Subway, Collaborative Matching, and Innovation*

Yumi Koh[a†], Jing Li[b‡], Jianhuan Xu[b§]

[a] *University of Seoul, Korea*

[b] *School of Economics, Singapore Management University, Singapore*

December 30, 2021

Abstract

Expansion of subway networks helps to enhance connectivity and matches of people by facilitating their mobility. Using rapid expansion of the Beijing subway from 2000 to 2018, we analyze its impact on collaborative matches in innovations. We find that an hour reduction in travel time between a pair of locations in Beijing brought a 15% to 38% increase in collaborated patents. Far-apart location pairs were more affected, and the local average causal response is approximately 35% to 82%. Such effect is mainly driven by increased matches among highly productive inventors due to complementarity between inventors' productivity and travel time. At the same time, we find that the entry of new inventors, relocation of existing inventors, and collaborations among low productive inventors also contribute to the increase in collaborative matches, especially in the long run.

Keywords: innovation; matching; patent collaboration; Beijing subway; transportation infrastructure

JEL classifications: D83, O18, O31.

---

*We thank Xiaofang Dong, Yuming Fu, Richard Hornbeck, Jeffrey Lin, and seminar and conference participants at the Pennsylvania State University, the University of Seoul, Singapore Management University (SMU), Southwestern University of Finance and Economics, the 2020 SMU Virtual Conference on Urban and Regional Economics, the 2021 European Meeting of the Urban Economics Association (UEA), the 2021 American Real Estate and Urban Economics Association (AREUEA) National Meeting, the 2021 Jinan-SMU-ABFER Conference on Urban and Regional Economics, the 2022 AREUEA Annual Meeting for their helpful comments. We thank Aaron Zhe Yang Lim for his tremendous help as a research assistant. Remaining errors are our own.

†163 Seoulsiripdaero, Dongdaemun-gu, Seoul, Korea. Phone: +82-2-6490-2057. E-mail address: ymkoh@uos.ac.kr.

‡90 Stamford Road, Singapore 178903. Phone: +65-6808-5454. E-mail address: lijing@smu.edu.sg.

§90 Stamford Road, Singapore 178903. Phone: +65-6808-0085. E-mail address: jhxu@smu.edu.sg.

1. Introduction

Whereas inter-city transport networks shape the formation of cities by changing the ease of access to *goods and services* (Helpman 1998; Krugman 1991; Ottaviano et al. 2002), the intra-city transport infrastructure shapes the internal structure of cities mainly by altering the cost of moving *people*.[1] Previous studies modelling the role of intra-city transport infrastructure have mainly focused on its impact on commuting between residences and workplaces (Ahlfeldt et al. 2015; Baum-Snow 2007; Baum-Snow et al. 2017; Heblich et al. 2020). However, an important but under-explored implication of facilitating mobility of people within a city is that it also enhances their connectivity and matches with one another. This aspect is particularly important in the context of innovations and innovation-based urban growth, since innovators with complementary expertise have much to gain from collaboration as knowledge and technology become more advanced and specialized (Jones 2009).

In this paper, we analyze the impact of improving intra-city transport infrastructure on collaborative matches in innovations. In particular, we focus on a unique episode of rapid expansions of the subway system in Beijing from 2000 to 2018 and quantify its impact on patent collaborations. During this time period, Beijing witnessed a dramatic extension of its total subway length from 54 *km* to 655 *km*, overtaking all other global metropolitan cities to own the world's longest and busiest rail transit networks. At the same time, patent collaborations within Beijing grew substantially in terms of both scale and geographic scope, as demonstrated in detail in Section 2. As the subway system expanded to create an intricate web of underground train lines within Beijing, we track how improved connectivity ultimately increased innovation activities and reshaped spatial collaboration patterns.

Our study adds to different strands of literature and makes contributions in the following aspects. First, it draws a unique linkage between intra-city transport infrastructure and collaborative matches in innovations to aid our understanding of innovation-based urban growth (Acemoglu 2008; Jones 2009; Carlino et al. 2007; Kerr and Robert-Nicoud 2020). Recent studies have focused on how *inter-city* transport infrastructure, such as cheap airline connections or high-speed rail developments, affects

[1] As of today, the London Underground rapid transit system serves up to 5 million passenger journeys a day (`https://tfl.gov.uk/corporate/about-tfl/what-we-do`). The Beijing Subway, one of the world's longest and busiest metro systems, is projected to serve 18.5 million trips every day by 2021 (`http://www.chinadaily.com.cn/china/2017-01/12/content_27931764.htm`).

research paper publication among academic scholars (Catalini et al. 2020; Dong et al. 2020). Our study extends the literature by focusing on an *intra-city* transport (i.e. a subway system) at a more detailed micro geographic level to assess its impact on patent collaboration among inventors. Moreover, we incorporate various mechanisms by allowing for spatial sorting in our conceptual framework and empirically assess the relative importance of each of the channels, as will become apparent in later text.

Second, our theme on collaborative matches in cities is further linked to the literature on the nature of agglomeration economies that facilitate innovation growth. Agglomeration economies in innovation hinge on the idea that innovation clusters bring in external increasing returns to scale which enhance productivity (Moretti 2019). Such increasing returns are likely achieved through the mechanisms of sharing, learning, and matching as modelled in Duranton and Puga (2004). While much is known about how different mechanisms generate externalities in the production of *goods and services* (Combes and Gobillon 2015; Rosenthal and Strange 2004), less is known about how they manifest in producing *inventions and innovations* (Carlino and Kerr 2015). We highlight the role of matching facilitated by reduced travel cost in enhancing innovation productivity. Although we do not model agglomeration economies directly in our analysis, the gains from reduced travel time that we document in this paper speak directly to the rationalization of underlying agglomeration forces and returns to urban density.

Third, our paper is also important from a policy perspective. Urban policy-makers have long considered rail transit system as an effective way to reduce congestion and have, hence, made huge investments on constructing extensive and complex transit networks.[2] Evidence thus far suggests that subway line extensions effectively increase road speed, save time from reduced congestion, and also reduce air pollution (Gendron-Carrier et al. 2018; Gu et al. 2019). Apart from these directly targeted outcomes, our paper shows that a subway network potentially entails much broader economic consequences, such as improving collaborative matching. Quantifying such indirect but important economic consequences has become crucial for policy-makers, as they are hard-pressed to make thorough and comprehensive net-benefit justifications given skyrocketing construction and maintenance costs of transportation infrastructures.[3] Our back-of-the-envelope calculations show that the increase

[2]By 2014, 171 cities worldwide have a subway system in operation (Gendron-Carrier et al. 2018).

[3]For instance, the estimated cost of the Long Island Rail Road project, known as the "East Side Access," has ballooned

in collaborative innovations in Beijing do account for a significant share of benefits arising from the build-up of the subway system.

One important aspect that needs to be addressed carefully when evaluating the impact of subway expansion is that the selection of locations into the treatment group may be nonrandom. Such correlation invalidates the standard ordinary least squares assumptions and renders the estimated coefficients biased. We address this issue with a collection of efforts. Specifically, our analysis fully exploits both cross-sectional and intertemporal variations in connectivity between locations, which allow us to control for two-way fixed effects that absorb additive location-pair-specific and time-specific unobserved characteristics. Accounting for such unobserved features, we estimate 1) a discrete difference-in-differences (DiD) specification; 2) a continuous "gravity-equation"-like specification; 3) a two-stage least squares specification by adopting an instrumental variable approach, and 4) a control-pair specification as proposed in Jaffe et al. (1993).

In our DiD estimation, we define a location-pair as treated if the pair-specific connectivity improves substantially as travel time decreases by more than 30 minutes along the fastest travel route. Then our aim is to characterize the extent to which collaborative matches in innovation, as measured by the number of collaborated patents, formed between treated location-pairs increase relative to those formed between other location-pairs that were not treated in the same period. Given the staggered treatment timing, we implement DiD with the Callaway and Sant'Anna (2021) estimation procedure to achieve a more interpretable coefficient. Building on this approach, we then adopt a "gravity-equation"-like specification to measure connectivity continuously. The idea is analogous to a "gravity model," in which we expect collaborative matches formed between two locations to be inversely proportional to the travel time between those locations. To further corroborate our findings, in our third approach, we instrument for connectivity between two locations using the interaction of their Euclidean distance and the citywide aggregate expansion of the subway network. The identification assumption is that the expansion of the subway system changes the connectivity of far-apart locations more than that of close-together locations and this differential impact leads to the differences in patent collaborations only through the channel of reduced travel costs.[4] Last, as another

to $12 billion, or nearly $3.5 billion for each new mile of track—seven times the average elsewhere in the world (`https://www.nytimes.com/2017/12/28/nyregion/new-york-subway-construction-costs.html`).

[4]This instrument is analogous to the traditional Bartik shift-share instrument in the sense that the exogenous distance between locations serve as the pre-determined Bartik weight that governs the differential exposure to a common aggregate

robustness check, we follow the matched control approach proposed in Jaffe et al. (1993) to compare the true collaboration pairs to matched control pairs in response to a change in connectivity of two locations. The identification assumes independence of unobserved errors conditional on observed matching characteristics.

Another challenge that we face in evaluating the impact of subway expansion is to distinguish productivity growth from reorganization of existing economic activities. That is, even if subway stations were randomly assigned to locations, it is still difficult to tell whether collaborative matches increased due to a location-specific productivity growth or a spatial reshuffling of innovation activities (Redding and Turner 2015).[5] To investigate the roles of growth versus reorganization, as well as other possible mechanisms, we lay out a simple matching model that explicitly accounts for spatial sorting and solves for innovators' collaboration, location, and occupation choices in response to improved connectivity between locations. The model conceptualizes how the build-up of the subway system produces spatial variations in returns to innovations and, hence, shapes the matching patterns between collaborators across different locations.

The model yields various hypotheses that allow us to empirically test for the presence of different mechanisms. Specifically, it suggests that collaborative matches can increase through four possible mechanisms. First, given the complementarity between travel time and inventors' productivity, the *High-Quality Matches Channel* shows that more collaboration matches would be formed, especially among highly productive inventors. Second, the *Marginal Matches Channel* shows that newly formed collaborative matches would mostly involve low-value inventor pairs on the margin due to reduced cost of collaboration. Third, the *Relocation Channel* suggests that inventors, especially the more productive ones, move to places that become more accessible and such relocation drives the formation of collaborative matches. Fourth, the *Marginal Inventors Channel* shows that as reduced travel time increases the returns to innovation and induces more people to become inventors, collaborative matches increase from a larger pool of inventors. Because each mechanism generates predictions for different subgroups of inventors, we conduct empirical tests to assess the relative importance of the channels.[6]

shock.

[5]This same issue of distinguishing growth from reorganization appears in the literature evaluating placed-based policies (Neumark and Simpson 2015)

[6]Our model differs from the modeling setup in Catalini et al. (2020) by allowing for inventors' spatial relocation decisions and occupation entry decisions. As we highlight in our empirical analysis later on, both channels are quantitatively important, especially in a longer time horizon.

We obtain the following findings. First, our DiD estimation shows a positive and statistically significant treatment effect, indicating a larger increase in the number of patents collaborated between locations whose connectivity improved. Second, we find that an improvement in connectivity between two locations, proxied using the total length of subway lines that connects the two locations along their shortest-time travel path, increases the number of collaborated patents between those two locations. The Wald estimators suggest that an hour reduction in travel time leads to an increase in collaborated patents that ranges from 15% to 38% on average, depending on travel speed assumptions. Third, the impact of subway expansion on patent collaborations is highly heterogeneous: the effect is much larger for far-apart location pairs. The instrumental variable approach reveals an estimated local average causal response among "complying" location pairs, which translates into Wald estimators ranging from 35% to 82%. Robustness checks using the control-pair approach also provide consistent findings.

To assess underlying mechanisms, we examine the extent to which collaborative matches increased in response to improvement in connectivity for different subgroups, created based on productivity, mobility, and entry criteria. For each group, we examine both the short-run and the long-run impact following the shock of a nearby subway station opening at both locations. Whereas all four channels are present in the long run, we find that the *High-Quality Matches Channel* is the most important mechanism driving our reduced-form evidence. Whereas this finding is consistent with Catalini et al. (2020), which identifies the complementarity between travel costs and the quality of a coauthor match in an academic setting, we further highlight the importance of the *Relocation Channel* and *Marginal Inventors Channel*.

The rest of the paper is organized as follows. We present the institutional background in Section 2 and discuss our empirical research designs in Section 3. We discuss data and variables in Section 4. Findings are presented in Section 5, followed by mechanism analysis in Section 6. Section 7 contains the back-of-the-envelope calculations and section 8 concludes. Remaining materials are in the Online Appendix.

2. Institutional Background

Beijing's planning committee proposed the construction of the subway system in 1953 (Sultana and Weber 2016). The initial purpose was to "ferry[ing] soldiers from their barracks on the outskirts to the city center" (Poon 2018), but the subway eventually became means of public transportation.[7] The subway system developed slowly until the 2000s and then went through a major expansion since 2008 for two main reasons. First, as the city won the bid to host the 2008 Summer Olympics, new lines were constructed to connect the main stadiums, to reduce traffic on circular freeway, and to connect the existing large residential areas (Yang et al. 2013). Second, with the central government's stimulus package during the global financial crisis of 2007-2008, the Beijing subway system further expanded to connect nearby suburban districts. Figure 1 shows that between 2000 and 2018, the cumulative count of subway stations grew by a factor of nine (from 41 to 379) and the total subway length increased by a factor of twelve (from 54.1 *km* to 655 *km*).

The extraordinary expansion of the Beijing subway system was accompanied by a remarkable increase in the number of collaborated patents and a geographical expansion of their scope. Panel (a) in Figure 2 illustrates locations of all collaborators of the patents produced in 2000-2001 in red x-marks. It also depicts the subway stations available back then, in black star marks, which were part of the east-west line and the inner loop line covering Beijing's CBD area. Vast majority of collaborators were centrally clustered around those two lines. However, the spatial layout of collaborators greatly expanded by 2017-2018, as shown in Panel (b). A striking feature is that the locations of collaborators closely overlap with the layout of the subway system, which now includes many lines forming loops or branching out to suburban areas. As a simple illustration on the development of spatial connections in patent collaborations across time, we now turn to Panel (c). This shows how the collaboration links evolved for a particular location that had the largest number of collaborated patents in 2000-2001. Yet, all collaborated patents originated in this location during 2000-2001 were formed with collaborators in another single location, and this link between two locations is shown by a red line. However, by 2017-2018, this location formed numerous collaboration links, as shown by a set of many gray lines connecting this location with other locations. A striking feature of these collaboration links is that

[7] `https://www.bloomberg.com/news/articles/2018-02-26/china-is-reining-in-its-subway-boom`

they closely overlap with the layout of the subway lines, as shown in black star marks.

Overall, Figures 1 and 2 show that subway stations and patents grew over time, especially with more collaboration links being formed in locations that were covered by the subway lines. Motivated by such striking graphical evidence, we now turn to empirical research designs that would allow us to investigate the causal impact of spatial connectivity on patent collaborations.

3. Empirical Research Designs

In the empirical framework, we start by estimating a two-way fixed effect (TWFE) model specified as:

$$CollabPatents_{ijt} = \beta Connect_{ijt} + \gamma_{ij} + \zeta_t + \upsilon_{ijt}, \tag{3.1}$$

where ij indexes a pair of grids and t indexes a year. Grids refer to locations of a specific size that we assign in the data and we elaborate on this in Section 4.2. $CollabPatents_{ijt}$ is the count of collaborated patents produced by collaborators in grids i and j in year t. Throughout, we use collaborated patents to proxy for collaborative matches in innovation. $Connect_{ijt}$ measures the subway build-up that enhances the connectivity between grids i and j in year t, which we proxy in two ways using either a binary variable or a continuous variable. The binary variable captures the treatment status in which connectivity between two grids dramatically improves as travel time reduces by more than 30 minutes along the fastest travel route due to the expansion of the subway. The continuous variable captures connectivity using the total length of the subway lines that connect the two grids along their shortest-time travel path. Lastly, γ_{ij}, ζ_t, and υ_{ijt} represent grid pair fixed effects, year fixed effects, and idiosyncratic error term, respectively.

The TWFE model has been conventionally adopted in the literature to identify the average treatment effect on the treated (ATT) in the absence of other endogeneity concerns. However, recent studies find that a TWFE model does not yield an interpretable causal parameter in a setting with variations in treatment timing and heterogeneous treatment effects (de Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021). This is applicable in our context, since the timing of subway station openings rolled out across years and effects are expected to be heterogeneous. Hence, we report the estimates from the TWFE model as a reference with such caveats in mind and we also turn to recent methodologies.

Namely, we estimate a staggered DiD model as proposed by Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). The idea is to group observations of grid-pairs into different cohorts based on the year in which they were first treated. A grid pair is defined to be treated if the travel time reduces by more than 30 minutes between the grids along the fastest travel route. We use the never-treated cohort as the control group as it provides a consistent set of controls for identifying the impact of all cohort-specific ATT. Then we compute the average treatment effect that aggregates the ATT across different cohorts and years by taking weighted averages as proposed by Sant'Anna and Zhao (2020). The underlying identifying assumption is that the evolution of the outcome in the absence of the treatment among the treated cohort would be the same as that of the never-treated cohort during the post treatment period. We discuss and evaluate the validity of this identifying assumption in detail later when we present our empirical findings.

While the above procedure delivers an interpretable coefficient, there may be concerns that the intertemporal layout of the subway system may not be purely exogenous. For instance, policymakers may have planned the phase-in of the subway construction in company with other concurrent place-based policies to promote innovations at particular locations. Two locations may have been linked via the build-up of a subway line, while policy incentives were also offered to attract inventors at such locations to be promoted as innovation hubs. In such a scenario, pairwise connectivity would be positively correlated with collaborative innovations driven by the presence of unobserved policy incentives at the location level. However, we argue that related concerns are unlikely to severely drive our results since our identification exploits the *exogenous timing* of the subway build-up. Beijing subway system is notorious for unexpected construction delays due to various technical obstacles.[8] To the extent that the actual construction completion year differs significantly from the planned completion year in a random way, the timing of subway construction may still be deemed as exogenous. We also find supporting evidence for the exogenous timing assumption in Figures 3 and 4. That is, the parallel trend assumptions are satisfied prior to a shock when a shock is defined at either the grid level or the grid-pair level. We provide more detailed discussions in Section 5.2.

Despite the strong prior, we adopt other research designs to further corroborate our findings. In

[8] See news reports in, for example, `https://wemp.app/posts/8977e79a-26af-429e-84f5-a910399cb516` and `https://m.news.cctv.com/2020/11/21/ARTImLRnVxYC55PyJAns0pc1201121.shtml`.

our second research design, we adopt a composite variable to instrument for $Connect_{ijt}$ by combining two sources of variations: 1) the aggregate development of the subway system over time, and 2) the time-invariant Euclidean distance between two locations. The first component reflects the "aggregate shift" as the subway system grows over time.[9] The second component reflects the "exposure" for a location pair to the subway system development. The identifying assumption is that differential exposure to the common shock (i.e. the development of the subway system) will affect the level of collaborated patents only through the channel of enhanced connectivity, and not through other potential confounding channels related to location selections. Note that our instrumental variable setup is somewhat analogous to the Bartik instrument (Bartik 1992) in that is shares the shift-share component structure. That is, we project the citywide aggregate subway development to the connectivity of location pairs, with their spatial distance being used as weights.[10]

The relevance of the instrumental variable in our context reflects the notion that pairs of far-apart grids will have a greater exposure to the subway expansion, given greater benefits from a larger reduction in travel time. If one were to consider an extreme opposite example of two adjacent grids, travel time would be irrelevant to the subway expansion. More generally, we argue that the Euclidean distance serves as a reasonable exposure measure in our context given the particular layout of subway lines and the general economic activities in Beijing. Namely, Beijing's subway is known for its "very orderly and grid-like" layout (Ingraham 2015).[11] To illustrate our point, we consider two types of layouts in the top panels of Figure 5. Dots, thin lines, and thick lines represent locations, roads, and subway lines, respectively. Example 1 has an orderly and grid-like layout compared to that in Example 2. Under the assumption that the subway reduces travel time by one fourth of the time it takes to travel by road, we compute the relationship between Euclidean distance and reduction in travel time due to subway operation for all 36 combinations of grid pairs given 9 location points. The corresponding bar graph for Example 1 shows that the average percent reduction in travel time increases with the Euclidean distance, whereas such relationship does not hold in Example 2. Indeed,

[9] We also construct an alternative leave-one-out aggregate shift to address the fact that including own observation may cause finite sample bias.

[10] One difference is that the Bartik instrument uses the inner product structure of an endogenous variable based on accounting identities (Goldsmith-Pinkham et al. 2020) and our instrument is not based on the inner product of different vectors.

[11] Source: `https://www.washingtonpost.com/news/wonk/wp/2015/01/16/quiz-can-you-name-these-cities-just-by-looking-at-their-subway-maps/`

given the orderly and grid-like layout of Beijing subway, we find that the magnitude of reduction in travel time between 2000 and 2018 increased with Euclidean distance for all grid pairs used in our empirical analysis (see Figure 6). Therefore, the Euclidean distance serves as a reasonable exposure measure in our context for Beijing, although that may not necessarily be the case in other cities.

One important identification issue that needs to be addressed carefully is that our exposure measure based on the Euclidean distance, despite being exogenous in nature, could be non-random. For example, Borusyak and Hull (2021) show that composite variables consisting of exogenous shocks but non-random exposure measures may still introduce omitted variable bias in either a reduced-form or an instrumental variable setting. We argue that our endogeneity concern on the selection of location pairs is unlikely to be correlated with the Euclidean distance between those locations.[12] However, exposure could be non-random in our context if grid pairs with large (small) Euclidean distance have a tendency to be in peripheral (center) areas of the city. Then the identification worry is that such spatial pattern may affect collaborative innovations through other confounding channels. To address such concerns, we investigate whether there is any evidence in the data that may compromise the exclusion restriction by adopting the following two approaches.

First, we take advantage of the pre-period present in our experimental design and follow Goldsmith-Pinkham et al. (2020) to test for parallel pre-trends as a way to assess the exogeneity of the exposure measure. Specifically, we interact the Euclidean distance between grid pairs with pre-period dummies up to 6 years prior to the treatment, while controlling for grid pair fixed effects and year fixed effects. Figure 7 shows that grid pairs with varying degrees of distance apart do not exhibit significantly different trends in collaborative patents prior to the treatment, suggesting our exposure measure is likely exogenous. Second, we investigate the relationship between the centrality of grid pairs and their Euclidean distance. For each grid pair, we calculate the Euclidean distance between each one of the two grids and the Tiannanmen Square (i.e. the central location of Beijing), respectively, and compute their average distance. Then we plot the kernel density of average distance to the city center for grid pairs, for different quintiles created based on the grid pair-specific Euclidean distance in the left panel

[12]If one were to consider potential agglomeration forces, nearby locations could be more likely to be selected at the same time to better internalize external benefits. In this case, the predictions would go against our argument as we claim that the *far-apart* grid pairs are more likely to be affected by the subway system expansion. To the extent that we show far-apart grids are associated with larger travel time reductions in the first stage and more collaborative matches between grids in the reduced-form, it suggests that the hypothetical scenario is unlikely to be true or at best contributes to a downward bias.

of Figure 8. We find that the 5th quintile seems to exhibit a centrality measure that is different from the rest, suggesting that the grid pairs with very large Euclidean distance are more concentrated in peripheral areas. The right panel of Figure 8 further breaks down the 5th quintile into the 9th decile and the 10th decile. Based on such evidence, we proceed with dropping either the top 10% or 20% of observations with large Euclidean distance in one of our robustness checks in Table 7.

Lastly, one additional caveat with the instrumental variable approach is that, in the event of heterogeneous causal responses, the two-stage least squares estimates produce the local average causal response. Depending on the extent of heterogeneity, the local average causal response could be very different from the average causal response which is of more general and inherent interest. Nevertheless, if the concern is on the potential endogeneity which may overstate the presence of the causal response, then the local average causal response at least identifies the presence of the causal response among the compliers. In our setting, the "complying grid pairs" are likely to be those that are far apart from each other based on the design of the instrument. As we expect the causal response to be mainly manifested through the far-apart grid pairs, the instrumental variable approach corrects the bias of the estimated average causal response for such far-apart "complying grid pairs."

4. Data and Variables

4.1. Data Overview

Our empirical analysis relies on two primary datasets. The first dataset contains information on Beijing-based patents from the China National Intellectual Property Administration (CNIPA). For each patent, we have unique patent identifier number, dates of application and publication, International Patent Classification (IPC) codes, names of applicant(s) and inventor(s), and an address. The second dataset contains information on subway stations in Beijing. From the Wikipedia, we track detailed geographic network of the Beijing subway system and its intertemporal development process.[13] For each subway station, we know its coordinates and opening year, as well as connectivity between stations at all phases of development. Our sample period is from 2000 to 2018, during which the Beijing subway system expanded rapidly.

[13] `https://en.wikipedia.org/wiki/Beijing_Subway`. The exact timing of each station opening during our sample period is summarized in the Supplementary Materials.

To better measure the implications of the subway expansion on collaborative matches in innovations, we focus on patent *applications*, as opposed to approved patents. The reasons are two fold. First, although not all applications end up being approved, submitting an application is the initial crucial step in seeking protection for innovations. It also signals the formation of a collaboration pair in the knowledge creation process. Second, there is usually a long and irregular delay before applications are finally approved. The application time better indicates when new knowledge was created and formalized.[14] We thus rely on the time when patent applications were submitted in drawing inference with the subway build-up time in our empirical analysis. Throughout this paper, we will refer "patent applications" as "patents" for short.

Table OA1 shows the summary statistics. During our sample period, there are approximately 42,877 patents in a year and the average number of collaborated patents in a year is 9,223. When we analyze 785,629 patents applied during 2000-2018, the average application year is 2013. This suggests that relatively more patents were applied towards the latter years of our sample period. The average number of applicants per patent is 1.283 and the collaborated patents contain a team of approximately 2.271 applicants on average. The last two sets of rows present summary statistics at particular geographical units, which we refer to as grids. The average logarithm of the Euclidean distance (*km*) between grids in a year is approximately 2.340 and the logarithm of the length of the subway lines (*km*) that connect two grids along their shortest-time travel path is 8.352 on average.

4.2. Location Grids

Our empirical analysis requires spatial units that correspond to a discrete set of locations. We thus divide the city of Beijing into 65,542 grids where each grid is 0.5 *km* by 0.5 *km*. For each of the subway stations, patents, and collaborators in our data, we assign corresponding grids using coordinates. For each year during our sample period, we track the presence of subway stations as they are built and added to the subway system. We also assign grids at the patent level, using the reported address for each patent application in our database. Lastly, our analysis also requires us to further track grids for each patent at the collaborator level. We use applicants as collaborators and assign the corresponding grid for each applicant in a patent. In the Online Appendix B, we provide a detailed

[14] This is also commonly followed in the literature, such as in Moretti (2019).

outline on how we assign applicants to grids using information in the patent database.

4.3. Collaborations

The main outcome of interest is patent collaborations across time and space. As the patent database lists the names of all applicants for a patent, we can easily identify which patents were produced by a collaborative team. We identify a patent with multiple applicants as a "collaborated patent," and that with a single applicant as a "solo patent."

To analyze spatial patterns of patent collaborations over time, we construct the key dependent variable, $CollabPatents$. This variable is grid pair- and year-specific. It is the total count of patents whose collaborators were located in each one of the paired grids in a given year. For collaborated patents, we track all possible pairwise collaborator combinations at the grid pair-level. For example, suppose a patent has 3 collaborators located in grids A, B, C, respectively, in year 2010. Then this patent would be counted in the variable $CollabPatents$ for the grid-pairs (A, B), (B, C), and (A, C) in year 2010. Given that we have 65,542 grids in total, the number of grid-pairs becomes exponentially large if we were to consider all possible combinations in each year. Thus, we construct $CollabPatents$ using only the grids that were ever located within 2 *km* from a subway station during our sample period in our baseline estimation. We later report spatial evidence as to why we chose 2 *km* as the cutoff in Section 5.1. In addition, we also check robustness of our main findings using alternative distance cutoffs and report them in Table OA4 in the Online Appendix.

4.4. Length and Travel Time

The key determinant of patent collaborations that we are interested in is the connectivity of a pair of grids in a year. As new subway stations and lines are built, the connectivity of collaborators between different grids is likely to improve over time. We capture this change using our key explanatory variable, $Length$. It measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. For instance, when the subway system was sparse in early years, the fastest route along which one travels from one grid to another likely involved no subway travels. In this instance, the $Length$ variable takes value 0. If in later years, more subway lines are laid up in-between those two grids and the fastest travel route between those two grids now involve 5 km of subway travel, the $Length$ variable takes value 5. We use the Geographic Information System (GIS) software to identify the route that minimizes the travel time between a pair of grids in a year based on different travel speed assumptions.

Although we use $Length$ as the key explanatory variable in our reduced-form regressions, we also consider an alternative variable, $TravelTime$, as a first-stage outcome. This variable captures the amount of time required to travel from the centroid of one grid to that of another grid in a given year along the fastest route. Apparently, this variable is also the object to minimize when we construct the $length$ variable. In calculating the $TravelTime$, we restrict the mode of transportation to only subway travels (to travel from one station to another) and offline travels (to travel between centroid of a grid to a station, or directly between centroid of the two grids by walking, cycling, or riding a bus). Although it would be most ideal to calculate travel time based on various combinations of transportation modes, we make this simplifying assumption due to lack of data on complete time-varying transport networks that allow for all modes of travel. We impose the assumption of 36 km/h for the subway speed and consider the range of 8 km/h to 12 km/h for the offline travel speed.[15] We further assume that individuals travel along the Euclidean distance between the centroid of a grid to the corresponding subway station, for both grids in a pair.[16] In our empirical analysis, we report results based on different travel speed assumptions.

While the calculation of travel time is sensitive to the imposed speed assumptions, the selection of the fastest travel route along the subway lines is highly robust. Less than 0.5% of the travel routes change when we vary the offline travel speed from 8 km/h to 12 km/h. Thus, we select the fastest travel route based on the 10 km/h offline speed and use the corresponding $length$ along this route as our key variable in a reduced-form setting. For this reason our reduced-form evidence (i.e. the impact of $length$ on $CollabPatents$) is more robust and reliable, but we also report the first-stage results (i.e. the impact of $length$ on $TravelTime$) to facilitate a more structural interpretation.

Figure 9 shows how segmented distance and travel time along the fastest route have changed

[15]We calculate the mean subway travel speed based on the 2018 train travel time averaged across all pairwise stations. We use 36 km/h for the subway speed despite that the speed of newly built subways is higher than that of old ones (`https://www.chinadaily.com.cn/china/2016-12/27/content_27792438.htm`). We take 8 km/h to 12 km/h as a reasonable range to capture the walking speed (or the average speed of bus travels or cycling in more realistic terms) to reach the stations along a straight line.

[16]The actual walking distance may differ from the Euclidean distance, given the layout of roads. Yet, the Euclidean distance can serve as a reasonable proxy.

over time with the expansion of the subway system. Panel (a) shows the average *Length* in the data increased over time, as more lines were built and connectivity improved. At the same time, the distance between the centroid of a grid to the corresponding station diminished for grids in a pair, as more stations were built. Given such changes, Panel (b) shows that the average travel time also decreased considerably over time.

5. Results

5.1. Initial Evidence at the Grid-Level

Although our focus is on patent collaborations at the grid-pair level, we begin by conducting a grid-level analysis to capture some initial evidence on whether subway station openings had an impact on overall patent growth. In Table OA2, we report results obtained from the standard TWFE model and the staggered DiD model as in Callaway and Sant'Anna (2021) using the never-treated as controls. The binary indicator $\mathbb{1}$(Treated) equals one starting from the year of opening of a subway station in a grid and zero otherwise. In Columns (1), (3), and (5), the dependent variable is a binary indicator that captures the existence of any collaborated patents, total patents, and solo patents in a grid in a year, respectively. In Columns (2), (4), and (6), the dependent variable is the actual count of patents to capture the magnitude.[17] Across all specifications, we find that the magnitude of estimates are quite comparable between the two methods.[18] Based on the DiD approach in Panel B, the ATT on collaborative patents is 19.3% and the ATT on total patents is 28.2%. In Figure 3, we plot the event study estimates based on Callaway and Sant'Anna (2021). In the left (right) panel, the dependent variable is collaborated patents (total patents) in its logarithmic-like inverse hyperbolic sine transformation. In both panels, the grid-specific patent counts change sharply with the opening of a subway station, which is in line with our assumption that the timings of the subway station openings

[17]Throughout our manuscript, whenever we have logarithm of patent count as a dependent variable, we use the logarithmic-like inverse hyperbolic sine transformation to address observations with zero values. Such transformation is a concave log-like transformation and allows retaining zero-valued observations. The coefficient estimate yields a similar interpretation to that of a standard logarithmic specification. See Bellemare and Wichman (2020) for a formal proof.

[18]The TWFE estimator is essentially a weighted average of pairwise DiD estimates based on a particular set of weights assigned to group-time cohorts (Goodman-Bacon 2021). The weights are variation hungry and are not the standard weights used for the ATT. Bias may also arise when the later treated cohorts are compared to "controls" which were treated earlier and potentially contaminated by treatment dynamics. Despite those concerns, we obtain very similar magnitude which suggests that the issue-prone comparisons may not carry high weights when generating the TWFE estimates.

are exogenous. All coefficients during the pre-treatment period are insignificantly different from zero, which supports the parallel trend identification assumption.

In addition, we conduct a complementary analysis at the grid-level to assess the spatial scope of the subway impact. In our grid-pair analysis in the following subsection, constructing every possible grid-pair using the sample of all grids is computationally infeasible as the number of possible combinations becomes exponentially large. Therefore, we identify spatial attenuation patterns and use the results to narrow down the number of grids that are used in constructing the grid-pair sample. More specifically, we estimate the following specification:

$$log(Patents)_{it} = \sum_{r=0}^{6} \rho^r \mathbb{1}(StatOpen)^r_{it} + \mu_i + \delta_t + \varepsilon_{it}, \tag{5.1}$$

where subscripts i and t denote grid and year, respectively. The superscript r denotes distance rings based on various distance cutoffs.[19] The dependent variable is the total patent counts at grid i in year t, in its logarithmic-like inverse hyperbolic sine transformation. $\mathbb{1}(StatOpen)^r_{it}$ is a binary indicator that takes value 1 if t is greater than or equal to the year when the first station in ring r comes into operation, and 0 otherwise. δ_t, μ_i, and ε_{it} represent year fixed effects, grid fixed effects, and the idiosyncratic error term, respectively. Figure 10 plots the corresponding coefficients ρ^r's along with their 95% confidence intervals. We find spatial attenuation patterns, with the impact of subway station openings on patent counts being positive and statistically significant up to ring 3 (i.e. approximately within 2 *km*) and such effects dying out beyond 2 *km*.[20] In the grid-pair analyses that follow, we thus use the sample of grid-pairs constructed based on all grids that were ever located within 2 *km* from a subway station during our sample period.[21]

5.2. Patent Collaborations at the Grid Pair-Level

We now discuss our main findings on patent collaborations based on evidence at the grid-pair level. We define a binary indicator $\mathbb{1}$(Treated) equals one starting from the year when there is a reduction in travel time by more than 30 minutes between a given pair of grids along the shortest

[19]There are 7 distance rings: ring 0 if the opening station is at the same grid; ring 1 if it is at 0-0.5 *km*; ring 2 if it is at 0.5-1 *km*; ring 3 if it is at 1-2 *km*; ring 4 if it is at 2-4 *km*; ring 5 if it is at 4-7 *km*; and ring 6 if it is at 7-10 *km*.

[20]The coefficient of ring 6 turns to be slightly negative, which potentially suggests that there is spatial reallocation of resources. We explore the channel of spatial reallocations later in a different set-up.

[21]In Table OA4, we find that out main findings are robust to changing the 2 *km* cutoff.

travel route and zero otherwise.[22] Panel A of Table 1 shows the estimates from the TWFE model. As for the dependent variable, we use a binary variable (continuous variable) to capture the existence (magnitude) of collaborated patents in Columns (1) and (2) (Columns (3) and (4)). The TWFE model estimates show that the reduction in travel time between a pair of grids increases both the probability of collaborated patents being created and the extent of such collaborated patents.

Panel B of Table 1 shows the estimates from the staggered DiD model as in Callaway and Sant'Anna (2021), using the never-treated cohort as the control group. We find the ATT to be positive and statistically significant for both the probability of having collaborated patents and the magnitude of collaborated patents. The ATT on collaborative patents is 2.94% and is statistically significant at 1%. As in the grid-level analysis, we again find that the TWFE and DiD estimates are quite comparable. In Figure 4, we plot the even-study estimates. To check the robustness of our treatment definition, we use the cutoff of reduction in travel time by more than 30 minutes (1 hour) in the left (right) panel. Both panels show supporting evidence for the parallel trend assumption.

Next, instead of using a binary treatment status, we adopt a continuous measure and estimate the specification in equation (3.1). The continuous measure *Length* captures the total length of the subway lines that connect two grids along their shortest-time travel path. Conditional on year fixed effects and grid fixed effects, Table 2 shows that a longer *Length* between a pair of grids significantly decreases the probability of collaborations in Columns (1) and the count of collaborated patents in Column (3). Consistent with our prior, the probability and extent of collaborations decreases as the distance of the fastest subway travel increases in the absence of grid-pair fixed effects. Once we control for the grid-pair fixed effects in Columns (2) and (4), we find that a larger *Length*, which implies greater connectivity due to more subway lines and stations being built, increases both the probability and extent of collaborations.

For ease of interpretations, we report the Wald estimators in Table 2 that show the extent to which an hour reduction in travel time affects the probability of patent collaborations in Columns (1) and (3), or the count of collaborated patents in Columns (2) and (4). Such estimates are derived based on the first-stage estimation of how subway expansions affect travel time, which is reported in Table OA5. The Wald estimators in Column (4) suggest that an hour reduction in travel time increases

[22] Since the control pairs include those that presumably also experience an improvement in connectivity but not as substantial as the treatment pairs, we identify a differential effect here.

collaborated patents by approximately 15% to 38% on average, depending on the offline travel speed assumptions.

To corroborate our findings, we further instrument for *Length* using the interactions of Euclidean distance for a given grid pair and the total cumulative length of subway lines in Beijing. Note that the instrumental variable approach identifies the local average causal response subject to the identifying assumptions (Imbens and Angrist 1994) when the extent of the returns from enhanced connectivity is heterogeneous across grid pairs. As an initial check, we first repeat our TWFE analyses but classify grid-pairs into either 'near" or "far" samples, using the mean Euclidean distance between grid-pairs as the cutoff. Results in Columns (1), (2), (5), and (6) of Table 3 show that the effects seem heterogeneous; the far grid-pairs experience a larger reduction in travel time and thus greater increase in collaborative patents than that of near grid-pairs.

In Columns (3), (4), (7), and (8) of Table 3, we present the local average causal response obtained from the IV approach. Note that in Columns (4) and (8), we alternatively construct a leave-one-out (LOO) aggregate shift since including own observation may cause finite sample bias. We find that improved connectivity significantly increases both the probability and extent of patent collaborations. More specifically, Column (8) shows that an hour reduction in travel time increases the number of collaborated patents by 31% to 73% on average, depending on different offline speed assumptions.

Lastly, we implement the control pair method proposed in Jaffe et al. (1993) as a robustness check. We examine whether a collaborator pair is more likely to form once connectivity improves with the subway expansion. The idea is to match the *true* collaboration pairs to *control* collaboration pairs based on observable characteristics of the patents they produced. For matching, we use application year and the first-listed IPC at the 3-digit level to generate control pairs. In Online Appendix E, we elaborate on the procedure that we follow. If potential endogeneity concern resides on the selection on observables, the matching procedure allows us to address such concern. Using matched *control* collaboration pairs, we estimate a linear probability model specified as the following:

$$\mathbb{1}(CollabPatents)_{ijpt} = \beta \log(Length)_{ijpt} + \theta_{ij} + \lambda_t + \xi_{ijpt}, \tag{5.2}$$

where $\mathbb{1}(CollabPatents)_{ijpt}$ is an indicator variable that equals 1 if a collaborator pair p is a *true* pair and equals 0 if it is a *control* pair; $Length_{ijpt}$ is a proxy for the extent of subway build-up between grids i and j in which the matched collaborators reside; θ_{ij} represents grid-pair fixed effects; λ_t represents year fixed effects; ξ_{ijpt} is the idiosyncratic error. Note that here we examine at the collaborator-pair level as opposed to the grid-pair level. The identification assumption is that, conditional on the matched characteristics of collaborated patents and location-pair and year fixed effects, the conditional independence assumption is satisfied (i.e. $\text{cov}(Connect_{ijpt}, \xi_{ijpt}) = 0$).

In Columns (1) and (2) of Table 4, we create one control pair for every actual pair (i.e. "One-to-One Match"). In Columns (3) and (4), we create two control pairs for every actual pair. We find that the results are robust regardless of whether we use one-to-one match or one-to-two match. Without controlling for grid pair fixed effects in Columns (1) and (3), applicants in a pair are less likely to collaborate as *Length* between their locations increases, compared with control applicant pairs. However, once we further control for grid pair fixed effects in Columns (2) and (4), enhanced connectivity captured by a longer *Length* improves the collaboration probability. These results are consistent with our previous findings.

6. Mechanism Analysis

6.1. Discussion of the Mechanisms

Online Appendix F presents a conceptual framework, summarized here, on specific channels through which enhanced subway connectivity increases collaborative matches in innovations. In the model, a city consists of N locations and a unit of potential inventors who are mobile across locations. To maximize utility, they first decide whether to innovate and where to locate. Then they learn their productivity levels and further decide whether to collaborate with other inventors. The model equilibrium generates an endogenous distribution of inventors and the extent of pairwise collaborations across locations. We solve the model following backward induction.

We assume that a pair of inventors, who are characterized by their productivity levels respectively, in two different locations encounter one another randomly. The encounter probability depends on the density of inventors in each of the two locations for a given level of productivity, and location pair-specific degree of interactions. Upon encountering, the probability of having a successful collaboration depends on inventors' productivity and travel time between their locations. Having a shorter travel time increases the likelihood of having a successful collaboration since inventors can have more frequent face-to-face interactions, which facilitates their coordination on exchanging information, skills, and ideas. Under the super-modularity assumption, the probability of forming successful collaborations increases more among highly productive inventors, as travel time decreases.

An inventor's location decision is determined by location-specific returns to collaborative and solo innovations. Expected returns to solo innovations at a location are higher if the location is easily accessible to other locations with a large measure of inventors. Similarly, expected returns to collaborated innovations also increase with connectivity of a location. Given the aggregate expected profits from collaborative and solo innovations and idiosyncratic preferences, utility-maximizing inventors choose either to locate in one of the N locations, or to simply take an outside option and not innovate.

The equilibrium gives the distribution of inventors across locations, and we can further solve for the closed-form solution by adopting certain functional form assumptions. Under characterization of the model, we derive the following channels to shed light on different mechanisms that give a rise to collaborative matches when travel time decreases.

High-Quality Matches Channel. – As travel time decreases, more collaboration pairs are formed due to the complementarity between travel time and inventors' productivities. In particular, more collaborative matches are formed among highly productive inventors.

Marginal Matches Channel. – A reduction in travel time causes more collaboration pairs to be formed on the margin. These are the low-value matches, which would not have been formed without a reduction in travel time.

Relocation Channel. – As a location becomes more accessible, it attracts inventors from elsewhere due to greater expected profits from collaborative and solo innovations. It attracts highly productive inventors in particular, and collaborations there will increase.

Marginal Inventors Channel.– When a location becomes more accessible, some of the potential inventors who previously chose the outside option will no longer be screened out. As these marginal inventors with relatively low productivity join the pool of active inventors, more collaborative matches are formed.

6.2. Empirical Evidence

Whereas the aforementioned four mechanisms are likely to coexist, we want to assess which one is quantitatively more important in our empirical context. Given that each mechanism places different emphasis on various subgroups of inventors, we use such predictions to guide our empirical analysis.

We first distinguish the *Marginal Inventors Channel* from the rest three channels. Whereas the other channels all commonly predict increased collaborations among *existing* (i.e. incumbent) inventors, the *Marginal Inventors Channel* highlights the role of *new inventors*. If we find that most collaborations are formed between existing inventors, then the *Marginal Inventors Channel* is not likely to be the main driving force.

Next, among existing inventors, we now distinguish the *Relocations Channel* from the *High-Quality Matches Channel* and the *Marginal Matches Channel*. Whereas the *Relocations Channel* emphasizes inventors moving to more accessible locations, the other two channels identify inventors remaining in their previous locations. If we find that most collaborations involve inventors who did not relocate, then the *Relocations Channel* is unlikely to be the main channel.

Lastly, among existing inventors who did not relocate, the *High-Quality Matches Channel* predicts that collaborations will be mostly driven by pairs of highly productive inventors, whereas the *Marginal Matches Channel* predicts the exact opposite. Therefore, depending on which productivity type of inventors become more active in forming collaboration pairs, we can further identify the dominating force of the two mechanisms.

In Tables 5 and 6, we carry out analyses as outlined above by looking into different sub-samples. In both tables, there are Panels A, B, and C. In Panel A, we first classify collaborated patents into those that were created by a pair of inventors who were previously both active (i.e. incumbent pairs) versus those created by pairs involving at least one new inventor. Table 5 presents the results from the TWFE and DiD model using the binary treatment variable and Table 6 shows results based on the IV approach using a continuous $Length$ measure. In Panel A of both tables, we find that the effect of enhanced connectivity is much more pronounced for incumbent pairs. This implies that the role of new inventors, as emphasized by the *Marginal Inventors Channel*, is less likely to be the dominant underlying force for the increase in collaborative matches.

In Panel B of Tables 5 and 6, we classify the sample of collaborated patents into those that were created by pairs of nonmovers versus pairs consisting of movers. Mobility of each applicant is identified by tracking address in the patent data across years. The nonmover pairs consist of all pairs of collaborators whose locations remained the same as before in a given year.[23] Panel B of Tables 5 and 6 shows that there was a greater increase in collaborated patents by nonmover pairs and thus the *Relocations Channel* does not seem to be the most dominant channel.

Lastly, in Panel C of Tables 5 and 6, we focus on collaborated patents created by nonmover pairs but further classify the sample into two groups based on the productivity type. In each year, we first track the cumulative count of patents created by each applicant. Then using the median value of the distribution in a year as the cutoff, we assign each applicant's productivity type as either high or low. In Columns (1) and (2), we use the sample of collaborated patents created by nonmover pairs who were both high types. Columns (3) and (4) contain the sample of collaborated patents created by nonmover pairs which include at least one applicant of the low type. The magnitude of effect from enhanced connectivity is much larger on collaborations among the H-H nonmover type, and thus the *High-Quality Matches Channel* seems to be the quantitatively most dominant channel.

To quantitatively compare the relative importance of the channels, we now classify all collaborative matches into one of the four exclusive categories: movers, entrants, H-types, and L-types. The movers (entrants) category consists of all collaborative matches for which at least one applicant was a mover (an entrant) in that year. Among non-movers and incumbents, the H-types category consists of all pairs that involve both applicants of the high type in that year, while the L-types category consists of pairs that involve at least one low type in that year. The categories of movers, entrants, and H-types correspond to the *Relocations Channel*, the *Marginal Inventors Channel*, and the *High-Quality Matches Channel*, respectively. However, the L-types category is slightly broader than what the *Marginal Matches Channel* represents, because it also includes patent collaborations among the low-productivity types that are not necessarily formed on the margin.

Table 8 reports our findings for each group. As before, we consider a treatment to exist if the travel time reduces by more than 30 minutes along the fastest travel route. We adopt separate treatment indicators for the short run (within 5 years) and long run (after 5 years), since the effect can differ across time. In Panel A, we find that the short run responses are only present among movers and

[23]In Online Appendix C, we elaborate on how we track movers and nonmovers in the data.

H-type non-movers, with the effect being more pronounced among the H-type non-movers. In the long run, however, a positive impact exists across all subgroups. That is, all four channels coexist in the long run, while the *High-Quality Matches Channel* contributes to approximately 75.65% of the total increase in patent collaborations in the long run. One may raise the possibility that some types of applicants face mobility or institutional restrictions and this may underestimate the relative importance of the *Relocation Channel*. In Panel B, we thus drop all pairs that involve universities and repeat our analysis to see whether this is the case. We find that restricting the sample in this way does not change our results much and our findings remain robust.

7. Back-of-the-Envelope Calculation

This section presents a simple back-of-the-envelope calculation. The goal is to compare the economic significance of our estimates with other benefits of subway infrastructure presented in the recent literature.

Gu et al. (2019) study the impact of subway build-up on congestion and travel time. They show that the subway system increases the road speed in Beijing by about 3%, which implies a 1.68 minutes decrease in average commuting time. Taking into consideration the monetary value of commute time and the number of commutes, the Beijing subway system brought total welfare increase of 1.6 billion yuan (249 million US dollars) per year due to faster commutes in 2016. Gendron-Carrier et al. (2018) study the impact of subway build-up on pollution reductions and show that, on average, a subway opening reduces about 2.0 $\mu g/m^3$ of PM10. According to Gendron-Carrier et al. (2018), studies in the literature find that a 1 $\mu g/m^3$ decrease in ambient PM10 averts about 10 infant deaths per 100,000 births. If we were to extrapolate this estimated effect to the context of Beijing with 0.8% of birth rate and 21 million regular residents, the subway system should avert about 33.6 infant deaths per year which translates into the value of averted infant deaths of 75 million US dollars per year.[24]

We document that an hour reduction in travel time increases patent collaborations by 35 to 82 percent, depending on travel speed assumptions. During our sample period, the average travel time between grid pairs decreased by 1.30 to 0.77 hours, which translate into increased patent collaborations

[24]The Beijing demographics are taken from `https://knoema.com/atlas/China/Beijing/Birth-Rate`. In addition, we take Viscusi and Aldy's (2003)'s estimate of 0.6 elasticity of the value of statistical life with respect to income, and impute Beijing's value of statistical life in 2019 US dollars from the U.S. value of $6 million in 2000 US dollars.

by 46 to 63 percent. Taking into consideration the average number of patent collaborations (116,654), the Beijing subway expansion during our sample period led to 53,661-73,492 patent collaborations, which are roughly equivalent to 22,909-31,376 approved collaborative patents after adjusting for the ratio of pairwise collaborations to total collaborative patents and the patent approval rate.[25] We take the estimates in Wei et al. (2021) on patent values in Beijing (3,291 - 6,191 yuan in 2008 values) and extrapolate into our setting based on the median of the range converted into 2019 values.[26] The corresponding values for the increased collaborative patents in our setting are roughly 143 million to 195 million yuan, which correspond to about 21 million to 28 million US dollars.[27]

A few caveats to keep in mind. First, the simple back-of-the-envelope calculation is based on the estimates of the local average causal response. To the extent that the local average causal response may be severely different from the average causal response, our calculations may be biased. Second, we took the estimated patent values in Beijing from Wei et al. (2021) to impute the values of collaborative patents. Despite that their study also relies on patents in Beijing during a similar sample period, Wei et al. (2021) focus on eight industries targeted by a Chinese pro-innovation subsidy program, whereas we focus on all industries in our setting.[28] Third, our estimates are based on contemporaneous estimates that do not account for long-term dynamic responses. Given the dynamic nature of innovation activities, the long-term implications are likely to be substantially larger.

8. Conclusion

In this paper, we analyze the extent to which development of intra-city transport infrastructure affects collaborative matches in innovations. Using rapid expansion of the subway system in Beijing from 2000 to 2018, we investigate the impact of enhanced connectivity on innovation activities and

[25] We calculate in our data that the ratio of total collaborated patents to the total number of pairwise collaborations is 0.6095. We calculate the collaborative patent approval rate as 71.15% based on sample between 2000 and 2013. We do not use the years after 2013 as it can take a few years for the approval status to be determined. Our sample stops in 2018 and patents approved afterwards are not captured. Given that we observe a steady patent approval rate until 2013 and a sharp decline afterwards due to sample truncation, we decide to use sample between 2000 and 2013 to calculate the average approval rate for the collaborative patents.

[26] The Reminbi inflation factor is taken from `https://www.in2013dollars.com/china/inflation`.

[27] The average exchange rate, 0.1448, is taken from `https://www.exchangerates.org.uk/CNY-USD-spot-exchange-rates-history-2019.html`.

[28] The eight industries are the pharmaceutical manufacture (CSIC 27), the special equipment manufacture (CSIC 36), the transportation equipment manufacture (CSIC 37), the communication equipment & computer manufacture (CSIC 40), the precision instrument manufacture (CSIC 41), computer service (CSIC 61), software service (CSIC 62) and environmental protection industry (CSIC 80).

spatial collaboration patterns. Our implementation of various research designs shows that the buildup of subway system facilitated patent collaborations considerably. Collaborated patents increased by 15% to 38% with an hour reduction in travel time, depending on travel speed assumptions. Far-apart locations benefited more from the buildup of subway, with the local average causal response being approximately 35% to 82%. A further investigation of the underlying mechanism shows that the increase in collaborative matches is largely driven by more matches among highly productive inventors due to positive assortative matching and the complementarity between productivity and connectivity. At the same time, entry of new inventors, relocation of existing inventors, and low productive inventors also contribute to the increase in collaborative matches in the long run.

Whereas this paper establishes an important link between connectivity and collaborative matches in innovations, there are several limitations and possible extensions to consider. First, we currently focus on the quantitative aspect of the outcome, as measured by the count of collaborated patents. Using data on patent citations, it would be interesting to further investigate whether highly cited patents with a greater impact on subsequent innovation activities are sparked by enhanced connectivity. Second, instead of looking at all possible pair-wise combinations of collaborators, one can further consider analyzing the network of collaborators. This can provide information on whether any systematic relationship exists between innovators who are at the center of a network and the degree of their accessibility based on their physical location. Lastly, as the gains from reduced travel time that we document speak directly to the underlying rationalization of agglomeration and returns to urban density, it would be interesting to directly model agglomeration economies and estimate their effect.

References

Acemoglu, D. (2008). *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica 83*(6), 2127–2189.

Bartik, T. J. (1992). *Who Benefits from State and Local Economic Development Policies?*, Volume 68.

Baum-Snow, N. (2007). Suburbanization and Transportation in the Monocentric Model. *Journal of Urban Economics 62*(3), 405–423.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, Q. Zhang, V. Henderson, M. A. Turner, and Q. Zhang (2017). Roads, Railroads and Decentralization of Chinese Cities. *Review of Economics and Statistics 99*(3), 435–448.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy 122*(3), 507–553.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61.

Borusyak, K. and P. Hull (2021). Non-Random Exposure to Exogenous Shocks: Theory and Applications. *NBER Working Paper Series*.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Carlino, G. and W. R. Kerr (2015). Agglomeration and Innovation. *Handbook of Regional and Urban Economics 5*, 349–404.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). Urban density and the rate of invention. *Journal of Urban Economics 61*, 389–419.

Catalini, C., C. Fons-Rosen, and P. Gaulé (2020). How do travel costs shape collaboration. *Management Science 66*(8), 3340–3360.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica 80*(6), 2543–2594.

Combes, P.-P. and L. Gobillon (2015). The Empirics of Agglomeration Economies. *Handbook of Regional and Urban Economics*, 247–348.

Davis, D. R. and J. I. Dingel (2019). A spatial knowledge economy. *American Economic Review 109*(1), 153–170.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996.

Dong, X., S. Zheng, and M. E. Kahn (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics 115*(November 2019).

Duranton, G. and D. Puga (2004). Micro-Foundations of urban agglomeration economies. *Handbook of regional and urban economics*.

Gendron-Carrier, N., M. Gonzalez-Navarro, S. Polloni, and M. Turner (2018). Subways and Urban Air Pollution. *National Bureau of Economic Research*.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Gu, Y., J. Zhang, and B. Zou (2019). Subways and Road Congestion. *SSRN Electronic Journal*.

Heblich, S., S. J. Redding, and D. M. Sturm (2020). The Making of the Modern Metropolis: Evidence from London. *Quarterly Journal of Economics 135*(4), 2059–2133.

Helpman, E. (1998). The Size of Regions. In Y. Z. D. Pines E. Sadka (Ed.), *Topics in Public Economics*. New York: Cambridge University Press.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics 108*(3), 577–598.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *Review of Economic Studies 76*, 283–317.

Kerr, W. R. and F. Robert-Nicoud (2020). Tech Clusters. *Journal of Economic Perspectives 34*(3), 50–76.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy 99*(3), 483–499.

Lee, J. and J. Xu (2020). Why do businesses grow faster in urban areas than in rural areas? *Regional Science and Urban Economics 81*(March 2018), 103521.

Moretti, E. (2019). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *Working paper*.

Neumark, D. and H. Simpson (2015). Chapter 18 - Place-Based Policies. In G. Duranton, J. V. Henderson, W. C. B. T. H. o. R. Strange, and U. Economics (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 1197–1287. Elsevier.

Ottaviano, G. I. P., T. Tabuchi, and J.-F. Thisse (2002). Agglomeration and trade revisited. *International Economic Review 43*(2), 409–436.

Redding, S. J. and M. A. Turner (2015). Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity. In G. Duranton, J. V. Henderson, W. C. B. T. H. o. R. Strange, and U. Economics (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 1339–1398. Elsevier.

Rosenthal, S. S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of regional and urban economics 4*, 2119–2171.

Sant'Anna, P. H. C. and J. Zhao (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics 219*(1), 101–122.

Sultana, S. and J. Weber (2016). *Minicars, Magleves, and Mopeds: Modern Modes of Transportation Around the World*. ABC-CLIO, LLC.

Yang, Z., J. Cai, H. F. L. Ottens, and R. Sliuzas (2013). Beijing. *Cities 31*, 491–506.

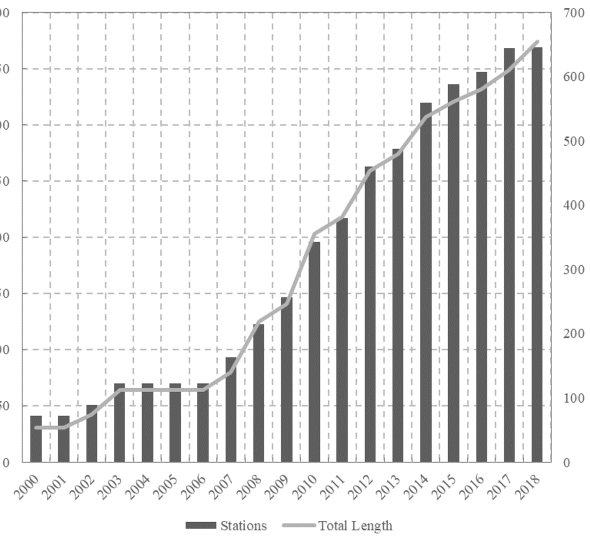
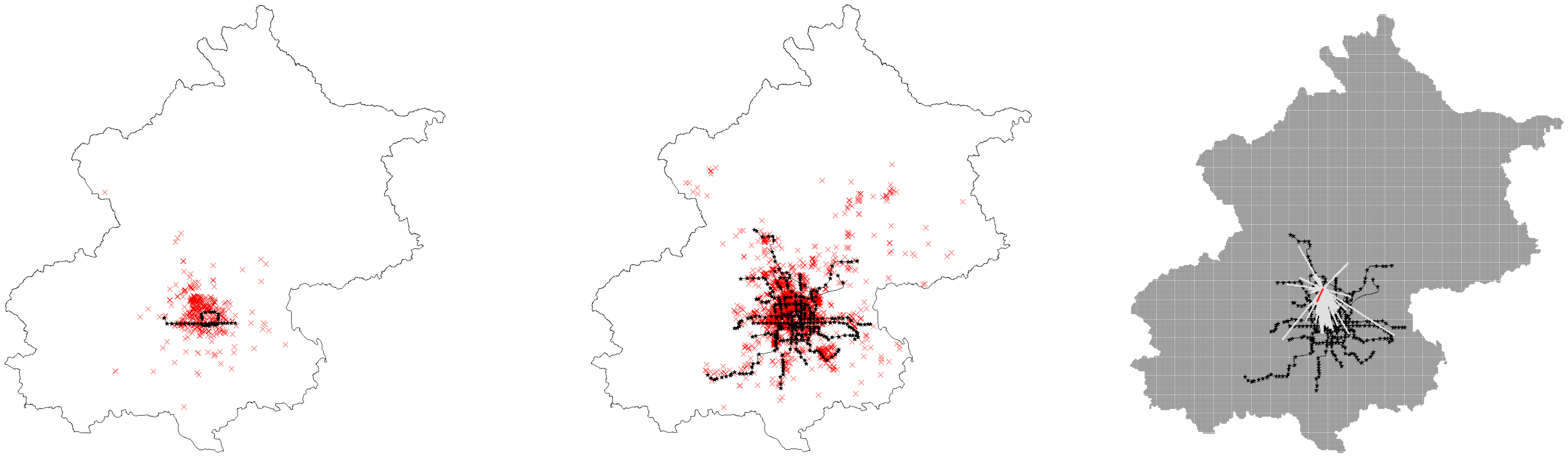


Figure 1: Cumulative Count of Subway Stations and Length of Subway Lines in Beijing

Notes: The cumulative count of subway stations in Beijing is indexed to the left axis and the total length of subway lines in Beijing is indexed to the right axis.



(a) 2000 and 2001 (b) 2017 and 2018 (c) Illustration

Figure 2: Spatial Patterns of Patents, Patent Collaborations, and Subway Stations in Beijing

Notes: Panel (a) shows the spatial distribution of patents applied in 2000 and 2001 (in red x-marks) and the subway stations available during the same period (in black star-marks). Similarly, Panel (b) shows the spatial distribution of patents and the subway stations in 2017 and 2018. Panel (c) shows collaboration links for the grid that had the largest number of collaborations during our sample period. The red line shows the unique collaboration link that was formed during 2000-2001 period, whereas the light gray lines show collaboration links formed during 2017-2018 period. The complete set of subway stations that came to operations during our sample period are shown in black star-marks.

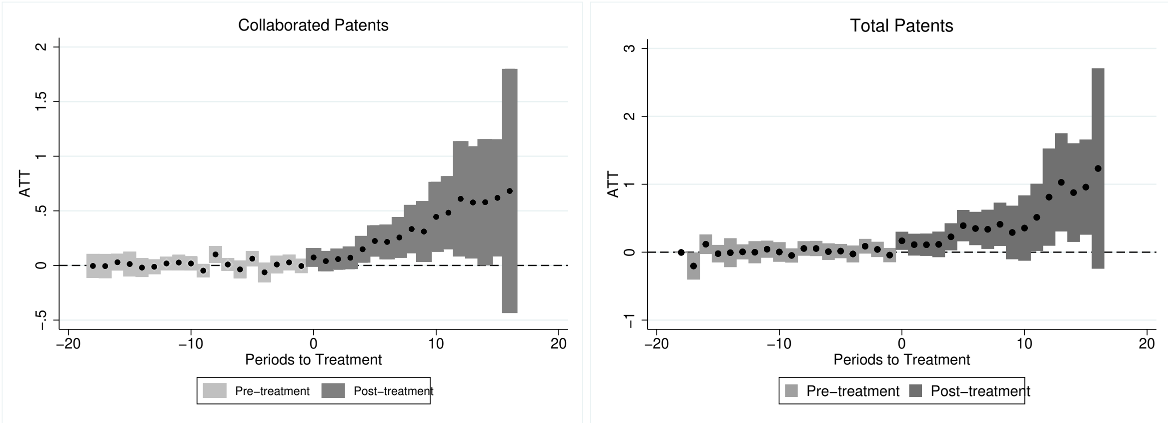


Figure 3: Time-Varying Effects of Improved Connectivity on Collaborated and Total Patents (Grid-Level)

Notes: The dynamic ATT's are plotted using the staggered DiD model, based on Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). The control group consists of the grids that were never treated during our sample period. In the left (right) panel, the dependent variable is collaborated patents (total patents) in its logarithmic-like inverse hyperbolic sine transformation.

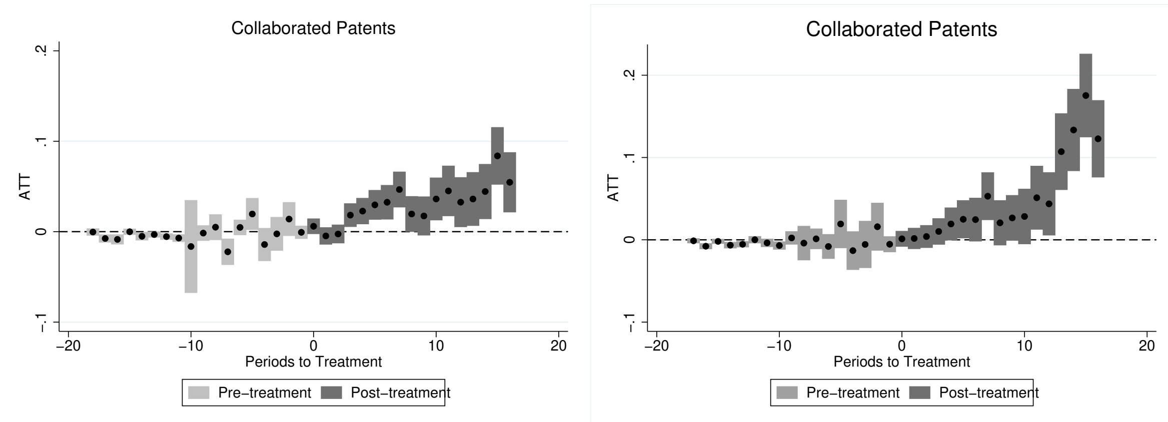


Figure 4: Time-Varying Effects of Subway Station Opening on Patent Collaborations (Grid Pair-Level)

Notes: The dynamic ATT's are plotted using the staggered DiD model, based on Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). The control group consists of the grids that were never treated during our sample period. The dependent variable is collaborated patents in its logarithmic-like inverse hyperbolic sine transformation. In the left (right) panel, we define treatment to be reduction in travel time by 30 minutes (1 hour) along the fastest travel route.

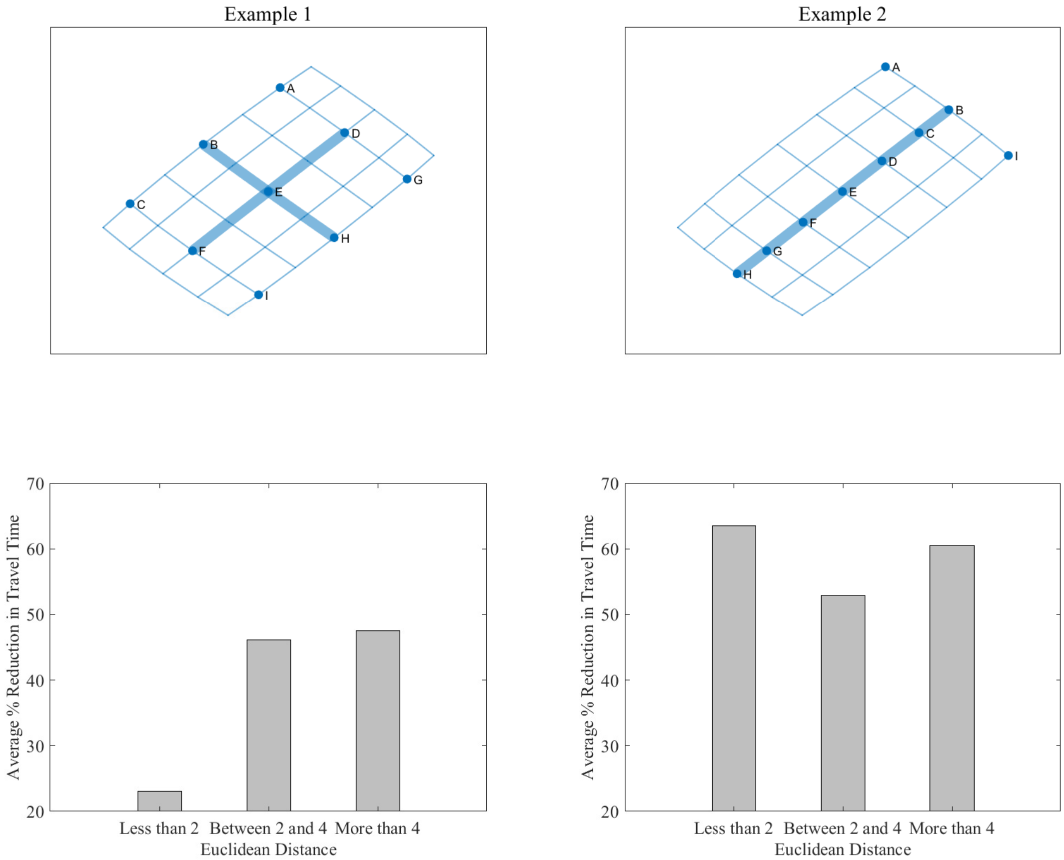


Figure 5: An Illustration

Notes: The upper panel in Example 1 assumes a symmetric and grid-like layout of 9 locations (Dots A-I) and the subway lines that connect some of those locations (thick blue line). We assume that travelling between locations take place along the grid lines only and the subway will increase the travel speed by a factor of four. The lower panel of Example 1 shows, in this scenario, how the average percent reduction in travel time changes with the Euclidean distance between all pairs of locations. The upper panel of Example 2 assumes an alternative layout of 9 locations that are less grid-like in nature (Dots A-I) and the subway lines that connect some of those locations (thick blue line). We impose the same assumption that travelling between locations take place along the grid lines only and the subway will increase the travel speed by a factor of four. The lower panel of Example 2 shows, in this scenario, how the average percent reduction in travel time changes with the Euclidean distance between all pairs of locations.

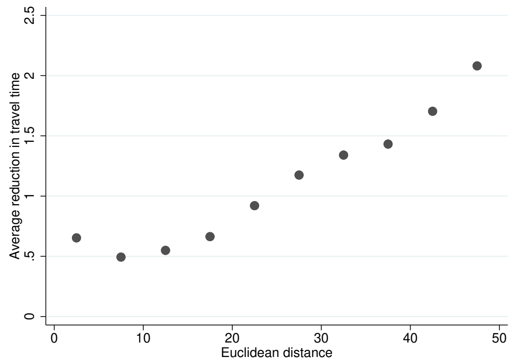


Figure 6: Euclidean Distance and Average Reduction in Travel Time in the Data

Notes: For all grid-pairs used in our analysis, we first calculate the Euclidean distance. We also calculate the reduction in travel time between the grids for each pair between 2000 and 2018. Then using 10 evenly spaced bins based on Euclidean distance, we compute the average reduction in travel time.

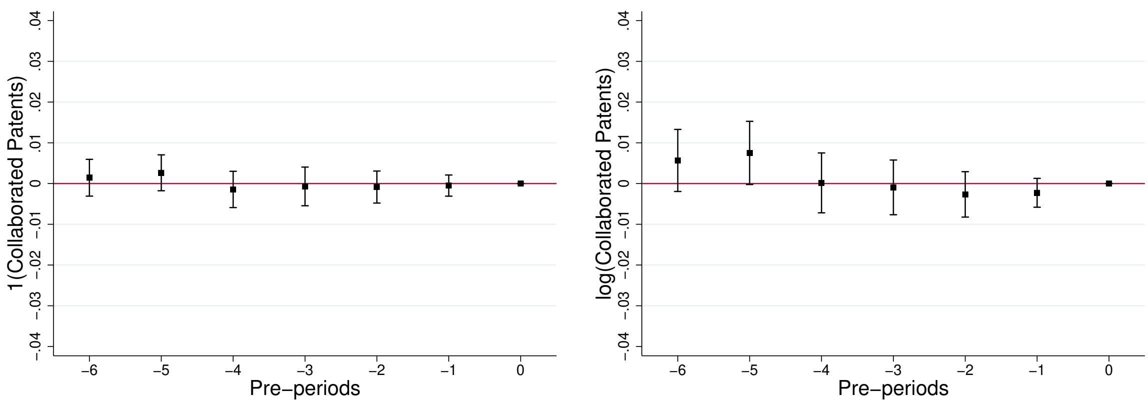


Figure 7: Pre-trends

Notes: The sample uses grid-pairs during the pre-periods (i.e. up to 6 years before the subway station opening). Period 0 is the year when the subway station opens, which is the omitted category. In the left and right panel, we use $\mathbb{1}$(Collab. patents) and log(Collab. patents) as the dependent variable, respectively. In the regression, we include the pre-period indicator interacted with logarithm of euclidean distance, origin fixed effects, destination fixed effects, and year fixed effects. The standard errors are clustered at the grid-pair level.

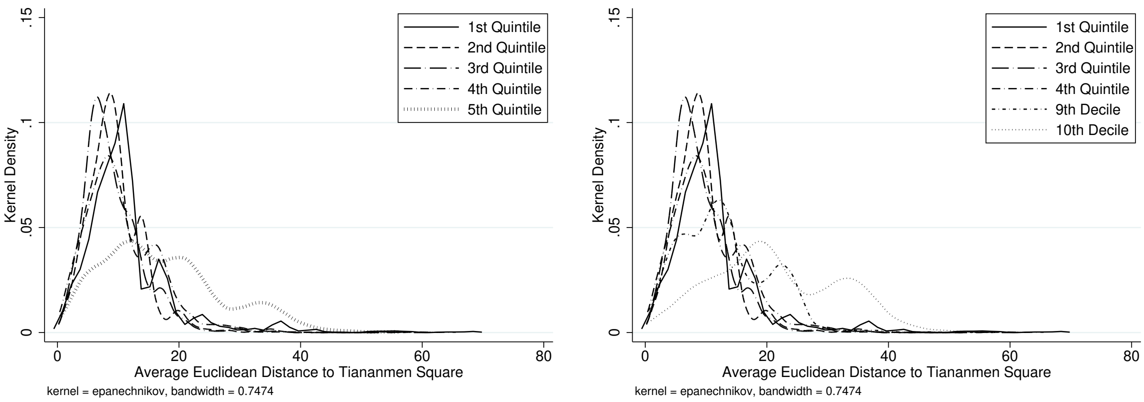
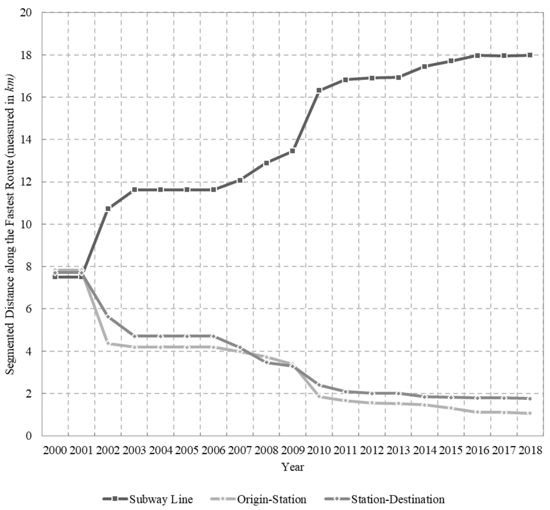
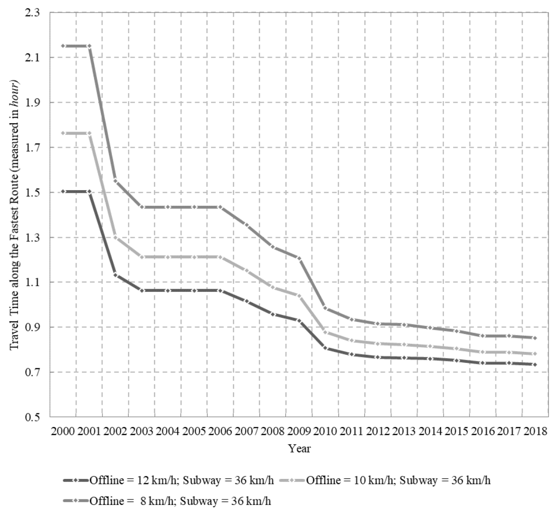


Figure 8: Kernel Density

Notes: For all grid-pairs used in our analysis, we first calculate the Euclidean distance between the centroid of the origin grid and Tiananmen Square, and that between the centroid of the destination grid and Tiananmen Square. Then we calculate their average distance and we do this for every grid-pair. Then we divide them into five quintiles and plot the kernel density for each one of the quintiles.



(a) Segmented Distance along the Fastest Route



(b) Travel Time along the Fastest Route

Figure 9: Subway Expansion and Travel Time

Notes: Panel (a) shows the segmented distance measures (*km*) along the fastest travel route. Panel (b) shows the travel time measures (*hour*) along the fastest travel route.

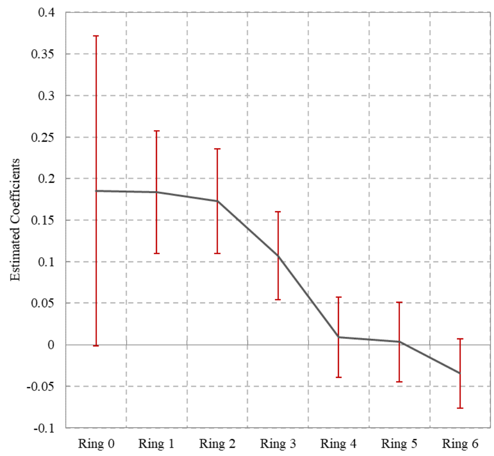


Figure 10: Ring-specific Coefficients

Notes: The estimated effects of rings at various distances from subway stations are plotted, along with bars representing the 95% confidence intervals. The ring number increases with the distance from a subway station. Corresponding estimates are reported in Column 7 of Table OA3.

Table 1: Treatment Effect on Patent Collaboration—Grid Pair-Level Analysis

	$\mathbb{1}$(Collab. patents)		log(Collab. patents) [asinh transf]	
Panel A.	(1)	(2)	(3)	(4)
$\mathbb{1}$(Treated)	0.0162***	0.0129***	0.0331***	0.0335***
	(0.0027)	(0.0031)	(0.0069)	(0.0078)
Observations	188,081	188,081	188,081	188,081
Method	TWFE	TWFE	TWFE	TWFE
Adjusted R^2	0.052	0.065	0.081	0.177
Year FE	Yes	Yes	Yes	Yes
Grid FE	Yes	No	Yes	No
Grid Pair FE	No	Yes	No	Yes

Panel B.		
$\mathbb{1}$(Treated)	0.0131***	0.0294***
	(0.0029)	(0.0059)
Observations	188,081	188,081
Method	DiD	DiD

Note: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018. In Panel B, the control group consists of the pairs of grids that were never treated, but were located within 2*km* from a subway station during our sample period. For a grid pair, $\mathbb{1}$(Treated) equals one for the first and the following years when the travel time between the grids gets reduced by more than 30 minutes compared to the previous year; and zero otherwise. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 2: Impact of Subway Expansion on Patent Collaboration—Grid Pair-Level Analysis

Dep. variable	𝟙(Collab. patents)		log(Collab. patents) [asinh transf]	
	(1)	(2)	(3)	(4)
log(Length)	-0.0025***	0.0026***	-0.0045***	0.0049***
	(0.0004)	(0.0006)	(0.0010)	(0.0015)
Wald Estimators Based on Travel Speed Assumptions of ...				
Offline = 12 km/h; Subway = 36 km/h	-0.1866	-0.2000	-0.3358	-0.3769
Offline = 10 km/h; Subway = 36 km/h	-0.2475	-0.1238	-0.4455	-0.2333
Offline = 8 km/h; Subway = 36 km/h	-0.4902	-0.0788	-0.8824	-0.1485
Year FE	YES	YES	YES	YES
Grid FE	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES
Observations	188,081	188,081	188,081	188,081
R-squared	0.070	0.114	0.098	0.221

Notes: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 3: Impact of Subway Expansion on Patent Collaboration—Grid Pair-Level Analysis
Heterogeneity and Local Average Causal Response

	𝟙(Collab. patents)				log(Collab. patents) [asinh transf]			
	(1) Near grid pairs	(2) Far grid pairs	(3) All grid pairs	(4) All grid pairs	(5) Near grid pairs	(6) Far grid pairs	(7) All grid pairs	(8) All grid pairs
log(Length)	0.0011*	0.0249***	0.0139**	0.0127**	0.0019	0.0558***	0.0432***	0.0376***
	(0.0007)	(0.0072)	(0.0061)	(0.0050)	(0.0016)	(0.0165)	(0.0151)	(0.0122)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-	-	-0.2648	-0.2476	-	-	-0.8229	-0.7329
Offline = 10 km/h; Subway = 36 km/h	-	-	-0.1691	-0.1568	-	-	-0.5255	-0.4642
Offline = 8 km/h; Subway = 36 km/h	-	-	-0.1124	-0.1033	-	-	-0.3492	-0.3059
Kleibergen-Paap rk Wald F statistic	-	-	239.508	369.948	-	-	239.508	369.948
Observations	105014	82915	188081	188081	105014	82915	188081	188081
Root MSE	0.251	0.264	0.251	0.251	0.430	0.406	0.416	0.415
Method	TWFE	TWFE	IV	IV (LOO)	TWFE	TWFE	IV	IV (LOO)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample selection criteria is based on the following: if the Euclidean distance between the grids of collaborators is below the median of the entire distribution, we assign the pair as a "Near grid pair." Otherwise, we assign the pair as a "Far grid pair." In Columns (3) and (7), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year interacted with the Euclidean distance between those two locations. In Columns (4) and (8), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year subtracting own pair length before interacted with the Euclidean distance between those two locations (Leave-one-out IV). Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 4: Impact of Subway Expansion on Probability of Collaboration—Collaborator Pair-Level Analysis

Dep. variable	𝟙(Actual collaborator pair)			
Sample	One-to-One Match		One-to-Two Match	
	(1)	(2)	(3)	(4)
log(Length)	-0.0489***	0.0066***	-0.0414***	0.0073***
	(0.0007)	(0.0025)	(0.0006)	(0.0024)
Year FE	YES	YES	YES	YES
Grid FE	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES
Observations	130,397	115,741	195,893	176,010
R-squared	0.286	0.742	0.237	0.720

Notes: Using the empirical methodology developed by Jaffe et al. (1993), we create the control group of non-collaborators by matching international patent classification (IPC) codes and application year of collaborated patents. For "One-to-One Match" sample, we use one randomly selected control pair for each actual pair of collaborators. Similarly, for "One-to-Two Match" sample, we use up to two randomly selected control pairs for each actual pair of collaborators. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote $p<0.01$, $p<0.05$, $p<0.1$ respectively.

Table 5: Impact on Different Types of Collaborators using TWFE and DiD Approaches —Grid Pair-Level Analysis

	(1)	(2)	(3)	(4)
Dep. variable	log(Collab. patents)		log(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
$\mathbb{1}$(Treated)	0.0301***	0.0275***	0.0036**	0.0022
	(0.0075)	(0.0055)	(0.0018)	(0.0023)
Observations	188,081	188,081	188,081	188,081
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
$\mathbb{1}$(Treated)	0.0330***	0.0296***	0.00006	-0.0001
	(0.1016)	(0.0059)	(0.0001)	(0.0003)
Observations	188,081	188,081	188,081	188,081
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
$\mathbb{1}$(Treated)	0.0296***	0.0299***	0.0043***	-0.00005
	(0.0784)	(0.0058)	(0.0013)	(0.0018)
Observations	188,081	188,081	188,081	188,081
Method	TWFE	DiD	TWFE	DiD
Year FE	YES		YES	
Grid Pair FE	YES		YES	

Notes: We use the sample from 2001 to 2018. We drop the initial year of 2000, since we use the previous year's information to determine incumbent, mobility, and productivity status. For a grid pair, $\mathbb{1}$(Treated) equals one for the first and the following years when the travel time between the grids gets reduced by more than 30 minutes compared to the previous year; and zero otherwise. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 6: Impact on Different Types of Collaborators using IV Approach
—Grid Pair-Level Analysis

	(1)	(2)	(3)	(4)
Dep. variable	log(Collab. patents)		log(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
log(Length)	0.0353**	0.0324**	0.0093***	0.0082***
	(0.0151)	(0.0128)	(0.0031)	(0.0026)
Observations	178,182	178,182	178,182	178,182
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
log(Length)	0.0386**	0.0354***	0.0003	0.0003
	(0.0156)	(0.0132)	(0.0003)	(0.0002)
Observations	178,182	178,182	178,182	178,182
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
log(Length)	0.0356**	0.0331**	0.0064**	0.0053**
	(0.0155)	(0.0130)	(0.0030)	(0.0025)
Observations	178,182	178,182	178,182	178,182
Method	IV	IV (LOO)	IV	IV (LOO)
Year FE	YES	YES	YES	YES
Grid Pair FE	YES	YES	YES	YES

Notes: We use the sample from 2001 to 2018. We drop the initial year of 2000, since we use the previous year's information to determine incumbent, mobility, and productivity status. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 7: Robustness Checks – Sample Selection

	(1)	(2)	(3)	(4)
Sample	90%	80%	90%	80%
Panel A.				
	$\mathbb{1}$(Collab. patents)		log(Collab. patents)	
log(Length)	0.0141***	0.0149***	0.0411***	0.0474***
	(0.004)	(0.0038)	(0.0113)	(0.009)
Observations	181,773	166,782	181,773	166,782
Panel B.				
	log(Collab. patents) by Incumbent Pairs		log(Collab. patents) by Non-Incumbent Pairs	
log(Length)	0.0363***	0.0460***	0.0084***	0.0063***
	(0.0118)	(0.0098)	(0.0024)	(0.0021)
Observations	172,206	158,004	172,206	158,004
Panel C.				
	log(Collab. patents) by Nonmover Pairs		log(Collab. patents) by Mover Pairs	
log(Length)	0.0443***	0.0479***	0.0004	0.0003
	(0.0144)	(0.0101)	(0.0002)	(0.0002)
Observations	172,206	158,004	172,206	158,004
Panel D.				
	log(Collab. patents) by H-H Nonmover Pairs		log(Collab. patents) by Other Nonmover Pairs	
log(Length)	0.0376***	0.0460***	0.0049**	0.0042**
	(0.0120)	(0.0010)	(0.0023)	(0.0020)
Observations	172,206	158,004	172,206	158,004
Method	IV (LOO)	IV (LOO)	IV (LOO)	IV (LOO)
Year FE	YES	YES	YES	YES
Grid Pair FE	YES	YES	YES	YES

Notes: In Panel A, we use the sample from 2000 to 2018. In Panels B to D, we drop the initial year of 2000 since we use the previous year's information to determine incumbent, mobility, and productivity status. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table 8: Dynamic Effects Across Years for Subgroups—Grid Pair-Level Analysis

Panel A.	Full Sample			
	(1)	(2)	(3)	(4)
Dep. variable: log(Collab. patents)	Movers	Entrants	H-types	L-Types
$\mathbb{1}$(within 5 years)	0.0018**	0.0031	0.0139*	0.0013
	(0.0008)	(0.0032)	(0.0083)	(0.0010)
$\mathbb{1}$(after 5 years)	0.0016**	0.0084**	0.0404***	0.0030**
	(0.0008)	(0.0034)	(0.0110)	(0.0013)
Observations	178182	178182	178182	178182
R-squared	0.067	0.256	0.353	0.083
Year FE	Yes	Yes	Yes	Yes
Grid Pair FE	Yes	Yes	Yes	Yes

Panel B.	Restricted Sample			
	(1)	(2)	(3)	(4)
Dep. variable: log(Collab. patents)	Movers	Entrants	H-types	L-Types
$\mathbb{1}$(within 5 years)	0.0017**	0.0046	0.0225***	0.0021**
	(0.0008)	(0.0030)	(0.0080)	(0.0009)
$\mathbb{1}$(after 5 years)	0.0016**	0.0075**	0.0324***	0.0031***
	(0.0008)	(0.0032)	(0.0105)	(0.0012)
Observations	178,182	178,182	178,182	178,182
R-squared	0.064	0.235	0.305	0.076
Year FE	Yes	Yes	Yes	Yes
Grid Pair FE	Yes	Yes	Yes	Yes

Note: The sample contains all grids ever located within 2 *km* from a subway station during 2001-2018. For a grid pair, $\mathbb{1}$(within 5 years) equals one for the first 5 years starting from the year when the travel time reduces by more than 30 minutes compared to the previous year along the fastest travel route; and zero otherwise. Similarly, $\mathbb{1}$(after 5 years) equals one starting from the sixth year onward since the treatment; and zero otherwise. In Panel B based on the "Restricted Sample," we drop all pairs that include universities. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote $p<0.01$, $p<0.05$, $p<0.1$ respectively.

Online Appendix

A. Beijing Subway Development

Line	Date	Stations
4	9/28/2009	Beigongmen, Weigongcun, Xidan, Haidianhuangzhuang, Gongyixiqiao, Caishikou, National Library, Zhongguancun, Beijing Zoo, East Gate of Peking University, Renmin University, Xiyuan, Ping'anli, Taoranting, Yuanmingyuan Park, Xuanwumen, Majiapu, Beijing South railway station, Jiaomen West, Xisi, Lingjing Hutong, Xinjiekou, Anheqiao North, Xizhimen
4	12/30/2010	Huangcun Xidajie, Xihongmen, Yihezhuang, Huangcun railway station, Gaomidian South, Gaomidian North, Qingyuanlu, Tiangongyuan, Zaoyuan, Xingong, Biomedical Base
5	10/7/2007	Puhuangyu, Liujiayao, Tiantandongmen, Beixinqiao, Huixinxijie Nankou, Hepingxiqiao, Tiantongyuan, Zhangzizhonglu, Yonghegong Lama Temple, Dongsi, Dongdan, Tiantongyuan South, Hepingli Beijie, Tiantongyuan North, Ciqikou, Dengshikou, Huixinxijie Beikou, Beiyuanlu North, Chongwenmen, Datunlu East, Lishuiqiao, Lishuiqiao South, Songjiazhuang
6	12/30/2012	Dongsi, Ping'anli, Dongdaqiao, Caofang, Jintailu, Hujialou, Chegongzhuang West, Baishiqiao South, Changying, Haidian Wuluju, Chegongzhuang, Beihai North, Nanluoguxiang, Dalianpo, Huangqu, Shilipu, Huayuanqiao, Qingnianlu, Chaoyangmen, Cishousi
6	12/28/2014	Tongzhou Beiguan, Beiyunhe West, Wuzixueyuanlu, Haojiafu, Lucheng, Dongxiayuan
6	12/30/2018	Tian Cun, Xihuang Cun, Yang Zhuang, Liaogong Zhuang, Beiyunhe East, Jin'anqiao

Line	Date	Stations
7	12/28/2014	Beijing Happy Valley, Jiulongshan, Guangqumenwai, Daguanying, Huagong, Guangqumennei, Hufangqiao, Nanlouzizhuang, Zhushikou, Jiaohuachang, Ciqikou, Beijing West Railway Station, Caishikou, Guang’anmennei, Dajiaoting, Shuanghe, Qiaowan, Baiziwan, Wanzi
7	12/30/2018	Fatou
7	12/28/2019	Lang Xin Zhuang, Hua Zhuang, Huang Chang, Wansheng Dong (East), Qun Fang, Hei Zhuang Hu, Gao Lou Jin, Shuangjing, Wansheng Xi (West)
8A	7/19/2008	South Gate of Forest Park, Beitucheng, Olympic Green
8A	10/9/2008	Olympic Sports Center
8A	12/31/2011	Yuxin, Lincuiqiao, Huilongguan Dongdajie, Xixiaokou, Yongtaizhuang, Huoying
8A	12/30/2012	Guloudajie, Anhuaqiao
8A	12/28/2013	Nanluoguxiang, Zhuxinzhuang, Shichahai, Yuzhilu, Pingxifu
8A	12/26/2015	Andelibeijie
8A	12/30/2018	National Art Museum
8B	12/30/2018	Huojian Wanyuan, Yinghai, Dahongmen Nan (South), Wufu Tang, Zhushikou, Heyi, Donggao Di, Muxi Yuan, Haihu Tun, Yongdingmenwai, Demao, Tian Qiao
9	12/31/2011	Beijing West railway station, Fengtainanlu, Fengtai Science Park, Liuliqiao, Liuliqiao East, Guogongzhuang, Qilizhuang, Keyilu
9	10/12/2012	Fengtai Dongdajie
9	12/30/2012	Baiduizi, Baishiqiao South, National Library
9	12/21/2013	Military Museum
10	7/19/2008	Agricultural Exhibition Center, Haidianhuangzhuang, Zhichunli, Jinsong, Taiyanggong, Liangmaqiao, Jiandemen, Anzhenmen, Jintaixizhao, Shuangjing, Sanyuanqiao, Beitucheng, Zhichunlu, Tuanjiehu, Suzhoujie, Mudanyuan, Huixinxijie Nankou, Hujialou, Shaoyaoju, Bagou, Xitucheng, Guomao
10	12/30/2012	Caoqiao, Shilihe, Dahongmen, Chedaogou, Fenzhongsi, Gongzhufen, Shiliuzhuang, Cishousi, Shoujingmao, Changchunqiao, Songjiazhuang, Lianhuaqiao, Xidiaoyutai, Jijiamiao, Jiaomen West, Huoqiying, Xiju, Liuliqiao, Panjiayuan, Chengshousi
10	5/5/2013	Fengtai Railway Station, Jiaomen East, Niwa
13	9/28/2002	Shangdi, Dazhongsi, Xi’erqi, Huilongguan, Huoying, Qinghe, Zhichunlu, Xizhimen, Wudaokou, Longze
13	1/28/2003	Guangximen, Dongzhimen, Shaoyaoju, Lishuiqiao, Wangjing West, Beiyuan, Liufang

Line	Date	Stations
14A	12/28/2014	Jintailu, Donghuqu, Futong, Shan'gezhuang, Laiguangying, Wangjing South, Jiangtai, Zaoying, Dongfengbeiqiao, Wangjing
14A	12/26/2015	Jiulongshan, Beigongda Ximen, Yongdingmenwai, Puhuangyu, Shilihe, Fangzhuang, Jingtai, Dawanglu, Beijing South railway station
14A	12/31/2016	Chaoyang Park
14A	12/30/2017	Pingleyuan
14B	5/5/2013	Zhangguozhuang, Dawayao, Qilizhuang, Xiju, Garden Expo Park, Dajing, Guozhuangzi
15	12/30/2010	China International Exhibition Center, Hualikan, Maquanying, Wangjing West, Sunhe, Wangjing, Houshayu, Cuigezhuang
15	12/31/2011	Shunyi, Fengbo, Shimen, Nanfaxin
15	12/28/2014	Olympic Green, Anlilu, Guanzhuang, Beishatan, Qinghuadongluxikou, Liudaokou
15	12/26/2015	Datunlu East
15	12/31/2016	Wangjing East
16	12/31/2016	Xibeiwang, Malianwa, Yongfeng South, Xiyuan, Daoxianghulu, Beianhe, Tundian, Wenyanglu, Yongfeng
16	12/30/2017	Nongdananlu
Ba Tong	12/27/2003	Guaan Zhuang, Li Yuan, Lin He Li, Gao Bei Dian, Jiu Ke Shu, Si Hui Dong(E), Tongzhou Beiyuan, Ba Li Qiao, Si Hui, Guo Yuan, Communication University of China, Tu Qiao, Shuang Qiao
Ba Tong	12/28/2019	Hua Zhuang
Capital Airport	7/19/2008	Dongzhimen, Terminal 3, Terminal 2, Sanyuanqiao
Changping	12/30/2010	Gonghuacheng, Shahe, Zhuxinzhuang, Nanshao, Shahe University Park, Xi'erqi, Life Science Park
Changping	12/26/2015	Changping, Changping Xishankou, Changping Dongguan, Ming Tombs, Beishaowa
Fangshan	12/30/2010	Liangxiangnanguan, Suzhuang, Daotian, Changyang, Liangxiang University Town West, Dabaotai, Libafang, Liangxiang University Town North, Liangxiang University Town, Guangyangcheng
Fangshan	12/30/2011	Yancun East
Fangshan	12/31/2011	Guogongzhuang
Yizhuang	12/30/2010	Xiaocun, Ciqu South, Songjiazhuang, Rongjingdongjie, Ciqu, Jinghailu, Rongchangdongjie, Tongjinanlu, Jiugong, Xiaohongmen, Wanyuanjie, Yizhuang Culture Park, Yizhuangqiao
Yizhuang	12/30/2011	Yizhuang Railway Station

B. Assignment of Applicants to Grids

For each patent in the CNIPA database, address is reported only for the first-listed applicant. If we consider a collaborated patent with three applicants, for example, address is reported only for the first-listed applicant but not for the remaining two applicants. This requires us to impute missing information on addresses of such collaborators who were not listed first.

We first create a search database, using all available addresses for applicants of solo patents and first-listed applicants of collaborated patents. Note that this search database contains the exact address of such applicants, as they are reported in the CNIPA data. Suppose a collaborator listed as the second applicant in patent A was either the only or the first applicant in patent B in the same year. Then we retrieve the address for that collaborator using patent B in our search database. To identify and match applicants across patents, we first extensively clean all names in a consistent format and assign unique identifier numbers to 77,716 unique names. In the event that there is no patent by the same collaborator in the same year, we instead search for other patents by the same collaborator but in a different year. We search patents in earlier and nearer years first. Suppose we are searching for the address of a collaborator in year 2010. Then we search for that collaborator's patents in our search database in the following order of years: 2010, 2009, 2008, ..., 2000, 2011, 2012, ..., and 2018. For some firm applicants, we further search their location using the search engine provided by Tianyancha database which contains exact addresses of 3,141 companies for year 2019. After completing this procedure, we have addresses for a total of 398,121 observations of applicants for 175,232 collaborated patents between 2000 and 2018. 175,232 (44.01%) observations are first-listed applicants, for whom address is reported in the data, and 205,689 (51.66%) observations are those with imputed address. As for remaining 17,300 (4.34%), we are unable to impute address information.

Despite these efforts, there are some caveats to note. First, there may be different applicants with the same name. Among 398,121 observations of collaborators, 92.82% are non-individuals, such as firms, universities, research institutes, hospitals, etc. Whereas it is less likely for institutions to share the same name, the likelihood is higher for individuals and this can result in ambiguous cases. For example, if we have two observations of the same applicant name but in two different locations in a year, it is unclear whether this is due to relocation of a single applicant or two different applicants existing in two different locations. We make the following simplification: as long as these two addresses are within a threshold of 2 *km*, we treat them as the same applicant and randomly select one of the two addresses to use. Second, the imputation is likely to be less accurate for the applicants who appear less in our search database. For example, suppose we are trying to impute a collaborator's address in 2010 but the only available information is in 2018. Whereas we assume that the applicant's location is the same as what it was in 2018, this may not necessarily be true if the applicant had moved between 2010 and 2018. This measurement error is likely to be more problematic if, for example, there is a large share of applicants who produce many patents but rarely appear as either solo or first-listed applicants. However, such problem is less likely to be severe given that we fail to identify address for only 4.34% of our observations. Moreover, 17,300 observations with unidentified addresses are scattered across 9,221 applicants, as opposed to being concentrated to only a small number of applicants.

C. Identifying Movers vs. Non-movers

In Tables 5 and 6, we classify whether collaborative patents in a year were produced by movers or non-movers. We first identify pairs of collaborators who were for sure both "non-movers" in that year, while placing all other remaining pairs in the "movers" category. For this aim, we first compile the entire list of all first-listed applicants using both solo and collaborated patents in our data. We use locations of such solo or first-listed applicants, since their addresses are precisely reported in the data without us having to impute information. For each applicant, we then compare the reported locations across patents applied in different years. If an applicant's location remained the same during a specific time period, then we mark that applicant as a "non-mover" during that time period. For example, suppose an applicant was observed at the same location in 2000, 2005, and 2007. Then that applicant is classified to be a non-mover from 2000 to 2007.[29] In Panel B of Tables 5 and 6, "Nonmover Pairs" refer to pairs of applicants who were identified to be both non-movers in that particular year. The 'Mover Pairs" refer to all other remaining pairs.

In Table OA6, we apply an alternative way of defining movers. We first identify the pairs which contain at least one applicant who moved for sure in that particular year. Such pairs are now classified as "Mover Pairs" in Panel B, while all remaining other pairs are classified as "Nonmover Pairs." We

[29]Although it is possible, for example, that the applicant moved elsewhere and then returned to the original location between 2001 and 2004, this possibility is unlikely to be high.

compare locations of each applicant across patents applied in different years. If location changed at some point, then we assign the applicant to be a mover in the first year when the applicant was observed in a different location.[30]

D. Tracking Patents after Movers' Relocation

Once we assign applicants for each collaborated patent in a year to corresponding grids as outlined in Appendix B, we further would like to identify whether that patent was produced by a collaborator after relocating to a different place. That is, for each patent given a collaborator, we identify whether the patent was produced in a new location that differs from the collaborator's previous location. Although some applicants may have moved multiple times during our sample period, we use the first year when an applicant was observed in a new location. Note that this year may not necessarily coincide with the actual moving year, since we observe an applicant only when a patent is produced.[31] That is, we identify the first year when a collaborative patent was produced after relocation took place. Note that because our analysis focuses on patent counts at either grid-level or grid-pair level, we are more interested in whether a patent was produced after relocation, as opposed to the exact year when a collaborator moved.

For each applicant, we first compare addresses reported in all patents produced by that applicant during our sample period. If there is a unique, consistent address across all patents, then we first rule out these applicants since we detect no relocation during our sample period. Next, there are cases in which an applicant is observed in multiple locations but consistently *across multiple years.* Most of these cases arise from the fact that an institution has multiple subdivisions located in different places (e.g., different departments of the same university or different branches of the same company). A few of such cases also involve individual applicants who happen to share the same name. Because we consistently observe different set of locations consistently across years, it is less likely that this is due to relocation. Thus, we also rule of such observations of applicants. Then we are left with the applicants who issued multiple patents and there is at least one change in the reported address during

[30] Suppose we have observations of an applicant in 2000, 2005, 2007, and 2010. If locations change from 2005 to 2007, then we identify that the applicant was a mover in year 2007. Note that because we do not observe applicants every year, it is difficult to pinpoint the exact relocation year.

[31] For example, suppose an applicant produced a patent in location A in year 2000 and another patent in location B in year 2005. We do not know exactly when the applicant moved, but we can conclude that the latter patent is the first patent that the applicant produced after moving.

sample period. By comparing address for an applicant across years, we identify the first year when a patent was produced by an applicant in a new location that differed from the previous location. Then we identify all patents produced by that applicant after (before) that identified year to be patents produced *after moving* (*before moving*).

E. Construction of Control Applicant Pairs

Jaffe, Trajtenberg and Henderson (1993) generate control patents by implementing the following procedure – "For each citing patent, we identified all patents in the same patent class with the same application year (excluding any other patents that cited the same originating patent). We then chose from that set a control patent whose grant date was as close as possible to that of the citing patent" (p.582, Jaffe et al. 1993).

To generate control pairs, we first identify all possible pairs of collaborators within the team of applicants for a collaborated patent.[32] We identify 77,749 pairs of applicants who worked together on a patent and whose respective locations are identified in the data. Then we extract the first-listed IPC code at the 3-digit level and application year from all collaborated patents. A patent typically has multiple IPC codes, but the first-listed IPC code represents the primary classification of the patent's content. Among collaborated patents, we identify the groups of patents that share the same 3-digit first-listed IPC-code and application year.[33] Finally, we construct control applicant pairs by taking one applicant from the original patent and one from a control patent. For example, if the actual pair of collaborators is (A, B), we create a matched control applicant pair (B, C), where C is an applicant of another collaborated patent that shares the same first-listed IPC code and application year as the patent invented by (A, B). As we have 8,529,862 control applicant pairs in total, we either randomly select one or two control pairs for a given pair of actual collaborators in our analyses.

[32]For instance, if there are 4 applicants for a collaborated patent, this gives us 6 possible pairs of collaborators.

[33]One could potentially use more detailed subclasses of an IPC code or several IPC codes to select control patents. Whereas this increases similarity between patents, it greatly reduces the available set of patents that can be used. We focus on collaborated patents only because collaborated patents could differ from patents with solo applicant in various unobservable aspects.

F. Theoretical Model

Consider the city of Beijing comprising a discrete set of N locations, where each location is endowed with 1 unit of land. Assume there is 1 unit of potential inventors, who are ex-ante homogeneous and mobile across locations. Based on their random preferences, they first jointly choose whether to innovate (occupation decision) and where to locate (location decision). Then they draw their productivity $z \geq 0$ from a distribution with density $f(z)$ and, based on that, choose whether to collaborate with other inventors (collaboration decision). We let g_n denote the endogenous measure of inventors at location $n = 1, 2, ..., N$, g_0 denote the endogenous measure of inventors who do not innovate, and $\sum_n g_n + g_0 = 1$. We solve the set of decisions, in the form of g_n and g_0, by following backward induction.

Collaboration Decision. – An inventor collaborates with another inventor if such collaboration is expected to yield positive net returns.

Inventors simultaneously search for collaborators and they encounter one another randomly. The probability of inventor i with productivity z_i at location m encountering inventor j with productivity z_j at location n is $\theta_{mn} f(z_i) f(z_j)$, where θ_{mn} is a location pair-specific parameter that captures exogenous factors determining the likelihood of the encounter.[34]

Upon encountering, collaboration is successfully formed when inventors are matched with probability $q(z_i, z_j, \tau_{mn})$, where τ_{mn} is the travel time between locations m and n. In this case, a net profit of $2\pi > 0$ is realized and is equally split between the two inventors. With probability $1 - q(z_i, z_j, \tau_{mn})$, inventors fail to match and there is no profit from collaborations. To make things more concrete, we specify the matching probability as follows:

$$q(z_i, z_j, \tau_{mn}) = \mu e^{-\left(b + z_i z_j\right) \gamma \tau_{mn}}, \tag{OA1}$$

where $\gamma > 0$ measures the elasticity of the matching probability with respect to travel time τ_{mn}; b is a constant capturing the decay rate of the matching probability when $z_i z_j = 0$ (i.e. the lowest quality match); $0 < \mu < 1$ is a constant to scale q.

[34] The assumption of exogenous encountering does not compromise the generality of our model since the collaboration returns are considered after the interaction of both encountering and matching upon encountering and we allow the matching probability to change with travel time as shown below.

Note that the matching probability increases as travel time decreases and such effect is more pronounced for more productive pairs of inventors with higher $z_i z_j$. This setup is consistent with important regularities in the literature. First, it suggests that collaborators' productivities are complements, which has been widely documented and is often imposed in various structural frameworks modeling idea exchange and knowledge spillover in the literature (Davis and Dingel 2019). Such complementarity determines that an inventor with high productivity is more likely to be matched to another highly productive inventor for collaboration. Second, the matching probability between productive inventors increases more as travel time decreases (Combes et al. 2012; Behrens et al. 2014; Davis and Dingel 2019; Lee and Xu 2020). For a more accessible location, we would thus expect to see the high-quality matches being more disproportionately formed (i.e. the "log super-modularity" condition).

Therefore, for inventor i at location m, the expected net value of being matched to an inventor at location n is

$$v_{mn}(z_i, \tau_{mn}) = \theta_{mn} \pi g_n \int q(z_i, z_j, \tau_{mn}) f(z_j) dz_j. \tag{OA2}$$

Consequently, the level of collaborative matches between locations m and n is

$$N_{mn} = \theta_{mn} g_m g_n \int \int q(z_i, z_j, \tau_{mn}) f(z_i) f(z_j) dz_j dz_i. \tag{OA3}$$

Location and Occupation Decisions. – An inventor's location decision is determined by location-specific returns to innovations, which come from both collaborative and solo innovations.

Based on Equation (OA2), the expected profit from collaborative innovations for inventor i at location m is

$$\int \sum_{n=1}^{N} v_{mn}(z_i, \tau_{mn}) f(z_i) dz_i. \tag{OA4}$$

We assume the expected profit from solo innovations for inventor i at location m is

$$\int Z_m \bar{\tau}_m^{-\alpha} z_i f(z_i) dz_i, \tag{OA5}$$

where Z_m is the location-specific productivity for solo innovations; $\alpha > 0$ measures the elasticity of solo innovations' profit with respect to location accessibility, $\bar{\tau}_m$ is defined as

$$\bar{\tau}_m = \sum_{n=1}^{N} e^{-\tau_{mn}} g_n. \tag{OA6}$$

Equations (OA5) and (OA6) show that a location that is closely connected to other locations with a large measure of inventors yields high expected profits from solo innovations. The setup captures the existence of knowledge spillovers in that more inventors contribute to a higher overall productivity and the parametric form embeds the property that the knowledge spillover decays with distance.

Therefore, inventor i's expected net value of choosing location m is

$$V_m = \eta_m^{-\xi} H \left[\int \sum_{n=1}^{N} v_{mn}\left(z_i, \tau_{mn}\right) f\left(z_i\right) dz_i, \int Z_m \bar{\tau}_m^{-\alpha} z_i f(z_i) dz_i \right], \tag{OA7}$$

where H is an aggregator which combines the expected profits from collaborative and solo innovations. If the two types of innovations were independent, H would be a simple summation of the two. But H does not have to be a linear aggregator as collaborative and solo innovations may affect each other (complements or substitutes). η_m is an observed feature of location m (e.g. disamenities) and we assume it to be exogenous. ξ captures the elasticity of η_m.[35]

An inventor chooses either to locate in one of the $m = 1, 2, \ldots, N$ locations to engage in innovation activities or not to innovate by taking the outside option. The goal is to maximize her expected value:

$$\max[V_1 \varepsilon_{i1}, V_2 \varepsilon_{i2}, \ldots, V_N \varepsilon_{iN}, K \varepsilon_{i0}], \tag{OA8}$$

where ε_{im} is inventor i's unobserved taste for location m. If she chooses not to innovate and takes the outside option, she receives $K\varepsilon_{i0}$. K is a common parameter across all inventors. ε_{i0} is inventor i's unobserved taste for the outside option. Both ε_{im} and ε_{i0} are drawn from a Frechet distribution with CDF $e^{-\varepsilon^{-\kappa}}$, where a larger κ implies a smaller dispersion of individual taste.

We express the measures g_m ($m = 1, 2, ..., N$) and g_0 as the following:

$$g_m = \frac{V_m^{\kappa}}{K^{\kappa} + \sum_{n=1}^{N} V_n^{\kappa}} \qquad \text{for } m = 1, 2, \ldots, N \tag{OA9}$$

[35]In principle, η_m may depend on τ. For simplicity, we ignore this effect in the model. We will handle this possibility in the empirical part.

$$g_0 = \frac{K^\kappa}{K^\kappa + \sum_{n=1}^N V_n^\kappa} \tag{OA10}$$

Equilibrium Definition. – An equilibrium is a sequence of $\{V_m, g_m\}_{m=1}^N$ and g_0, where g_m is the measure of inventors at location m and g_0 is the measure of innovators who choose the outside option, such that Equations (OA7), (OA9), and (OA10) hold.

To further characterize the equilibrium and discuss the mechanisms, we impose the following assumptions.

Assumption 1 *H is a Cobb-Douglas aggregator,*

$$H = \left[\int \sum_{n=1}^N v_{mn}(z_i, \tau_{mn}) f(z_i) dz_i\right]^\beta \left[\int Z_m \bar{\tau}_m^{-\alpha} z_i f(z_i) dz_i\right]^{1-\beta}, \qquad 0 < \beta < 1. \tag{OA11}$$

Assumption 2 *z follows a Bernoulli distribution, where $z = 0$ with probability λ and $z = z_H (> 0)$ with probability $1 - \lambda$.*

Assumption 3 *The level of solo innovations at location m, defined as I_m, is proportional to the total value of solo innovations at location m:*

$$\ln I_m \propto \left[\int Z_m \bar{\tau}_m^{-\alpha} z_i f(z_i) dz_i\right] \tag{OA12}$$

Under the three assumptions, we obtain a closed-form solution of the model. Assumption 1 allows collaborative and solo innovations to be substitutable and regulates the elasticity of substitution to be one. Assumption 3 allows us to obtain the level of solo innovations conveniently and can be rationalized by considering that the value of each solo innovation is similar. Thus total innovation value is proportional to the counts of solo inventions.[36]

Under Assumption 2, there are two types of matches: a high-quality match ($z_i z_j = z_H^2$) and a low-quality match ($z_i z_j = 0$). We adopt the notation $s \in \{0, 1\}$, such that $s = 1$ ($s = 0$) refers to the high-quality (low-quality) match. Then, based on Equation (OA3), the levels of collaborative innovations conducted by the high-quality and low-quality matches, respectively, are as follows:

$$\ln N_{mn,s=1} \propto -\left(b + z_H^2\right) \gamma \tau_{mn} + \ln g_m + \ln g_n + \ln \theta_{mn}, \tag{OA13}$$

[36] Readers may think that the solo invention per inventor $\frac{I_m}{g_m}$ is proportional to the value of solo innovation instead of total invention counts. However, the term g_m will be absorbed in the equation (OA16) later.

$$\ln N_{mn,s=0} \propto -b\gamma\tau_{mn} + \ln g_m + \ln g_n + \ln\theta_{mn}. \tag{OA14}$$

Both Equations (OA13) and (OA14) indicate that the level of collaborative innovations increases as travel time decreases, but the increase is more pronounced for the high-quality matches due to the log complementarity between productivities as shown in Equation (OA1).

The total level of collaborative innovations between locations m and n can be expressed as:

$$\ln N_{mn} \propto \underbrace{-\ln\frac{N_{mn,s=0}}{N_{mn,s=0}+N_{mn,s=1}}}_{\text{High-Quality Matches Channel}} - \underbrace{b\gamma\tau_{mn}}_{\text{Marginal Matches Channel}} + \underbrace{\ln g_m + \ln g_n}_{\text{Distribution Effect}} + \ln\theta_{mn}, \tag{OA15}$$

which shows that N_{mn} increases when the share of low-quality matches ($\frac{N_{mn,s=0}}{N_{mn,s=0}+N_{mn,s=1}}$) decreases, when the reduction in travel time (τ_{mn}) induces an increase among low-quality matches, and when the measure of inventors in corresponding locations (g_m, g_n) increases. N_{mn} also increases if the exogenous parameter θ_{mn} governing the efficiency of encountering probability improves. We label the three channels highlighted in Equation (OA15) as High-Quality Matches Channel, Marginal Matches Channel, and the Distribution Effect, correspondingly.

Using Equations (OA7), (OA9) and (OA10), we further decompose the distribution effect, consisting of the $\ln g_m + \ln g_n$ terms, as follows:

$$\ln g_m \propto \underbrace{\kappa\delta\ln\left[\sum_{n=1}^{N} N_{mn}\right] + \kappa\rho\ln I_m}_{\text{Relocation Channel}} - \underbrace{\ln g_0}_{\text{Marginal Inventors Channel}} - \kappa\xi \underbrace{\ln\eta_m}_{\text{Disamenities}} + \varepsilon_m, \tag{OA16}$$

where δ, ρ and ξ are the elasticities of g_m with respect to each component. The first under-bracket term which is proportional to the value from innovations (aggregator H) shows that, as increased total innovations at location m make the location more attractive and g_m increases as inventors relocate. The second under-bracket term shows that g_m increases when the measure of inventors choosing the outside option g_0 decreases. The third under-bracket term captures other exogenous characteristics of location m. ε_m is an idiosyncratic error term.

Therefore, we consider the impact of travel time on the total level of collaborative innovations arising from the following four channels.

High-Quality Matches Channel. – With a reduction in travel time, more collaborations take place since the matching probability increases. In particular, there will be more high-quality matches being formed due to the log super-modularity condition.

Marginal Matches Channel. – With a reduction in travel time, a larger number of collaborative matches will be formed on the margin. Such matches formed on the margin are those with low-values, which would have not taken place without a reduction in travel time. Both the High-Quality Matches Channel and Marginal Matches Channel focus on the changes in the matching probability $q(z_i, z_j, \tau)$ given distribution of inventors across locations.

Relocation Channel. – The first under-bracket term in equation (OA16) shows that when the total level of solo and collaborative innovations at a location increases, more inventors from elsewhere will be attracted to that location. Inflow of inventors, especially those who are highly productive, will hence further spur greater collaborative innovations as a location becomes more accessible.

Marginal Inventors Channel. – When a location becomes more accessible, some of the inventors who previously chose the outside option will no longer be screened out. These marginal inventors who now engage in innovations have a lower productivity than other inventors. With a larger pool of active inventors, collaborative innovations will increase in more accessible locations. Both the Relocation Channel and the Marginal Inventors Channel focus on the distribution of inventors, by analyzing the impact of travel time on measures of inventors across locations.

G. Additional Results

Table OA1: Summary Statistics

Unit of Analysis	Variable	Mean	SD	No. Obs
Year level				
	No. patents in a year	42,876.55	36,431.05	19
	No. collaborated patents a year	9,222.73	9,144.18	19
Patent level				
	application year	2013.534	3.980	785,629
	No. applicants	1.283	0.606	785,629
	No. applicants of collaborated patents	2.271	0.625	175,232
Grid year level				
	No. patents	5.961	70.853	139,900
	No. of collaborated patents	1.317	48.179	139,900
Grid pair year level				
	No. of patents for a pair of grids in a year	0.485	11.555	240,300
	log(Length)	8.352	3.044	240,300
	log(Euclidean distance)	2.340	0.959	240,300

Notes: For the grid pair-year level analysis, we focus on the sample of 12,015 grid pairs that ever had at least one collaborated patent during the 19 years of our sample period.

Table OA2: Treatment Effect on Patents—Grid Level Analysis

	𝟙(Collab. patents)	log(Collab. patents)	𝟙(Total patents)	log(Total patents)	𝟙(Solo patents)	log(Solo patents)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
𝟙(Treated)	0.086***	0.182***	0.037	0.306***	0.028	0.268***
	(0.020)	(0.054)	(0.023)	(0.092)	(0.024)	(0.087)
Observations	132905	132905	132905	132905	132905	132905
Method	TWFE	TWFE	TWFE	TWFE	TWFE	TWFE
Adjusted R^2	0.292	0.401	0.346	0.503	0.337	0.488
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Grid FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel B.						
𝟙(Treated)	0.081***	0.193***	0.055*	0.282***	0.036	0.231**
	(0.026)	(0.058)	(0.030)	(0.096)	(0.031)	(0.093)
Observations	132,905	132,905	132,905	132,905	132,905	132,905
Method	DiD	DiD	DiD	DiD	DiD	DiD

Note: The sample contains all grids during 2000-2018. In Panel B for the staggered DiD analysis, we use the never-treated as the control group. Treatment equals one for the first and the following years when a subway station in a grid opens; and zero otherwise. In Columns (2), (4) and (6), we use the logarithmic-like inverse hyperbolic sine transformation. In Panel A for the TWFE model, standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA3: Impact of Station Opening on Patents —Grid-Level Analysis (Spatial Decay)

Dep. variable	log(Patents) [asinh transf]						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Central Grid	0.3063***	0.2985***	0.2139**	0.1863**	0.1862**	0.1853**	0.1853**
	-0.0919	(0.0919)	(0.0932)	(0.0932)	(0.0931)	(0.0932)	(0.0932)
At 0 - 0.5 *km*	-	0.3003***	0.2166***	0.1839***	0.1840***	0.1831***	0.1833***
	-	(0.0334)	(0.0361)	(0.0369)	(0.0369)	(0.0369)	(0.0369)
At 0.5 - 1 *km*	-	-	0.2163***	0.1736***	0.1738***	0.1728***	0.1727***
	-	-	(0.0297)	(0.0312)	(0.0316)	(0.0315)	(0.0315)
At 1 - 2 *km*	-	-	-	0.1060***	0.1071***	0.1071***	0.1071***
	-	-	-	(0.0230)	(0.0266)	(0.0266)	(0.0266)
At 2 - 4 *km*	-	-	-	-	-0.0024	0.0085	0.0089
	-	-	-	-	(0.0212)	(0.0241)	(0.0241)
At 4 - 7 *km*	-	-	-	-	-	-0.0184	0.0032
	-	-	-	-	-	(0.0194)	(0.0241)
At 7 - 10 *km*	-	-	-	-	-	-	-0.0346*
	-	-	-	-	-	-	(0.0210)
Year FE	YES	YES	YES	YES	YES	YES	YES
Grid FE	YES	YES	YES	YES	YES	YES	YES
Observations	132,905	132,905	132,905	132,905	132,905	132,905	132,905
R-squared	0.529	0.532	0.533	0.534	0.534	0.534	0.534

Notes: The dependent variable is the count of all patents in grid i in year t, in its logarithmic-like inverse hyperbolic sine transformation. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA4: Impact of Subway Expansion on Patent Collaboration—Alternative Distance Thresholds for Sample Selection

Dep. variable	𝟙(Collab. patents)		log(Collab. patents) [asinh transf]	
	(1)	(2)	(3)	(4)
Panel A.				
Sample: All grids ever located within **1.5** *km* from a subway station during 2000-2018				
log(Length)	-0.0026***	0.0026***	-0.0045***	0.0051***
	(0.0004)	(0.0006)	(0.0010)	(0.0016)
Observations	177,327	177,327	177,327	177,327
R-squared	0.070	0.115	0.098	0.224
Panel B.				
Sample: All grids ever located within **2.5** *km* from a subway station during 2000-2018				
log(Length)	-0.0025***	0.0024***	-0.0046***	0.0047***
	(0.0004)	(0.0006)	(0.0010)	(0.0015)
Observations	193,781	193,781	193,781	193,781
R-squared	0.070	0.113	0.098	0.219
Panel C.				
Sample: All grids ever located within **3** *km* from a subway station during 2000-2018				
log(Length)	-0.0025***	0.0024***	-0.0046***	0.0046***
	(0.0004)	(0.0006)	(0.0010)	(0.0014)
Observations	197,239	197,239	197,239	197,239
R-squared	0.070	0.113	0.098	0.218
Year FE	YES	YES	YES	YES
Grid FE	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES

Notes: Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA5: Impact of Subway Expansion on Travel Time —Grid Pair-Level Analysis

Dep. variable	Travel Time (Unit: hour)					
Assumptions	Offline = 12km/h Subway = 36km/h		Offline = 10km/h Subway = 36km/h		Offline = 8km/h Subway = 36km/h	
	(1)	(2)	(3)	(4)	(5)	(6)
log(Length)	0.0134***	-0.0130***	0.0101***	-0.0210***	0.0051***	-0.0330***
	(0.0008)	(0.0018)	(0.0010)	(0.0022)	(0.0012)	(0.0028)
Year FE	YES	YES	YES	YES	YES	YES
Grid FE	YES	NO	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES	NO	YES
Observations	188,081	188,081	188,081	188,081	188,081	188,081
R-squared	0.776	0.800	0.759	0.779	0.740	0.756

Notes: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA6: Impact on Different Types of Collaborators using IV Approach
—Grid Pair-Level Analysis - Robustness with Alternative Definition of Movers

	(1)	(2)	(3)	(4)
Dep. variable	log(Collab. patents)		log(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
log(Length)	0.0353**	0.0324**	0.0093***	0.0082***
	(0.0151)	(0.0128)	(0.0031)	(0.0026)
Observations	178,182	178,182	178,182	178,182
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
log(Length)	0.0286***	0.0241***	0.0154	0.0159
	(0.0059)	(0.0049)	(0.0146)	(0.0124)
Observations	178,182	178,182	178,182	178,182
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
log(Length)	0.0277***	0.0233***	0.0011*	0.0009*
	(0.0058)	(0.0048)	(0.0006)	(0.0005)
Observations	178,182	178,182	178,182	178,182
Method	IV	IV (LOO)	IV	IV (LOO)
Year FE	YES	YES	YES	YES
Grid Pair FE	YES	YES	YES	YES

Notes: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.