High-Speed Railways and Collaborative Innovation

Douglas Hanley[*], Jiancheng Li[†], Mingqin Wu[‡]

July 2021

Abstract

This paper investigates the impact of infrastructure on innovation collaboration between enterprises in different locations using the introduction of high-speed railways in China. Applying an instrumental variable approach to control for the endogeneity of high-speed railways, and using a large enterprise census that includes both manufacturing and service sectors in China, we find that high-speed railways can improve innovation collaboration substantially at the city level. More importantly, we match city pairs based on high-speed railway routes and calculate the amount of time saved by a high-speed railway for each city pair. The empirical results suggest that the innovation collaboration also increases significantly at the city-pair level. Innovation quality, measured by patent citations between cities, also increases. Further evidence on spatial heterogeneity, industry heterogeneity and ownership heterogeneity suggests that the impact of high-speed railways is more significant for collaboration in less developed regions, in the service sector and in domestic enterprises.

Keywords: High-Speed Railway, Patent, Innovation Collaboration, China
JEL Classification: O31 R41 O18

[*]The University of Pittsburgh
[†]Sun Yat-Sen University
[‡]South China Normal University

1 Introduction

Technological and scientific progress propels substantial economic growth and accelerates the transformation of economies. Local economic outcomes increasingly depend on local idea and innovation generation (Davis and Dingel, 2019). Knowledge and technology are embedded in individual scientists, research institutes, private sector entities and government agencies (Acemoglu et al., 2016). New innovations and ideas are built based on past achievements and cooperation between the participants. Cooperation among different sources of knowledge can stimulate knowledge creation, diffusion and new innovations. Eliminating obstacles to mobility is key to technology improvement and innovation ability, since the mobility of individuals has been proven to be an important mechanism for knowledge diffusion (Rosenkopf and Almeida, 2003; Singh, 2005). Moreover, the movement of people fosters the movement of capital, which is one of the key factors for innovation (Campante and Yanangizawa-Drott, 2018). Physical proximity is closely related to increased communication and intellectual interaction (Audretsch and Feldman, 2004; Charlot and Duranton, 2004; Arzaghi and Henderson, 2008; Davis and Dingel, 2019). Reducing transportation costs can increase the circulation of people in a region and facilitate knowledge flows, diffusion and spillover, thus increasing the likelihood of innovation collaboration, which has become a widespread phenomenon, particularly in industries with rapid technological development (König et al., 2019). Despite the literature on studying reductions of transportation costs and trade cost, we lack studies that can explain the extent to which transportation infrastructure affects knowledge flows and innovation collaboration between enterprises. This paper examines whether and how high-speed transportation, i.e., high-speed railways, improves the innovation collaboration between enterprises located in different places.

Transportation costs affect the location, production and agglomeration of economic activities and knowledge creation. In this paper, we use high-speed railways in China to examine the roles and benefits of high-speed transportation in knowledge creation and collaboration between enterprises located in different cities. We first identify the impact of high-speed railway on collaborative patents between firms across China at the city level: the number of innovation collaborations increases significantly after the introduction of high-speed railways. Moreover, we further match city pairs connected by high-speed railways and calculate the specific amount of time saved by high-speed railways for each city pair, and we find that collaborative patents increase significantly at the city-pair level. Patent citations between cities increase substantially due to the reduction in transportation costs. These findings not only highlight the role of infrastructure in facilitating collaborations between enterprises and promoting knowledge exchanges but also touch the heart of innovation and growth in the literature (Aghion and Howitt, 1992; Acemoglu and Akcigit, 2012; Agrawal et al., 2017).

China provides an appropriate context for this study. First, China is on a rapid ascent in the world's science league (Hu, 2020). According to the 2019 World Intellectual Property Indicators, China has a leading position on multiple indicators of intellectual property rights. Second, China currently hosts the world's largest high-speed railway network. The first high-speed railway in Mainland China was completed in 2008. Until July 2020, the total length of high-speed railways had reached more than 36,000 kilometers[1]. China's high-speed rail system, which provides passengers fast, reliable and comfortable travel, has resulted in major changes in people's personal and working lives, promoted economic development along the lines, and moved China further toward being a "high-speed rail society"[2].

Using merged data from a large-scale enterprise census and collaborated information on patents registered with the State Intellectual Property Office (SIPO), which was replaced by China National Intellectual Property Administration (CNIPA) in August 2018[3], we construct a rich longitudinal database and examine whether and how enterprises in different cities collaborate on patents. Since high-speed railways reduce transportation costs, they make it more convenient for the employees of different enterprises to meet, discuss, exchange ideas, attend conferences and cooperate, especially on innovation that requires more than one type of expertise. From 2005 to 2015, if both cities are connected by high-speed railway, the annual growth rate of patent collaboration is 3.34% between cities. The patent collaboration increases significantly after the high-speed railway was put into use in 2008.[4]. A major identification challenge in estimating the effect of high-speed railway on innovation collaboration is the endogeneity issue affecting high-speed railways due to reverse causality or omitted variables. To address this problem, we follow a growing literature that focuses on the impact of the U.S. interstate road system, and exploits instrumental variable

[1] http://politics.people.com.cn/n1/2020/0810/c1001-31815882.html

[2] http://www.gov.cn/xinwen/2015-01/25/content_2809770.htm

[3] The website is: http://epub.cnipa.gov.cn/

[4] The average annual growth rate for the number of collaborated patents between cities is 0.92% if they are not connected to the high-speed railway line. If one city is connected to the high-speed railway and the other city is not, the annual growth rate for the number of collaborated patents is 1.17% between these two cities.

for high-speed railways (Baum-Snow, 2007; Michaels, 2008; Duranton & Turner, 2012; Duranton et al., 2014; Agrawal et al., 2017). We employ the Chinese railway network in 1934 as the instrument for the presence of high-speed railways. This historical instrumental variable is related with the transportation infrastructure at present, and it is uncorrelated with the residuals in the regression.

We examine the impact of high-speed railways on innovation collaboration not only at the city level but also by building the bilateral high-speed railway network connecting each city pair and calculating the time that is saved by comparing the travel time of high-speed and ordinary railway. The reduction in travel time gives enterprises more opportunities to exchange ideas and communicate with other enterprises in another city, thus increasing patent collaborations. A one percent increase in the time saved by high-speed railway translates into a 0.19% increase in the total number of patent collaborations, and a more than 0.95% increase in the patent citations between cities. Similar effects are found for extensive margin (i.e., the number of cooperated parties for the patents beween two cities) and intensive margins (i.e., the number of collaborated patents for each pair of cooperating party). Moreover, the impact of high-speed railways on innovation collaboration is nonlinear. We find that the most efficient distance between cities for high-speed railways to promote patent collaboration is within 200 kilometers. In addition, high-speed railways not only facilitate knowledge exchange within the manufacturing industry but also increase cooperation cross sectors, and the later effect is even larger. The impact of high-speed railways is more significant for the western than for the eastern and central regions in China. Finally, the effect of high-speed railways is more pronounced for domestic firms.

This paper is connected to the literature on the determinants of innovation and the impacts of transportation infrastructure on regional economic growth. We contribute to the literature in two ways. First, this paper contributes to the literature on knowledge creation, spillover and transmission by investigating the innovation collaboration between enterprises. Romer (1986), Krugman (1991a, b), and Grossman and Helpman (1991) focused on the roles of knowledge spillover across individuals and firms. Berliant and Fujita (2008) highlighted the history of meetings and how their content is important for knowledge creation, and they presented a micromodel of knowledge creation through interactions. Singh (2005) investigated how interpersonal networks determine knowledge diffusion patterns. Berliant et al. (2006) established a microfoundation to explain the patterns and implications of knowledge exchange. Davis and Dingel (2019) proposed that idea exchange outside the boundaries of enterprises can promote knowledge creation and spatial agglomeration since the gains from the exchange of ideas are greater in locations where conversation partners are more numerous. Faria et al. (2010) emphasized the importance of partners in innovation activities. We extend the literature by proposing one way to increase the frequency with which meetings and communications can feasibly occur, i.e., high-speed railway. Better and more efficient infrastructure can promote knowledge flows and creation through frequent meetings between enterprises and increase the likelihood of innovation collaboration.

Second, this paper contributes to the literature on estimating the economic impacts of transportation infrastructure projects. A number of studies have documented that knowledge spillovers are constrained by geography. Recent studies have examined the broad range of economic effects of transport infrastructure, such as urban employment (Duranton and Turner, 2012; Lin, 2017), long-term GDP growth (Banerjee et al., 2020), the presence of human capital (Glaeser and Saiz, 2004), urban structure (Baum-Snow et al., 2017), skill premia in the labor market (Michaels, 2008), gains from trade (Donaldson, 2018), and climate and other amenities (Rappaport, 2007). Agrawal et al.(2017) estimated the effect of interstate highways on regional innovation. Compared to the existing literature, this paper examines the effect of high-speed railways on knowledge creation and innovation collaboration, which has rarely been studied in the literature.

One paper that is related to this study is Dong et al. (2020), who examined the impacts of the transportation network on teamwork in research paper publication and citations. They found that reducing transportation costs can facilitate research cooperation between professors at top universities and researchers with complementary skills in another city. Our paper is also closely related with Wang and Cai (2020), who investigated the effect of transportation infrastructure on research collaborations cross cities in China. They find that research collaborations occur mainly between poorly innovative cities and highly innovative cities. Another paper that is related to our study is Gao and Zheng (2020), who provided direct evidence for the innovation hypothesis, i.e., transportation infrastructure can energize innovation by creating opportunities to accomplish old things well or do something new. They found that high-speed railways promote firm innovation and elaborate the underlying mechanisms using waves of innovation surveys on manufacturing firms in China.

There are several differences that distinguish our paper from the above literature. First, patents are usually for new innovations or improved equipments and methods, mainly related to technology applications. Patents are intangible firm property that can enable a firm to occupy a favorable position in market competition. By contrast, a research paper is an academic contribution that contains new scientific research results or innovative insights, mainly for reading, communicating or discussing at academic conferences or for publication in academic journals. Moreover, research papers are published to be cited and evaluated by peers, while patents depend on their practical implementation and transfer. Second, we calculate the saved time by comparing the timetables of high-speed railways and ordinary railways, which has rarely been used in the literature, to investigate the extent to which high-speed railways improve efficiency and knowledge diffusion. This provides a more accurate estimate for the effect of transportation infrastructure on innovation. Third, we investigate the impact of high-speed railways on the quality of innovation, measured by patent citations, which has seldomly been examined in the literature because of the insufficiency and unavailability of citation information for Chinese patent filings (Zhu et al., 2019). Fourth, we consider the intensive margin and extensive margin for innovation collaboration at both the city level and the city-pair level. This provides underlying mechanisms for knowledge diffusion and innovation collaboration. Finally, this paper is able to examine the heterogenous effect across space, industry, ownership structure and distance based on the rich information of the economic census and patent database.

The remainder of the paper is organized as follows. In section 2, we briefly review the background regarding patents and high-speed railways in China. In section 3, we present the empirical strategy and the dataset. In section 4, we report the empirical results, robustness checks, and heterogeneous tests. The final section concludes the paper.

2 Background

Patents. The Chinese patent system has some similarities and differences with the United States system. In the U.S., there are three different patent types: utility, design, and plant patents. China does not recognize plant patents. Instead, China has three patent types: invention, utility, and design patents. Invention patents in China are very similar to utility patents in the U.S. China has issued various laws and policies to promote innovations (Liu et al., 2011; Wang et al., 2017). According to Chinese Patent Law, for inventions completed by two or more units or individuals, unless agreed otherwise, the right to apply for a patent is joint. The number of collaborated patents takes 6.40% of all patents. Around 90.13% of collaborated patents are completed by two parties. Only a small proportion, i.e., 9.87% of the cooperated patents, are completed by three parties or more.

In Figure 1, we plot the collaborative patents on a map of China within each province in the upper panel A. The points in this figure are quite dense because they are based on firms' geographical locations. In panel B, we first pick up three representative cities in the eastern, central and western regions, i.e., Shanghai, Zhengzhou and Chongqing, then we plot the innovation collaborations for each city. Most of the collaborations within provinces are in more developed provinces in the coastal regions. Most of the collaborations across provinces are between inland provinces and coastal provinces. This further sheds light on our argument that high-speed railways reduce transportation costs and accelerate innovation collaborations between firms in different cities, enhancing knowledge creation and spillover.

[Insert Figure 1 here]

High-speed railway construction. China planned its high-speed railway construction in 2003. In 2004, the State Council issued the *Medium- and Long-Term Railway Development Plan*[5], and set the objective that by 2020, the national railway operating length would reach 100,000 kilometers; more than 12,000 kilometers of high-speed railway lines would be in operation, and the speed of these high-speed railway lines would be 200 kilometers per hour or above. This plan designed a national high-speed railway structure composed of "Four north-south vertical high-speed railway networks" and "Four east-west horizontal high-speed railway networks". This plan was revised in 2008, when the National Development and Reform Commission approved the *Medium- and Long-Term Railway Development Plan (revised in 2008)*[6]. It proposed that more than 16,000 kilometers of high-speed railway lines should be constructed by 2020. In July 2016, the National Development and Reform Commission, the Ministry of Communications and the China Railway Corporation jointly issued the *Medium- and Long-Term Railway Network Plan (2016-2030)*, which outlined the new era of "Eight north-south vertical high-speed railway networks" and "Eight east-west horizontal

[5] http://www.ndrc.gov.cn/fzgggz/fzgh/ghwb/gjjgh/200709/t20070913_709844.html
[6] http://www.ndrc.gov.cn/zcfb/zcfbqt/200906/t20090605_284525.html

high-speed railway networks", and proposed that more than 30,000 kilometers of high-speed railway lines would be constructed by 2020.

In 2014, China surpassed Japan, and became the country with the greatest high-speed railway transportation capacity. By the end of 2018, the national railway operating length had reached more than 131,000 kilometers, of which more than 29,000 kilometers were high-speed railways. Figure 2 describes the expansion of the high-speed railway network, defined as the lines traveling at 300 kilometers per hour or above from 2008 to 2015. The mileage of high-speed railway increases substantially each year.

[Insert Figure 2 here]

3 Empirical strategy

3.1 Dataset

Our analysis draws on six sources of data, which span from 2005 to 2015. The first main dataset used in this study comes from the second wave of the economic census of 2008, conducted by the National Bureau of Statistics of China. All legal entities in secondary and tertiary industries were surveyed in this census. The industries that are covered by the census include mining, manufacturing, and service sectors. The census data contain enterprises' complete basic information, such as address, industry, ownership, number of employees, output, assets, profit, etc. The advantage of using the census over the Annual Survey of Industrial Firms (ASIF) often used in the literature (Hsieh and Klenow, 2009; Brandt et al., 2017) is that it covers not only manufacturing firms but also firms in service sectors, that are seldomly examined in the literature. The number of original observations in the 2008 census is 4.95 million. We exclude financial firms, non-profit organizations and social groups in the final sample.

The second main dataset is the patent filing dataset, which includes the applicant name, address, abstract, industry, and content of each patent in each year since 1986. This information can be accessed at the China National Intellectual Property Administration website (http://epub.cnipa.gov.cn/)[7]. However, for the patent applicant, it only provides basic characteristics such as the name and address of the patent applicant. To obtain detailed information on patent applicants, we merge the

[7]SIPO has been changed to China National Intellectual Property Administration in 2018.

census data with the patent data using the full name of the firms (He et al., 2013; Holmes et al., 2015;) and validate our matching with firms' location information. To perform the match, we compute the Levenshtein distance (also known as the edit distance) between pairs of firm names appearing in the patent data and census data. If this distance is small enough, we deem the pair to be valid, meaning they are thought to refer to the same firm. Unfortunately, the total number of possible pairs is extremely large. There are approximately 3 million distinct firms in the census data and approximately 0.8 million distinct names in the patent data, meaning that there are almost 2 trillion possible name pairs. Even at 1 million comparisons per second, this would take on the order of one month. To get around this issue, we employed a technique known as locality sensitive hashing (Charikar, 2002; Manku et al., 2007) to substantially narrow down the set of possible match pairs, thus making a solution to the problem feasible. The details of the implementation are discussed in Appendix B.

After merging the patent data with the census data, we can observe the patent filings of each firm since 1986, together with the location, industry and other characteristics of each firm. Note that firms in the sample are constrained within the firms that appear in the 2008 economic census[8]. Since we have the geographical location of each firm, we can summarize the number of patents at the city level and city-pair level. If there is more than one applicant for a patent, we define this patent as a collaborative patent. In Appendix Table F1, we show the total number of collaborative patents across industries from 2008 to 2015. We can see that the number of collaborative patents is the largest in Software (industry code: 62), followed by Education (industry code:84) and Research and Experimental Development (industry code:75). For manufacturing industries, Raw Chemical Materials and Chemical Products (industry code: 26) ranks first in collaborative patent applications.

Moreover, we try to get the citation for each patent. Due to the large number of patents, it is hard to calculate the number of citations by hand. The citations for each patent can be accessed from the website *https://patents.google.com/*. We search and calculate the number of citations using a Python program to crawl the data in this website. Since we know the location of each firm, we can obtain the citations in city i from city j in each year.

[8]We admit that we cannot observe the entry and exit of firms using 2008 economic census. If more firms are involved with innovation collaboration across cities or more innovative firms enter the city, our estimate would be the lower bound for the true effect.

The third main dataset covers the high-speed railway network, where high-speed trains are defined as those operating at an average speed of 300 kilometers/h or more. We mainly rely on the China Railway Yearbook from 2006 to 2016 to obtain the names and locations of high-speed railway stations, construction starting dates, opening dates, speed and total length. Then, we know whether city i has a high-speed railway station in year t and whether city i and city j are connected by a high-speed railway line in year t.

The timetable for railway transportation comes from the official website *https://www.12306.cn/index/* and the annual "*National Railway Timetable*" published by the China Railway Publishing House. We can check all the railway lines across China, and identify whether and when a city is connected to high-speed railways in each year. In general, the first letter of railway lines includes "G", "C", "D", "Z", "T", "K", or numbers from 1001 to 7598. We regard railway lines with speeds of 300 kilometers/h or more as high-speed railways, which include the first letter "G" and some railway lines with the first letter "C". We compile the city pairs (i, j) based on the high-speed railway networks. In total, we have 81702 city pairs. By comparing the time difference between high-speed railways and ordinary railways, we calculate the amount of time that is saved by taking high-speed railway.

The fourth dataset describes China's railway lines in 1934, which was obtained from the book "*The latest Chinese situation at a glance*". In November 1928, the Ministry of Railways of the Nanjing National Government of China was formally established and began to systematically plan the national railway network. In January 1929, the Ministry of Railways proposed funding sources for railway construction. In 1931, the Nanjing National Government drafted a new "Five-year Plan for Railway Construction" and expanded the national railway plan of 1929. It planned to build four major railway networks throughout the country within five years, with a total length of 8,000 kilometers. However, at the outbreak of the Anti-Japanese War in 1937, only nine railways had been built nationwide, with a total length of 3,795 kilometers.

In the book, we can see various types of railways that differ in their funding sources in the 1920s and 1930s. In general, there are five types of railways: (1) railways funded by private enterprise, (2) railways funded by foreign firms, (3) railways funded by the Department of Transportation of China, (4) railways funded by joint ventures of foreign capital and local governments, and (5) railways funded by both the Department of Transportation and joint ventures. Some city names differ between 1934 and the present, and we transfer the city name in 1934 into the standard nomenclature of 2008. The distribution of railways in 1934 is shown in Appendix Figure A1.

The fifth dataset comes from City Statistical Yearbooks and Regional Economic Statistical Yearbooks issued by the National Bureau of Statistics of China from 2006 to 2016. These yearbooks provide basic statistics on cities, such as GDP per capita, population, FDI, fixed investment, education expenditures, number of passengers by highway, and the number of industrial firms, etc. In addition, we checked whether each city has an airport by hand. The marketization index is at the province level, and it comes from the annual report "*Marketization Index of China's Provinces*" (Wang et al., 2006-2016).

The final dataset is the China Stock Market & Accounting Research (CSMAR) Database, which provides detailed data on China stock markets and the financial statements of China's listed companies. The name of the head company and subsidiary company can be seen in the dataset, thus we can find the patents that are developed through collaboration by the head company and its subsidiaries. In the end, we obtain 47,975 patents which are collaborated patents between head and subsidiary companies. The details for the data construction are discussed in the Section 4.4.

The sample period in this paper is 2005 to 2015 and the summary statistics for the main variables are described in Table 1. In panel A, we report the summary statistics for the variables at the city level. The average number of collaborative patents in each city is 51.50, the average number of collaborative partnerships in a city is 10.96, and the average number of collaborative patents for each partnership is 2.31. As the key independent variable at the city level, we use two measures of high-speed railway. The first is, *HSR*, a dummy indicating whether city i is connected to high-speed railway network, and the second is, *HSRDegree*, the number of cities that are connected to city i by high-speed railways, which is 1.71. If we only consider the sample of cities that are connected to the high-speed railway network, the average number of cities is connected to city i is 12. The summary statistics for the control variables at the city level are reported in Appendix Table C1.

In Panel B of Table 1, we report the summary statistics at the city-pair level. The average number of collaborative patents for each city pair is 0.19. The average number of collaborative partnerships for each city pair is 0.04, and the average number of collaborative patents for each partnership is 0.06. The mean value for citations between cities is 7.224. Moreover, the average time that is saved by a high-speed railway compared to a normal railway is 1.40 minutes in the sample period. This number is small because from 2005 to 2008, it takes a value of 0 if there were no high-speed railways between cities. If we only compare the cities that are connected by high-speed railways, the average time that is saved is 248 minutes. The summary statistics for the control variables at city-pair level are reported in Appendix Table C2. Finally, we also compare the variables between cities that are connected by high-speed railways and cities that are not connected at the city level and city-pair level in Table C3 and the upper panel of Table C4. In general, cities that are connected by high-speed railways have more patent collaborations and better economic performance. If we compare the differences between cities that were connected by railways in 1934 and cities that were not in the lower panel of Table C4, we find that there are almost no differences in innovation collaboration.

[Insert Table 1 here]

3.2 Empirical specification

We construct the empirical specification separately at the city level and city-pair level in this section.

City level specification. At the city level, we examine the effect of high-speed railways on the patent collaboration for each city c. The baseline econometric specification is a difference-in-difference specification:

$$Y_{ct} = \alpha_1 HSR_{ct} + \alpha_2 X_{ct-1} + \lambda_t + \pi_c + e_{ct} \tag{1}$$

We have three measures for the dependent variable, Y_{ct}. The first is "$Total$", the summarized number of collaborative patents for firm in city c with firms in other cities in year t. The second measure for Y_{ct} is "$Extensive$", the summarized number of partnerships in city c, which can be considered as the extensive margin. In addition, the third measure is "$Intensive$", the number of collaborations for each pair of collaborating parties in city c. This can be considered as the intensive margin. We take the logarithmic forms for the dependent variables.

The main explanatory variable is HSR_{ct}, whether city c was first connected to the high-speed railway network in year t, and HSR_{ct} is equal to 1 if city c has a high-speed railway station in year t and afterwards, and it is 0 before year t. We expect α_1 to be positive, since being connected to high-speed railway lines can increase the opportunities for knowledge flow and transmission, thus increasing the possibility of collaboration and the actual number of patent collaborations. X_{ct-1} refers to the control variables at city level. We control for city-level economic variables, such as GDP per capita, population, FDI, fixed investment, government education expenditures, the number of passengers by highway, whether the city has an airport, number of noncollaborative patents, and the number of industrial firms. All the control variables are lagged to avoid the simultaneity bias. The year fixed effect λ_t controls for all yearly factors that are common to cities, such as macro-level shocks. The city fixed effect π_c captures all time-invariant differences across cities.

To address the potential heteroskedasticity and serial correlation, we cluster the standard errors at city level, following the suggestion by Bertrand et al. (2004). We also replace λ_t with *region-by-year* fixed effects, λ_{rt}, to address any region-specific yearly shocks to local economic conditions. *Region* refers to the eight main regions in Mainland China: Northeast, Northern Coastal, Eastern Coastal, Southern Coastal, Southwest, Northwest, the central regions of the Yangtze River and the central regions of Yellow River. The empirical results are still consistent, and the empirical results are presented in the appendix in section G.

An alternative way to measure high-speed railways is the number of cities connected by high-speed railway lines to city i in year t, $HSRDegree_{it} = \sum_{j=1}^{N} HSR_{ij,t}$, and it takes value 0 if there is no railway connected to city c in year t. The specification is changed to,

$$Y_{ct} = \beta_1 HSRDegree_{ct} + \beta_2 X_{ct-1} + \lambda_t + \pi_c + \varepsilon_{ct} \tag{2}$$

Similar to equation (1), we expect β_1 to be positive. Connecting to more cities through high-speed railway lines can increase the opportunities for innovation cooperation, thus increasing the number of patent collaborations with other cities. To address potential heteroskedasticity and serial correlation, we cluster the standard errors at the city level. Similar to the above specification, we also replace λ_t with region-by-year fixed effects, λ_{ct}, to address any region-specific yearly shocks to local economic conditions. The empirical results using this alternative fixed effects are presented in the appendix in section G, and the findings are still consistent.

City-pair level specification. We examine the impact of high-speed railway on patent collaboration at the city-pair level, i.e., the collaboration between city i and city j. The first dependent variable, $Y_{ij,t}$, is "*Total*", the summarized number of collaborative patents for firms in city pair

(i, j) in year t. The second dependent variable is "*Extensive*", the number of collaborating parties in city pair (i, j), which can be considered the extensive margin. The third dependent variable is "*Intensive*", the number of collaborations for each pair of collaborating parties in city pair (i, j), which can be considered the intensive margin. In addition, we also examine the patent citations between firms in city i and firms in city j, and this is the fourth dependent variable, which measures the quality of innovations.

$HSRtime_{ij,t}$ is the time difference between ordinary and high-speed railways. We expect the impact of $HSRtime_{ij,t}$ on patent collaboration to be positive, since being connected directly by a high-speed railway reduces travelling time between cities. $HSRtime_{ij,t}$ is a continuous variable, which can take the value of 0 or any positive number. With the help of this variable, we can consider not only the impact of the dummy variable, i.e., whether two cities are connected by a high-speed railway, but also compare the impacts of high-speed railway on innovation collaborations even if city pairs are all connected by high-speed railway. Thus, we have the difference-in-difference specification (3):

$$Y_{ij,t} = \gamma_1 HSRtime_{ij,t} + \gamma_2 X_{ij,t-1} + \lambda_t + \pi_{ij} + \epsilon_{ij,t} \tag{3}$$

$X_{ij,t-1}$ are control variables for city pair $(i,\ j)$, which include GDP per capita, population, government education expenditure, average fixed assets, FDI, the number of passengers by highway, the presence of an airport, the number of industrial firms, the number of noncollaborative patents, and the marketization index. These control variables are calculated at city-pair level using the format "$\log(X_i + X_j)$", while the number of passengers by highway and whether the city has an airport are controlled at the city level (Dong et al., 2020). The last control variable, the marketization index, is controlled at province-pair level by calculating the absolute value of the difference between two provinces in year t if city i and city j are not in the same province. All the control variables are lagged to avoid the simultaneity bias.

The year fixed effect λ_t and city pair fixed effect π_{ij} are also included. To address the potential heteroskedasticity and serial correlation, we cluster the standard errors at the city-pair level. We also replace λ_t with region pair-by-year fixed effects, $region\ pair_{m,n} \times Year$, where $region\ pair$ is the bilateral pair between any two regions $(m,\ n)$, to address any region-specific yearly shocks to local economic conditions. The empirical results are still consistent, and the empirical results are presented in the appendix in section G.

The high-speed railway routes are determined by the central government, since most railway routes usually travel across provinces. Usually, the high-speed railway plan is proposed by National Development and Reform Commission, Ministry of Transport and China Railway Company and approved by the State Council. The funding for high-speed railway mainly comes from the budget of the central government through the lending of state-owned banks and financial institutions, and the remaining capital is from the bond issued by the Ministry of Railway. A small proportion would come from local government mainly through compensation for land use (Lin et al., 2020). Thus, for local enterprises, the construction of high-speed railways is exogenous.

However, it is possible that one city may contribute to the high-speed railway construction to be connected to the high-speed railway network, thus increasing its human capital accumulation, physical capital accumulation, and knowledge communications in order to obtain faster economic growth. This reverse causality may cause the estimates for α_1, β_1 and γ_1 to be biased. From Figure 2, we see that the distribution of high-speed railways is not random across cities. It is denser in coastal regions than inland regions, similar as the argument in Dong et al. (2020). According to *Medium- and Long-Term Railway Development Plan* issued by State Council in 2004, the placement of China's high-speed railway is centrally managed, planned and financed by the government, taking economic development, population and resource distribution, national security, environmental concerns and social stability into considerations (Lin et al., 2020). Thus, it is less likely for the central government to build high-speed railway randomly.

To address this concern, we estimate an instrumental variable regression. A valid instrument should be a good predictor of high-speed railway and be orthogonal to the structural equation error term. Following Duranton and Turner (2012), Duranton et al. (2014), Agrawal et al. (2017), Möller and Zierer (2018), and Wang and Cai (2020), we use the historical railway lines in 1934 as the instrument for the high-speed railway lines in our sample period. The railroads in 1934 is an appropriate instrument rests on the length of time since these railroads were built, and the fundamental changes in the nature of the economy in the intervening years (Duranton and Turner, 2012;). First, the railways used in the 1930s were mainly built by central government, domestic private enterprises and western countries. Western countries provided much of the financing and had substantial influence over the placement of railways, and their objectives were to promote western economic and military interests in China (Banerjee et al., 2020). The main objective of state-owned railways is to transport agricultural goods, mineral goods and manufacturing goods. Second, during the Anti-Japan war in World War II, some of the railways were destroyed, for instance, railways of ChaoZhou-ShanTou and XinHui-TaiShan were destroyed in 1938, and railway of ZhuZhou-LanChuan was torn down in 1937. Part of railways, such as ZhangZhou-GuangZhou, was torn down due to the construction of highway between ZhangZhou and ChongYu. Third, People's Republic of China, a completely new government, was created in 1949 after World War II. The institutional arrangement, economic systems, city functions in 1930s were completely different with the cases at present. As the argument of Wang and Cai (2020), who use the railway route in 1961 as instrumental variable for high-speed railway, on the one hand, historical railway routes are related to current railway lines; on the other hand, it is difficult for historical railway routes to influence the current research collaboration and innovation through channels other than the high-speed railway connection. Thus, the historical railways in 1934 are unlikely to be correlated with a firm's innovation activity in the 2000s through other channels, but only impact the innovation behaviors through the construction of high-speed railway.

In particular, the instrument for HSR_c is whether city c had a railway station in 1934, $Railway_{c,1934}$. It takes value 1 if the city had a railway station in 1934 and 0 otherwise. The instrument for $HSRDegree_c$ is $RailwayDegree_{c,1934}$, which is the number of cities connected to city c in 1934, and it takes value 0 if there is no railway connected to city c. The instrument for $HSRtime_{ij}$ is $Railway_{ij,1934}$, which is equal to 1 if there is a railway between city i and city j in 1934, 0 otherwise.

As the instrument is time-invariant, it enters the panel instrumental variable estimation interacted with time trend, ϕ_t, so that the instrument can be time-varying. Hence, the first stage of the panel instrumental variable estimation at the city level is as follows,

$$HSR_{ct} = \theta_1 Railway_{c,1934} \times \phi_t + \theta_2 X_{ct-1} + \lambda_t + \pi_c + e_{ct} \tag{4}$$

$$HSRDegree_{ct} = \delta_1 RailwayDegree_{c,1934} \times \phi_t + \delta_2 X_{ct-1} + \lambda_t + \pi_c + \varepsilon_{ct} \tag{5}$$

The first stage of the panel instrumental variable estimation at the city-pair level is

$$HSRtime_{ij,t} = \eta_1 Railway_{ij,1934} \times \phi_t + \eta_2 X_{ij,t-1} + \lambda_t + \pi_{ij} + \epsilon_{ij,t} \quad (6)$$

Another possible identification challenge is the timing of city i being connected to a high-speed railway. The opening time of a high-speed railway is determined by construction progress, which depends on engineering difficulties (Lin, 2017). Usually, high-speed railway construction will be slower if the line is longer, has more stations and has higher bridge and tunnel ratios. These factors are exogenous to firms' innovation collaboration. Thus the opening time t is not endogenous in our specification.

4 Empirical results

4.1 The innovation collaboration at the city level

In this section, we apply the specification in equation (1) to examine the impact of high-speed railways on innovation collaboration at the city level. In Column∞ (1) of Table 2, we present the OLS estimates for total number of collaboration of city c in the upper panel, and all control variables and fixed effects are included in the regression. The coefficient of $HSRstation$, which refers to the key variable HSR_{ct} in equation (1) is 0.21, which is positive and statistically significant. In the middle panel, the dependent variable, *Extensive*, is the extensive margin, i.e., the number of firms in city c that have cooperated with firms in other cities. The coefficient of $HSRstation$ is smaller than that for total effect, but still positive and significant, which suggests that the presence of high-speed railways encourages innovation collaboration. In the lower panel, the dependent variable is the intensive margin, i.e., the number of patent collaborations for each partnership in city c, and the effect of high-speed railway on intensive margin is not significant.

Since the dependent variable is log of the (1+ collaborated patents). The total number of collaborated patents for cities after high-speed railway connection increases by 22.94%, and extensive margin, i.e., the number of partnerships at city level, increases by 16.66%, based on the OLS estimation[9].

[Insert Table 2 here]

[9] We calculate the implied growth using exp(coefficient)-1 for all the regression coefficients which are larger than 0.1. Thus, total number of patent increases by (exp(0.2065)-1)=22.94%, and extensive margin, i.e., the number of partnerships at city level, increases by (exp(0.1541)-1)=16.66%.

The empirical results in Column (1) of Table 2 indicate that high-speed railways can promote innovation collaboration. As we noted in the last section, the construction of high-speed railways can be endogenous and thus bias the estimates. To solve this problem, we conduct instrumental variable regression in Column (2). The instrument for $HSRstation$ is whether city c had a railway station in 1934 multiplied by the time trend. The dependent variables are the same as Column (1). In the upper panel of Column (2), we can see that the presence of high-speed railways substantially increases the total number of collaborations of city c. The results for the first stage is also reported in Table 2. The instrument variable is significant and positively correlated with the endogenous variable. In Column (2), instrumental variable regression suggests that high-speed railway can significantly increase the total number of patent collaborations in city c to a larger extent. The extensive margin is also significantly increasing with the construction of high-speed railways. However, the result is still not significant for the intensive margin when we use instrumental variable regressions.

Comparing OLS and IV estimates, we can see that the estimated effect of high-speed railway on innovation collaboration increases when we correct for the endogeneity issue. This means that high-speed railway assigned to city at random is associated with more innovation collaboration than a high-speed railway by the prevailing process (Duranton and Turner, 2012). One possible reason is that the existence of omitted variables, such as the price of construction materials and labor cost, could be negatively correlated with the high-speed railway construction. Moreover, developing cities whose innovation collaboration is less active can intentionally chose to improve their transportation infrastructure (Gao and Zheng, 2020). The issues of omitted variables and reverse causality would lead to a downward bias in OLS estimate. Finally, it may be that high-speed railway construction serves as a substitute for social assistance, and that high-speed railways are built where land and labor are cheap rather than in the places where they are in short supply (Duranton and Turner, 2012), thus the estimates from instrumental variable regression are much larger than OLS estimates (Dong et al, 2020; Duranton and Turner, 2012; Agrawal et al., 2017).

Specifically, the coefficients in Column (2) of Table 2 imply that the introduction of an HSR connection increases the number of collaborated patents by 4.03 times, compared to the control cities without high-speed railway or cities that are connected to high-speed railway network at a later stage[10]. In the second panel of Table 2, we can see that the extensive margin increases by 3.02 times. Given that the sample mean for total number of collaborated patents at city level is 33.81 in the pretreatment period and 10.33 for extensive margin, the increase associated with being connected to high-speed railway corresponds to an increase in the number of innovation collaboration of 136 patents per city-year in the posttreatment period, and an increase of 31.20 partnerships for extensive margin.

Moreover, we estimate specification (2) to examine the degree of high-speed railway connections on innovation collaboration. Theoretically, if one city is connected with more cities through high-speed railway lines, firms in this city would have more opportunities to communicate with other firms, thus have a higher probability of obtaining more collaborative patents. The OLS empirical results are shown in Column (1) of Table 3. When one additional city is connected with city c by high-speed railway, the total number of collaborative patents would increase, and similar findings can be observed for the extensive margin. The results for the intensive margin are not significant.

[Insert Table 3 here]

As we explained above, the degree of high-speed railway connections can be endogenous due to reverse causality or omitted variables; thus, we use the degree of a city in the railway network in 1934 as the instrument for the degree of high-speed railway connections at present. The instrumental variable regression results are reported in Column (2) of Table 3. Compared with the results in Column (1), the instrumental results using specification (2) reveal that the magnitude for the effect of high-speed railway degree is significant and larger. We find that if one more high-speed railway route is connected at city level, the total number of patent collaboration increases by 6.71%, and it is 5.86% for extensive margin based on IV estimates in Column (2). The results are not significant for the intensive margin.

The intensive margin at city level is not significant in Table 2 or Table 3; in addition, the magnitude of high-speed railway is smaller for intensive margin than total effect and extensive margin at city level. Intensive margin at city level measures the average number of collaborated patents of each cooperated partnership. Once a city is connected to the high-speed railway network, there are more opportunities for collaboration for this city, thus the total number of innovation collabo-

[10] According to Column (2) of Table 2, after HSR is connected, total effect at city level increases by (exp(1.6153)-1)=4.03, and extensive margin increases by (exp(1.3920)-1)=3.02.

ration can increase. However, the increase in the number of potential collaborated partners, does not necessarily mean that the number of collaborated patents would increase for each partnership. For instance, firms may concentrate on several important or key partners rather separating the innovation resources to all possible partners even after high-speed railway is connected. We can only infer the impact of high-speed railway on intensive margin at city-pair level, which will be presented in next section.

This magnitude at city level based on IV estimate in Column (2) of Table 2 is much larger than the estimates at city level in the literature (Wang and Cai, 2020; Gao and Zheng, 2020), due to the significant increases in patent applications in this period. In Dong et al. (2020), they find that once a city is connected by high-speed railway, the academic production of academic papers increases by 30.1% in quantity. Our estimates at city level suggest that HSR connection increases the number of collaborated patents by 4.03 times. When we consider the impact of high-speed railway degree on innovation collaboration in Column (2) of Table 3, we find that there is 6.71% increase in total patent collaboration if one more high-speed railway route is added at city level.

4.2 Innovation collaboration at the city-pair level

In this section, we examine innovation collaboration at the city-pair level, i.e., the impact of high-speed railways on patent collaboration between city i and city j. We analyze the impact of high-speed railway on innovation collaboration at both city level and city-pair level. The main reason is that they reflect different contents and directions of innovation collaboration. In particular, city-pair results tell us the specific cooperating parties, the direction of knowledge flows, plus we can know the impact of saved time generated by high-speed railway only at city-pair level. For results at city level, we can see the total impacts of high-speed railway on innovation collaboration for each city, together with the time trend for total innovation collaboration.

We estimate equation (3), and the empirical results are reported in Table 4. In Columns (1), we show the OLS estimates after controlling for control variables, city pair fixed effect and year fixed effect. The coefficients for the key explaining variable, $HSRtime_{ij,t}$, which is represented by $\Delta TravelTime$ between city i and j in Table 4, are statistically significant and positive for total effect in the upper panel. In the middle panel, we examine the effect of saved time on extensive margin, i.e., the number of collaborative partnership between city i and city j, and the its effect is much smaller than total effect. In the lower panel, we test the effect of saved time on intensive margin, which is the number of patents for each collaborative partnership. The time saved by high-speed railways can also significantly increase the number of collaborative patents for each pair of collaborating parties.

[Insert Table 4 here]

As we noted in the last section, the construction a high-speed railway between city i and j can be endogenous, thus bias the OLS estimates. To solve this problem, we estimate an instrumental variable regression in Column (2) of Table 4. The instrument for $HSRtime_{ij,t}$ is the time trend multiplied by $Railway_{ij,1934}$, which is equal to 1 if there is a railway between city i and city j in 1934, 0 otherwise. The result for the first stage is also presented in Table 4, which implies that the instrument variable is closely related with the endogenous variable.

The coefficients in Column (2) of Table 4 are much larger than the coefficients in Column (1) because the endogeneity biased the estimates downward as we discussed in last section. When we include control variables, year fixed effects and city pair fixed effects in the instrumental variable regression, the coefficient of $HSRtime_{ij,t}$, which is the logarithm of saved time, is 0.187 for the total number of collaborations of each city pair, 0.136 for the measure of the extensive margin, and 0.097 for the measure of the intensive margin. Since high-speed railway increases travel speed and saves travel time, one percentage increases in saved time between city i and j increases the total collaborated patents by 0.19% for the city pair (i and j), and 0.14% for extensive margin, i.e., the number of partnerships at city-pair level, and 0.10% for intensive margin, i.e., average collaborated patents of each partnership.

Using the dummy variable whether the city i and j are connected by high-speed railway, literatures have examined the impact of high-speed railway on research collaboration. For instance, Wang and Cai (2020) find that high-speed railway connection can increase collaborated patents at city-pair by 2% to 12%. Gao and Zheng (2020) show that high-speed railway connection can increase innovation of any type by approximately 15% after controlling for the endogeneity of high-speed railway. Our estimate is a little smaller, because we consider the impact of saved time between city i and j which is more accurate and specific, but rarely used in the literature.

From Table 4, we can see that high-speed railways can significantly facilitate patent collaborations at city-pair level, especially when we correct for the endogenous issue regarding the construction of high-speed railways. The impacts of high-speed railway are reflected on both extensive margin and intensive margin.

4.3 Citations at city-pair level

In this section, we examine how high-speed railways affect the number of citations across cities. Citations have been widely used in the literature to measure knowledge flows and the quality of patents (Jaffe et al., 1993; Jaffe and Trajtenberg, 2002; Singh, 2005). After two cities i and j have been connected, it is more likely for firms in city i to cite firms in j or vice versa because convenient transportation promotes knowledge spillover and provides more opportunities for communication and cooperation. The dependent variable is the logarithm of the number of citations of city i from city j in Table 5. To build the citation measure, we first calculate the citations for each firm f in city i and then summarize the total number of citations of city i from city j. In Column (1), we present the OLS estimation result. The saved travel time due to high-speed railway significantly increases patent citations between cities. To control for the endogeneity of travel time, similar to Table 4, we estimate an instrumental variable regression in Column (2), and the instrumental variable regressions show that high-speed railways substantially improve patent citations, and we have an even larger coefficient, which is 0.95. One percent increase in saved time by high-speed railway translates into an 0.95% increase in patent citations. These results suggest that innovation quality increases substantially due to the reduction in transportation time across cities.

[Insert Table 5 here]

4.4 Robustness checks

Parallel trends test. The identifying assumption for the estimation equation (1), (2) and (3) is that the outcome variables followed similar time trends before high-speed railway is constructed. Meanwhile, right after high-speed railway is put into use, innovation collaboration rises statistically significant and gradually increases at city level or city-pair level. The difference in innovation collaboration between cities (city pairs) with high-speed railway and cities (city pairs) without high-speed railway begins to diverge after high-speed railway is put into actual use.

In this section, we test the following equation (7) and examine the coefficients of α_n and α_m on Y_{ct}, which is the number of collaborative patents for firms in city c with firms in other cities in year t. $FirstHSRconnect$ is equal to 0 if city c is not connected to the high-speed railway network, and it is 1 when city c is firstly connected to the high-speed railway network and thereafter. $FirstHSRconnect_{c,t-n}$ is the nth lag and $FirstHSRconnect_{c,t+m}$ is the mth lead.

In Figure 3, we plot the time trends for the coefficients of α_n and α_m. This enables us to trace the effect of high-speed railways on innovation collaboration before and after city i is connected to the network. Figure 3, where we use *time= -5* as the omitted reference group, shows that there are no anticipatory effects before a high-speed railway opens in city c, and the innovation collaboration at city c increases substantially after high-speed railway is in operation.

$$Y_{ct} = \sum_{n=1}^{n=4} \alpha_n FirstHSRconnect_{c,t-n} + \sum_{m=0}^{m=5} \alpha_m FirstHSRconnect_{c,t+m} + \alpha_2 X_{ct-1} + \lambda_t + \pi_c + e_{ct} \quad (7)$$

[Insert Figure 3 here]

To see the possible consequences of anticipation effects, we exclude the sample 5 years before the opening of the high-speed railway at city level. Note that the opening date to high-speed railway is different in different cities. Suppose city c is firstly connected to the high-speed railway network in year T, we exclude the 5 years before year T for city c. The empirical result at city level is presented in panel A of Table D1. Comparing the coefficients with Table 2, we find that the magnitudes of coefficients in panel A of Table D1 is only slightly larger than the coefficient in Column (2) of Table 2, suggesting the anticipation effect is not a big issue at city level.

Moreover, we examine the following equation (8) to check the parallel trends of innovation collaboration at city-pair level. $FirstHSRconnect\_time_{ij,t}$ is the saved time after city i and j are firstly connected. $FirstHSRconnect\_time_{ij,t-n}$ is the nth lag and $FirstHSRonnect\_time_{ij,t+m}$ is the mth lead. In Figure 4 we plot the time trends for the coefficients of γ_n and γ_m. This enables us to trace the effect of high-speed railways on innovation collaboration before and after city i and city j are connected by high-speed railway. From Figure 4, where *time= -5* is the omitted group, we can see that all coefficients are not significant before time 0, and the magnitudes of the coefficients keep increasing after high-speed railway is constructed between cities and most of the coefficients are statistically significant after time 0. Meanwhile, we have to admit that innovation collaboration at city-pair level is trending upward prior to the actual construction, although this effect is not significant. This implies that the innovation collaboration might have started increasing in anticipation of high-speed railway's opening. This anticipation effect may bias our estimates.

$$Y_{ij,t} = \sum_{n=1}^{n=4} \gamma_n FirstHSRconnect\_time_{ij,t-n} + \sum_{m=0}^{m=5} \gamma_m FirstHSRonnect\_time_{ij,t+m} + \gamma_2 X_{ij,t-1} + \lambda_t + \pi_{ij} + \epsilon_{ij,t} \quad (8)$$

[Insert Figure 4 here]

To see the possible consequences of anticipation effects at city-pair level, we exclude the sample 5 years before city i and j are firstly connected to high-speed railway network. Note that the opening date to high-speed railway for each city pair (i, j) is different. Suppose city pair (i, j) is firstly connected to the high-speed railway network in year T, we exclude 5 years before year T for this city pair. The empirical result at city-pair level is presented in Column (1) of Table 6. Obviously, the estimates are also sligtly larger than the results in Column (2) of Table 4. The results for extensive margin and intensive margin are reported in Panel A of Table E1 in the appendix, showing that the significance and magnitudes of the coefficients are also a little larger than those in Table 4. Thus, we conclude that the anticipation effect may exist before the construction of high-speed railway, however, its impact is quite small. In most cases, it is not clear for the specific route for the construction of each high-speed railway, which is based on the geographical conditions, economic factors, political reasons, etc. We have also tested the robustness by excluding the sample 4 years, 3 years, 2 years, before the opening of the high-speed railway, and the empirical results are similar, suggesting that anticipation effect is not a main issue.

[Insert Table 6 here]

Remove the municipality and capital city. Cities in China differ in hierarchical significance. In particular, our regression sample includes four municipalities directly under central government control. A municipality is a city with a uniform jurisdiction, which has the same administrative level as a province, and this is clearly higher than an ordinary city. Currently, there are 27 provinces and four municipalities in Mainland China, i.e., Beijing, Tianjin, Shanghai and Chongqing. Besides, we have 31 provincial capital cities, which are political, economic, scientific, educational, cultural, and transportation center of each province (Banerjee et al., 2020; Faber, 2014; Gao and Zheng, 2020).

We admit that bigger and faster growing cities were the ones that were usually first connected by high-speed railway. These cities are likely the ones that would have had more innovative activity growth even without high-speed railways. In this case, the OLS estimates are biased upward. To ensure that our results are not driven by these large cities, we exclude the four centrally administered municipalities and provincial capital cities from the sample.

The empirical results using instrumental variable at city-pair level are presented in Column (2) and (3) of Table 6. In Column (2), we only remove municipalities, and in Column (3), we remove both municipalities and capital cities. The estimates for extensive margin and intensive margin at city-pair level are shown in Panel B and C in Table E1. Compared with the results in Table 4, the coefficients of $HSRtime_{ij,t}$ is slightly smaller, but still statistically significant. Even if large cities benefit more from innovation collaboration driven by the high-speed railway network, smaller cities can also benefit from innovation collaboration by entering the high-speed railway network. In addition, we control for the total number of noncollaborative patents in at city-pair level, which can mitigate this issue.

The empirical results at city level are reported in Panel B and C in Table D1. We find that the magnitudes of the coefficients are similar as the coefficients in the baseline estimates in Table 2 when we exclude municipalities, while the coefficients are much smaller if we exclude both municipalities and provincial capital cities.

An alternative instrument variable. In order to check the robustness of the instrument variable, we use an alternative instrument following Duranton et al. (2014), who used the planned highway network as the instrument for modern network of interstate highways in U.S.. This is similar as Gao and Zheng (2020) and Hornung (2015), who used a straight-line strategy to construct instrument for actual high-speed railway connections. The primary objective for high-speed railway is to shorten the travel time between large cities, and straight line is the best design to fulfill this aim. We first draw straight lines between the starting city and ending city based on the construction plan of "Four north-south vertical high-speed railway networks" and "Four east-west horizontal high-speed railway networks" which was initially proposed in 2004. We construct a dummy variable at city level which takes 1 if the city is located on any of these straight lines, and 0 otherwise, multiplied by time trend. At city-pair level, the instrument is a dummy variable which is equal to 1 if two cities are located on any of these straight lines at the same time, and 0 otherwise, multiplied by time trend. This new variable is used as an alternative instrument for high-speed railway.

The distribution of "Four north-south vertical high-speed railway networks" and "Four east-west horizontal high-speed railway networks" is shown in Figure A2. The empirical results using this new instrument are reported in Table D3 at city level and Table E2 at city-pair level in the appendix. The coefficient of the key explaining variable is still statistically significant and positive, suggesting that high-speed railways indeed improve the innovation collaboration at either city level or city-pair level. We admit that the magnitudes of the coefficients for the measures of high-speed railway in this section are much larger than the baseline estimates in Table 2 and Table 4, but the significance and the sign of the high-speed railway still remains. Thus, the results are qualitatively similar even if they are not quantitatively similar when we use this alternative instrument variable.

Innovation collaboration between headquarter and subsidiary. Large firms are more likely to have multiple branches outside their headquarter cities. Headquarters tend to be located in larger cities that are more likely to be connected by high-speed railway than branches are. If the innovation collaborations occurred between headquarters and branches which were located in two cities, they were counted as the innovation collaboration at both cities for the analysis at city level. To deal with the potential issue that the innovation collaboration between headquarter and subsidiary can exist even if there is no high-speed railway, or the impact of high-speed railway can be overestimated, we try to exclude the effect of head-subsidiary innovation collaborations from the sample.

First, we merge the patent data with listed companies and subsidiary companies. The name of head company and subsidiary company comes from CSMAR database. We obtained 47,975 collaborated patents between head company and subsidiary companies based on this merged dataset. Second, we merge our main dataset with the patent between head company and subsidiary companies based on the patent application number. We obtain 8285 patents which are collaborated patents between head and subsidiary company and this takes 5.51% of the total number of collaborated patents[11]. Finally, we exclude these collaborated patents between head and subsidiary company from our sample, and examine the impact of high-speed railway on innovation collaboration.

The empirical result is presented in Column (4) of Table 6 for city-pair level and Panel D of Table D1 for city level. The estimates for extensive margin and intensive margin at city-pair level are shown in Panel D of Table E1. These results show that our regressors of interest barely changed

[11] The total number of collaborated patents is 150341, and this ratio is calculated by 8285/150341.

in their significances and magnitudes either at city level or city-pair level, compared to the baseline estimates. Thus, the empirical findings are still robust after we consider the effect of innovation collaboration between headquarter and subsidiary of listed companies.

The effect of financial crisis. China reacted to the global financial crisis with a massive fiscal stimulus. In 2008, the State Council of China announced a package amounted to be 4,000 billion yuan RMB, which is approximately 590 billion U.S. dollars. These stimulus plans were implemented immediately and focused on investment. This surge in investment was achieved by injecting financial resources into state-owned firms and local infrastructure projects. The stimulus package helps China to escape the Great Recession (Wen and Wu, 2019; Ouyang and Peng, 2015; Huang et al., 2020). It can affect firm's innovations as well, for instance, through government subsidies or bank credits to firms. They can also have a direct impact on the construction of high-speed railways, which is one of the most important ways for infrastructure construction. Omitting the impact of this stimulus package can generate biased estimation for the effect of high-speed railway on innovation collaboration.

In this section, we attempt to exclude the possible effects of such government investments in high-speed railway construction and innovation by excluding the observations in 2008 and 2009 from the sample. The instrumental variable regression result at city-pair level is reported in Panel A of Table E3 in the appendix. The coefficients of $HSRtime_{ij,t}$ for total patents, the extensive margin and the intensive margin at city-pair level do not change much compared to the results in Table 4, suggesting that the financial crisis did not have a significant effect in our framework. We also consider the effect of financial crisis at city level, and the regression results can be found in Panel A of Table D2 in the appendix. Estimates at city level also suggest that our results are robust to the effect of financial crisis.

Restrict the sample to cities that are connected or planned to be connected by high-speed railway. It is possible that the construction of a high-speed railway is determined by the past economic growth of cities, expected future growth and other unobserved factors. If this is the case, the estimation assumption that the innovation collaboration has parallel trend between cities connected with high-speed railway and cities that are not connected may not hold in reality (Lin, 2017). To exclude the possibility that cities that are connected (or planned to be connected) and those that are not connected are fundamentally different, we restrict the sample to the cities that are either connected or planned to be connected by high-speed railway in 2020, and exclude cities that are not connected to high-speed railways.

Once we confine the sample to cities that are connected or planned to be connected by high-speed railway, the number of cities is 119 and the number of city pairs is 2287 and this sample size is much smaller compared to the baseline estimation. The empirical results at city-pair level are reported in Panel B of Table E3 in the appendix, and results at city level is in Panel B of Table D2. Compared with the baseline result at city-pair level in Table 4, the coefficients of the key independent variable, $HSRtime_{ij,t}$, are all statistically significant and positive, and the magnitudes of coefficients are slightly larger in Table E3. For the results at city level, the impact of $HSRstation$ is also a little larger than the baseline estimate in Table 2, although the significance level drops for the total effect in Table D2.

Redefine the opening time of high-speed railway. Given the fact there can be a time lag between the construction of high-speed railway and its application on innovation collaboration, we redefine the opening date of high-speed railway. For instance, if the high-speed railway is put into use in December in year t, its real effect will take place in year $t+1$; and if it is completed in January in year t, its effect can start in year t. If we do not adjust the date, it may underestimate the actual effect of high-speed railway. Since we do have the actual date for each high-speed railway line, we redefine the opening time for high-speed railway as follows: at city-pair level, if the opening date of high-speed railway connecting city i with city j is July or later of year t, we redefine its opening date to be $t+1$; If the opening date is in or before June of year t, we define its opening year to be t. Similar definition applies at city level.

Panel C of Table E3 in the appendix reports the regression results using this new definition of opening time at city-pair level and results at city level is in Panel C of Tabel D2 in the appendix. We can see that the coefficients are slightly larger than the coefficients in Table 4 for results at city-pair level and the coefficients in Table 2 for results at city level. But the sign and significance of all the main results are preserved when we redefine the opening time of high-speed railway.

Restrict the sample to cities without airports. Air travel is less popular than railway or high-speed railway in China (Lin, 2017), however, the existence of air travel may induce the problem that the coefficient of high-speed railway is overestimating the impact of high-speed railway on innovation collaboration if air travel and high-speed railway are close substitutes.

In this section, we exclude cities that have airports. Using this small sample, we report the estimates at city-pair level in Panel D of Table E3 and results at city level in Panel D of Table D2. The coefficient for the key explaining variable is not significant but still positive at citiy level due to the small sample. Across all specifications at city-pair level, we find a strong, significant and positive association between the saved time and patent collaboration, but the magnitudes turns to be smaller than those in Table 4. Even if most cities in this small sample are not as developed as those with airports, we can nevertheless observe a positive and significant effect on innovation collaboration across cities.

Balanced sample. The composition of cities changes across years due to the missing values of some control variables, such as FDI and the number of industrial firms. In order to make sure that our main findings are not driven by these cities with missing values, in this section, we examine the impact of high-speed railway on innovation collaboration using the balanced sample at both city level and city-pair level.

The estimation results at city level are reported in Panel E of Table D1. Comparing with the baseline estimation in Table 2, we find that the magnitude of the coefficient for *HSR* is only slightly smaller, and the significance keeps the same. The robustness check at city-pair level is presented in Panel E of Table E3, and the coefficient for the key explaining variable is similar as the main finding in Table 4, with a slightly smaller coefficient. Thus, our findings that high-speed railway promotes innovation collaboration is robust.

4.5 Heterogeneous effects

In the above analysis, we have studied the impact of high-speed railway on innovation collaboration and proved the robustness of this effect. In this section, we will examine the heterogenous effect of high-speed railway on innovation collaboration based on the specific characteristics of the cooperating cities and firms, in particular, we focus on the spatial heterogeneity, industry heterogeneity, ownership heterogeneity and heterogeneity effect based on distance across cities.

Spatial heterogeneity effects. In this section, we consider the spatial heterogeneity for the impact of high-speed railways on innovation collaboration. Economic growth is unbalanced in China. It is more developed in the eastern and coastal regions and less developed in inland regions, i.e., the western region and central region. To understand the heterogeneous impacts of high-speed railway on innovation collaboration in different regions, we examine its impact separately in eastern, central and western regions. The results at city-pair level are reported in Panel A of Table 7. In panel A, we report the results for the total number of collaborated patents at city-pair level. Column (1) reports the result for the eastern region, Column (2) reports the results for central region, and Column (3) is the result for western region. The results for the extensive margin is reported in Panel A of Table E4 in the appendix, and Panel A of Table E5 shows the results for the intensive margin.

It is clear that the coefficient of the western region is the largest for total number of collaboration, the extensive margin, and the intensive margin because the connections of western cities to other cities through high-speed railway network increases their communication and cooperation opportunities more substantially than for cities in other regions. This is similar to the findings of Zheng and Kahn (2013), who showed that high-speed railways stimulated the development of second-tier and third-tier cities. The innovation collaboration in eastern and central regions can be less sensitive to high-speed railways because most innovation collaborations are generated internally, and knowledge is more likely to circulate within their boundaries, given their geographical advantages and better economic performance. This evidence suggests that the impact of high-speed railways depends on the characteristics of local environment, such as inventor quality, technological specialization, local human capital and economic development (Agrawal et al., 2017).

[Insert Table 7 here]

Industry heterogeneity effects. In this section, we examine whether high-speed railways have different impacts for different industries at the city-pair level. We divide industries into three main types, i.e., manufacturing, services and other industries. Enterprises can collaborate with other enterprises in another city within their own industry, or they can collaborate with enterprises in different industries in other cities. The empirical results for total effect are shown in Panel B of Table 7. Column (1) reports the result for manufacturing industry, Column (2) is the results for service sector, and Column (3) shows the result for other industries. In addition, we also examine the impact of high-speed railways on the innovation collaboration across industries, i.e., the collaboration between manufacturing and service industry in Column (4), the collaboration between service and other industry in Column (5), and the collaboration between manufacturing and other industries in Column (6). We report the results for the extensive margin in Panel B of

Table E4 and intensive margin in Panel B of Table E5 in the appendix.

For the total number of patents, the coefficients of $HSRtime_{ij,t}$ are much larger for collaboration within the service sector and the collaboration between manufacturing and service sector. This indicates that a reduction in transportation costs can not only increase collaboration within sectors but also strengthen collaboration across sectors, especially between manufacturing and service sector. We obtain similar findings for the extensive margin and intensive margin.

The cooperation within the manufacturing industry, within other industries, between service and other industries, and between manufacturing and other industries are also experiencing positive and significant effects, although the coefficients are slightly smaller. These findings provide evidence to support the finding of Jacbos (1969) that the exchange of complementary knowledge across industries is central to the creation of new knowledge.

Ownership heterogeneity effects. In this section, we examine how high-speed railways impacted innovation collaboration for different ownerships. We have three types of ownership: state-owned enterprises, foreign enterprises and private enterprises. In Panel C of Table 7, Columns (1), (2) and (3) report the results examining the impacts of high-speed railways on innovation collaboration for state-owned enterprises, foreign firms, and private firms, respectively. We report the results for the extensive margin in Panel C of Table E4 and intensive margin in Panel C of Table E5 in the appendix.

The coefficients of $HSRtime_{ij,t}$ are both positive and significant for state-owned enterprises and private enterprises. In addition, the coefficient of $HSRtime_{ij,t}$ is larger for state-owned enterprises than private enterprises. For foreign firms, the innovation collaboration is negative, but not statistically significant.

We further consider the collaboration between different types of ownership in Columns (4), (5) and (6). The coefficients of $HSRtime_{ij,t}$ are all positive and significant. In particular, the coefficient is much larger for the cooperation between private enterprises and state-owned enterprises in Column (5). Thus, high-speed railways can strengthen innovation collaboration and knowledge spillover across ownership types, especially for private firms and state-owned firms.

Heterogeneity effects based on distance. High-speed railway has a comparative advantage over airplanes, normal train and other transportation modes within a certain distance. If the travel distance is too long, most people may want to switch from high-speed railway to air travel, while if the distance is too short, people may want to take a bus or drive themselves. Similar as Wang and Cai (2020), we test how geographical distance affect innovation collaboration between firms in Table 8. Results for the extensive margin and intensive margin are presented in Table E6 in the appendix. We generate several dummy variables based on the straight-line distance between city pairs, i.e., the distance smaller than 100 kilometers, the distance between 100 and 200 kilometers, the distance between 200 and 300 kilometers and the distance larger than 300 kilometers. Then, we examine the impact of high-speed railway based on distance differences. These empirical results show that, compared to the distance (200, 300), the impact of high-speed railway is more effective in shorter distance, i.e., less than 200 kilometers, on innovation collaboration. High-speed railway is not a complete substitute for airlines, especially for longer trips. In addition, face-to-face communication is still playing key role in innovation collaboration, even in the age of internet (Wang and Cai, 2020).

[Insert Table 8 here]

5 Conclusion

The rapid rise of innovation in China, as measured by patents, has received considerable attention. The extant literature has mostly studied the possible factors to increase innovation. There has been no rigorous empirical study using micro firm-level data to examine innovation collaboration from the perspective of high-speed railways in China. In addition, the existing literature mainly focuses on developed countries, such as the U.S. and Europe. The current study is, to the best of our knowledge, the first study to investigate innovation collaboration using matched firm data and patent data in developing countries. We use China's high-speed railway network construction as a source variation and examine its impact on innovation collaboration between firms located in different cities in China. The difference-in-difference estimation results suggest that once a city is connected to a high-speed railway network, the number of collaborative innovations increases statistically significantly at both city level and city-pair level. The patent quality, measured by the number of citations, is also improved substantially. We find that high-speed railways have encouraged enterprises in different cities to cooperate more often on innovation on both the extensive margin and the intensive margin.

To address the endogeneity of high-speed railways, we estimate an instrumental variable regression using the historical railway lines in 1934 as the instruments for the high-speed railway lines in our sample period. After controlling for the endogeneity of high-speed railways, we find that high-speed railways can substantially facilitate innovation collaboration. The impact of high-speed railways on innovation collaboration varies across different regions. Its effect is more significant in western than in eastern and central regions. High-speed railways have larger effects on innovation collaboration in service sector than in the manufacturing sector. Finally, high-speed railways can boost the innovation collaboration between state-owned enterprises and domestic private enterprises. Our research contributes to the literature on evaluating the impact of high-speed infrastructure construction on innovation collaboration. Our findings can enrich the understanding of the rapid development of innovation and economic growth in China.

The findings that high-speed railway improved the quantity and quality of collaborated patents have important implications for policymakers. First, transportation construction within a city or between cities which can increase the frequency for face-to-face meetings should be improved further to facilitate knowledge flow and research collaboration. Although the primary objective for high-speed railway construction is to improve railway services, promote industrial transformation, urbanization, and coordinated development of regional economy, it also serves as an important way to facilitate knowledge flow and transmission. Second, due to high-speed railway, more innovation collaborations come from the cooperation between more developed and less developed regions. For less developed regions, more infrastructure investments are required to attract more resources for innovation, such as human capital. In the future plan of transportation construction, more investments can be put in less developed regions. Therefore, our findings also have policy implications for the regional disparities across China. High-speed railways have effectively improved communication and cooperation between cities and contributed to more balanced development across regions in China. Finally, supplementary polices, such as regional policies, industrial polices and financial policies, should be combined with infrastructure construction to improve the innovation abilities of enterprises and increase the cooperation across regions and industries.

References

[1] Acemoglu, D., U. Akcigit, 2012. Intellectual property rights policy, competition and innovation. Journal of European Economic Association 10, 1-42.

[2] Acemoglu, D., U. Akcigit, W. R. Kerr, 2016. Innovation network. Proceeding of National Academy of Science 113(41), 11483-11488.

[3] Aghion, P., P. Howitt, 1992. A model of growth through creative destruction. Econometrica 60, 323-351.

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

[5] Arzaghi, M., J. V. Henderson, 2008. Networking off Madison Avenue. Review of Economic Studies 75 (4), 1011–38.

[6] Audretsch, D. B., M. P. Feldman, 2004. Knowledge spillovers and the geography of innovation. In Handbook of Regional and Urban Economics, Vol. 4, edited by J. Vernon Henderson and Jacques-Francois Thisse, 2713–39. Amsterdam: Elsevier.

[7] Banerjee, A., E. Duflo., N. Qian, 2020. On the road: Access to transportation infrastructure and econmic growth in China. Journal of Development Economics 145, 102442.

[8] Baum-Snow, N., 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122(1), 775-805.

[9] Baum-Snow, N., V. Henderson, L. Brandt, M. Turner, Q. Zhang, 2017. Roads, railroads and decentralization of Chinese Cities. Review of Economics and Statistics 99(3), 435-448.

[10] Berliant, M., M. Fujita, 2008. Knowledge creation as a square dance on the Hilbert Cube. International Economic Review 49(4), 1251-1295.

[11] Berliant, M., R. R. Reed III, P. Wang, 2006. Knowledge exchange, matching, and agglomeration. Journal of Urban Economics 60, 69-95.

[12] Bertrand, M., E. Duflo, S. Mullainathan, 2004. How much should we trust Differences-in-Differences dstimates. The Quarterly Journal of Economics 119(1), 249-275

[13] Brandt, L., Van Biesebroeck, J., Wang, L., Zhang, Y., 2017. WTO accession and performance of Chinese manufacturing firms. American Economic Review 107(9), 2784-2820.

[14] Campante, F. and D. Yanagizawa-Drott, 2018. Long-range growth: Economic development in the global network of air links. The Quarterly Journal of Economics, 1395-1458.

[15] Charikar, M. S., 2002. Similarity estimation techniques from rounding algorithms. In Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, 380-388.

[16] Charlot, S., G. Duranton, 2004. Communication externalities in cities. Journal of Urban Economics 56(3), 581-613.

[17] Davis, D.R., J. I. Dingel, 2019. A spatial knowledge economy. American Economic Review 109(1), 153-170.

[18] Donaldson, D., 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. American Economic Review 108(4-5), 899-934.

[19] Dong, X., S. Zheng, M. Kahn, 2020. The role of transportation speed in facilitating high skilled teamwork. Journal of Urban Economics, 115, 103212.

[20] Duranton, G., P. Morrow, M. A. Turner, 2014. Roads and trade: Evidence from the US. Review of Economic Studies 81(2014), 681-724.

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

[22] Faber, B. 2014. Trade integration, market size, and industrialization: evidence from China's national trunk highway system. The Review of Economics Studies, 81(3), 1046-1070.

[23] Faria, P. de, F. Lima, R. Santos, 2010. Cooperation in innovation activities: The importance of partners. Research Policy 39, 1082-1092.

[24] Gao Y., J. Zheng, 2020. The impact of high-speed rail on innovation: An empirical test of the companion innovation hypothesis of transportation improvement with China's manufacturing firms. World Development, 127, 104838.

[25] Glaeser, E. L., A. Saiz, 2004. The rise of the skilled city. Brookings-Wharton Papers on Urban Affairs, 47-94.

[26] Grossman, G. M., E. Helpman, 1991. Trade, knowledge spillovers, and growth. European Economic Review 35(2-3), 517-526.

[27] He, Z., Tong, T. W., He, W., Zhang, Y., Lu, J., 2013. Chinese patent database user documentation: Matching sipo patents to chinese publicly-listed companies and subsidiaries. User Documentation, Chinese Patent Data Project September.

[28] Holmes, T. J., McGrattan, E. R., Prescott, E. C., 2015. Quid pro quo: Technology capital transfers for market access in china. The Review of Economic Studies 82(3), 1154-1193.

[29] Hornung, E., 2015. Railroads and growth in Prussia. Journal of European Economic Association, 13(4), 699-736.

[30] Hsieh, C. T., Klenow, P. J., 2009. Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics 124(4), 1403-1448.

[31] Hu, A. G. Z., 2020. Public funding and the ascent of Chinese science: Evidence form the National Natural Science Foundation of China. Research Policy, forthcoming.

[32] Huang Y., M. Pagano, U. Panizza, 2020. Local crowding-out in China. Journal of Finance, 75(6), 2855-2898.

[33] Jacobs, J., 1969. The Economy of Cities. New York: Random House.

[34] Jaffe, A. B., M. Trajtenberg, 2002. Patents, Citations and Innovations: A Window on the Knowledge Economy. MIT Press, Cambridge, MA.

[35] Jaffe, A. B., M. Trajtenberg, R. Henderson, 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108, 578–598.

[36] König, M. D., X. Liu, Y. Zenou, 2019. R&D networks: Theory, empirics, and policy implications. Review of Economics and Statistics 101(3), 476-491.

[37] Krugman, P., 1991a. Increasing returns and economic geography. Journal of Political Economy 99(3), 483-499.

[38] Krugman, P., 1991b. Geography and trade, Cambridge (MA): MIT Press.

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

[40] Lin, Y., Y. Qin, J. Sulaeman, J. Yan and J. Zhang, 2020. Expanding footprints: fhe impact of passenger transportation on corporate locations. Working paper at SSRN: https://ssrn.com/abstract=3418227

[41] Liu, F., D. Simon, Y. Sun, C. Cao, 2011. China's innovation policies: Evolution, institutional structure and trajectory. Research Policy 40, 917-931.

[42] Manku, G. S., Jain, A., Das Sarma, A., 2007. Detecting near-duplicates for web crawling. In Proceedings of the 16th international conference on World Wide Web, 141-150.

[43] Möller J., M. Zierer, 2018. Autobahns and jobs: A regional stuy using historical instrumental variables. Journal of Urban Economics, 103, 18-33.

[44] Hong, Maoxian, 1930. The latest Chinese situation at a glance, by Eastern Yudi Academy Society.

[45] Michael, G., 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. Review of Economics and Statistics 90(4), 683-701.

[46] Ouyang, M., Y. Peng, 2015. The treatment-effect estimation: A case study of the 2008 economic stimulus package of China. Journal of Econometrics 188, 545-557.

[47] Qin, Y. 2017. "No county left behind?" The distributional impact of high-speed rail upgrades in China. Journal of Econoimc Geography, 17(3), 489-520.

[48] Rappaport, J., 2007. Moving to nice weather. Regional Science and Urban Economics 37(3), 375-398.

[49] Romer, Paul M., 1986. Increasing returns and long run growth. Journal of Political Economy 94, 1002-1037.

[50] Rosenkopf, L., P. Almeida, 2003. Overcoming local search through alliances and mobility. Management Science 49(6), 751–766.

[51] Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. Management Science 51(5), 756-770.

[52] Wang J., S. Cai, 2020. The construction of high-speed railway and ubran innovation capacity: Based on the perspective of knowledge spillover. China Economic Review 63, 101539

[53] Wang, Y., J. Li, J. L. Furman, 2017. Firm performance and state innovation funding: Evidence from China's innofund program. Research Policy 46, 1142-1161.

[54] Wang, X., G. Fan, L. Hu. Marketization index of China's provinces, 2006-2016. Social Sciences Academic Press (China)

[55] Wen Y., J. Wu, 2019. Withstanding the great recession like China. The Manchester School 87 (2), 138-182.

[56] Zheng, S., M. Kahn, 2013. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceeding of National Academy of Science 110(14), E1248-E1253.

[57] Zhu, J., Y. Fan, X. Deng, L. Xue, 2019. Low-carbon innovation induced by emissions trading in China. Nature Communications, 10, 4088.