during our study period (column (2)). This reinforces our view that bullet trains facilitate the circulation of ideas even in the absence of scholar migration.

*** Insert Table 4 about here ***

Secondary cities may benefit more when they are connected to a mega city featuring brilliant scholars and advanced labs at the leading universities. In columns (3) - (5) we explore the heterogeneous effect of HSR connection on three different types of city pairs: mega city - secondary city, secondary city with secondary city, mega city with mega city. The subgroup of mega and secondary city pairs receives the largest benefit from bullet train connection. The connection between two mega cities creates little effect, perhaps because they have already been well connected by other transportation networks like airline and highway and had intensive academic interactions.

In Table 5 we decompose the effect between pairs of scholars who collaborate both before and after HSR connection (the intensive margin pairs) versus pairs of scholars who collaborate after the connection (the extensive margin pairs). We find a stronger effect for intensive margin pairs (column (1) and (2)), consistent with the idea that even a few, additional face-to-face communications could have a tangible impact on the rate of academic collaboration. At the same time, the HSR connection also enables experimentation in the form of new collaborations of academics in different cites (column (3) and (4)), but this extensive margin is smaller than the intensive margin.

*** Insert Table 5 about here ***

HSR enables scholars to move faster across space. The HSR service has a comparative advantage over air travel for journeys of up to three hours or 750 km.[10] Here we test whether those cities located in this travel time "sweet spot" to mega cities benefit more from the intellectual spillovers. Those secondary cities become "closer" (measured in travel time) to the mega cities because bullet trains connect them. We construct a set of dummies to measure a city's travel time by train to the closest mega-city.[11] Table 6 shows that this publication premium due to the HSR connection exhibits a clear decay pattern as the secondary city is further away from the mega city. The premium is the largest for secondary cities within a 1.5-hour travel distance, followed by those in a 1.5-hour to 3-hour travel distance, and lose its significance for the cities outside the 3-hour radius.

*** Insert Table 6 here ***

4.3 Researcher Migration Patterns

We focus on a subsample of researchers who have migrated and track which cities they move to. This subsample has 58,460 movers and each of them face a choice set of 286 prefecture-level cities (all 287 prefecture-level cities minus the origination city), so the expanded sample size is huge. We randomly pick up a 1% subsample to run a conditional logit model. Let $s = 1,2, \dots, N$ denote moved researchers, $i = 1,2, \dots, I$denote researcher k's origination city, and $j = 1,2 \dots,287$denote the choice set of destination cities (including all 287 prefecture-level

---

[10]According to the World Bank Working Paper "High-Speed Rail: The Fast Track to Economic Development?" (No. 55856), a high-speed rail service has a time advantage over air travel for journeys of up to three hours or 750 km. For short journeys, up to 100 km, the private vehicle is the bullet train's main competitor.

[11]For each city, we identify its closest mega cities by its distance using ArcGIS. Here, *travel time=Min(travel time by ordinary train, travel time by bullet train)*

cities). We estimate the conditional logit model based on the utility function presented in equation (4):

$$U_{kij}^{*} = \beta_1 distance_{ij} + \beta_2 distance_{ij}^{2} + \beta_3 mega_j + \beta_4 HSR_{mega_j} + \beta_5 pop_j + \varepsilon_{kij} \quad (4)$$

In equation (4) U_{kij}^{*} is the latent variable of researcher k's utility from moving from city i to j. This researcher will compare her utility from each possible destination and select the utility-maximizing choice. $distance_{ij}$ measures the travel distance by ordinary train between city i and city j. $mega_j$ is a dummy indicator of whether city j is a mega city. HSR_mega_j is a dummy of whether city j has been connected to a mega city by direct bullet train in the year that this researcher makes the decision to move. We also control for city j's population size. We estimate this conditional logit model for the 1% sample, and then by two sub-groups – highly productive researchers, and others (impact factor weighted number of publications above or below the sample's median value).

Table 7 presents the estimation results. The likelihood of migrating decreases nonlinearly with respect to the travel distance between origination city and the destination city (column (1)). Researchers do favor mega cities. Controlling for travel distance and whether the destination is a mega city, they also have a strong preference of moving to the secondary cities that are directly connected to mega cities by HSR (the default group includes other cities). In column (2), we further control for city size, and the patterns are similar. Using the two coefficients of *HSR_mega* and log(*pop*), our calculation shows that the "attractiveness" a secondary city gains after being connected by high speed rail is equivalent to a population growth by 80%. We then divide this sample into highly productive researchers and other researchers (column (3) and (4)). Comparing the two marginal effects of *mega* and *HSR_mega*, we see that high-productively researchers have a slightly higher probability of moving to mega cities. This is suggestive evidence of researchers' sorting. HSR-connected secondary cities provide a new choice for the researchers who have chosen not to move to mega cities but still want to access the star researchers there.

*** Insert Table 7 here ***

5. Conclusion

In the classic monocentric model, increases in within-city travel speeds lower real estate prices at the center and cause the city to spread out further into the suburban fringe (Wheaton 1977). Such within city speed facilitates increased trade, matching and learning (Prud'homme and Lee 1999). Singapore's adoption of road pricing is an example of a policy that increases such speeds.

In the classic Roback (1982) system of cities model, migration costs are assumed to be zero but workers do not have the option of working in one city and living in another city. However, this might be a very attractive option to young people in cities featuring high productivity and high housing prices. Increases in travel speeds across cities increases the set of possible joint work and residential decisions.

In this paper we have studied a special segment of the labor market: academics in China. Such scholars are footloose with flexible work schedules that allow them to travel to work face to face with collaborators. Cross-city transportation time and financial costs limit the extent of these possibilities.

This paper argues that China's investment in High Speed Rail creates an integrated, regional system of cities close enough to travel by fast train but far enough to not be car friendly. We have studied the productivity impacts of cross-city transport improvements by focusing on publication and citation patterns of China's university researchers.

The empirical results in this paper show that once a city is connected into the HSR network, the researchers in that city will experience significant productivity increase in terms of quantity and quality of journal publications. We find that travel speed facilitates matching and idea flows between two HSR-connected cities. Larger productivity gains are observed for the secondary cities close enough to the mega cities to access them by HSR. We find larger productivity effects for social scientists and for the incumbent coauthors (the intensive margin). For the subsample of migrants, we find that they are more likely to choose those secondary cities that are directly connected with mega cities by HSR, compared to other secondary cities. These empirical findings bolster our claim that cross-city speed facilitates learning and matching across cities.

This finding has implications both for efficiency and for equity in the modern Chinese economy. In a human capital based economy, high speed rail induced reductions in transportation costs increase regional productivity by improving matching and lowering the cost of face-to-face interaction. At the same time, in China where the best universities and most productive people are concentrated in a handful of cities, there is the possibility of extreme spatial income inequality. China’s richest cities are towards the east. Cities such as Beijing and Shanghai are home to the nation’s most talented workers and most productive firms. These cities feature very high pay and real estate prices.[12]

[12]While there is not complete free mobility in China, the loosening of the domestic passport system varies by city type such that smaller cities are home to workers and firms with less human capital and lower overall productivity.

The Chinese Communist Party is deeply concerned about such inequality. The provision of transportation infrastructure is one of the major policies that China's central and local governments implemented to spur regional economic growth through its effect on productivity, employment and investment. We have argued that highly skilled researchers in secondary cities in China become more productive and their partnerships with star scientists in mega cities yield a greater quantity and quality of research as cross-city travel time declines . This flattening of the hierarchy of human capital distribution in Chinese cities helps to promote both economic growth and reduces within region human capital inequality. Therefore, our findings also have macro-implications for the long-run sustainability of China's economic growth.

Our main finding that faster cross-city commuting speeds enhance productivity extends 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. The boundary of a city's agglomeration area is endogenous and hinges on transportation speed. If new technologies such as high speed rail effectively make nearby cities "closer" to superstar cities (through moving at a faster speed), then agglomeration benefits spread out further across space.

References

Agrawal A, Galasso A, Oettl A. Roads and innovation. Review of Economics and Statistics, 2017, 99(3): 417-434.

Arzaghi M, Henderson J V. Networking off madison avenue. The Review of Economic Studies, 2008, 75(4): 1011-1038.

Azoulay P, Graff Zivin J S, Wang J. Superstar extinction. The Quarterly Journal of Economics, 2010, 125(2): 549-589.

Baum-Snow N, Brandt L, Henderson J V, et al. Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics, 2017, 99(3): 435-448.

Catalini C, Fons-Rosen C, Gaule P. Did cheaper flights change the direction of science?. 2016.

Claudel M, Massaro E, Santi P, et al. An exploration of collaborative scientific production at MIT through spatial organization and institutional affiliation. PloS one, 2017, 12(6): e0179334.

Duranton G, Turner M A. Urban growth and transportation. Review of Economic Studies, 2012, 79(4): 1407-1440.

Duranton G, Morrow P M, Turner M A. Roads and Trade: Evidence from the US. Review of Economic Studies, 2014, 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, 81(3): 1046-1070.

Freeman RB, Huang W. China's "Great Leap Forward" in Science and Engineering. InGlobal Mobility of Research Scientists 2015 (pp. 155-175).

Gaspar J, Glaeser E L. Information technology and the future of cities. Journal of urban economics, 1998, 43(1): 136-156.

Glaeser E L, Scheinkman J A, Shleifer A. Economic growth in a cross-section of cities. Journal of monetary economics, 1995, 36(1): 117-143.

Glaeser E L, Mare D C. Cities and skills. Journal of labor economics, 2001, 19(2): 316-342.

Hanson G H. Market potential, increasing returns and geographic concentration. Journal of international economics, 2005, 67(1): 1-24.

Harris C D. The, Market as a Factor in the Localization of Industry in the United States. Annals of the association of American geographers, 1954, 44(4): 315-348.

Helsley R W, Strange W C. Matching and agglomeration economies in a system of cities. Regional Science and urban economics, 1990, 20(2): 189-212.

Henderson R, Jaffe A B, Trajtenberg M. Universities as a source of commercial technology: a detailed analysis of university patenting, 1965–1988. Review of Economics and statistics, 1998, 80(1): 119-127.

Jaffe AB, Trajtenberg M, Henderson R. Geographic localization of knowledge spillovers as evidenced by patent citations. the Quarterly journal of Economics. 1993 Aug 1;108(3):577-98.

Lin Y. Travel costs and urban specialization patterns: Evidence from China’s high speed railway system. Journal of Urban Economics, 2017, 98: 98-123.

Moretti E. Human capital externalities in cities (Chapter 51). Handbook of Regional and Urban Economics. Elsevier B.V. 2004:2243-2291.

Mortensen D T, Pissarides C A. New developments in models of search in the labor market. Handbook of labor economics, 1999, 3: 2567-2627.

Prud'homme R, Lee C W. Size, sprawl, speed and the efficiency of cities. Urban Studies, 1999, 36(11): 1849-1858.

Quan W, Chen B, Shu F. Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). Aslib Journal of Information Management, 2017, 69(5): 486-502.

Rauch J E. Productivity gains from geographic concentration of human capital. 1991.

Roback J. Wages, rents, and the quality of life. Journal of political economy. 1982 Dec 1;90(6):1257-78.

Rosenthal S S, Strange W C. Geography, industrial organization, and agglomeration. review of Economics and Statistics, 2003, 85(2): 377-393.

Rosenthal S S, Strange W C. Evidence on the nature and sources of agglomeration economies. Handbook of regional and urban economics. Elsevier, 2004, 4: 2119-2171.

Rossi-Hansberg E, Sarte PD. Firm fragmentation and urban patterns. International Economic Review. 2009 Feb 1;50(1):143-86.

Wheaton W C. Income and urban residence: An analysis of consumer demand for location. The American Economic Review, 1977, 67(4): 620-631.

Zheng S, Kahn M E. China’s bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceedings of the National Academy of Sciences, 2013, 110(14): E1248-E1253.

Zheng S, Sun W. and Kahn M E. Investor Confidence as a Determinant of China’s Urban Housing Market Dynamics. Real Estate Economics, 2016, 44.4: 814-845.

Figure 1 The Geographic Distribution of China's Best Universities

Figure 1-A The Spatial Distribution of China's Top Universities ("211 Program" universities)

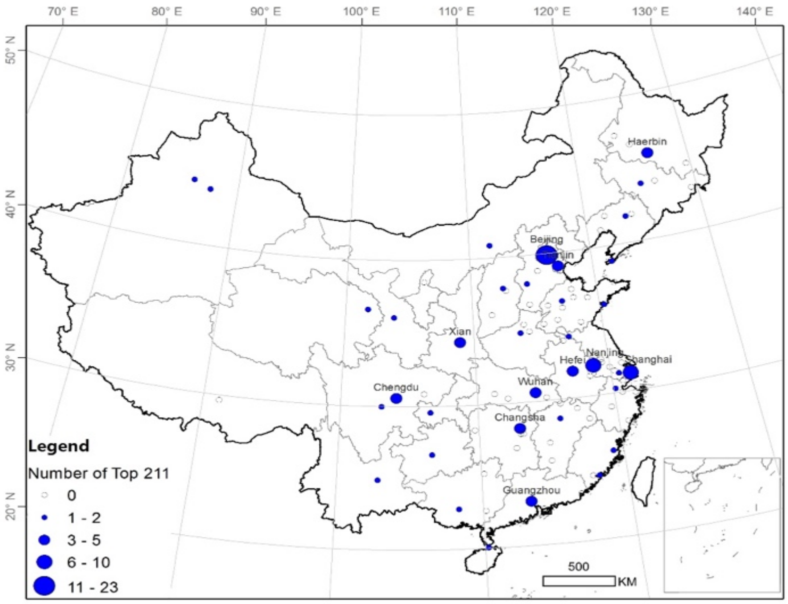
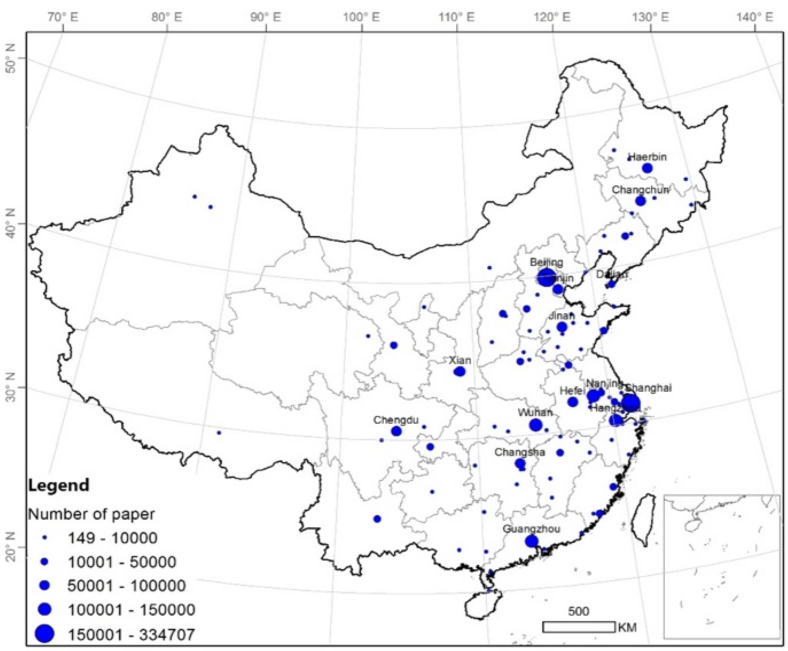


Figure 1-B The Count of International Journal Papers[13]



[13] Source: Science citation index & social science citation index (2001-2015)

Figure 2 The Geography of HSR and Co-Authored Papers

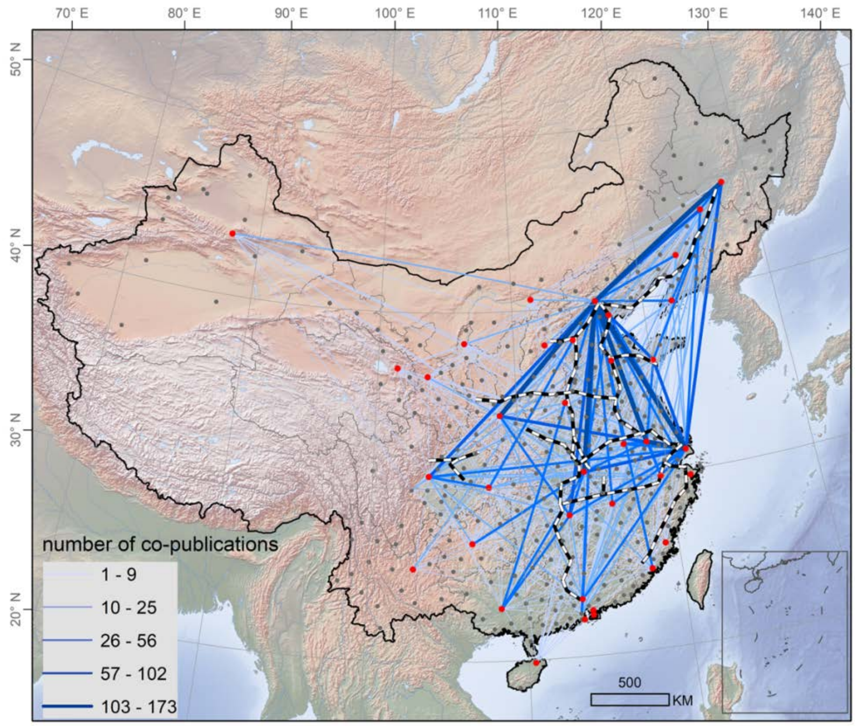


Table 1: Summary Statistics

Variable	Explanation	Obs.	Mean	Std. Dev.	Min	Max
Panel A: by city by year						
year	year	15,633	2011	3.163	2006	2015
papers	# of papers in SCI&SSCI journals	15,633	167.142	961.529	0	26279
weighted_papers	journal impact factor weighted papers	15,633	453.677	3033.306	0	99162.55
connect	Dummy, = 1 after the city is connected by HSR; = 0 otherwise	14,333	0.207	0.405	0	1
highway	log(highway ridership+1)	13,936	8.569	1.053	0	12.57
airport	log(air ridership +1)	12,409	5.088	2.213	0	10.25
fund	log(size of research fund allocated by government)	14,010	9.5	1.55	0.122	14.873
gdp_pc	log(GDP per capita +1)	13,736	10.234	0.797	7.922	13.108
population	log(population+1)	13,953	5.855	0.699	0.039	8.125
Panel B: by city, long-difference (2006 to 2015)						
connect_2006,2015	Whether the city is connected by HSR during 2006 to 2015	300	0.427	0.495	0	1
iv_hist	IV: = 1 if this city had a rail station in 1961; = 0 otherwise	242	0.364	0.482	0	1
iv_military	IV: = 1 if this city was a major military deployment place in 2005; = 0 otherwise	242	0.331	0.471	0	1
Panel C: City-pair by year						
year	year	85,990	2011	3.162	2006	2015

papers	# of papers in SCI&SSCI journals published by the coauthors in these two cities.	85,990	5.749	47.55	0	3248
weighted_papers	Journal impact factor weighted papers published by the coauthors in these two cities.	85,990	13.33	133.51	0	11583.57
connect	Dummy, = 1 after the two cities is connected by HSR; = 0 otherwise	17,468	0.302	0.459	0	1
gdp	log(the sum of the two cities' GDP per capita+1)	55,631	11.405	0.638	9.004	13.578
population	log(the sum of the two cities' population+1)	59,022	6.987	0.43	5.004	8.479

Notes: The data is collected from Web of Science, 12306 HSR website (and manual collecting of the news of China High Speed Railway program), the Chinese City Yearbook.

Table 2: HSR connection and Local Academic Productivity

	(1)	(2)	(3)	(4)	(5)	(6)
	All cities		Non-movers		Social science	Science & Technology
Dependent variables	*papers*	*weighted_papers*	*papers*	*weighted_papers*	*weighted_papers*	*weighted_papers*
Panel A: OLS						
connect_1	0.078*	0.063	0.125***	0.123***	0.103**	0.036
	(0.040)	(0.044)	(0.047)	(0.054)	(0.045)	(0.056)
# of obs.	10,525	10,525	10,525	10,525	4,210	6,315
R-squared	0.804	0.796	0.764	0.753	0.585	0.917
Controls	GDP per capita, population, air and highway ridership					
Fixed effects	City, field, year					
Panel B: ZINB						
connect_1	0.339***	0.308***	0.283***	0.319***	0.383***	0.357***
	(0.055)	(0.060)	(0.056)	(0.065)	(0.123)	(0.073)
# of obs.	10,525	10,525	10,525	10,525	4,210	6,315
Controls	GDP per capita, population, airport and highway ridership					
Fixed effects	Province, field, year					

Notes: All regressions include a constant. Standard errors are clustered at the province level. *,**,***: indicate statistical significance at the 10%,5%, 1% level. The dependent variable in Panel A is the logarithm of "the count of papers (or the count of impact-factor weighted papers) plus 1". The dependent variable in Panel B is the count of papers (or the count of impact-factor weighted papers). All regressions control for the city GDP per capita, population, airport and highway ridership, field FEs and year FEs. City fixed effects for Panel A and province fixed effects for Panel B are included. The dependent variable in Panel A is the logarithm of the count of papers (or the count of weighted papers) plus 1. "Non-movers" are defined as those first authors who publish at least two papers in our study period and do not change the city they work in.

Table 3: HSR Connection and the Long Difference Effect on Local Academic Productivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All cities		Non-movers		Social science		Science & Technology	
Dependent variable	*papers*	*weighted_ papers*	*papers*	*weighted_ papers*	*papers*	*weighted_ papers*	*papers*	*weighted_ papers*
Panel A:OLS								
connect_2006,2015	0.058	0.037	0.056	0.040	0.154**	0.176**	-0.030	-0.065
	(0.053)	(0.064)	(0.063)	(0.076)	(0.071)	(0.085)	(0.073)	(0.090)
# of obs.	1,090	1,090	1,090	1,090	436	436	654	654
R-squared	0.538	0.544	0.624	0.616	0.505	0.470	0.377	0.277
Panel B: IV								
connect_2006,2015	0.301**	0.309**	0.622***	0.688***	0.420***	0.486***	0.156	0.184
	(0.107)	(0.129)	(0.134)	(0.161)	(0.140)	(0.160)	(0.144)	(0.182)
# of obs.	1,090	1,090	1,090	1,090	436	436	654	654
R-squared	0.528	0.537	0.597	0.591	0.489	0.452	0.370	0.268
Controls	Initial value of the dependent variable, city GDP per capita growth, city population growth, city's latitude and longitude, province capital dummy, historical colonial status dummy							
Fixed effects	Province, field							

Notes: All regressions include a constant. *, **,***: indicate statistical significance at the 10%, 5%, 1% level. "*connect_0615"* indicates whether the city is connected by HSR during 2006 and 2015. The dependent variable is change in the logarithm of "the count of papers (or the count of impact-factor weighted papers) plus 1", from 2006 to 2015. All regressions control for the initial value of the dependent variable in 2006, the growth of city GDP per capita, the growth of city population, and a group of geographic and political characteristics (the city's latitude and longitude, province capital dummy, historical colonial status dummy). "Non-movers" are defined as those first authors who publish at least two papers in our study period and do not change the city they work in. "Social Science" includes the fields of Arts & Humanities and Social Science, and "Science and Technology" includes the fields of Life Sciences and Biomedicine, Physical Sciences, and Technology.

Table 4: Testing for Variation in the HSR Productivity Effect by City Type

Dependent variable	(1) All	(2) Non -movers	(3) Mega-secondary	(4) Secondary-secondary	(5) Mega-mega
Panel A: OLS, log (*weighted_papers* +1)					
connect_1	0.048	0.115**	0.134	-0.025	-0.041
	(0.063)	(0.058)	(0.084)	(0.090)	(0.108)
# of obs.	9,599	9,599	3,540	5,894	165
R-squared	0.851	0.820	0.874	0.745	0.970
Controls	Sum of two cities' GDP per capita, population, air and highway ridership, respectively.				
Fixed effects	City-pair, year				
Panel B: ZINB, *weighted papers*					
connect_1	0.481***	1.071***	0.531***	0.402***	0.482***
	(0.069)	(0.093)	(0.084)	(0.110)	(0.074)
# of obs.	9,599	9,599	3,540	5,894	165
Controls	Sum of two cities' GDP per capita, population, air and highway ridership, respectively.				
Fixed effects	Province-pair, year				

Notes: All regressions include a constant. *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Standard errors are clustered at the province-pair level. The "City-pair controls" include the sum of two cities' GDP per capita, city population, airport and highway ridership. "Non-movers" are defined as the author pairs that both coauthors do not move in our study period and co-publish at least two papers in two different years. "Mega cities" are defined as Beijing, Shanghai, Nanjing, Guangzhou, Wuhan, Tianjin, Chengdu, Changsha and Xi'an. Cities other than mega cities are defined as "secondary cities".

Table 5: Testing for HSR's Effect on the Intensive and Extensive Productivity Margins

	(1)	(2)	(3)	(4)
	Intensive margin		Extensive margin	
Dependent variable	*papers*	*weighted_papers*	*papers*	*weighted_papers*
Panel A: OLS				
connect_1	0.132***	0.107**	0.049	0.027
	(0.036)	(0.051)	(0.052)	(0.073)
# of observations	13,551	13,551	8,518	8,518
R-squared	0.907	0.858	0.857	0.820
Controls	Sum of two cities' GDP per capita, population, air and highway ridership, respectively.			
Fixed effects	City-pair, year			
Panel B: ZINB				
connect_1	0.721***	0.803***	0.575***	0.578***
	(0.028)	(0.029)	(0.024)	(0.025)
# of observations	13,551	13,551	8,518	8,518
Controls	Sum of two cities' GDP per capita, population, air and highway ridership, respectively.			
Fixed effects	Province-pair, year			

Notes: All regressions include a constant. *, **, ***: significant at 10%, 5%, 1%. Standard errors are clustered at province-pair level. The dependent variable in Panel A is the logarithm of the count of papers (or the count of weighted papers) plus 1. "Extensive margin" measures the HSR effect on those new coauthor pairs formed after the two cities are connected by HSR; "intensive margin" measures the HSR effect on those incumbent coauthor pairs before HSR connection.

Table 6: Testing for HSR Heterogeneous Treatment Effects as a Function of Cross-City Commute Times

	(1)	(2)		(3)	(4)
	City-pair			City-level	
Dependent variable	*papers*	*weighted_ papers*		*papers*	*weighted_ papers*
connect	0.010	0.074	*connect_mega*	-0.060	-0.074
	(0.044)	(0.057)		(0.045)	(0.056)
*connect *hour<=1.5*	0.130**	0.123	*connect_mega *hour<=1.5*	0.189**	0.181*
	(0.065)	(0.082)		(0.086)	(0.095)
*connect *1.5<hour<=3*	0.005	0.010	*connect_mega *1.5<hour<=3*	0.134**	0.157**
	(0.067)	(0.084)		(0.058)	(0.061)
connect 3<hour<=5*	-0.024	-0.065	*connect_mega *3<hour<=5*	0.121	0.149
	(0.059)	(0.074)		(0.134)	(0.113)
Controls	The sum of two cities GDP per capita, population, air and highway ridership, respectively			GDP per capita, population, air and highway ridership	
Fixed effects	City-pair, year			City, field, year	
Observations	10,955	10,955		6,246	6,246
R-squared	0.858	0.845		0.895	0.881

Notes: All regressions include a constant. *, **, *** indicate the statistical significance at 10%, 5%, 1% level. The dependent variable is the logarithm of the count of "papers (or the count of impact-factor weighted papers) plus 1". *"hour"* in column (1) and (2) measures the actual travel time by HSR between the city pair, and in column (3) and (4) it measures the actual travel time by HSR to the closest mega city. Column (1) and (2) show the estimation result for city-pair level analysis, and the default category is the city pairs with HSR travel time longer than 5 hours. Column (3) and (4) are for city-level analysis. We only include secondary cities, and the default category is the secondary cities with HSR travel time longer than 5 hours to its closest mega city. In column (3) and (4), "*connect_mega*" equals 1 after this city is connected to the closest mega city by HSR. The default category is the secondary cities with HSR travel time to the closest mega city longer than 5 hours. "*Mega cities*" are defined as Beijing, Shanghai, Nanjing, Guangzhou, Wuhan, Tianjin, Chengdu, Changsha, and Xi'an.

Table 7: Migration Discrete Choice Model Estimates

Dependent variable: *move*	(1) All movers	(2) All movers	(3) Highly productive researchers	(4) Other researchers
distance	-0.249***	-0.249***	-0.206	-0.273***
	(0.080)	(0.089)	(0.172)	(0.104)
distance^2	0.037***	0.033**	0.017	0.041**
	(0.013)	(0.016)	(0.031)	(0.018)
mega	3.803***	2.865***	2.520***	3.108***
	(0.113)	(0.121)	(0.180)	(0.164)
HSR_mega	0.949***	0.816***	0.684***	0.962***
	(0.136)	(0.137)	(0.245)	(0.174)
Log(pop)		1.343***	1.559***	1.230***
		(0.094)	(0.159)	(0.116)
Observations	90,014	90,014	29,580	60,434

Notes: All regressions are estimated using the "clogit" command in Stata. *, **, *** indicate the statistical significance at 10%, 5%, 1% level. Column (1) and (2) show the estimation results for all movers. Column (3) and (4) are the results for high productive and low productive researchers, respectively (in our study period, researchers who publish more papers than the median number of papers per scholar are defined as the highly productive researchers).

Appendix 1: The construction of the two instrumental variables for determining the HSR's placement

Following the transportation economics literature (Duranton and Turner 2012; Duranton el al.,2014), the first instrument variable we use is based on the nation's historical railway network. Baum-Snow et al. (2017) 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 to construct an IV to investigate HSR's effect on housing prices. They collect information based on historical Chinese railroad maps indicating whether the city had at least one station on the railway map (connected by the train network) in 1961 ($rail1961_i$). The second instrument variable we use is the spatial distribution of major military troop deployments in 2005($military_i$). China's central government built the network to ship non-war military troops and light equipment in case of emergency.

The validity of the instruments relies on the assumption that, conditional on controls, factors that do not directly affect city academic performance determine both the instruments distribution and current HSR network, and these factors are the only channel through which HSR connection affects city research productivity. However, the exclusion restriction could be violated if locations among the historical rail route in 1961 or the military deployment are correlated with economic city characteristics due to history or geography. We therefore estimate regressions including a set of additional controls that could be correlated with the instruments while also affecting the bullet train connection.

Appendix 2: Robustness Checks for the Lag Structure of the HSR connection

	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
	Current year					2 years lag			
	City level		City-pair			City level		City-pair	
Dependent variables	*papers*	*weighted_ papers*	*papers*	*weighted_ papers*		*papers*	*weighted_ papers*	*papers*	*weighted_ papers*
Panel A: OLS									
connect	0.071**	0.054	0.028	0.081	*connect_2*	0.092*	0.090*	0.058	0.071
	(0.034)	(0.038)	(0.039)	(0.049)		(0.045)	(0.047)	(0.059)	(0.069)
Fixed effects	City, field, year		City-pair, year			City, field, year		City-pair, year	
# of obs.	10,530	10,530	10,955	10,955		9,255	9,255	8,245	8,245
R-squared	0.804	0.796	0.858	0.845		0.810	0.803	0.867	0.857
Panel B: ZINB									
connect	0.360***	0.386***	0.495*	0.508***	*connect_2*	0.335*	0.341***	0.657*	0.610***
	(0.050)	(0.055)	(0.052)	(0.059)		(0.068)	(0.076)	(0.081)	(0.092)
Fixed effects	Province, year		Province-pair, year			Province, year		Province-pair, year	
# of obs.	10,530	10,530	10,955	10,955		9,255	9,255	8,245	8,245

Notes: All regressions include a constant. *,**,***: significant at 10%, 5%, 1%. The dependent variable in Panel A is the logarithm of the count of papers (or the count of weighted papers) plus 1. Standard errors are clustered at province level for city regression in column (1), (2), (5) and (6). Standard errors are clustered at province-pair level for city pair regression in column (3), (4), (7) and (8). The city level controls include city GDP per capita, population, airport and highway ridership for each city, and the city pair level controls include the sum of two cities GDP per capita, population, air and highway ridership, respectively.

Appendix 3: University Investment and City HSR Connection

Dependent variable	(1) *log(university fund)*		(2) *log(university fund)*
connect	0.053	*connect_1*	0.064
	(0.058)		(0.049)
# of observations	10,530		10,275
R-squared	0.941		0.951
Controls	GDP per capita, population, air and highway ridership		
Fixed effects	City, field, year		

Notes: All regressions include a constant. Standard errors are clustered at the provincial level. *, **, *** indicate statistical significance at 10%, 5%, 1% level. All regressions control for the city GDP per capita, population, air and highway ridership. The variable "*log(university fund)*" measures the amount of research fund (in logarithm) by city-year, which is collected from the Chinese education statistical year book.