12-2021

Subway, collaborative matching, and innovation

Yumi KOH
Singapore Management University, ymkoh@smu.edu.sg

Li JING
Singapore Management University, lijing@smu.edu.sg

Jianhuan XU
Singapore Management University, jhxu@smu.edu.sg

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*

June 26, 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 treatment effect 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 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, the 2020 SMU Virtual Conference on Urban and Regional Economics, the 2021 European Meeting of the Urban Economics Association, and the 2021 American Real Estate and Urban Economics Association National Meeting for their helpful comments. We thank Aaron Zhe Yang Lim for his tremendous help as 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 we show 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. It is widely recognized that research and development (R&D) and innovations play a central role in advancing the technology frontier and promoting economic growth (Acemoglu 2008). As technologies become

[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`).

more specialized and sophisticated, innovations relying on collaboration networks become essential to achieve effective destructive breakthroughs (Jones 2009). As a result, innovations are increasingly clustered in large metropolitan cities where networks of high-quality human capital and infrastructures are present (Carlino et al. 2007; Kerr and Robert-Nicoud 2020). As collaborations rely heavily on close in-person interactions, our paper contributes to the understanding of how intra-city transport infrastructure fosters collaborative matches which enable cities to better embrace path-breaking innovations.[2]

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.[3] 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

[2]Note that there are recent studies on how inter-city transport infrastructure affects collaborative innovations. For example, studies have analyzed the impact of cheap airline connections (Catalini et al. 2020) or high-speed rail (Dong et al. 2020) on research paper publication among academic scholars. Whereas our study shares the similar spirit, we extend 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 investigate various underlying mechanisms beyond what have been documented previously and assess their relative importance.

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

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.[4]

One important aspect that needs to be addressed carefully when evaluating the impact of subway expansion is that 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 cross-sectional and intertemporal unobserved features, we estimate 1) a discrete difference-in-differences 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 difference-in-differences estimation, we define a location-pair as treated if a nearby subway station has been opened at both locations. 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. 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 the Euclidean distance between those locations 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.[5] Last, as another

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

[5]This instrument is analogous to the traditional Bartik shift-share instrument in the sense that the exogenous distance

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).[6] 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.[7]

between locations serve as the pre-determined Bartik weight that governs the differential exposure to a common aggregate shock.

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

[7] 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

We obtain the following findings. First, our difference-in-differences 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 but is statistically insignificant for nearer location pairs. The instrumental variable approach reveals an estimated local average treatment effect among "complying" location pairs, which translates into Wald estimators ranging from 34.92% to 82.29%. 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. We find that the *High-Quality Matches Channel* is the most important mechanism driving our reduced-form evidence. This finding is consistent with Catalini et al. (2020), which highlight complementarity between travel costs and the quality of a coauthor match in an academic setting. To further assess the relative importance of the channels quantitatively, we classify all collaborative matches into four mutually exclusive groups: matches involving new entrants, matches involving movers, matches between high-productive non-moving incumbents, and matches involving low-productive non-moving incumbents. 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 and result in an increase in collaborative patents, approximately 70% of the contribution comes from the *High-Quality Matches Channel*; and 21% is due to the *Relocation Channel* and *Marginal Inventors Channel* combined. When we drop some types of inventors that may face mobility or institutional restrictions (e.g. universities) from our sample, the relative importance of the *Relocation Channel* and *Marginal Inventors Channel* increases.

The rest of the paper is organized as follows. We present the institutional background in Section

important, especially in a longer time horizon.

2 and the conceptual framework in Section 3. We discuss our empirical research designs in Section 4 and present discussions on data and variables in Section 5. Findings are presented in Section 6. Section 7 concludes. Remaining materials are in the Appendix.

2. Institutional Background

Construction of the Beijing subway was first proposed by the city's planning committee as early as 1953 but its dramatic expansion started only in recent decades.[8] The initial purpose of the Beijing subway was to "ferry[ing] soldiers from their barracks on the outskirts to the city center" (Poon 2018), but it eventually became means of public transportation.[9] 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 with a goal to "establish a rail service for all residents, who will never be more than 1 *km* from a station on the city's 19 line subway network" (p.302, Clark (2009)). As shown by the bar graph in Figure 1, the cumulative count of subway stations grew by a factor of nine between 2000 and 2018 (from 41 to 379) and the speed of expansion clearly accelerated since 2008. The line graph in Figure 1 shows that the total subway length in Beijing increased by a factor of twelve between 2000 and 2018 (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. Back then, the subway system only included an east-west line and an 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). Moreover, one striking feature from the figure is that the locations of collaborators closely overlap with the layout of

[8]For more detailed background of the Being subway, see Sultana and Weber (2016).

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

the subway system that now includes many lines forming loops or branching out to suburban areas.

Stylized facts further indicate that patent collaborations increased between more distant locations as their connectivity improved from the build-up of subway system. Panel (a) of Figure 3 tracks the average distance between all pairs of patent collaborators. The average distance remained about the same until 2008, but then increased rapidly afterwards, which coincides with the rapid expansion of the subway system since 2008. Panel (b) shows that the growth rate of collaborated patents was higher for collaborators who were located far apart from each other during our sample period. As a simple illustration, Figure 4 shows how the collaboration links evolved for a particular location (i.e. a grid) that had the largest number of collaborated patents in 2000-2001. All collaborated patents originated in this grid 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 grid formed numerous collaboration links elsewhere, as shown by a set of many gray lines connecting this location with other locations. A striking feature of these collaboration links is that they closely overlap with the layout of the subway lines, as shown in black star marks.

Overall, Figures 1, 2, 3, and 4 show how 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 next formalize our conceptual model which guides our empirical analysis of causal inference and identification of underlying mechanisms.

3. Conceptual Framework

In this section, we lay out the conceptual framework to understand specific channels through which the build-up of the subway system produces spatial variations in returns to innovations and, hence, shapes the collaborative matches and the level of innovations across different locations. We first introduce the model setup in Section 3.1 and then characterize the model's equilibrium and mechanisms in Section 3.2.

3.1. Model Setup

Consider the city of Beijing comprising discrete N locations, where each location is endowed with 1 unit of land. There are two types of agents, homogeneous workers and heterogeneous inventors, who are mobile across locations. Assume that the number of workers is significantly larger than the number of inventors and so is the corresponding demand for land. Therefore, the price of land at each location is determined by the population density of workers. We begin by solving for the workers' location choices to endogenously determine the land prices, which will consequently affect inventors' location decisions.

Homogeneous Workers: The utility of a worker living at location n is

$$u_n = max_{c,h} c^{\beta} h^{1-\beta} A_n \varepsilon_n \tag{3.1}$$
$$\text{s.t. } c + p_n h = w.$$

In the expression, c represents the consumption good which is freely traded and whose price is normalized to 1; h is the quantity of land demanded by a worker; p_n is the unit price of land at location n; w is the wage of workers. As we assume a perfectly competitive labor market, wage w is the same across all locations and we further assume that it is exogenous for simplicity. A_n is location n's amenity, which is observed and reflects the accessibility of location n. ε_n captures an individual worker's idiosyncratic preference for location n and is drawn from a Type II Extreme Value distribution. After observing preference shocks across all locations, a worker chooses the location that maximizes his or her utility. Then the population density of location n is

$$L_n = \frac{A_n \left(p_n\right)^{-(1-\beta)}}{\sum_{m=1}^{N} A_m \left(p_m\right)^{-(1-\beta)}}, \tag{3.2}$$

and the land price at location n is

$$p_n = (1-\beta)\, w L_n. \tag{3.3}$$

Heterogeneous Inventors: Inventors are heterogeneous in their productivity z, where $z \in [\underline{z}, \bar{z}]$. We use $f(z)$ to denote the PDF of z, which is exogenous. Inventors maximizes their utility by making optimal decisions on collaboration, location, and occupation.

Collaboration Decision. – An inventor collaborates with another inventor if such collaboration results in positive net returns. Inventors simultaneously search for other collaborators and they encounter one another randomly. Let $g_n(z)$ denote the PDF of productivity z at location n. Then the probability of inventor i at location m encountering inventor j at location n is $\theta g_m(z_i) g_n(z_j)$, where θ measures the meeting probability between the two inventors.[10] We specify inventor i's revenue from collaborating with inventor j as $R(z_i, z_j; \tau_{mn})$ and the collaboration cost as $C(\tau_{mn})$, where τ_{mn} represents the travel time between locations m and n.[11] Hence, for inventor i at location m, the expected net value of finding a collaborator from location n is

$$v_{mn}(z_i) = \int_{\underline{z}}^{\bar{z}} \theta \max\left[R(z_i, z_j; \tau_{mn}) - C(\tau_{mn}), 0\right] g_n(z_j)\, dz_j. \tag{3.4}$$

Therefore, if v_{mn} is positive, inventor i at location m will search to collaborate with inventors at location n. If v_{mn} is non-positive, inventor i will not search collaborators in location n. The level of collaboration between locations m and n is

$$\int\int_{R\left(z_i, z_j; \tau_{mn}\right) \geq C(\tau_{mn})} \theta g_m(z_i) g_n(z_j)\, dz_i dz_j. \tag{3.5}$$

Location Decision. – An inventor's location decision is determined by location-specific returns to innovation and land price. The returns to innovation come from both collaborated and solo innovations. The expected net value from collaborations for a given location is derived in Equation (3.4). If inventor i engages in solo innovations, the expected profit from solo work at location m is $Z_m z_i$, where Z_m is a location-specific productivity for solo innovations which is positively correlated with the accessibility of location m. Putting together, inventor i's expected net value of locating at m is

$$V_m(z_i) = \sum_{n=1}^{N} v_{mn}(z_i) + Z_m z_i - \xi p_m, \tag{3.6}$$

The first two terms capture the net values from collaborated and solo innovations, respectively. The last term represents the land price, where ξ captures the quantity of land demanded by an inventor. The inventor chooses the location that maximizes $V_m(z_i) e_{im}$, where e_{im} is inventor i's unobserved taste for location m and is drawn from a Type II Extreme Value distribution.

[10] For simplicity, we assume that the meeting probability θ is exogenous and constant. However, the effect of an endogenous meeting probability on the net value could be incorporated into the revenue function R.

[11] As the travel time directly affects collaboration efficiency, revenue $R\left(z_i, z_j; \tau_{mn}\right)$ is also a function of τ_{mn}.

Occupation Decision. – An inventor's occupational choice specifies whether to engage in innovation activities or not. If the inventor does not innovate, he or she can take the outside option with value K. Hence if the expected net value from innovating at the optimal location is lower than K, he or she chooses not to invent at all. The inventor's optimized value function is determined by

$$\max\left\{E[\max_{m} V_m(z_i)\, e_{im}], K\right\}, \tag{3.7}$$

where $E[.]$ is the expectation value over the random variable e_{im}. We denote the occupation choice from the above problem as $\chi(z_i)$ such that $\chi(z_i) = 1$ if inventor i chooses to invent and $\chi(z_i) = 0$ otherwise.

Equilibrium Definition: An equilibrium is a quadruple $\{p_m, g_m(z), k_m(z), \chi(z)\}$ for any location m, where p_m is the land price at location m, $g_m(z)$ is the density of inventors with productivity z at location m, $k_m(z)$ is the probability that an inventor with productivity z chooses to locate at m, and $\chi(z)$ is the occupation choice such that

1. homogeneous workers maximize their utilities and sort into different locations given p_m;
2. heterogeneous inventors maximize their net values and sort into their preferred locations given $g_m(z)$ and p_m; and
3. market clears for any location m and any productivity z (i.e., $k_m(z) f(z) \chi(z) = g_m(z)$).

3.2. Characterization of the Equilibrium

Having laid out the model structure, we impose three additional assumptions to further characterize the equilibrium and discuss the mechanisms.

Assumptions: The revenue from collaboration $R(z_i, z_j; \tau)$ satisfies the following:

1. $\frac{\partial^2 R(z_i, z_j; \tau)}{\partial z_i \partial \tau} < 0$,
2. $\frac{\partial^2 R(z_i, z_j; \tau)}{\partial z_i \partial z_j} > 0$,
3. $\frac{\partial^3 R(z_i, z_j; \tau)}{\partial z_i \partial z_j \partial \tau} < 0$.

Assumption 1 suggests that the connectivity between the collaborators' locations and inventor i's own productivity are complementary in generating inventor i's revenue from collaboration. That is, collaboration revenue from greater productivity decreases with τ (Combes et al. 2012; Behrens et al. 2014; Davis and Dingel 2019; Lee and Xu 2020), which suggests that more productive inventors disproportionately seek more accessible locations in the equilibrium.

Assumption 2 suggests that two collaborators' productivities are complements. Such complementarity has been widely documented in the literature and is often imposed in various structural frameworks modeling idea exchange and knowledge spillovers (Davis and Dingel 2019). Given the presence of complementarity, there exists a positive assortative matching based on the productivity of collaborators in the equilibrium.

Assumption 3 implies that the marginal benefit of forming a collaboration pair between more productive inventors decreases with travel time τ. For instance, for two productivity levels $z_H > z_L$, $[R(z_H, z_H, \tau) - R(z_H, z_L, \tau)]$ increases when τ decreases. The assumption equivalently implies that an inventor is more likely to be matched to another inventor with similar productivity in the equilibrium, given a shorter travel time. This is consistent with Ganguli et al. (2020). In their model, there exists a search cost to find collaborators and when such cost decreases, productive inventors are more patient to wait till finding another productive inventor.

Mechanisms: We now analyze the comparative statics generated by the model. Based on the above assumptions, the model suggests that reducing travel time increases the number of collaboration pairs formed through four possible mechanisms. We label the four mechanisms as *High-Quality Matches Channel, Marginal Matches Channel, Relocation Channel*, and *Marginal Inventors Channel*.

High-Quality Matches Channel. – This channel captures the aspect that more collaboration pairs will be formed as travel time decreases, given the complementarity between travel time and inventors' productivities (Assumptions 2 and 3). With a smaller τ_{mn}, the number of collaborative matches formed between locations m and n increases since collaboration becomes more profitable. In particular, such an increase in the number of matches is driven by more active collaborations among highly productive inventors. This is because the revenue from collaborations between highly productive inventors increases more with a smaller τ_{mn} (Assumption 3).

Marginal Matches Channel. – This channel shows that with reduced travel time, there will be more collaboration pairs formed on the margin which would not have been formed otherwise. Inventors collaborate as long as their revenue exceeds the collaboration cost (i.e. $R(z_i, z_j; \tau_{mn}) \geq C(\tau_{mn})$). Therefore, if τ_{mn} sufficiently reduces $C(\tau_{mn})$, even low-value matches will be formed between locations m and n. In this case, the increase in collaboration pairs is due to a larger number of low-value matches formed on the margin given reduced travel time.

Relocation Channel. – This channel focuses on the aspect that more accessible locations attract inventors from elsewhere and thus result in more collaborations. Accessibility of a location affects inventors' returns through both collaborative and solo innovations. This is because for collaborations, location accessibility and inventor's productivity are complements in generating collaboration revenue (Assumption 2). Returns to solo innovations are also higher in more accessible locations, as shown by $Z_m z_i$ in Equation (3.6). Inventors will thus be attracted to more accessible locations, especially those with higher productivity, and collaborations there will increase.

Marginal Inventors Channel.– This channel highlights increased collaborations given matches from a larger pool of inventors. For some inventors, choosing the outside option is more profitable than choosing to innovate. However, when locations become more accessible, some of these inventors will no longer be screened out. These marginal inventors, who now choose to innovate given a reduction in travel time, have lower productivity than the average inventors who previously existed. With a larger pool of active inventors, there will be a greater number of collaboration matches formed in more accessible locations.

4. Empirical Research Designs

We now present the empirical research designs to identify the causal impact of the subway build-up on innovations. Throughout, we use collaborated patents to proxy for collaborative matches in innovation. We first lay out the identification strategies for reduced-form evidence in Section 4.1, which is then followed by a discussion on how we empirically assess each mechanism in Section 4.2.

4.1. Patent Collaborations

Based on our conceptual model, the build-up of the subway system reduces travel time between two locations and such reduction in collaboration cost spurs collaborations formed between those locations. In the empirical framework, we start by establishing the reduced-form causal relationship between subway build-up and patent collaborations. Our regression on collaborated patents is specified as follows:

$$CollabPatents_{ijt} = \beta Connect_{ijt} + \gamma_{ij} + \zeta_t + \upsilon_{ijt}, \tag{4.1}$$

where ij indexes a pair of locations (location i and location j) and t indexes a year; $CollabPatents_{ijt}$ represents the count of collaborated patents produced by collaborators in locations i and j in year t; $Connect_{ijt}$ measures the subway build-up that enhances the connectivity between locations i and j in year t —we proxy the extent of the subway build-up in two ways: one is by a discrete dummy variable that captures the treated period in which connectivity of locations i and j dramatically improves due to the openings of nearby subway stations and the other is by the total length of the subway lines that connect locations i and j along their shortest-time travel path; γ_{ij}, ζ_t, and υ_{ijt} represent location-pair fixed effects, year fixed effects, and the idiosyncratic error term; β captures the impact of the subway build-up on innovation collaborations.

Despite controlling for an extensive set of fixed effects that absorb unobserved time invariant factors at the location-pair level and unobserved intertemporal variations universal to all location pairs, there may still exist other unobserved factors in the error term that are correlated with our key regressor $Connect_{ijt}$. For instance, the planning and the intertemporal layout of the subway system may not be purely exogenous if high potential for collaborations between two locations reversely causes the subway system to be laid out in a way that disproportionately enhances connectivity between those two locations. It could also be the case that policy-makers plan the phase-in of the subway construction in company with other policies that aim to promote innovations in some particular locations.

We argue that the related concerns are less likely to severely drive our results since the construction of the subway stations and lines is notorious for unexpected delays in Beijing due to various technical obstacles. To the extent that the actual construction completion year differs significantly from the planned completion year in an unsystematic way, the timing of subway construction may still be deemed as exogenous. In addition, we conduct various checks in an event study setting in Section 6 to verify the parallel trend assumption and find supporting evidence for the exogeneity assumption. Nevertheless, to further alleviate concerns on potential endogeneity, we adopt other research designs to corroborate our findings.

Our second research design is to instrument for the subway build-up between two locations in a year using the total cumulative length of the subway lines in Beijing in that year interacted with the

Euclidean distance between those two locations. With this instrument, the extent of the subway build-up that enhances the connectivity between two locations is explained by two sources of exogenous variations: 1) the aggregate development of the subway system over time, and 2) the time-invariant geographic distance between those two locations. Both factors are likely to be exogenous and uncorrelated with potential unobserved factors, once we control for year fixed effects and location-pair fixed effects. The relevance of the instrument is governed by the fact that the development of the subway system over time affects the connectivity of two far-apart locations more than that of two close-together locations. The extreme case would be that if two locations are right next to each other, the overall development of the citywide subway system is irrelevant to the travel time between those two locations. Such instrument is analogous to the Bartik instrument proposed in Bartik (1992) in that we project the citywide aggregate subway development to the connectivity of location pairs based on their Euclidean distance. The spatial distance between two locations is, hence, analogous to the Bartik weight.[12]

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

4.2. Empirical Test of Mechanisms

The conceptual framework produces four mechanisms which commonly predict that more collaborative matches will be formed between locations as they become better connected. Whereas all four

[12]The Bartik approach has been widely used across many fields in economics. A number of related studies define the Bartik instrument as the interaction of the national industry employment growth rate and local industry employment share, and use it to predict the local employment growth rate (e.g., Diamond (2016)).

mechanisms are likely to co-exist, we want to assess which channel is quantitatively more important in our empirical context. As the mechanisms generate different predictions for various subgroups of inventors, we separately identify these predictions to shed light on the relative importance of each mechanism.

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* who now begin to innovate as locations become more accessible. If we find that most collaborations form between previously 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. Therefore, if we find that most collaborations take place among inventors who did not relocate, then the *Relocations Channel* is unlikely to be the main channel.

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

5. Data and Variables

5.1. 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 the exact 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.

To better measure the implications of the subway expansion on collaborative matches in innovations, we focus on collaborated patent *applications*, as opposed to approved patents. The reasons are two fold. First, although not all applications end up being patented, 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 of patent applications 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 throughout this paper.

Table A1 shows the summary statistics. During our sample period, there are approximately 42,876 patents in a year and the average number of collaborated patents in a year is 9,144. When we analyze 785,629 patents applied during 2000-2018, the average application year is 2013. This suggests that there were relatively a larger number of patents being applied in the latter years of our sample period. The average number of applicants for a patent is approximately 1.283 and the collaborated patents contain a team of approximately 2.271 applicants on average. The last two sets of rows in Table A1 present summary statistics at particular geographical units. As we elaborate in the next subsection, we divide the city of Beijing into specific location grids. Given a large number of grids, we find that the average number of patents and collaborated patents in a grid in a year equal 5.961 and 1.317, respectively. However, a large standard deviation suggests that grids differ greatly in terms of the number of patents being produced and this is particularly so across years as well. We also analyze the data at the grid pair-year level, which contains a sample of 12,015 grid pairs that ever had at least one collaborated patent during our sample period. 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.

[13]https://en.wikipedia.org/wiki/BeijingSubway.

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

5.2. Location Grids

Our empirical analysis requires spatial units that correspond to a discrete set of "locations" specified in our conceptual framework. 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 Appendix B, we provide a detailed outline on how we track applicants' locations using information in the patent database.

5.3. 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 would become 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 4.1. In addition, we also check robustness of our main findings using alternative distance cutoffs and report them in the Appendix.

5.4. 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 regression, 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 from origin to a station, a station to destination, or directly from origin to destination by walking, cycling, or riding a bus). Although it is possible that travel time may further be reduced if we are more specific about 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

[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.

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.

Although 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. It is also for this reason that our reduced-form evidence (i.e. the impact of $length$ on $CollabPatents$) is more robust and reliable, although we also report the first-stage results (i.e. the impact of $length$ on $TravelTime$) to facilitate a more structural interpretation.

Figure 5 shows how segmented distance and travel time along the fastest route have changed over time with the expansion of the subway system. Panel (a) shows the average $Length$ in the data for each year increased over time, as more lines were built and connectivity improved. In particular, such trend became more evident since 2008. At the same time, the distance between the centroid of a grid to the corresponding station diminished, for both origin and destination grids in a pair. This reflects more stations being built across locations. Given such changes, Panel (b) shows that the average travel time also decreased considerably over time.

6. Results

As an initial investigation, we begin by analyzing whether the subway station openings led to an increase in patents in more accessible locations by conducting a grid-level analysis in Section 6.1. Then we report our main results on patent collaborations at the grid-pair level in Section 6.2. Finally, we discuss and evaluate the underlying mechanisms in Section 6.3.

6.1. Initial Evidence at the Grid-Level

Although our focus is on patent collaborations at the grid-pair level, we conduct a grid-level analysis first to capture some initial evidence on the impact of subway expansions on overall patent

[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.

growth. This analysis also fulfills two additional purposes that facilitate our investigations at the grid-pair level in Section 6.2.

The first purpose is to assess the spatial scope of the subway impact so as to narrow down the number of grids that we use to construct grid pairs later. To achieve this goal, we assess the spatial attenuation effect of subway station openings on patent counts by estimating the following specification:

$$log(Patents)_{it} = \sum_{r=0}^{6} \rho^r \mathbb{1}(StatOpen)^r_{it} + \mu_i + \delta_t + \varepsilon_{it}, \tag{6.1}$$

where subscripts i and t denote location and year, respectively. The superscript r denotes distance rings based on various distance cutoffs.[17] The dependent variable is the logarithm of total patent counts in location i and year t.[18] $\mathbb{1}(StatOpen)^r_{it}$ is an indicator variable that takes value 1 if t is greater than or equal to the year when the first station in ring r comes into operation; it takes value 0 otherwise. δ_t, μ_i, and ε_{it} represent year fixed effects, location fixed effects, and the idiosyncratic error term, respectively. The coefficients of interest are the ρ^r's which are identified off the intertemporal and cross-sectional variations in the timing of subway station openings during the sample period.

Table A2 reports the estimated coefficients when we gradually add ring-specific treatment variables. The last column, hence, contains the full set of ring-specific treatment variables to capture the treatment effect up until 10 *km* from the stations that come into operations. Figure 6 plots the corresponding coefficients ρ^r's along with their 95% confidence intervals. Whereas we find spatial attenuation patterns, the impact of subway station openings on patent counts remains to be positive and statistically significant up to Ring 3 (i.e. approximately within 2 *km*). Overall, Figure 6 provides evidence that the subway expansion affected the creation of patents from a spatial perspective and

[17]More specifically, we have 7 distance rings:

$$r = \begin{cases} 0, & \text{if the opening station is at the same grid;} \\ 1, & \text{if the opening station is at 0-0.5 } km \text{ distance;} \\ 2, & \text{if the opening station is at 0.5-1 } km \text{ distance;} \\ 3, & \text{if the opening station is at 1-2 } km \text{ distance;} \\ 4, & \text{if the opening station is at 2-4 } km \text{ distance;} \\ 5, & \text{if the opening station is at 4-7 } km \text{ distance;} \\ 6, & \text{if the opening station is at 7-10 } km \text{ distance.} \end{cases}$$

[18]We approximate the logarithm transformation using either $log(Patents+1)$ or the commonly used alternative, the inverse hyperbolic sine transformation. The inverse hyperbolic sine (arcsinh) transformation converts a random variable x into $\tilde{x} = arcsinh(x) = \log(x+\sqrt{x^2+1})$ (Bellemare and Wichman 2020). Both produce consistent estimates with small quantitative differences.

such effects tend to die out beyond 2 *km*.[19] For this reason, for the grid-pair analyses that follow, we will use the sample of all grid-pairs that were ever located within 2 *km* from a subway station during our sample period.[20]

The second purpose of our grid-level analysis is to check on the exogeneity of the timing of a subway station opening, which also serves as a critical assumption to validate our two-way fixed effects model in identifying the subway impact on patent collaborations. We achieve this focus by assessing the dynamic effect. We use the non-parametric event-study setup and examine the patterns of patent counts for each of the years before and after the opening of a subway station. More specifically, we estimate:

$$log(Patents)_{it} = \sum_{\substack{k=-5 \\ k\neq -1}}^{10} \eta_k \mathbb{1}(StatOpen)_{itk} + \mu_i + \delta_t + \varepsilon_{it}, \tag{6.2}$$

where subscripts i and t denote location and year, respectively. The dependent variable is the logarithm of total patent counts in location i in year t.[21] $\mathbb{1}(StatOpen)_{itk}$ takes value 1 if year t is k $(-k)$ years after (before) the first station comes into operation, and 0 otherwise. δ_t, μ_i, and ε_{it} represent year fixed effects, location fixed effects, and idiosyncratic error, respectively. Our focus lies on the η_k's, the coefficients on time indicators relative to the station opening year. One year before the opening of a subway station is the omitted category.

Panel (a) of Figure 7 plots the coefficients η_k's along with their 95% confidence intervals. From the figure, we can visually assess how the location-specific *total patent counts* appear to have changed sharply around the timing of a station opening. Similarly, we also plot in panel (b) of Figure 7 the estimated coefficients when *collaborated patent counts* are used as the dependent variable. In both specifications, all coefficients for the years prior to the entry of a subway station are insignificantly different from zero. Once a subway station opens, we find a statistically significant increase in the location-specific total patent counts and collaborated patent counts for each of the years. The impact does not seem to be transient, as the magnitude of effects tends to grow over time. A reliable causal

[19]Note that 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.

[20]We also experimented with varying this criteria and find consistent and robust evidence.

[21]We approximate the logarithm transformation using either $log(Patents+1)$ or the commonly used alternative, the inverse hyperbolic sine transformation. The inverse hyperbolic sine (arcsinh) transformation converts a random variable x into $\tilde{x} = arcsinh(x) = \log(x+\sqrt{x^2+1})$ (Bellemare and Wichman 2020). Both produce consistent estimates with small quantitative differences.

inference in this and the following grid-pair-level analysis relies on the identifying assumption that conditional on station opening during our sample period and the included controls, the timing of the station opening is exogenous and, hence, there are no pre-trends.[22] This assumption is largely supported by our initial evidence at the grid level.

6.2. Main Results - Patent Collaborations between Grids

6.2.1. Two-way Fixed Effects Model Approach

We now conduct grid-pair level analyses to investigate the effects of subway build-up on collaborative matches. Using variations in the timing of treatment across grid-pairs, we first estimate a difference-in-differences specification as denoted in Equation 4.1. We restrict our sample to all grid pairs based on the grids that were ever located within 2 *km* from a subway station during our sample period, in light of findings in Section 6.1. In columns (1) and (2) of Table 1, the dependent variable is a binary indicator of whether at least one collaborated patent was created between that grid pair in a year. In columns (3) and (4), we use the logarithm of collaborated patent counts in a year at the grid-pair level to capture the extent of the magnitude. The treatment status $\mathbb{1}$(Treated) equals to 1 starting from the year in which nearby subway stations open in both of the two grids in a pair; it equals to 0 otherwise.[23] Across all specifications, we find that the treated grid pairs show a higher probability of patent collaborations, as well as an increase in the number of collaborated patents, compared to the controlled grid pairs.[24]

The identifying assumption is the exogeneity of the timing of station openings which, more specifically in this case, requires that the timing of station opening does not respond to time-varying unobserved features at the grid-pair level. The exogenous nature of the intertemporal variations in the subway expansion was highlighted in Figure 7. To further test the pre-trend assumption at the grid-pair level, we plot the lead and lag coefficients in Figure 8 based on a similar exercise as in column

[22]Among other things, a station opening that is preceded by an increase in local innovation activities, or a station opening caused by the rising demand from patent collaborations would violate this assumption.

[23]For example, if grid *A* and grid *B* experience an opening of a nearer station, respectively, in year 2008, then that grid pair is defined to be treated from 2008 until the end of our sample period.

[24]In Table A3, we consider an alternative definition of the treatment status based on a reduction of travel time. That is, we classify a grid pair to be treated starting from the year onwards when the travel time reduces by more than 0.5 hours. We again find statistically significant and positive treatment effect across all specifications based on this alternative treatment definition.

(4) of Table 1, but using a more detailed event study set-up. We again find that there does not appear to exist any pre-trend and, hence, the exogeneity assumption is likely satisfied.

Next, instead of using a binary treatment status, we estimate the "gravity-equation"-like specification as defined in equation (4.1). Table 2 summarizes our findings, in which we use the same sample and dependent variables. Conditional on year fixed effects, origin fixed effects, and destination fixed effects, a longer *Length* between a pair of grids decreases the probability of collaborations in columns (1) and the count of collaborated patents in column (3). Such effects are statistically significant at 1% level. These estimates are consistent with our intuition that, without controlling for grid-pair fixed effects, as the distance of the fastest subway travel route (i.e. *Length*) becomes longer, the probability and extent of collaborations would decrease. Once we control for origin-destination fixed effects in columns (2) and (4), we find that a larger *Length* increases both the probability and extent of collaborations. This is because conditional on the origin-destination fixed effects, a longer *Length* implies greater connectivity due to more subway lines and stations being built.

For ease of interpretations, we also report first-stage evidence on the impact of subway expansions on travel time in Table A4. We adopt three different sets of assumptions on travel speed as elaborated earlier. We find consistent evidence that a longer *Length* is associated with longer travel time, but once the origin-destination fixed effects are controlled for, longer *Length* reduces the travel time between locations. Based on the first-stage result, we report the Wald estimators that show the extent to which an hour reduction in travel time contributes to either the probability of patent collaborations (i.e. columns 1 and 3) or the count of collaborated patents (i.e. columns 2 and 4). We find that an hour reduction in travel time increases collaborated patents by approximately 15% to 38% on average in column (4), depending on the offline travel speed assumptions.

6.2.2. Instrumental Variable Approach

Our findings based on the difference-in-differences specification and the "gravity-equation"-like specification show consistent results and are validated by the parallel trends observed *prior to the treatment*. However, there may still be concerns on whether unobserved factors *after the treatment* are correlated with variations in the timing of the treatment. To further alleviate such concerns, we now implement the instrumental variable approach where we instrument for *Length* for a grid pair using the interactions of their Euclidean distance and the total cumulative length of subway lines in Beijing.

Before reporting the results from the instrumental variable approach, we first investigate the heterogeneous nature of the treatment effect. That is, we examine whether the effect from improved connectivity seems to have differed across grid pairs depending on their distance. If the extent of the returns from enhanced connectivity is heterogeneous across grid pairs, our instrumental variable approach then identifies the local average treatment effect subject to the identifying assumptions (Imbens and Angrist 1994). In columns (1), (2), (4), and (5) of Table 3, we repeat our two-way fixed effects model but classify grid pairs as either 'near" or "far." This distinction is based on a threshold, which is the mean Euclidean distance between a pair of grids in our data. We expect that far grid pairs are more likely to experience a larger reduction in travel time and thus greater benefits. The estimates from the two-way fixed effects model indeed confirm that this is the case. The effect is much larger and statistically significant at 1% for far grid pairs, compared to that for near grid pairs.

In columns (3) and (6) of Table 3, we now present the local average treatment effect estimated using the instrumental variable approach. Consistent with our previous findings in Section 6.2.1, improved connectivity increases both the probability and extent of patent collaborations. More specifically, column (6) shows that an hour reduction in travel time increases the number of collaborated patents by 35% to 82% on average, depending on different offline speed assumptions. Note that the magnitude of estimated impact is much larger than what we found earlier in Table 2 using the two-way fixed effects model. This may suggest that the "complying" grid pairs are disproportionately located farther apart from each other than that of the average location pairs in the data, which is also consistent with our prior.

6.3. Mechanisms

To reveal more on the mechanisms, we start by identifying collaborated patents created by a pair of incumbent inventors (I-I type) versus patents collaborated with at least one new inventor (other type). This exercise helps to assess the importance of the *Marginal Inventors Channel* relative to the other channels. We report our findings in Table 4, using the same instrumental variable approach but conducting analyses separately for different subgroups based on types. Specifically, we use samples of collaborator pairs which both inventors are incumbents in columns (1) to (3), and the collaborator pairs of other types in columns (4) to (6). Evidence suggests that the impact of subway expansion on patent collaborations are mainly driven by the incumbent pairs. Specifically, an hour reduction in travel time leads to a 29% to 68% increase in patent collaborations among the I-I type, while the corresponding impact for other types is relatively small. This suggests that perhaps the effect from the *Marginal Inventors Channel* is limited.

Next, to shed light on the relative importance of the rest of the channels, we create subgroups of collaborator pairs based on the productivity types. We first track the cumulative number of patents created by each applicant in a year. Then we classify each applicant's productivity type as either high or low, using the median cutoff value in each year. In Table 5, the H-H type in columns (1) to (3) refers to collaborator pairs for which both applicants are the high productivity type, and other types in columns (4) to (6) refer to all other collaborator pairs. Our two-way fixed effects model in columns (1), (2), (4), and (5) again provides the consistent finding that the effect is much more pronounced for far grid pairs, as opposed to near grid pairs. We further estimate the local average treatment effect in columns (3) and (6). The magnitude of effect is approximately 5.5 times larger for patents collaborated by H-H types, as opposed to those by other types. The evidence reveals that the reduced-form evidence of subway expansion on patent collaborations is mainly driven by either the *High-Quality Matches Channel* or the *Relocation Channel*.

In an attempt to further tell apart the *High-Quality Matches Channel* from the *Relocation Channel*, in Table 6, we estimate the local average treatment effect using four subgroups of collaborator pairs. The subgroups are defined based on two productivity types (i.e. H-H type vs other type) and two mobility types (i.e. non-movers vs. movers) of collaborators. Productivity type is defined in the same way as in Table 5. As for mobility, we track each applicant's location across time in our patent data to see whether relocation ever occurred during our sample period. The non-movers type consists of all collaborator pairs whose locations remained the same as before in a given year. The movers type consists of pairs for which at least one collaborator moved in that year. In Appendix C, we provide a detailed description on how we assign movers and non-movers. Table 6 contains eight columns where the first four columns are based on one method of identifying movers first, and the remaining four columns are based on an alternative method of identifying non-movers first. A consistent finding across both methods is that the local average treatment effect is most pronounced for the H-H non-movers group, followed by other non-mover group. An hour reduction in travel time increases the number of patent collaborations by H-H non-movers by 22% - 68% on average depending on travel speed assumptions, whereas the effect is by 0.8% to 12% for other non-movers. As for the other two groups involving movers, we find that the effect is either mostly statistically insignificant or marginally significant with very small magnitude. The evidence suggests that the *High-Quality Matches Channel* dominates the *Relocation Channel* in explaining the impact of subway expansion on patent collaborations.

The evidence so far suggests that the *High-Quality Matches Channel* seems to be the 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 at least one applicant who is highly productive in that year, while the L-types category consists of all applicants being the low productive type in that year. Note that 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 7 reports our findings for each group, considering dynamic responses. Once a subway station opens and lines are continually added to the subway system, the intensity of treatment as well as its impact vary dynamically over time. Therefore, we divide the treatment period into short run and long run using the first five year as the cutoff. Across columns (1) to (4), we do not find any statistically significant impact during the first five years of the treatment period.[25] In the long run (i.e. after 5 years), however, statistically significant effect is present across all groups. More specifically, we find that the *High-Quality Matches Channel* in column (3) contributes to approximately 70% of the total increase in patent collaborations, while the *Relocations Channel* and the *Marginal Inventors Channel* in columns (1) and (2) account for approximately 21% of the increase.

[25]Note that this result is also consistent with the evidence presented in Figure 7.

In column (5) to (8), we further consider the aspect that some types of applicants may face mobility or institutional restrictions and this may underestimate the relative importance of the *Relocation Channel* and the *Marginal Inventors Channel*. Thus, we drop all pairs that involve universities and repeat our analysis to see whether this is the case. Columns (6) and (7) show that now there is some statistically significant effect that supports the presence of the *High-Quality Matches Channel* and the *Marginal Inventors Channel* in the short run. As before, we find the evidence that all four channels exist in the long run. Yet, as we expect, the relative contribution from the *High-Quality Matches Channel* now becomes smaller and is reduced to 65%. In contrast, the *Relocations Channel* and the *Marginal Inventors Channel* in columns (5) and (6) jointly account for approximately 25% of the increase.

6.4. Control Pair Approach

We adopt an alternative research design to verify robustness of our main findings. Using the control pair method proposed in Jaffe et al. (1993) and adopted widely in follow-up studies such as Agrawal et al. (2017) and Ganguli et al. (2020), 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, such as application year and similarity of the technological space. 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:

$$\mathbb{1}(CollabPatents)_{ijpt} = \beta \log(Length)_{ijpt} + \theta_{ij} + \lambda_t + \xi_{ijpt}, \tag{6.3}$$

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 locations i and j in which the matched collaborators reside; θ_{ij} represents location-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 8, 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 origin-destination 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 origin-destination 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.

In Table 9, we further examine whether a collaborator pair is more likely to form given an improvement in connectivity using different subsamples. We use the subsample of H-H pair type in columns (1) to (4) and the remaining other pair type in columns (5) to (9), where each column uses a different threshold for identifying the high vs. low productivity type. Controlling for origin-destination fixed effects across all specifications, we find that enhanced connectivity improves the collaboration probability for the H-H type and this effect is consistent and statistically significant at 1% across columns (1) to (4). However, when we look at the other type in columns (5) to (8), effect is no longer statistically significant. This finding is again highlights the complementarity between travel time and inventors' productivities. In Table A6, we repeat the same analyses as in Table 9, but instead use one-to-two matching as opposed to one-to-one matching. Using a larger number of observations, we similarly find that the impact of connectivity on the collaboration probability is more consistent and stronger for the H-H type pair subsample, compared to that for the other type pair subsample.

7. 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 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 treatment effect being approximately 34.92% to 82.29%. 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, we also find that the entry of new inventors, relocation of existing inventors, and low productive inventors also contribute to the increase in collaborative matches, especially in the long run.

Whereas this paper establishes an important link between connectivity and collaborative matches in the context of 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 that are deemed to have 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.

Agrawal, A., A. Galasso, and A. Oettl (2017). Roads and Innovation. *Review of Economics and Statistics 99*(3), 417–434.

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.

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.

Clark, G. (2009). Recession, Recovery and Reinvestment : the role of local economic leadership in a global crisis. Technical report, OECD.

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.

Diamond, R. (2016). The determinants and welfare implications of US Workers' diverging location choices by skill: 1980-2000.

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*.

Ganguli, I., J. Lin, and N. Reynolds (2020). The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences. *American Economic Journal: Applied Economics 12*(2), 278–302.

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

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.

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.

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.

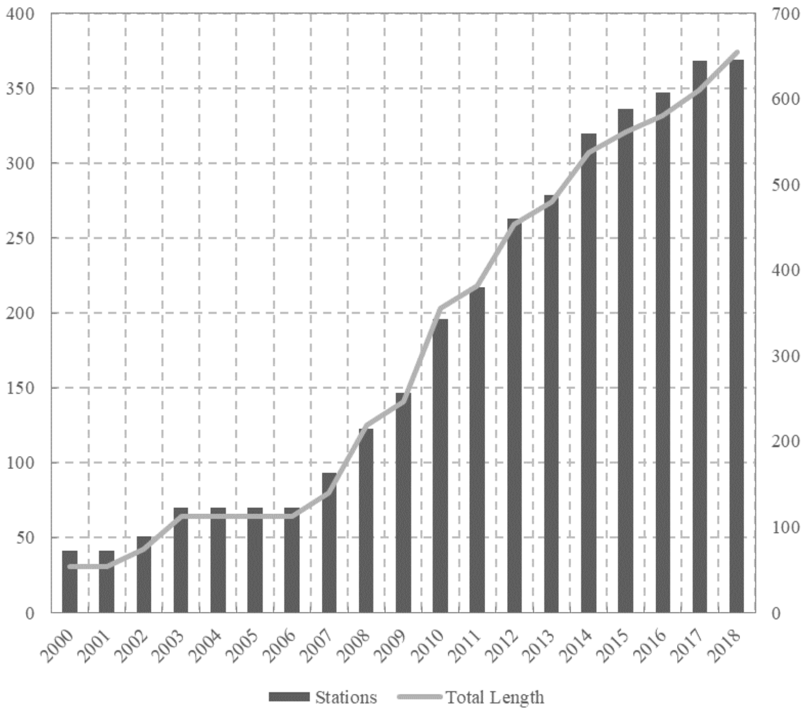
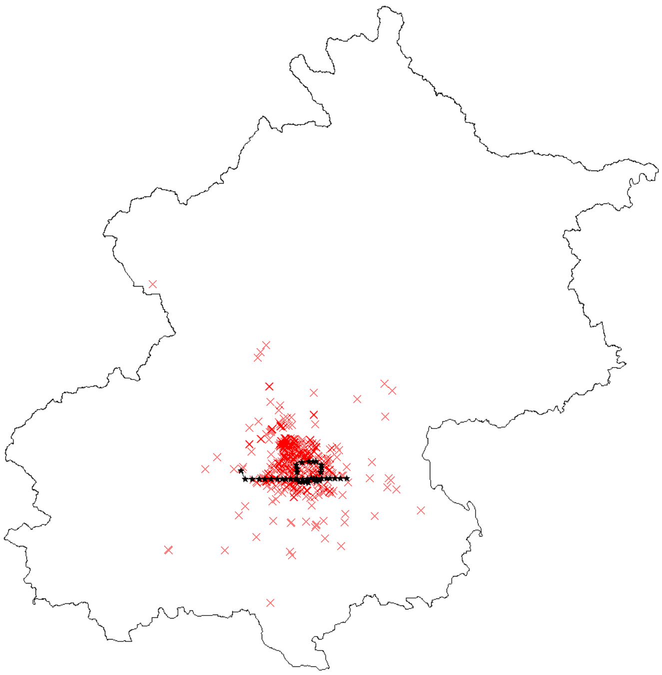
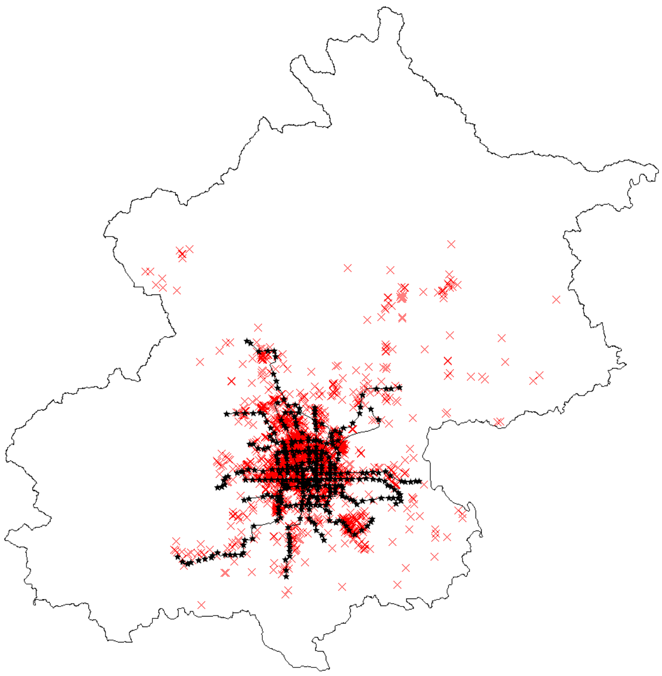


Figure 1: Cumulative Count of Subway Stations and Length of Subway Lines in Beijing

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.



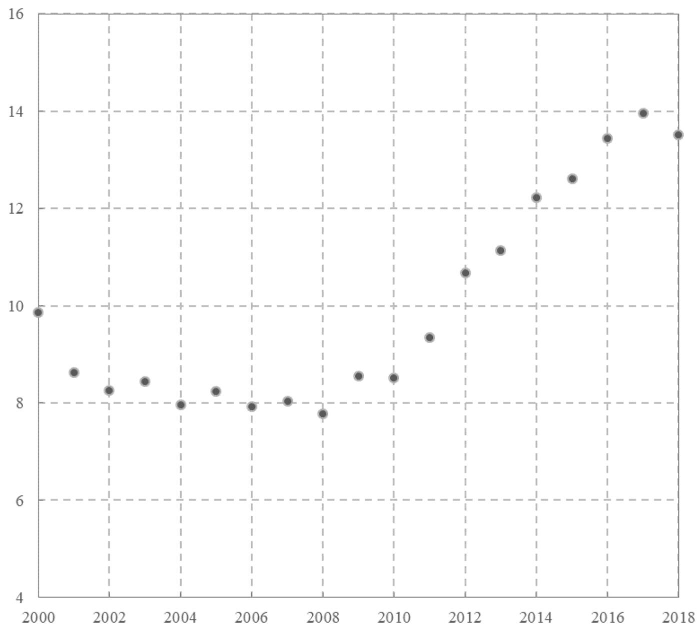
(a) 2000 and 2001



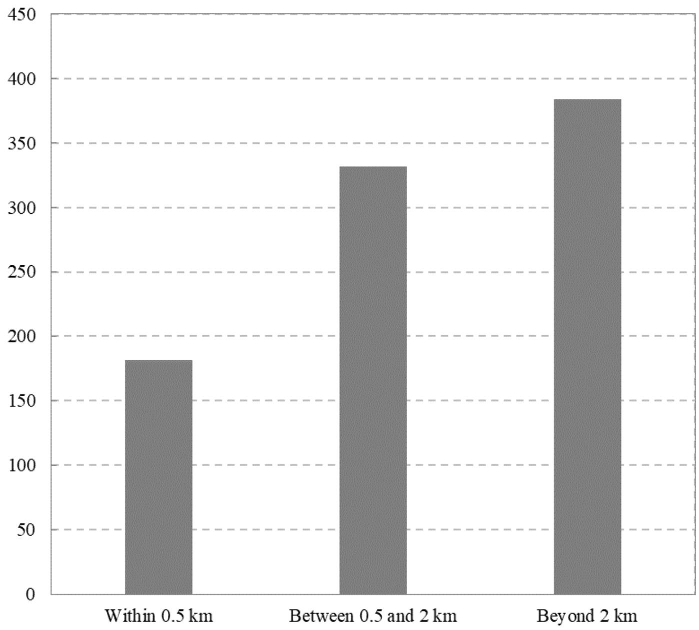
(b) 2017 and 2018

Figure 2: Spatial Patterns of Patents 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.



(a) Average Distance between Collaborators



(b) The Growth Rate of Collaborated Patents by Distance Bands

Figure 3: Spatial Patterns of Collaboration 2000-2018

Notes: In panel (a), the average distance (*km*) between all pairs of collaborators who jointly produced a patent in that year are plotted. Panel (b) shows the growth rate of cumulative count of collaborated patents from 2000 to 2018, conditional on different thresholds of distances between collaborators.

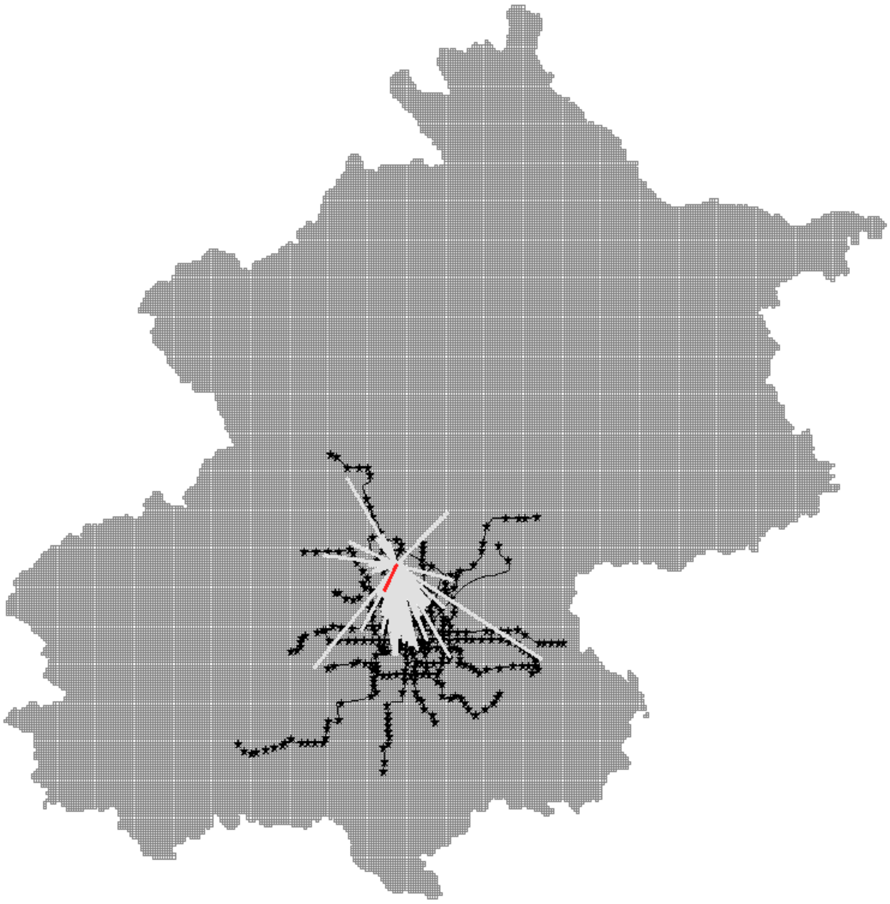
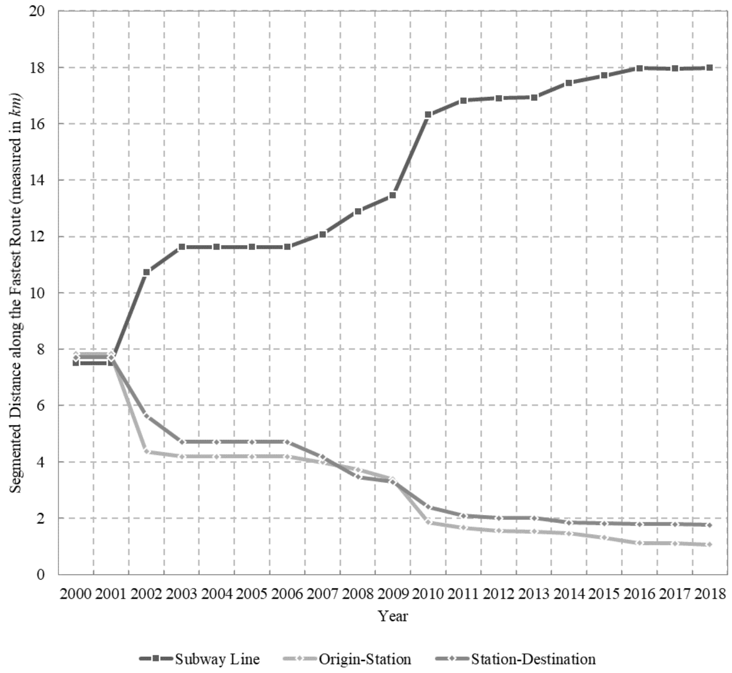
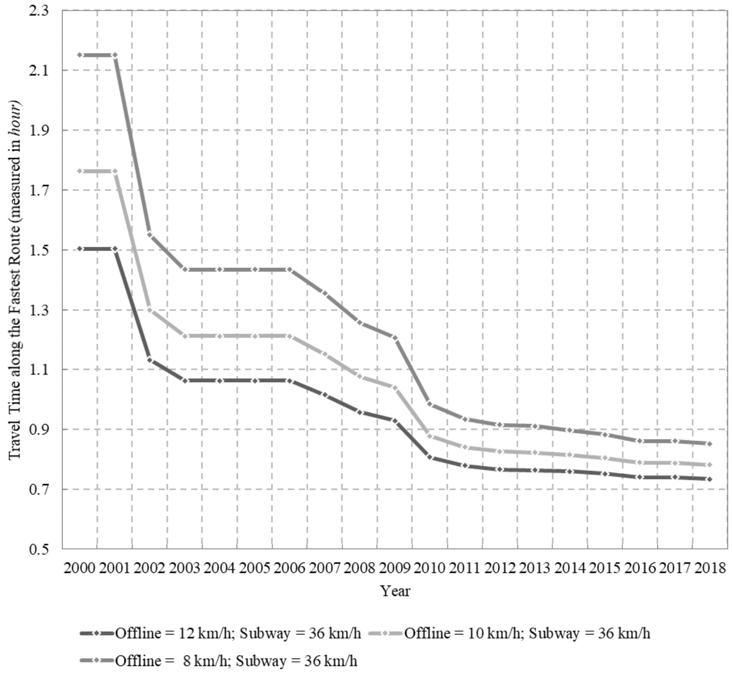


Figure 4: Time-Varying Collaboration Patterns

Notes: The figure 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.



(a) Segmented Distance along the Fastest Route



(b) Travel Time along the Fastest Route

Figure 5: 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.

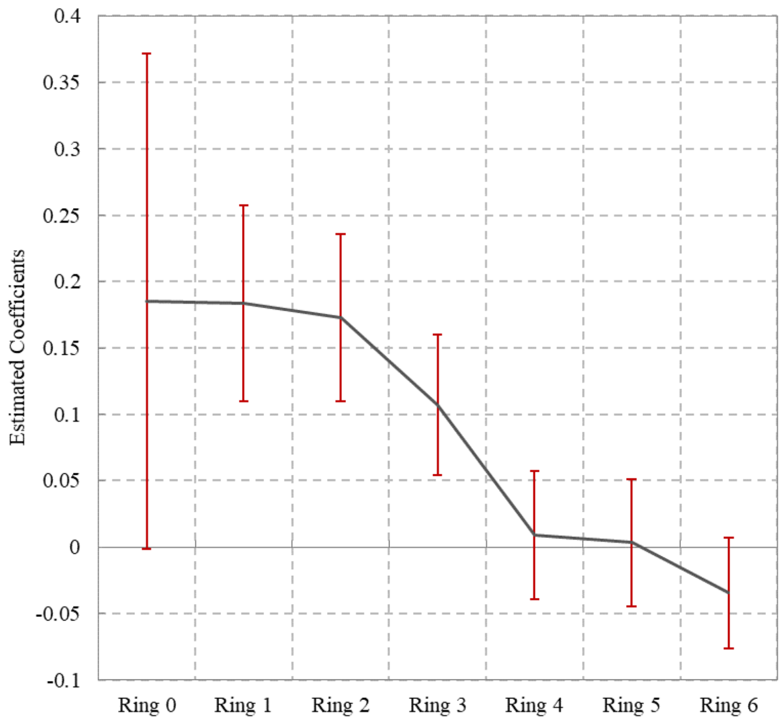
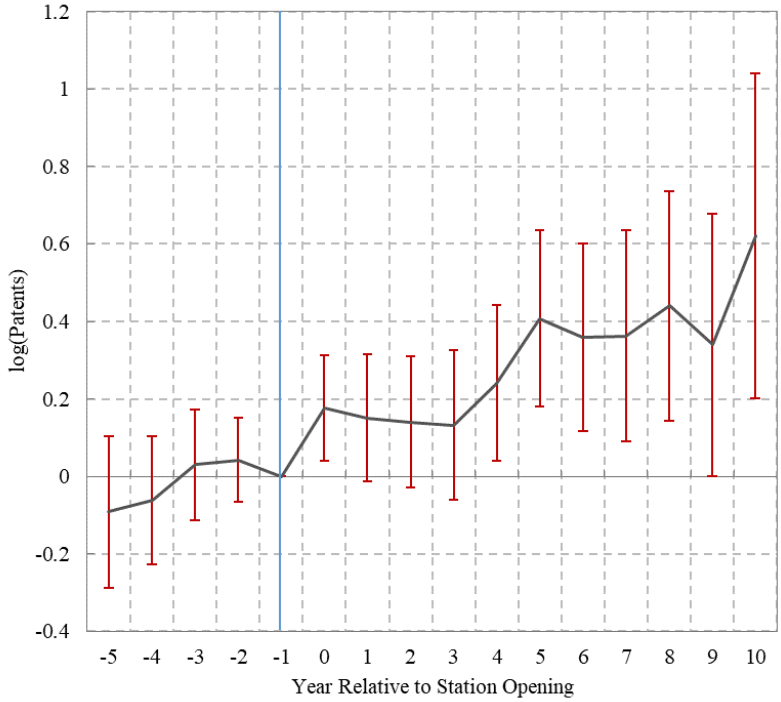
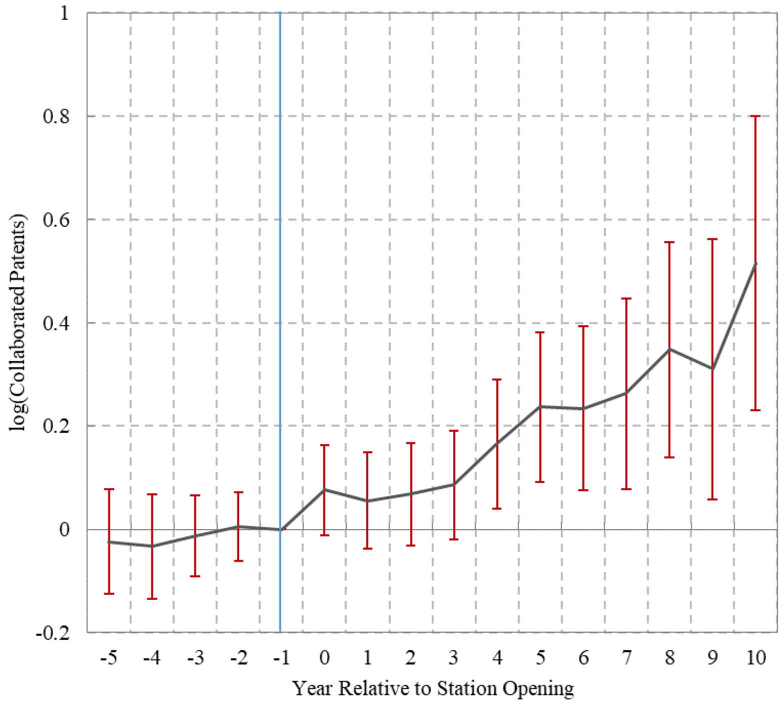


Figure 6: Ring-specific Coefficients

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 A2.



(a) Total Patent Counts



(b) Collaborated Patent Counts

Figure 7: Time-Varying Effects of Subway Station Opening on Patents and Patent Collaborations (Grid Level)

Notes: The estimated effects of years relative to subway station opening are plotted, along with bars representing the 95% confidence intervals. One year before the opening of a subway station is the omitted category.

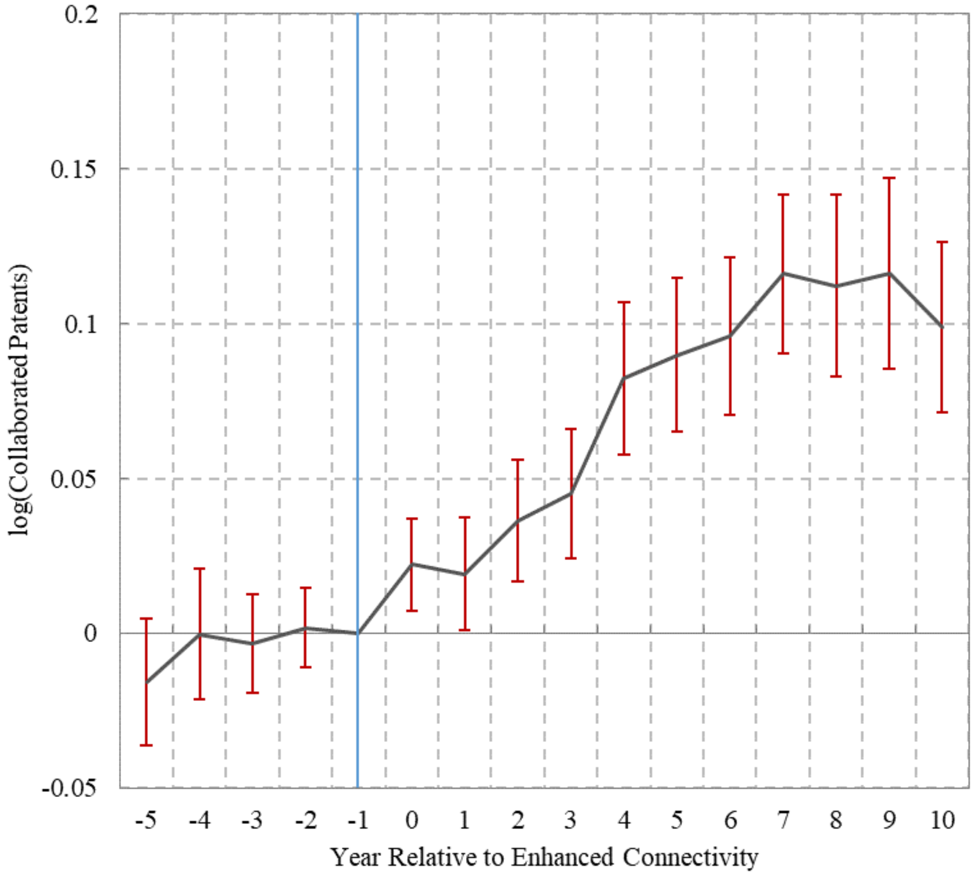


Figure 8: Time-Varying Effects of Subway Station Opening on Patent Collaborations (Grid-Pair Level)

Notes: The estimated effects of years relative to subway enhanced connectivity are plotted, along with bars representing the 95% confidence intervals. The year when connectivity for a grid pair enhances is defined to be the year when a subway station that is closer by is built in each of the two grids. One year before the enhanced connectivity is the omitted category.

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

Dep. variable	𝟙(Collab. patents)		log(Collab. patents)	
	(1)	(2)	(3)	(4)
𝟙(Treated)	0.0173***	0.0092***	0.0361***	0.0207***
	(0.0025)	(0.0031)	(0.0063)	(0.0069)
Year FE	YES	YES	YES	YES
Origin FE	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO
Origin-Destination 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. For a grid pair, 𝟙(Treated) equals one for the first and the following years when a subway station that is closer by is built in each of the two grids; 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	$\mathbb{1}$(Collab. patents)		log(Collab. patents)	
	(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
Origin FE	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO
Origin-Destination 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 Treatment Effect (LATE)

Dep. variable	𝟙(Collab. patents)			log(Collab. patents)		
Sample	Near grid pairs	Far grid pairs	All grid pairs	Near grid pairs	Far grid pairs	All grid pairs
Estimation	Two-way FE	Two-way FE	LATE (IV)	Two-way FE	Two-way FE	LATE (IV)
	(1)	(2)	(3)	(4)	(5)	(6)
log(Length)	0.0011*	0.0249***	0.0139**	0.0019	0.0558***	0.0432***
	(0.0007)	(0.0072)	(0.0061)	(0.0016)	(0.0165)	(0.0151)
Wald Estimators Based on Travel Speed Assumptions of ...						
Offline = 12 km/h; Subway = 36 km/h	-	-	-0.2648	-	-	-0.8229
Offline = 10 km/h; Subway = 36 km/h	-	-	-0.1691	-	-	-0.5255
Offline = 8 km/h; Subway = 36 km/h	-	-	-0.1124	-	-	-0.3492
Kleibergen-Paap rk Wald F statistic	-	-	239.508	-	-	239.508
Year FE	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES
Observations	105,014	82,915	188,081	105,014	82,915	188,081
Root MSE	0.251	0.264	0.251	0.430	0.406	0.416

Notes: 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 (6), 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. 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 Different Types of Collaborators—Grid Pair-Level Analysis
Heterogeneity and Local Average Treatment Effect (LATE)

Dep. variable	log(Collab. patents by I-I type)			log(Collab. patents by other types)		
Sample	Near grid pairs	Far grid pairs	All grid pairs	Near grid pairs	Far grid pairs	All grid pairs
Estimation	OLS	OLS	LATE (IV)	OLS	OLS	LATE (IV)
	(1)	(2)	(3)	(4)	(5)	(6)
log(Length)	0.0022	0.0698***	0.0353**	0.0006	-0.0001	0.0093***
	(0.0019)	(0.0169)	(0.0151)	(0.0004)	(0.0041)	(0.0031)
Wald Estimators Based on Travel Speed Assumptions of ...						
Offline = 12 km/h; Subway = 36 km/h	-	-	-0.6723	-	-	-0.1771
Offline = 10 km/h; Subway = 36 km/h	-	-	-0.4294	-	-	-0.1131
Offline = 8 km/h; Subway = 36 km/h	-	-	-0.2854	-	-	-0.0752
Kleibergen-Paap rk Wald F statistic	-	-	249.244	-	-	249.244
Year FE	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES
Observations	95,529	80,970	178,182	95,529	80,970	178,182
Root MSE	0.415	0.382	0.395	0.182	0.159	0.167

Notes: We define an inventor as an "incumbent" in a year if the inventor has ever created a patent in the past. Otherwise, we define the inventor as a "new" inventor. We label the "incumbent-incumbent" inventor collaboration pairs as the "I-I type" and the rest of the collaborative matches as "other types." The sample selection criteria is based on the following: if the *Length* between the grids of collaborators is below the median of the entire distribution, we assign the pair as "Near grid pair." Otherwise, we assign the pair as "Far grid pair." 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 of Subway Expansion on Different Types of Collaborators—Grid Pair-Level Analysis
Heterogeneity and Local Average Treatment Effect (LATE)

Dep. variable	log(Collab. patents by H-H type)			log(Collab. patents by other types)		
Sample	Near grid pairs	Far grid pairs	All grid pairs	Near grid pairs	Far grid pairs	All grid pairs
Estimation	Two-way FE	Two-way FE	LATE (IV)	Two-way FE	Two-way FE	LATE (IV)
	(1)	(2)	(3)	(4)	(5)	(6)
log(Length)	0.0026	0.0726***	0.0357**	0.0001	-0.0034	0.0065**
	(0.0019)	(0.0174)	(0.0155)	(0.0003)	(0.0036)	(0.0030)
Wald Estimators Based on Travel Speed Assumptions of ...						
Offline = 12 km/h; Subway = 36 km/h	-	-	-0.6800	-	-	-0.1238
Offline = 10 km/h; Subway = 36 km/h	-	-	-0.4343	-	-	-0.0791
Offline = 8 km/h; Subway = 36 km/h	-	-	-0.2886	-	-	-0.0525
Kleibergen-Paap rk Wald F statistic	-	-	249.244	-	-	249.244
Year FE	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES
Observations	95,529	80,970	178,182	95,529	80,970	178,182
Root MSE	0.419	0.377	0.395	0.165	0.168	0.161

Notes: We define the type of an inventor based on the cumulative number of patents created by the inventor in the past years. We assign the productivity type of an inventor as "high" if the cumulative number of patents exceed the median of the distribution. Otherwise, we classify the productivity of an inventor as "low." The "H-H" type thus consists of inventors whose productivity types are both high, while the rest of collaborative matches are classified as "other types." 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 "Near grid pair." Otherwise, we assign the pair as "Far grid pair." 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 of Subway Expansion on Collaborators of Different Productivity and Mobility—Grid Pair-Level Analysis
Local Average Treatment Effect (LATE)

Grouping method	Method 1: Identify non-Movers First				Method 2: Identify Movers First			
Dep. variable	log(Collab. patents)							
	by H-H non-Movers	by Other non-Movers	by H-H Movers	by Other Movers	by H-H non-Movers	by Other non-Movers	by H-H Movers	by Other Movers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Length)	0.0277***	0.0011*	0.0125	0.0056*	0.0356**	0.0064**	0.0002	0.0002
	(0.0058)	(0.0006)	(0.0145)	(0.0029)	(0.0155)	(0.0030)	(0.0002)	(0.0002)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-0.5276	-0.0210	-0.2381	-0.1067	-0.6781	-0.1219	-0.0038	-0.0038
Offline = 10 km/h; Subway = 36 km/h	-0.3370	-0.0134	-0.1521	-0.0681	-0.4331	-0.0779	-0.0024	-0.0024
Offline = 8 km/h; Subway = 36 km/h	-0.2239	-0.0089	-0.1011	-0.0453	-0.2878	-0.0517	-0.0016	-0.0016
Kleibergen-Paap rk Wald F statistic	249.244	249.244	249.244	249.244	249.244	249.244	249.244	249.244
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	178,182	178,182	178,182	178,182	178,182	178,182	178,182	178,182
Root MSE	0.178	0.031	0.359	0.158	0.395	0.161	0.01134	0.01041

Notes: In columns (1) to (4), we first identify all pairs of inventors whose locations always remained the same during our sample period as "non-Movers," and the rest as "Movers." In columns (5) to (8), we alternatively first identify all pairs that include at least one inventor who collaborated after moving as "Movers," and the rest as "non-Movers." If each inventor's cumulative number of patents exceeds the median value of patent count distribution, we classify the pair as the "H-H" type. Otherwise, we label the pair as the "Other" type. 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: Dynamic Effects Across Years for Subgroups—Grid Pair-Level Analysis

Dep. variable	log(Collab. patents)							
Sample	Full Sample				Restricted Sample			
	Movers	Entrants	H-types	L-types	Movers	Entrants	H-types	L-types
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}$(within 5 years)	0.0008	0.0038	0.0027	0.0009	0.0007	0.0056**	0.0106*	0.0013
	(0.0006)	(0.0028)	(0.0064)	(0.0009)	(0.0006)	(0.0027)	(0.0059)	(0.0008)
$\mathbb{1}$(after 5 years)	0.0018**	0.0062*	0.0269***	0.0033***	0.0015**	0.0053*	0.0176*	0.0025**
	0.0008	(0.0032)	(0.0098)	(0.0012)	(0.0007)	(0.0031)	(0.0091)	(0.0011)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	178,182	178,182	178,182	178,182	178,182	178,182	178,182	178,182
R-squared	0.057	0.052	0.272	0.070	0.057	0.056	0.277	0.071

Notes: 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 a subway station that is closer by is built in each of the two grids; and zero otherwise. Similarly, $\mathbb{1}$(after 5 years) equals one starting from the sixth year onward since the treatment; and zero otherwise. In the "Restricted Sample" in columns (5) to (8), 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.

Table 8: Impact of Subway Expansion on Probability of Collaboration—Collaborator Pair-Level Analysis

Dep. variable	$\mathbb{1}$(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
Origin FE	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO
Origin-Destination 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 9: Impact of Subway Expansion on Probability of Collaboration—Collaborator Pair-Level Analysis by Productivity Types

Dep. variable	𝟙(Actual collaborator pair)							
	Method: One-to-One Match							
Sample	H-H type				Other type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cutoff for Productivity	50%	75%	90%	99%	50%	75%	90%	99%
log(Length)	0.0074***	0.0082***	0.0082***	0.0229***	0.0234	-0.0039	0.0074	0.0035
	(0.0027)	(0.0027)	(0.0031)	(0.0070)	(0.0218)	(0.0086)	(0.0045)	(0.0026)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	110,823	106,471	97,445	61,638	1,843	6,003	14,696	50,978
R-squared	0.746	0.749	0.759	0.826	0.815	0.818	0.805	0.753

Notes: In columns (1) to (4), we look at collaborative patents produced by H-H type pairs. To define the H-H productivity type, we use different cutoffs: 50%, 75%, 90%, and 99%. In columns (5) to (9), we look at collaborative patents produced by other pairs which are not the H-H types. We use the "One-to-One Match" sample, for which we use one randomly selected control pair 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.

Appendix

A. Additional Results

Table A1: 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 A2: Impact of Station Opening on Patents —Grid-Level Analysis (Spatial Decay)

Dep. variable	log(Patents)						
	(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 $log(patents_{it} + \sqrt{patents_{it}^2 + 1})$, where $patents_{it}$ captures the number of all patent applications in year t whose reported address lies within grid i. 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 A3: Treatment Effect on Patent Collaboration—Alternative Definition of Treatment Status based on Travel Time

Dep. variable	𝟙(Collab. patents)		log(Collab. patents)	
	(1)	(2)	(3)	(4)
𝟙(Treated)	0.0162***	0.0129***	0.0331***	0.0335***
	(0.0027)	(0.0031)	(0.0069)	(0.0078)
Year FE	YES	YES	YES	YES
Origin FE	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO
Origin-Destination 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. For a grid pair, 𝟙(Treated) equals one for the first and the following years when there is a reduction in travel time by more than 0.5 hours; 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 A4: 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
Origin FE	YES	NO	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO	YES	NO
Origin-Dest. 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 A5: Impact of Subway Expansion on Patent Collaboration—Alternative Distance Thresholds for Sample Selection

Dep. variable	𝟙(Collab. patents)		log(Collab. patents)	
	(1)	(2)	(3)	(4)
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
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
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
Origin FE	YES	NO	YES	NO
Destination FE	YES	NO	YES	NO
Origin-Destination 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 A6: Impact of Subway Expansion on Probability of Collaboration—Collaborator Pair-Level Analysis by Productivity Types

Dep. variable	𝟙(Actual collaborator pair)							
	Method: One-to-Two Match							
Sample	H-H type				Other type			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cutoff for Productivity	50%	75%	90%	99%	50%	75%	90%	99%
log(Length)	0.0080***	0.0106***	0.0081***	0.0187***	0.0377*	0.0132	0.0101**	0.0025
	(0.0026)	(0.0026)	(0.0028)	(0.0064)	(0.0218)	(0.0107)	(0.0042)	(0.0025)
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Origin-Destination FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	168,207	161,200	147,395	93,052	3,358	9,817	23,051	78,430
R-squared	0.722	0.724	0.732	0.802	0.851	0.835	0.819	0.738

Notes: In columns (1) to (4), we look at collaborative patents produced by H-H type pairs. To define the H-H productivity type, we use different cutoffs: 50%, 75%, 90%, and 99%. In columns (5) to (9), we look at collaborative patents produced by other pairs which are not the H-H types. We use the "One-to-Two Match" sample, for which 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.

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 Table 6, we classify whether collaborative patents in a year were produced by movers or non-movers. We use two methods to identify the "movers" and "non-movers" category. The first approach relies on first identifying 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.[26] Based on this first approach, "non-Movers" pair in columns (1) and (2) of Table 6 refer to a pair of applicants who were identified to be both non-movers in that particular year. The "movers" pair in columns (3) and (4) of Table 6 refer to all other remaining pairs.[27]

[26]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.

[27]Such pairs include cases in which at least one applicant being a mover in that year, having lack of information to conclude since there is only a single observed patent by an applicant, etc.

In the second approach, we now alternatively first identify the pairs which contain at least one applicant who moved for sure in that particular year. Such pairs are now classified as "movers" in columns (7) and (8) of Table 6, while all remaining other pairs are classified as "non-movers" in columns (5) and (6). As before, this second method also relies on comparing 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.[28]

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.[29] 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

[28] 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.

[29] 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.

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 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*).