Subway, Collaborative Matching, and Innovation*

Yumi Koh†, Jing Li‡, Jianhuan Xu§

Abstract

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 14.85% to 37.69% increase in collaborated patents. Far-apart location pairs were more affected, and the local average causal response is approximately 34.92% to 82.29%. The effect is mainly driven by increased matches among highly productive innovators. The entry of new innovators, relocation of existing innovators, and collaborations among low-productive innovators also contribute to the increase in collaborative matches, especially in the long run.

Keywords: innovation; collaborative matching; patent; Beijing subway; transportation

JEL classifications: D83, O18, O31.

---

*We thank Co-Editor, Professor Daniel Yi Xu, and the four anonymous referees for their insightful and constructive comments. We are also grateful to Xiaofang Dong, Rui Du, Yuming Fu, Richard Hornbeck, Jungho Lee, Jeffrey Lin, Sunha Myong, Binzhe Wang, and Daniel Da Cunha for their helpful comments and suggestions. This work was supported by the Singapore Ministry of Education (MOE) Academic Research Fund Tier 1 Grant (C244/ MSS20E002). Jing Li also gratefully acknowledges the financial support from Singapore Management University under the Lee Kong Chian Fellowship. Aaron Zhe Yang Lim provided excellent research assistance. Remaining errors are our own.

†School of Economics, University of Seoul; ymkoh@uos.ac.kr.

‡School of Economics, Singapore Management University; lijing@smu.edu.sg.

§School of Economics, Singapore Management University; 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), 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 mostly 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 connectivity and matches among people will also be affected. 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. 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's total subway length increased from 54 *km* to 655 *km* and became the world's longest and busiest rail transit network. At the same time, patent collaborations within Beijing increased substantially in terms of both scale and geographic scope, as we discuss further 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.

One important empirical aspect to address carefully is that selection of locations for building subway stations may be nonrandom. Such correlation invalidates the standard ordinary least squares assumptions and results in biased estimated coefficients. We address this issue

[1] 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 is projected to serve 18.5 million trips every day by 2021 (`http://www.chinadaily.com.cn/china/2017-01/12/content_27931764.htm`).

with a collection of efforts. We fully exploit both cross-sectional and intertemporal variations in connectivity between locations and control for two-way fixed effects that absorb additive location-pair-specific and time-specific unobserved characteristics. Accounting for such unobserved features, we estimate 1) a discrete difference-in-differences (DiD) specification; 2) a continuous "gravity-equation"-like specification; and 3) a two-stage least squares specification by adopting an instrumental variable (IV) approach.

In our DiD estimation, we define a location-pair to be treated if the pair-specific connectivity improves substantially due to the subway expansion—a reduction in travel time by more than 30 minutes along the fastest travel route. Our aim is to characterize the extent to which the number of collaborated patents formed between treated location-pairs increases relative to those formed between location-pairs that were not treated. Given the staggered treatment timing, we implement the DiD estimation as in Callaway and Sant'Anna (2021). 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 locations to be inversely proportional to the travel time between those locations. In our third approach, we instrument for connectivity between two locations using the interaction of their Euclidean distance and the citywide aggregate expansion of the subway network. The identification assumption is that the subway expansion changes the connectivity of far-apart locations more than that of nearby locations and this differential impact affects patent collaborations only through reduction in travel costs.[2] Lastly, as robustness checks, we adopt 1) the control pair approach following Jaffe et al. (1993) and 2) a counterfactual design using planned but unbuilt routes, as in Duranton and Turner (2011) and Donaldson (2018).

Another challenge in evaluating the impact of subway expansion is to distinguish productivity growth from reorganization of existing economic activities. Even if subway stations were randomly assigned to locations, it is still difficult to distinguish whether collaborative matches increased due to a location-specific productivity growth or a spatial reshuffling of innovation ac-

[2]This instrument is analogous to the traditional Bartik shift-share instrument in the sense that the exogenous distance between locations serves as the pre-determined Bartik weight that governs differential exposures to a common aggregate shock.

tivities (Redding and Turner 2015).[3] To investigate the roles of growth versus reorganization, as well as other possible mechanisms, we present a matching model that explicitly accounts for spatial sorting and solves for innovators' collaboration, location, and occupation choices in response to connectivity between locations. The model conceptualizes how the subway expansion produces spatial variations in returns to innovations and affects matching between innovators across different locations.

The model suggests that collaborative matches increase in response to improved connectivity and also yields the presence of the following mechanisms that we can test empirically. First, the *High-Quality Matches Channel* suggests that more collaboration matches would be formed, especially among highly productive innovators, due to complementarity between travel time and their productivity. Second, the *Marginal Matches Channel* shows that a lower cost of collaboration would induce new collaborative matches to be formed, especially those that involve low-value innovator pairs on the margin. Third, the *Relocation Channel* suggests that innovators, especially the more productive ones, move to places that become more accessible and such relocation drives the formation of collaborative matches. Fourth, the *Marginal Innovators Channel* shows that as reduced travel time increases the returns to innovation and induces entry of new innovators, collaborative matches increase from a larger pool of innovators. Because each mechanism generates predictions for different subgroups of innovators, we conduct empirical tests to assess the relative importance of the channels.

We obtain the following findings. First, our DiD estimation shows a positive and statistically significant treatment effect, indicating a larger increase in the number of patents collaborated between locations whose connectivity improved. Second, an improvement in connectivity between two locations, proxied using the total length of subway lines that connect a pair of locations along their shortest-time travel path, increases the number of patents collaborated between those two locations. The Wald estimators suggest that an hour reduction in travel time leads to an increase in collaborated patents from 15% to 38% on average, depending on travel speed assumptions. Third, the impact of subway expansion on patent collaborations is highly

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

heterogeneous, with the effect being much larger for far-apart location pairs. The IV approach reveals an estimated local average causal response among "complying" location pairs, which translates into Wald estimators ranging from 34.92% to 82.29%. As for our mechanism analysis, we find that while all four channels are present in the long run, the *High-Quality Matches Channel* is quantitatively most important.

Our study contributes to the literature in the following aspects. First, it draws a unique linkage between intra-city transport infrastructure and collaborative matches in innovations to enhance our understanding of economic integration in driving innovation-based urban growth (Acemoglu 2008; Carlino et al. 2007; Jones 2009; Kerr and Robert-Nicoud 2020). Recent studies have focused on how *inter-city* transport infrastructure (e.g. cheap airline connections, high-speed rail developments, bridging islands) affects research paper publication among academic scholars and patent collaborations among inventors (Bernard et al. 2020; Catalini et al. 2020; Dong et al. 2020). Our study extends the literature by further looking into the role of an *intra-city* transport (i.e. a subway system). Although both contexts share many similarities, there are clear distinctions and novelties.

To begin with, majority of collaborations take place within a city rather than across different cities. Inventive activities are highly concentrated in cities and distance has not become irrelevant (Paunov et al. 2019; Carlino et al. 2007). Specifically, "[t]he difficulty of sharing complex tacit knowledge and the advantages of colocation that innovators critically need (finance, qualified human capital, etc.) may explain why innovation is so concentrated in top cities" (Paunov et al. 2019, p.5). In an intra-city setting, it is much easier and cheaper for innovators to move around and to engage in more frequent face-to-face interactions. Therefore, we believe our intra-city transport analysis provides an ideal empirical setting to quantify the magnitude of impact from frequent in-person interactions on collaborative innovations. For instance, in Catalini et al. (2020) using an inter-city context, copublications at an individual-pair-year level equaled 0.11 and entry of a low-cost airline increased the number of collaborations by 0.3 and 1.1 times. In our intra-city setting, an average of 1.317 collaborated patents were produced within 0.5*km* by 0.5*km* sized piece of land within Beijing in a single year, and an hour reduction in travel time between a pair of locations brought 14.85% to 37.69% increase in collaborated patents. In another study based on labs on the Jussieu campus of Paris, Catalini (2018) shows that colocation increases the likelihood of joint collaboration by 3.5 times, suggesting the importance of geographic proximity and clustering. Although empirical contexts vary across studies and more careful investigation is needed to make a feature-by-feature comparison, the number of collaborative innovations generated in response to infrastructure investment seems to be substantial in an intra-city context.

In addition, intra-city transportation is less about moving goods but more relevant for moving people compared to inter-city transportation. As summarized in Redding and Turner (2015) (p. 1342), "[r]ailroads affect production more than the population . . . , while the spatial organization of the population is more sensitive to roads and subways than to railroads". This implies that the build-up of an intra-city transportation will have more direct impact on the distribution of innovators and subsequent spatial collaboration patterns. Our study adds value to the literature by tracking time-varying spatial distribution of innovation activities at a more fine-grained geographical level and shows how a city develops as as an innovation hub. On top of the channels residing on the complementarity between connectivity and productivity that were previously highlighted in an inter-city context, we also incorporate additional channels more suited to the intra-city context, such as the mobility of innovators.

Second, our theme on collaborative matches in cities is 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).[4] 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 in the context of *inventions and innovations* (Berkes and Gaetani 2021; 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

[4]The positive spillover effects are also documented for clusters of universities and research institutions (Hausman 2022; Jaffe 1989; Kantor and Whalley 2014; Li et al. 2021; Liu 2015).

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.[5] 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. 2022; Gu et al. 2021). Apart from these directly targeted outcomes, our paper shows that a subway network potentially entails much broader economic consequences, such as improving collaborative matching. Quantifying such indirect but important economic consequences has become crucial for policy-makers, as they are hard-pressed to make thorough and comprehensive net-benefit justifications given skyrocketing construction and maintenance costs.[6] Our back-of-the-envelope calculations show that the increase in collaborative innovations in Beijing do account for a significant share of benefits arising from the build-up of the subway system.

2. Institutional Background

Beijing's planning committee proposed subway construction in 1953 (Sultana and Weber 2016). The initial purpose was to transport soldiers to the city center, but the subway eventually became means of public transportation.[7] The subway system developed slowly until the 2000s and then went through a major expansion since 2008. 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 residential areas (Yang et al. 2013). Moreover, with

[5]By 2014, 171 cities worldwide have subway in operation (Gendron-Carrier et al. 2022).

[6]For instance, the estimated cost of the Long Island Rail Road project 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`).

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

the central government's stimulus package during the global financial crisis of 2007-2008, the Beijing subway system further expanded to connect suburban districts. Figure 1 shows that between 2000 and 2018, the cumulative count of subway stations grew from 41 to 379 and the total subway length increased from 54.1 *km* to 655 *km*.

The extraordinary expansion of the Beijing subway was accompanied by a remarkable increase in collaborated patents, as well as an expansion of their geographical scope. Panel (a) in Figure 2 illustrates the locations of all collaborators of patents between 2000-2001 in gray x-marks. The subway stations available back then are marked in black stars. The vast majority of collaborators were centrally clustered around the east-west line and the inner loop line covering Beijing's CBD area. However, the spatial layout of collaborators greatly expanded by 2017-2018, as shown in Panel (b). A striking feature is that the locations of collaborators closely overlap with the layout of the subway system. To illustrate development of spatial connections in patent collaborations across time, Panel (c) shows how the collaboration links evolved for a particular location that had the largest number of collaborated patents in 2000-2001. Yet, all collaborated patents originated in this location during 2000-2001 were formed with collaborators in another single location, and this link between two locations is shown by a gray line. However, by 2017-2018, this location formed numerous collaboration links, as shown by a set of many gray lines connecting this location with other locations. Again, we find that these collaboration links closely overlap with the layout of the subway lines, as shown in black star marks. Motivated by this graphical evidence, we now discuss empirical research designs that allow us to investigate the causal impact of spatial connectivity on patent collaborations.

3. Empirical Research Designs

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

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

where ij indexes a pair of grids and t indexes a year. Grids refer to locations of a specific size that we assign in the data and we explain this further in Section 4.1. $CollabPatents_{ijt}$ is the count of collaborated patents produced by innovators in grids i and j in year t. Throughout, we use collaborated patents to proxy for collaborative matches in innovation. $Connect_{ijt}$ measures the subway build-up that enhances the connectivity between grids i and j in year t, which we proxy using either a binary variable or a continuous variable. The binary variable captures the treatment status in which connectivity between two grids improves as travel time reduces by more than 30 minutes along the fastest travel route due to the subway expansion. The continuous variable captures connectivity using the total length of the subway lines that connect the two grids along their shortest-time travel path. Lastly, γ_{ij}, ζ_t, and υ_{ijt} represent grid pair fixed effects, year fixed effects, and idiosyncratic error term, respectively.

The TWFE model has been conventionally adopted in the literature to identify the average treatment effect on the treated (ATT). However, recent studies find that a TWFE model does not yield an interpretable causal parameter in a setting with staggered treatment timing and heterogeneous treatment effects (de Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021). In our context, the timing of subway station openings rolled out across years and effects are expected to be heterogeneous. Hence, we report the estimates from the TWFE model while turning to other recent methodologies.

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

Whereas the above procedure delivers an interpretable coefficient, there may be concerns that the intertemporal subway layout may not be purely exogenous. For instance, policymakers may have planned the phase-in of the subway construction in company with other concurrent place-based policies to promote innovations at particular locations. Two locations may have been linked via the build-up of a subway line, while policy incentives were also offered to attract innovators at such locations. In such a scenario, pairwise connectivity would be positively correlated with collaborative innovations driven by the presence of unobserved policy incentives at the location level. However, we argue that related concerns are unlikely to severely drive our results since our identification exploits the *exogenous timing* of the subway build-up. Beijing subway system is notorious for unexpected construction delays due to various technical obstacles.[8]. To the extent that the actual completed construction year differs significantly from the planned completion year in a random manner, the timing of subway station opening may still be deemed as exogenous. We also find supporting evidence for the exogenous timing assumption in Figure 3. The parallel trend assumptions are satisfied during the pre-treatment period, regardless of whether we look at the impact of subway station opening at the grid level or the impact of improved connectivity at the grid pair level.

Nevertheless, we adopt other research designs to further corroborate our findings. In our second research design, we adopt a composite variable to instrument for $Connect_{ijt}$ by combining two sources of variations: 1) the aggregate development of the subway system over time, and 2) the time-invariant Euclidean distance between two locations. The first component reflects the "aggregate shift" as the subway system grows over time.[9] The second component reflects the "exposure" for a grid pair to the subway system development. The identifying assumption is that differential exposure to the common subway development shock affects collaborated patents only through the channel of enhanced connectivity, and not through other potential confounding channels related to location selections. Our IV setup is somewhat analogous to the Bartik instrument (Bartik 1992) in that it shares the shift-share component structure. That is, we project the citywide aggregate subway development to the connectivity of grid pairs,

[8]See news reports: `https://wemp.app/posts/8977e79a-26af-429e-84f5-a910399cb516` and `https://m.news.cctv.com/2020/11/21/ARTImLRnVxYC55PyJAns0pc1201121.shtml`. More detailed discussions can be found in Beijing Metro Operation Co., Ltd. (2011)

[9]We also construct an alternative leave-one-out aggregate shift measure since including own observation may cause finite sample bias.

with their spatial distance being used as weights.[10]

The relevance of the IV in our context reflects the notion that far-apart grid pairs will be more affected by the subway expansion, given a larger reduction in travel time. If one were to consider an extreme example of two adjacent grids, travel time would be irrelevant to the subway expansion. More generally, we argue that the Euclidean distance serves as a reasonable exposure measure, given the particular layout of subway lines and the general economic activities in Beijing. Namely, Beijing's subway is known for its "very orderly and grid-like" layout (Ingraham 2015).[11] To illustrate our point, we consider two types of layouts in examples (a) and (b) in Figure OA2 in the Online Appendix. Points, thin lines, and thick lines represent locations, roads, and subway lines, respectively. Example 1 has an orderly and grid-like layout compared to that in Example 2. Under the assumption that the subway reduces travel time by one fourth of the time it takes to travel by road, we compute the relationship between Euclidean distance and reduction in travel time due to subway operation for all 36 combinations of location pairs given 9 location points. The corresponding bar graph for Example 1 shows that the average percentage reduction in travel time increases with the Euclidean distance, whereas such relationship does not hold in Example 2. Indeed, given the orderly and grid-like layout of Beijing subway, Panel (c) in Figure OA2 shows that the magnitude of reduction in travel time between 2000 and 2018 increased with Euclidean distance for all grid pairs in our sample. Therefore, the Euclidean distance serves as a reasonable exposure measure in our context for Beijing, although that may not necessarily applicable to other cities.

One important identification issue that needs to be addressed is that our exposure measure using Euclidean distance, despite being exogenous in nature, could still be non-random. Borusyak and Hull (2021) show that composite variables consisting of exogenous shocks but non-random exposure measures may still introduce omitted variable bias in either a reduced-

---

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

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

form or an IV setting. We argue that endogeneity concern on the selection of grid pairs is unlikely to be correlated with the Euclidean distance between those grids.[12] However, exposure could be non-random if, for example, grid pairs with large (small) Euclidean distance have a tendency to be in peripheral (center) areas of the city. Then the identification worry is that such spatial patterns may affect collaborative innovations through other confounding channels. To address such concerns, we conduct the following two analyses.

First, we use the pre-treatment period in our empirical setting and follow Goldsmith-Pinkham et al. (2020) to test for parallel pre-trends to assess the exogeneity of the exposure measure. Specifically, we interact the Euclidean distance between grid pairs with pre-treatment period dummies up to 6 years prior to the treatment, while controlling for grid pair fixed effects and year fixed effects. We find supporting evidence that our exposure measure is likely to be exogenous (see Figure OA3 in the Online Appendix). Second, we further investigate the relationship between the centrality of grid pairs and their Euclidean distance. For each grid pair, we calculate the Euclidean distance between each one of the two grids and the Tiannanmen Square (i.e. the central location of Beijing), respectively, and compute their average distance. Then we plot the kernel density of average distance to the city center for grid pairs. We find that the top quintile seems to exhibit a centrality measure that is different from the rest, suggesting that the grid pairs with very large Euclidean distance tend to be concentrated in peripheral areas (See Figure OA4 in the Online Appendix). Based on such evidence, we drop either the top 10% or 20% of observations with large Euclidean distance as robustness checks in Online Appendix E.1 and find that results are still consistent in Table OA9 .

One additional caveat with the IV approach is that, in the event of heterogeneous causal

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

responses, the two-stage least squares estimates produce the local average causal response. Depending on the extent of heterogeneity, the local average causal response could be very different from the average causal response which is of more general and inherent interest. Nevertheless, if the concern is on the potential endogeneity which may overstate the presence of the causal response, then the local average causal response at least identifies the presence of the causal response among the compliers. In our setting, the "complying grid pairs" are likely to be the far-apart grid pairs based on the design of the instrument. Thus, the IV approach corrects the bias of the estimated average causal response for such far-apart "complying grid pairs."

Last, we adopt another two empirical designs as robustness checks. We implement the matched control approach proposed in Jaffe et al. (1993) to compare the actual collaboration pairs to matched control pairs and see how they respond to changes in connectivity. The identification assumes independence of unobserved errors conditional on observed matching characteristics. In addition, we also exploit the fact that a small set of stations were planned but were not built during our sample period, which allow us to construct "counterfactual" stations in a DiD setup (Donaldson 2018; Duranton and Turner 2011). To the extent that the selection criteria were incorporated into the original plan of subway networks and the construction for some planned stations was delayed unexpectedly, such unbuilt stations could serve as a valid control set. Both approaches produce results that are consistent with our main findings and we present detailed discussions in Online Appendix Sections E.2 and E.3, respectively.

4. Data and Variables

Our empirical analysis relies on two 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 have information on its coordinates, opening year, and 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 capture the effects of the subway expansion on collaborative matches in innovations, we focus on patent *applications*, as opposed to approved patents.[14] Although not all applications end up being approved, submitting an application is the initial crucial step in seeking protection for innovations. It also signals the formation of a collaboration pair in the knowledge creation process. Moreover, there is usually a long and irregular delay before applications are finally approved. The time of application submission better indicates when new knowledge was created and formalized.[15] We thus rely on the time when patent applications were submitted in drawing inference with the subway build-up time in our empirical analysis. Throughout this paper, we will refer "patent applications" as "patents" for short.

4.1. Location Grids

Our empirical analysis requires spatial units that correspond to a discrete set of locations. We divide 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 of each patent in our database.

4.2. Innovators and Collaborations

An innovator is "a person or company that introduces new things, ideas or ways of doing something" (Oxford Learners' Dictionaries). In the patent database, we have applicants and inventors listed for each patent. Inventors are always individuals, whereas applicants can be

---

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

[14] As a robustness check, we conduct our empirical analysis using approved patents only in Online Appendix E.7.

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

firms, research institutes, individuals, and/or other entities. In our empirical analysis, we use applicants to represent innovators. One reason is because patents assigned to firms usually take up a dominant share in recent decades (Akcigit et al. 2022). Another reason arises from the data constraint, as locations of inventors are unidentified in the data. However, locations of applicants can be identified and we provide a detailed discussion in Online Appendix C on how we assign applicants to grids using information in the patent database. Later in Section 6.3, we also discuss implications of focusing on applicants as opposed to inventors.

The main outcome of interest is patent collaborations across time and space. In the data, we identify a patent with multiple applicants as a "collaborated patent." Our key dependent variable $CollabPatents_{ijt}$ is the count of patents whose applicants were located in each one of the paired grids (i, j) in a given year (t). Since this variable is grid pair specific, we create all possible pairwise applicant combinations for each collaborated patent. For example, suppose a patent has 3 applicants located in grids A, B, C, respectively, in year 2010. Then this patent would be counted in the variable $CollabPatents$ for the grid pairs (A, B), (B, C), and (A, C) in year 2010. Given that we have 65,542 grids in total, the number of grid pairs becomes exponentially large if we 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. In Section 5.1, we report spatial evidence as to why we chose 2 *km* as the cutoff.

4.3. 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 the subway system expands, connectivity of innovators across different grids changes 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 grid likely involved no subway travels and thus $Length$ equals to 0. If more subway lines are added such that the fastest travel route involves 5 *km* of subway travel, $Length$ equals to 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.

Although we use *Length* as the key explanatory variable in our reduced-form regressions, we consider an alternative variable, *TravelTime*, as a first-stage outcome. *TravelTime* captures the amount of time required to travel from the centroid of one grid to that of another grid in a year along the fastest route. Apparently, this variable is the object to minimize when we construct the *Length* variable. To calculate *TravelTime*, we restrict the mode of transportation to only subway travels (to travel from one station to another) and "offline travels." By offline travels, we refer to various ways (e.g. by foot, cycling, bus) to travel between centroid of a grid to a station, or directly between centroid of the two grids. Although it would be ideal to calculate travel time based on all possible transportation modes, we do not have complete time-varying transport networks that include all modes of travel. Instead, we adopt various travel speed assumptions: 36 km/h for the subway speed and a range of 8 km/h to 12 km/h for the offline travel speed.[16] We further assume that individuals travel along the Euclidean distance between the centroid of a grid to the corresponding subway station.[17]

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

[16]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.

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

a structural interpretation.

Lastly, in Figure OA1 in the Online Appendix, we illustrate how segmented distance and travel time along the fastest route have changed over time with the expansion of the subway system. We find that the average *Length* increased over time and average travel time decreased, as more lines were built and connectivity improved.

4.4. Summary of Statistics

In Table 1, we present summary of statistics. On average, there were 42,877 patents in a year and approximately 9,223 of them were collaborated patents with multiple applicants. Among 785,629 patent applications that we have between 2000-2018, the average application year is 2013. This suggests that the number of patents grew across years during the sample period. The average number of applicants per patent is 1.283 and collaborated patents contain a team of approximately 2.271 applicants on average. The average logarithm of the Euclidean distance between grids in a year is approximately 2.340 *km* and the logarithm of the length of the subway lines that connect two grids along their shortest-time travel path is 8.352 *km* on average. In addition, we find that State Grid Corporation of China, produced the largest number of collaborated patents (i.e. 33,016) during our sample period. China Petroleum & Chemical Corporation or Sinopec produced the next largest number of collaborated patents (i.e. 31,153), followed by BOE Technology Group Co., Ltd. which produced 10,979 collaborated patents. Therefore, we find that the top applicants were all firms.

5. Results

5.1. Initial Evidence at the grid level

Although our focus is on patent collaborations at the grid pair level, we present some initial evidence on whether subway station openings had an impact on patents at the grid level. We define a grid to be treated starting from the year when a subway station opens in the grid and zero otherwise. The estimates from the standard TWFE model and the staggered DiD model both suggest that the subway station opening had a positive and statistically significant impact on both the probability and extent of having patents. For instance, based on the DiD approach, we find that the ATT on collaborated patents is 19.3%. Detailed results can be found in Table OA1 in Online Appendix.

In Figure 3, we plot the event study estimates. In the left (right) figure in Panel (a) Grid Level, the dependent variable is collaborated patents (total patents) in its inverse hyperbolic sine transformation (hereinafter IHS) transformation. We use the logarithmic-like inverse hyperbolic sine transformation to address observations with zero values.[18] In both figures, the grid-specific patent counts change sharply with the opening of a subway station, which is in line with our assumption that the timings of the subway station openings are exogenous. All coefficients during the pre-treatment period are insignificantly different from zero, which supports the parallel trend identification assumption.

In addition, we conduct a complementary analysis at the grid level to assess the spatial scope of the subway impact. In our grid pair analysis that follows in Section 5.2, constructing every possible combination of grid pairs is computationally infeasible. Therefore, we identify spatial attenuation patterns and use the results to restrict the set of grids to be used in constructing the grid pair sample. More specifically, we estimate the following specification:

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

where subscripts i and t denote grid and year, respectively. The superscript r denotes distance rings based on various distance cutoffs.[19] The dependent variable is the total patent counts at grid i in year t, in its IHS transformation. $\mathbb{1}(StatOpen)_{it}^r$ is a binary indicator that takes value 1 if t is greater than or equal to the year when the first station in ring r comes into operation, and 0 otherwise. δ_t, μ_i, and ε_{it} represent year fixed effects, grid fixed effects, and the idiosyncratic

[18]This concave log-like transformation allows retaining zero-valued observations. The coefficient estimate yields a similar interpretation to that of a standard logarithmic specification (Bellemare and Wichman 2020).

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

error term, respectively. Figure 4 plots coefficients ρ^r with their 95% confidence intervals. We find spatial attenuation patterns, with the impact of subway station openings on patent counts being positive and statistically significant up to ring 3 (i.e. approximately within 2 *km*) and such effects dying out beyond 2 *km*.[20] Thus, we construct grid pairs using all grids ever located within 2 *km* from a subway station during the sample period, but also check robustness of our main findings using other distance cutoffs in Appendix E.4.

5.2. Patent Collaborations at the Grid Pair-Level

We now present our main findings at the grid pair level. We use two outcomes for the dependent variable. One is a binary indicator that captures the existence of collaborated patents in a grid pair in a year. The other one is the count of collaborated patents in a grid pair in a year in its IHS transformation. The binary indicator $\mathbb{1}$(Treated) equals one starting from the year when there is a reduction in travel time by more than 30 minutes between a grid pair along the fastest travel route, and zero otherwise.[21] Panel A of Table 2 shows the estimates from the TWFE model. As for the dependent variable, we use a binary variable (continuous variable) to capture the existence (magnitude) of collaborated patents in Columns (1) and (2) (Columns (3) and (4)). The TWFE model estimates show that the treatment effect increases both the probability of having collaborated patents and the number of collaborated patents.

Panel B of Table 2 shows the estimates from the staggered DiD model as in Callaway and Sant'Anna (2021), using the never-treated cohort as the control group.[22] We find the ATT to be positive and statistically significant across specifications, and the ATT on collaborative

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

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

[22]One may question whether the never-treated cohort serves as a reasonable counterfactual. However, since our sample of grid pairs consists of those that were ever located within 2 *km* from a subway station during our sample period, the never-treated cohort is likely to share many similarities with the treated cohort in various aspects.

patents is 2.94%. Moreover, the TWFE and DiD estimates are quite comparable.[23] In Panel (b) Grid Pair Level in Figure 3, we plot the even-study estimates. To check the robustness of our treatment definition, we define treatment to be a reduction in travel time by more than 30 minutes (1 hour) in the left (right) panel. Both figures show support for the parallel trend assumption.

Next, we estimate Equation (3.1) using the continuous *Length* variable. Conditional on year fixed effects and grid fixed effects, Table 3 shows that a longer *Length* between a pair of grids significantly decreases the probability of collaborations in Columns (1) and the count of collaborated patents in Column (3). Consistent with our prior, the probability and extent of collaborations decreases as the distance of the fastest subway travel increases in the absence of grid pair fixed effects. Once we control for the grid pair fixed effects in Columns (2) and (4), we find that a longer *Length*, which implies greater connectivity due to more subway lines and stations being built, increases both the probability and extent of collaborations.

For ease of interpretations, we report the Wald estimators in Table 3 that show the extent to which an hour reduction in travel time affects the probability of patent collaborations in Columns (1) and (3), or the count of collaborated patents in Columns (2) and (4). Such estimates are derived based on the first-stage estimation of how subway expansions affect travel time, which is reported in Table OA3 in Online Appendix. The Wald estimators in Column (4) suggest that an hour reduction in travel time increases collaborated patents by 14.85% to 37.69% on average, depending on the offline travel speed assumptions.[24] In Online Appendix

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

[24]The Wald estimator is the ratio of the impact of *Length* on collaborative patents (Table 3) and the impact of *Length* on travel time (Table OA3).

E.7, we repeat our analysis using the sample of *approved* collaborated patents and confirm consistency of our findings.

To corroborate our findings, we further instrument for *Length* using the interaction of Euclidean distance for a given grid pair and the total cumulative length of subway lines in Beijing. The IV approach identifies the local average causal response subject to the identifying assumptions, when the extent of returns from enhanced connectivity is heterogeneous across grid pairs (Imbens and Angrist 1994). As an initial check, we repeat our TWFE analyses but classify grid pairs into either "near" or "far" samples, using the mean Euclidean distance between grid pairs as the cutoff. Results in Columns (1), (2), (5), and (6) of Table 4 show that the effects are heterogeneous; the far apart grid pairs experience a larger reduction in travel time and show a larger increase in collaborated patents.[25]

In Columns (3), (4), (7), and (8) of Table 4, we present the local average causal response obtained from the IV approach. We find that improved connectivity significantly increases both the probability and extent of patent collaborations. For instance, Column (7) shows that an hour reduction in travel time increases the number of collaborated patents by 34.92% to 82.29% on average, depending on different offline speed assumptions. In Columns (4) and (8), we alternatively construct a leave-one-out (LOO) aggregate shift since including own observation may cause a finite sample bias. We obtain similar findings based on this alternative approach.

In Online Appendix E.2, we discuss the implementation and results of the matched control approach as proposed in Jaffe et al. (1993). In Online Appendix E.3, we use a small set of planned but unbuilt stations to construct "counterfactual" stations in a DiD framework (Donaldson 2018; Duranton and Turner 2011). In both analyses, we find consistent results.

[25]Although the evidence is documented at a granular geographic scale within a city, the pattern is consistent with the cross-city study of Perlman (2016) showing that transportation has a significant effect especially for counties that were not well connected previously.

6. Mechanism Analysis

6.1. Discussion of the Mechanisms

Based on our main findings, we further analyze where the increase in collaborated patents is coming from. Are certain types of matches being generated more in response to improvements in connectivity? If so, what are the characteristics of such matches and their underlying cause? To answer such questions, we present a formal conceptual model in Online Appendix B, but summarize the model and mechanisms in words here.

In the model, a city consists of N locations and a unit of potential innovators who are mobile across locations. To maximize utility, they first decide whether to innovate and where to locate. Then they learn their productivity levels and further decide whether to collaborate with other innovators. A pair of innovators, who are characterized by their productivity levels respectively, in two different locations encounter one another randomly. The encounter probability depends on the density of innovators in each of the two locations for a given level of productivity, and location pair-specific degree of interactions. Upon encountering, the probability of having a successful collaboration depends on innovators' productivity and travel time between their locations. Based on important regularities on positive assortative matching, highly productive innovators are more likely to be matched among themselves in the model (Bernard et al. 2020; Davis and Dingel 2019). Having a shorter travel time increases the likelihood of a successful collaboration since innovators can have more frequent face-to-face interactions, which facilitates their coordination on exchanging information, skills, and ideas (Behrens et al. 2014; Combes et al. 2012; Davis and Dingel 2019; Lee and Xu 2020).[26] Under the super-modularity

[26]Another explanation is that improved connectivity increases location-specific market access which spurs innovation collaborations (Melitz and Redding 2021). However, we think this is less relevant in our within-city context because the market for innovative products usually exceeds narrowly defined geographic locality. Moreover, Perlman (2016) explicitly controls for market access in studying the impact of railroad network connections on innovation activities and finds modest evidence that market access explains the increase in patent activity, but most of the relationship is explained by other variables correlated with transportation access.

assumption, the probability of a successful collaboration increases more among highly productive innovators, as travel time decreases.

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

The equilibrium generates an endogenous distribution of innovators and the extent of pairwise collaborations across locations. We can solve for the closed-form solution under some functional form assumptions. Given characterization of the model, we derive the following channels that give a rise to collaborative matches as travel time decreases.

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

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

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

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

6.2. Empirical Evidence

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

We first distinguish the *Marginal Innovators Channel* from the three channels. Whereas other three channels all commonly predict increased collaborations among *existing* (i.e. incumbent) innovators, the *Marginal Innovators Channel* highlights the role of *new* innovators. If we find that most collaborations are formed between existing innovators, then the *Marginal Innovators Channel* is unlikely to be the main driving force.

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

Lastly, among existing innovators who did not relocate, the *High-Quality Matches Channel* predicts most of collaborations arising from highly productive pairs of innovators, whereas the *Marginal Matches Channel* predicts the exact opposite. Therefore, depending on which productivity type of innovators become more active in forming collaboration pairs, we can further identify the dominant mechanism.

In Tables 5 and 6, we carry out analyses as outlined above by looking into different subsamples. Table 5 presents the results from the TWFE and DiD model using the binary treatment variable and Table 6 shows the results based on the IV approach using the continuous *Length* measure. In Panel A of each table, we first classify collaborated patents into those that were created by previously existing applicant pairs (i.e. incumbent pairs) versus those created by pairs involving at least one new entrant applicant. In both tables, Panel A shows that the effect of enhanced connectivity is much more pronounced for incumbent pairs. This implies that the *Marginal Innovators Channel* is not the most important channel.

In Panel B of Tables 5 and 6, we divide collaborated patents into those that were created by pairs of nonmovers versus pairs consisting of movers. Mobility of each applicant is identified by tracking address in the patent data across years. The nonmover pairs consist of all pairs of collaborators whose locations remained the same as before in a given year.[27] Panel B of Tables 5 and 6 shows that there was a greater increase in collaborated patents by nonmover pairs, suggesting that the *Relocations Channel* is not quantitatively most important.

Lastly, in Panel C of Tables 5 and 6, we focus on collaborated patents created by nonmover pairs but further classify the sample into two groups based on the productivity type. In each year, we first track the cumulative count of patents created by each applicant. Then using the median value of the distribution in a year as the cutoff, we assign each applicant's productivity type as either high or low. In Columns (1) and (2), we use the sample of collaborated patents created by nonmover pairs who were both high types. Columns (3) and (4) contain the sample of collaborated patents created by nonmover pairs which include at least one low type. The magnitude of effect is much larger among the high productivity nonmover type, suggesting that the *High-Quality Matches Channel* is quantitatively most important. Such evidence is consistent with either increases in the number of high productive matches on the extensive margin or increases in the productivity of high productive matches on the intensive margin (Bernard et al. 2020).[28]

We conduct additional robustness checks by adopting different definitions of mobility and productivity, respectively. Instead of identifying non-movers first and then assigning the rest to be movers, we alternatively identify movers first and then assign the rest as non-movers in Appendix E.5. As for an alternative definition of productivity, we use the overall patent approval rate between 2000 and 2015 for each applicant in Appendix E.6. Whereas our productivity measure in the main text focused on the quantity aspect of past patent production, this measure is informative of how successful an applicant is in converting applications to final patents to gain exclusive legal right. In both robustness checks, we find that the *High-Quality Matches Channel* seems to be quantitatively most important.

To compare the relative importance of the channels across time, we now classify all collaborative matches into one of the four exclusive categories: movers, entrants, H-types, and

[27] In Online Appendix D, we explain how we identify applicants as movers and nonmovers.

[28] In Online Appendix Table OA6, we provide detailed estimates for all other remaining subgroups created based on the joint criteria of mobility and productivity.

L-types. The movers (entrants) category consists of all collaborative matches for which at least one applicant was a mover (an entrant) in that year. Among non-movers and incumbents, the H-types category consists of all pairs that involve both applicants of the high type in that year, while the L-types category consists of pairs that involve at least one low type in that year. The categories of movers, entrants, and H-types correspond to the *Relocations Channel*, the *Marginal Innovators Channel*, and the *High-Quality Matches Channel*, respectively. The L-types category is slightly broader than what the *Marginal Matches Channel* represents, since it also includes matches that are not necessarily formed on the margin.

Table 7 reports our findings for each subgroup. The binary treatment is again defined as a reduction in travel time by more than 30 minutes along the fastest travel route, but we adopt separate indicators for the short run (within 5 years) and the long run (after 5 years). In Panel A, the short run responses are only present among movers and H-type non-movers, with the effect being more pronounced among the H-type non-movers. In the long run, however, a positive impact exists across all subgroups. One may raise the possibility that some types of applicants face mobility restrictions due to their innate features and this may underestimate the relative importance of the *Relocation Channel*. In Panel B, we thus drop all pairs that involve universities and repeat our analysis. We find that the results still remain consistent as before.

6.3. Other Possible Channels and Policy Implications

As we clarified in Section 4.2, the concept of innovators was empirically captured using applicants, given the importance of firms in innovations market and data limitations. In such a case, there may be concerns on whether using applicants instead of inventors invalidates the claim on relative importance of the channels and/or their interpretations. For instance, it is possible that non-individual applicants experience a faster increase in amenities once accessibility of their locations improves, which leads to a faster growth in their employment size or human capital. This may result in having more collaborative matches among applicants but this may be mostly driven by scale effects rather than the *High-Quality Matches Channel*.

To address this concern, we analyze whether non-individual applicants experience a faster growth in their pool of inventors due to subway expansion, compared to their counterparts that are not treated. For this aim, we use data on solo patents, by compiling the list of unique inventor names for a non-individual applicant in a year. Note that for solo patents, affiliations of all inventors for a patent are clearly identified as there is only a single listed applicant. We find that having a subway station open in its vicinity has no statistically significant impact on an applicant's pool size of inventors. The results are shown in Table OA4 in Online Appendix. Thus, there is lack of evidence that applicants in more accessible places experienced a faster growth in its inventor pool size in response to the subway expansion.

In addition, we test a new specification in which we take our main specification but further separately control for the grid fixed effect interacted with the treatment dummy for each of the two grids in a grid pair. The fixed effects capture unobserved differences in patent collaborations before and after opening of a station at both locations. Such unobserved differences may arise from relocation of applicants in response to improved accessibility of a location once a subway station opens, as we modeled in the paper. Other possibilities include relocation of inventors which results in a growth of location-specific employment size (Monte et al. 2018) and a screening effect that results in an increased productivity at a location (Behrens et al. 2014). We find that the estimated coefficients are slightly smaller in magnitude compared to our baseline estimates but are still statistically significant at conventional levels, as can be seen in Table OA5 in Online Appendix. Thus, the estimates show that even after controlling for those channels, we still obtain clear evidence on the *High-Quality Matches Channel*.

Lastly, we believe our analysis of heterogeneous responses to improved connectivity across different subgroups provides useful policy implications on efficient resource allocations. Given that the most significant increases in matches arise among highly productive incumbent innovators, policy makers may consider designing subway routes that better connect existing clusters of productive innovators if their aim is to promote innovation. Our analysis implies that this would have a larger and more immediate impact on innovations than connecting subway lines that aim at encouraging entry and matching among new innovators. Our analysis also provides compositional implications of infrastructure investment. Although infrastructure investment in our context targets specific geographic locations, we observe that it nevertheless has different degrees of impact on various subgroups of innovators. An accurate assessment on the overall composition of collaborative matches formed from infrastructure investment could complement better designs of other complementary policies aimed at increasing innovations (e.g. providing tax incentives, subsidizing R&D, offering low interest loans).

7. Back-of-the-Envelope Calculation

To compare the economic significance of our estimates with other benefits of subway infrastructure presented in the recent literature, we conduct a simple back-of-the-envelope calculation. We begin by reviewing various benefits from subway documented in recent studies. Gu et al. (2021) study the impact of subway construction on congestion and travel time. They show that the subway system increases the road speed in Beijing by approximately 3%, which implies a decrease of 1.68 minutes on average in commute time. Considering the monetary value of commute time and the number of commutes, the Beijing subway system brought a welfare increase of 265 million in 2019 US dollars due to faster commutes (Gu et al. 2021). In a different study, Gendron-Carrier et al. (2022) analyze the impact of subway construction on pollution reductions using a panel of different cities. They show that a subway opening reduces about 2.0 $\mu g/m^3$ of PM10 on average and argue that a 1 $\mu g/m^3$ decrease in ambient PM10 averts approximately 10 infant deaths per 100,000 births based on estimates from the previous literature. If we were to extrapolate this estimated effect to Beijing with 0.8% of birth rate and 21 million regular residents, the subway system averts approximately 33.6 infant deaths (or equivalently valued at 75 million in 2019 US dollars) per year.[29]

Our estimates show that an hour reduction in travel time increases patent collaborations by 34.92% to 82.29% in Beijing, depending on travel speed assumptions as shown in column (7) in Table 4. During our sample period, the average travel time between grid pairs decreased by 1.30 to 0.77 hours, resulting in an increase in patent collaborations by 45.40% (34.92%$\times$1.30) to 63.36% (82.29%$\times$0.77). Given the average number of patent collaborations (116,654), the Beijing subway expansion during our sample period led to 52,961-73,912 patent collaborations.

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

This roughly equals to 22,967-32,053 approved collaborative patents after adjusting for the ratio of pairwise collaborations to total collaborative patents and the patent approval rate.[30] We take the estimates in Wei et al. (2021) on patent values in Beijing (3,291 - 6,191 yuan in 2008 values) and extrapolate to our setting based on the median of the range converted into 2019 values.[31] The corresponding values for the increased collaborative patents in our setting are roughly 140 million to 195 million yuan, which correspond to about 20 million to 28 million in 2019 US dollars.[32]

A few caveats to keep in mind. First, the simple back-of-the-envelope calculation is based on the estimates of the local average causal response. To the extent that the local average causal response differs significantly from the average causal response, our calculations may be biased.[33] Second, we use the estimated patent values from Wei et al. (2021) to impute the values of collaborative patents. Although their study relies on patents in Beijing during a similar sample period, it only focuses on eight industries targeted by a Chinese pro-innovation

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

[31]The Reminbi inflation factor is taken from `https://www.in2013dollars.com/china/inflation`. According to this source, 100 RMB in 2008 is equivalent to 128.12 RMB in 2019.

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

[33]Such concern is not limited to our study, but applicable to all empirical analyses using the IV approach in which treatment effect is likely to be heterogeneous. Given the possibility that the effect may be overestimated, if we use the TWFE estimates in Table 2, we find that the range of economic gains lie between 8.5 million and 13 million in 2019 US dollars.

subsidy program, whereas we use patents from all industries.[34] Third, our estimates are based on contemporaneous estimates that do not account for long-term dynamic responses. Given the dynamic nature of innovations, the long-term implications are likely to be substantially larger.

8. Conclusion

In this paper, we analyze the extent to which development of intra-city transport infrastructure affects collaborative matches in innovations. Using rapid expansion of the Beijing subway from 2000 to 2018, we find that enhanced connectivity 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, with the local average causal response being 34.92% to 82.29%. Moreover, we find that the increase in collaborative matches is largely driven by more matches among highly productive innovators due to positive assortative matching and the complementarity between productivity and connectivity. At the same time, entry of new innovators, relocation of existing innovators, and innovators with low productivity also contribute to the increase in collaborative matches in the long run.

Whereas this paper establishes an important link between connectivity and collaborative matches in innovations, there are possible extensions to consider. First, using data on patent citations, it would be interesting to further investigate qualitative aspects to see whether highly cited patents are sparked by enhanced connectivity. Second, one can further consider analyzing the network of collaborators to see whether any systematic relationship exists between innovators who are at the center of a network and degree of their accessibility based on their location. Third, we believe the implications of our findings should be applicable to other metropolitan cities with high population density, congestion problems, and zoning restrictions. Nevertheless, the quantitative magnitude of impact may differ, so it would be interesting to explore other em-

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

pirical contexts in future work. Lastly, as the gains from reduced travel time that we document speak directly to the underlying rationalization of agglomeration and returns to urban density, it would be interesting to directly model agglomeration economies and estimate their effect.

References

Acemoglu, D. (2008). *Introduction to modern economic growth.* Princeton, NJ: Princeton University Press.

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

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva (2022). Taxatopm and Innovation in the 20th Century. *Quarterly Journal of Economics 137*(1), 329–385.

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.

Beijing Metro Operation Co., Ltd. (2011). *Responsibility and mission: Beijing Metro development history*. Beijing: Beijing Press.

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.

Berkes, E. and R. Gaetani (2021). The geography of unconventional innovation. *The Economic Journal 131*(636), 1466–1514.

Bernard, A., A. Moxnes, and Y. Saito (2020). The geography of knowledge production: connecting islands and ideas. *Working paper*.

Borusyak, K. and P. Hull (2021). Non-random exposure to exogenous shocks: Theory and applications. *NBER Working Paper Series*.

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

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

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

Catalini, C. (2018). Microgeography and the direction of inventive activity. *Management Science 64*(9), 4348–4364.

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

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

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. *Handbook of Regional and Urban Economics*, 247–348.

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

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

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

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*.

Duranton, G. and M. A. Turner (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review 101*(6), 2616–2652.

Gendron-Carrier, N., M. Gonzalez-Navarro, S. Polloni, and M. A. Turner (2022). Subways and urban air pollution. *American Economic Journal: Applied Economics 14*(1), 164–196.

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

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

Gu, Y., J. Zhang, and B. Zou (2021). Subways and road congestion. *American Economic Journal: Applied Economics 13*(2), 83–115.

Hausman, N. (2022, 07). University Innovation and Local Economic Growth. *The Review of Economics and Statistics 104*(4), 718–735.

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. (1989). Real effects of academic research. *American Economic Review 79*(5), 957–970.

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.

Kantor, S. and A. Whalley (2014). Knowledge spillovers from research universities: Evidence from endowment value shocks. *Review of Economics and Statistics 96*(1), 171–188.

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.

Li, J., S. Liu, and Y. Wu (2021). Identifying knowledge spillovers from universities: Quasi-experimental evidence from urban China. *Working paper*.

Liu, S. (2015). Spillovers from universities: Evidence from the land-grant program. *Journal of Urban Economics 87*, 25–41.

Melitz, M. J. and S. J. Redding (2021). Trade and Innovation. *NBER Working Paper*, 1–41.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018). Commuting, migration, and local employment elasticities. *American Economic Review 108*(12), 3855–3890.

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.

Paunov, C., D. Guellec, N. El-Mallakh, S. Planes-Satorra, and L. Nüse (2019). On the concentration of innovation in top cities in the digital age. (85).

Perlman, E. R. (2016). Connecting the periphery: Three papers on the developments caused by spreading transportation and information networks in the nineteenth century United States.

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, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 1339–1398. Elsevier.

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

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

Sultana, S. and J. Weber (2016). *Minicars, magleves, and mopeds: Modern modes of transportation around the world*. ABC-CLIO, LLC.

Wei, S.-j., J. Xu, G. Yin, and X. Zhang (2021). Losses from trade: The case of China' s Pro-innovation subsidy program. *Working paper*.

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

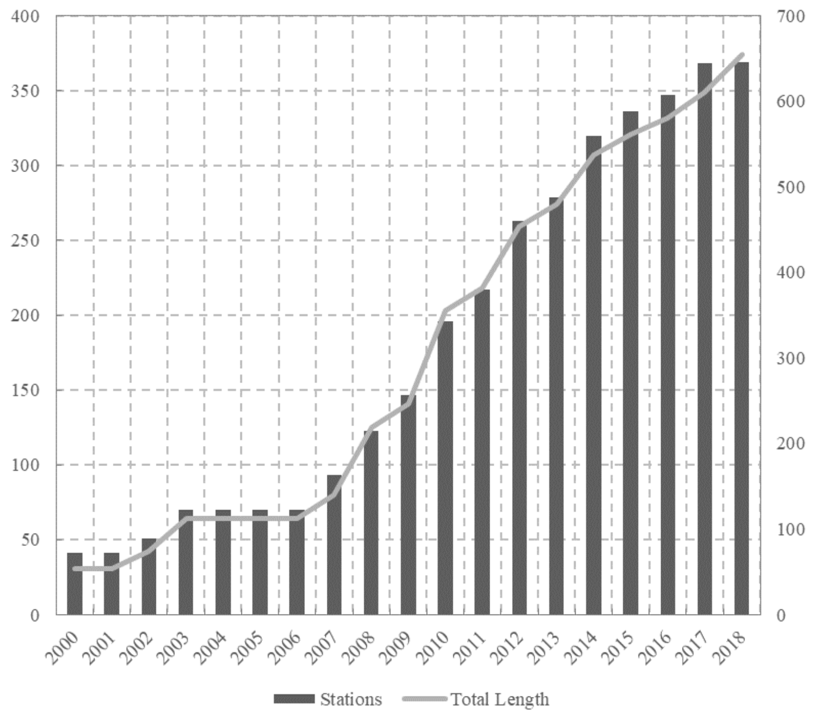
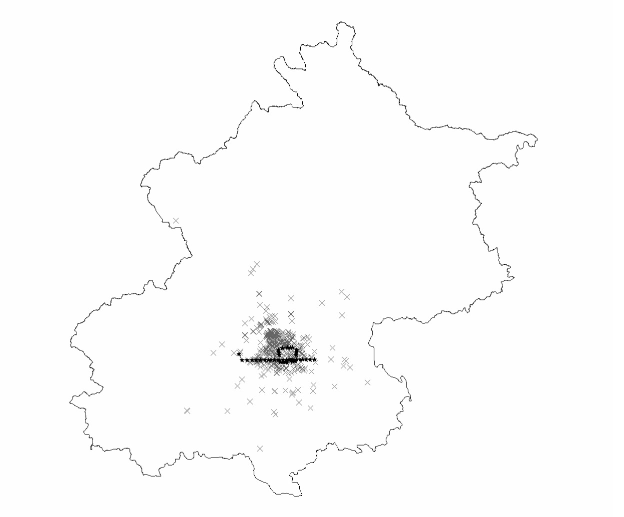
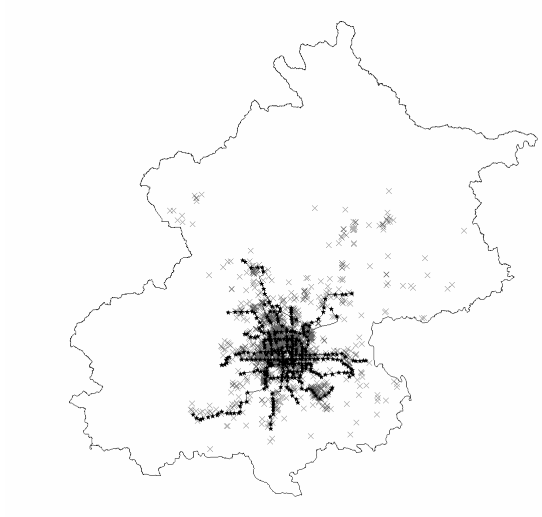


Figure 1: Number of Subway Stations and Length of Subway Lines in Beijing

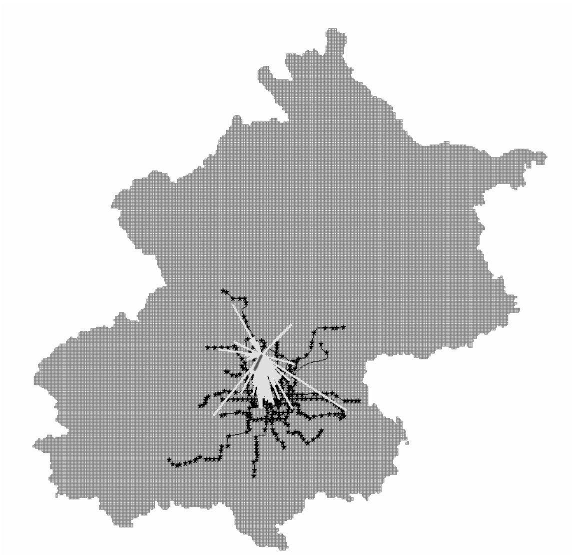
Notes: The cumulative count of subway stations in Beijing is indexed to the left axis and displayed using a bar graph. The total length of subway lines in Beijing is indexed to the right axis and displayed using a line graph.



(a) 2000 and 2001



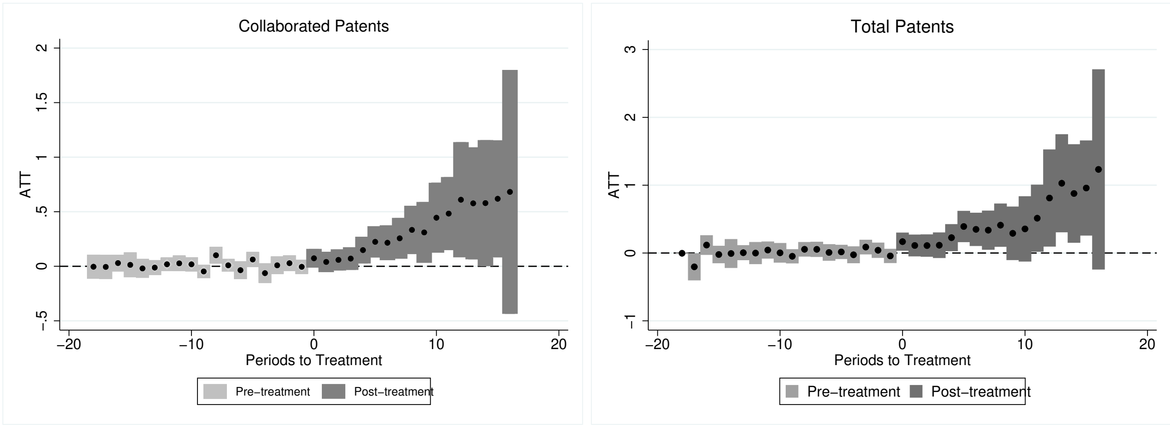
(b) 2017 and 2018



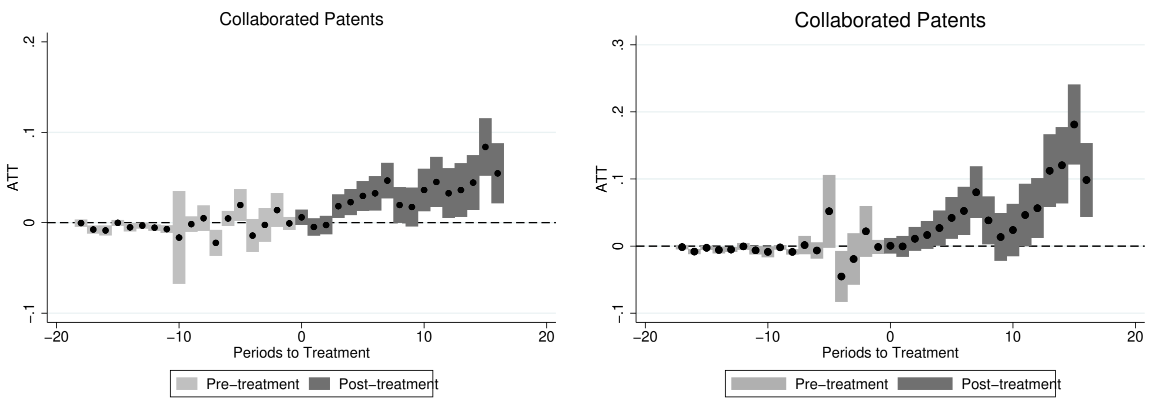
(c) An Illustration of Collaboration Links for a Grid

Figure 2: Spatial Distribution of Patents, Patent Collaborations, and Subway Stations

Notes: Panel (a) shows the spatial distribution of collaborated patents applied in 2000 and 2001 (in gray x-marks) and the subway stations available during the same period (in black star-marks) in Beijing. Similarly, Panel (b) shows the spatial distribution of collaborated patents and the subway stations in 2017 and 2018. Panel (c) shows collaboration links for the grid that had the largest number of collaborations during our sample period. The gray line shows the the only collaboration link that was formed during 2000-2001 period, whereas the gray lines show the 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) Grid Level



(b) Grid Pair Level

Figure 3: Time-Varying Effects of Improved Connectivity on Collaborated and Total Patents

Notes: The dynamic ATT's are plotted using the staggered DiD model, based on Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). In the Panel (a) Grid Level, the control group consists of the grids that were never treated during our sample period. In the top left (right) panel, the dependent variable is collaborated patents (total patents) in its IHS transformation. In Panel (b) Grid Pair Level, the control group consists of the grid pairs that were never treated during our sample period. The dependent variable is collaborated patents in its IHS transformation in both figures in Panel (b). In the bottom left (right) panel, we define treatment to be a reduction in travel time by 30 minutes (1 hour) along the fastest travel route.

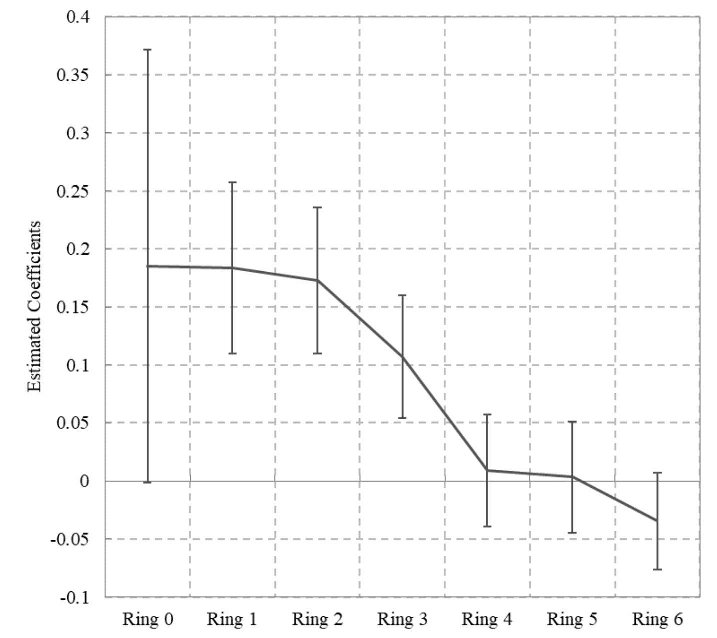


Figure 4: Ring-specific Coefficients

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 OA2 in Online Appendix.

Table 1: Summary Statistics

Unit of Analysis	Variable	Mean	SD	Number of Observations
Year				
	Number of patents in a year	42,876.55	36,431.05	19
	Number of collaborated patents a year	9,222.73	9,144.18	19
Patent				
	application year	2013.534	3.980	785,629
	Number of applicants	1.283	0.606	785,629
	Number of applicants of collaborated patents	2.271	0.625	175,232
Grid year				
	Number of patents	5.961	70.853	139,900
	Number of of collaborated patents	1.317	48.179	139,900
Grid pair year				
	Number of of patents for a pair of grids in a year	0.480	11.724	228,285
	log(Length)	8.325	3.061	228,285
	log(Euclidean distance)	2.340	0.959	228,285
Applicant				
	Number of patents in a year	7.695	72.385	131,063
	Number of collaborated patents in a year	3.038	55.603	131,063
	Number of collaborated patents in a year by firms	4.289	74.137	71,918
	Number of collaborated patents in a year by individuals	0.614	1.465	46,553
	Number of collaborated patents in a year by institutions	4.849	27.326	12,592

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

Table 2: Treatment Effect on Patent Collaborations—A Grid Pair-Level Analysis

	𝟙(Collab. patents)		IHS(Collab. patents)	
Panel A.	(1)	(2)	(3)	(4)
𝟙(Treated)	0.0162***	0.0129***	0.0331***	0.0335***
	(0.0027)	(0.0031)	(0.0069)	(0.0078)
Observations	188,081	188,081	188,081	188,081
Method	TWFE	TWFE	TWFE	TWFE
Adjusted R^2	0.052	0.065	0.081	0.177
Year FE	Yes	Yes	Yes	Yes
Grid FE	Yes	No	Yes	No
Grid Pair FE	No	Yes	No	Yes

Panel B.		
𝟙(Treated)	0.0131***	0.0294***
	(0.0029)	(0.0059)
Observations	188,081	188,081
Method	DiD	DiD

Notes: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018. In Panel B, the control group consists of the grid pairs that were never treated during our sample period. For a grid pair, 𝟙(Treated) equals one for the first and the following years when the travel time reduces by more than 30 minutes along the fastest travel route; 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 3: Impact of Subway Expansion on Patent Collaboration—A Grid Pair-Level Analysis

Dep. variable	$\mathbb{1}$(Collab. patents)		IHS(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
Grid FE	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES
Observations	188,081	188,081	188,081	188,081
R-squared	0.070	0.114	0.098	0.221

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

Table 4: Impact of Subway Expansion on Patent Collaboration—A Grid Pair-Level Analysis
Heterogeneity and Local Average Causal Response

	$\mathbb{1}$(Collab. patents)				IHS(Collab. patents)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	Near grid pairs	Far grid pairs	All grid pairs	All grid pairs	Near grid pairs	Far grid pairs	All grid pairs	All grid pairs
log(Length)	0.0011*	0.0249***	0.0139**	0.0128**	0.0019	0.0558***	0.0432***	0.0376***
	(0.0007)	(0.0072)	(0.0061)	(0.0050)	(0.0016)	(0.0165)	(0.0151)	(0.0122)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-	-	-0.2648	-0.2476	-	-	-0.8229	-0.7329
Offline = 10 km/h; Subway = 36 km/h	-	-	-0.1691	-0.1568	-	-	-0.5255	-0.4642
Offline = 8 km/h; Subway = 36 km/h	-	-	-0.1124	-0.1033	-	-	-0.3492	-0.3059
Kleibergen-Paap rk Wald F statistic	-	-	239.508	369.948	-	-	239.508	369.948
Observations	105,014	82,915	188,081	188,081	105,014	82,915	188,081	188,081
Root MSE	0.251	0.264	0.251	0.251	0.430	0.406	0.416	0.415
Method	TWFE	TWFE	IV	IV (LOO)	TWFE	TWFE	IV	IV (LOO)

Notes: All specifications include year fixed effects and grid pair fixed effects. If the Euclidean distance between a pair of grids is below (above) the median value of the entire distribution, we assign the pair as a "Near grid pair" ("Far grid pair"). In Columns (3) and (7), we instrument for the *Length* variable for a grid pair 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 grids. In Columns (4) and (8), we use the leave-one-out ("LOO") instrument. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote $p<0.01$, $p<0.05$, $p<0.1$ respectively.

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

	(1)	(2)	(3)	(4)
Dep. variable	IHS(Collab. patents)		IHS(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
$\mathbb{1}$(Treated)	0.0301***	0.0275***	0.0036**	0.0022
	(0.0075)	(0.0055)	(0.0018)	(0.0023)
Observations	188,081	188,081	188,081	188,081
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
$\mathbb{1}$(Treated)	0.0335***	0.0296***	0.000003	-0.0001
	(0.1014)	(0.0059)	(0.0002)	(0.0003)
Observations	188,081	188,081	188,081	188,081
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
$\mathbb{1}$(Treated)	0.0328***	0.0299***	0.0010	-0.00005
	(0.0773)	(0.0058)	(0.0016)	(0.0018)
Observations	188,081	188,081	188,081	188,081
Method	TWFE	DiD	TWFE	DiD
Year FE	YES		YES	
Grid Pair FE	YES		YES	

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

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

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

Notes: All specifications include year fixed effects and grid pair fixed effects. We use the sample from 2001 to 2018. We drop the initial year of 2000, since we use the previous year's information to determine incumbent, mobility, and productivity status. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. In Columns (1) and (3), we instrument for the *Length* variable for a grid pair 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 grids. In Columns (2) and (4), we use the leave-one-out ("LOO") instrument. 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—A Grid Pair-Level Analysis

Panel A.	Full Sample			
	(1)	(2)	(3)	(4)
Dep. var.: IHS(Collab. patents)				
	Movers	Entrants	H-types	L-Types
𝟙(within 5 years)	0.0018**	0.0031	0.0139*	0.0013
	(0.0008)	(0.0032)	(0.0083)	(0.0010)
𝟙(after 5 years)	0.0016**	0.0084**	0.0404***	0.0030**
	(0.0008)	(0.0034)	(0.0110)	(0.0013)
Observations	178,182	178,182	178,182	178,182
R-squared	0.067	0.256	0.353	0.083

Panel B.	Restricted Sample			
	(1)	(2)	(3)	(4)
Dep. var.: IHS(Collab. patents)				
	Movers	Entrants	H-types	L-Types
𝟙(within 5 years)	0.0017**	0.0046	0.0225***	0.0021**
	(0.0008)	(0.0030)	(0.0080)	(0.0009)
𝟙(after 5 years)	0.0016**	0.0075**	0.0324***	0.0031***
	(0.0008)	(0.0032)	(0.0105)	(0.0012)
Observations	178,182	178,182	178,182	178,182
R-squared	0.064	0.235	0.305	0.076

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

Online Appendix

Contents

A. Beijing Subway Development

Line	Date	Stations
4	9/28/2009	Beigongmen, Weigongcun, Xidan, Haidianhuangzhuang, Gongyixiqiao, Caishikou, National Library, Zhongguancun, Beijing Zoo, East Gate of Peking University, Renmin University, Xiyuan, Ping'anli, Taoranting, Yuanmingyuan Park, Xuanwumen, Majiapu, Beijing South railway station, Jiaomen West, Xisi, Lingjing Hutong, Xinjiekou, Anheqiao North, Xizhimen
4	12/30/2010	Huangcun Xidajie, Xihongmen, Yihezhuang, Huangcun railway station, Gaomidian South, Gaomidian North, Qingyuanlu, Tiangongyuan, Zaoyuan, Xingong, Biomedical Base
5	10/7/2007	Puhuangyu, Liujiayao, Tiantandongmen, Beixinqiao, Huixinxijie Nankou, Hepingxiqiao, Tiantongyuan, Zhangzizhonglu, Yonghegong Lama Temple, Dongsi, Dongdan, Tiantongyuan South, Hepingli Beijie, Tiantongyuan North, Ciqikou, Dengshikou, Huixinxijie Beikou, Beiyuanlu North, Chongwenmen, Datunlu East, Lishuiqiao, Lishuiqiao South, Songjiazhuang
6	12/30/2012	Dongsi, Ping'anli, Dongdaqiao, Caofang, Jintailu, Hujialou, Chegongzhuang West, Baishiqiao South, Changying, Haidian Wuluju, Chegongzhuang, Beihai North, Nanluoguxiang, Dalianpo, Huangqu, Shilipu, Huayuanqiao, Qingnianlu, Chaoyangmen, Cishousi
6	12/28/2014	Tongzhou Beiguan, Beiyunhe West, Wuzixueyuanlu, Haojiafu, Lucheng, Dongxiayuan
6	12/30/2018	Tian Cun, Xihuang Cun, Yang Zhuang, Liaogong Zhuang, Beiyunhe East, Jin'anqiao
7	12/28/2014	Beijing Happy Valley, Jiulongshan, Guangqumenwai, Daguanying, Huagong, Guangqumennei, Hufangqiao, Nanlouzizhuang, Zhushikou, Jiaohuachang, Ciqikou, Beijing West Railway Station, Caishikou, Guang'anmennei, Dajiaoting, Shuanghe, Qiaowan, Baiziwan, Wanzi
7	12/30/2018	Fatou
7	12/28/2019	Lang Xin Zhuang, Hua Zhuang, Huang Chang, Wansheng Dong (East), Qun Fang, Hei Zhuang Hu, Gao Lou Jin, Shuangjing, Wansheng Xi (West)
8A	7/19/2008	South Gate of Forest Park, Beitucheng, Olympic Green
8A	10/9/2008	Olympic Sports Center
8A	12/31/2011	Yuxin, Lincuiqiao, Huilongguan Dongdajie, Xixiaokou, Yongtaizhuang, Huoying

Line	Date	Stations
8A	12/30/2012	Guloudajie, Anhuaqiao
8A	12/28/2013	Nanluoguxiang, Zhuxinzhuang, Shichahai, Yuzhilu, Pingxifu
8A	12/26/2015	Andelibeijie
8A	12/30/2018	National Art Museum
8B	12/30/2018	Huojian Wanyuan, Yinghai, Dahongmen Nan (South), Wufu Tang, Zhushikou, Heyi, Donggao Di, Muxi Yuan, Haihu Tun, Yongdingmenwai, Demao, Tian Qiao
9	12/31/2011	Beijing West railway station, Fengtainanlu, Fengtai Science Park, Liuliqiao, Liuliqiao East, Guogongzhuang, Qilizhuang, Keyilu
9	10/12/2012	Fengtai Dongdajie
9	12/30/2012	Baiduizi, Baishiqiao South, National Library
9	12/21/2013	Military Museum
10	7/19/2008	Agricultural Exhibition Center, Haidianhuangzhuang, Zhichunli, Jinsong, Taiyanggong, Liangmaqiao, Jiandemen, Anzhenmen, Jintaixizhao, Shuangjing, Sanyuanqiao, Beitucheng, Zhichunlu, Tuanjiehu, Suzhoujie, Mudanyuan, Huixinxijie Nankou, Hujialou, Shaoyaoju, Bagou, Xitucheng, Guomao
10	12/30/2012	Caoqiao, Shilihe, Dahongmen, Chedaogou, Fenzhongsi, Gongzhufen, Shiliuzhuang, Cishousi, Shoujingmao, Changchunqiao, Songjiazhuang, Lianhuaqiao, Xidiaoyutai, Jijiamiao, Jiaomen West, Huoqiying, Xiju, Liuliqiao, Panjiayuan, Chengshousi
10	5/5/2013	Fengtai Railway Station, Jiaomen East, Niwa
13	9/28/2002	Shangdi, Dazhongsi, Xi'erqi, Huilongguan, Huoying, Qinghe, Zhichunlu, Xizhimen, Wudaokou, Longze
13	1/28/2003	Guangximen, Dongzhimen, Shaoyaoju, Lishuiqiao, Wangjing West, Beiyuan, Liufang
14A	12/28/2014	Jintailu, Donghuqu, Futong, Shan'gezhuang, Laiguangying, Wangjing South, Jiangtai, Zaoying, Dongfengbeiqiao, Wangjing
14A	12/26/2015	Jiulongshan, Beigongda Ximen, Yongdingmenwai, Puhuangyu, Shilihe, Fangzhuang, Jingtai, Dawanglu, Beijing South railway station
14A	12/31/2016	Chaoyang Park

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

B. Theoretical Model

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

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

Innovators simultaneously search for collaborators and they encounter one another randomly. The probability of innovator i with productivity z_i at location m encountering innovator j with productivity z_j at location n is $\theta_{mn} f(z_i) f(z_j)$, where θ_{mn} is a location pair-specific parameter that captures exogenous factors determining the likelihood of the encounter.[1]

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

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

where $\gamma > 0$ measures the elasticity of the matching probability with respect to travel time τ_{mn}; b is a constant capturing the decay rate of the matching probability when $z_i z_j = 0$ (i.e. the lowest quality match); $0 < \mu < 1$ is a constant to scale q. Having a shorter travel time increases the likelihood of having a successful collaboration since innovators can have more frequent face-to-face interactions, which facilitates their coordination on exchanging information, skills, and ideas.[2]

Note that the matching probability increases as travel time decreases and such effect is more pronounced for more productive pairs of innovators with higher $z_i z_j$. This setup is consistent with

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

[2] Another explanation is that improved connectivity increases location-specific market access which further spurs innovations and innovation collaborations (Melitz and Redding 2021). However, we think this is less likely to apply in our within-city context given that the market for innovative products usually exceed narrowly defined geographic locality. Moreover, Perlman (2016) explicitly controls for market access in studying the impact of railroad network connections on innovation activities and finds modest evidence that market access explains the increase in patent activity, but most of the relationship seems to be explained by other variables correlated with transportation access.

important regularities in the literature. First, it suggests that collaborators' productivities are complements, which has been widely documented and is often imposed in various structural frameworks modeling idea exchange and knowledge spillover in the literature (Davis and Dingel 2019). Such complementarity determines that an innovator with high productivity is more likely to be matched to another highly productive innovator for collaboration.[3] Second, the matching probability between productive innovators increases more as travel time decreases (Combes et al. 2012; Behrens et al. 2014; Davis and Dingel 2019; Lee and Xu 2020). For a more accessible location, we would thus expect to see the high-quality matches being more disproportionately formed (i.e. the "log supermodularity" condition).

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

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

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

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

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

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

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

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

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

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

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

Equations (OA5) and (OA6) show that a location that is closely connected to other locations with a

[3] Using data, we checked evidence on positive assortative matching (PAM) by calculating the following ratio: the share of match for a given type observed in the data divided by the share of match for a given type that would arise under random matching. If the value of ratio is greater than 1 between the same type but less than 1 across different types, that would provide supporting evidence on PAM. We try two alternative definitions of productivity. In the first method, when we use the count of patents in the past year to define productivity, we get the following ratios of each of the three types: 1.0165 for H-H type, 10.1334 for L-L type, and 0.6106 for the mixed type. In the second method, when we use the approval rate to define productivity, we get the following ratios of each of the three types: 2.2773 for H-H type, 1.0367 for L-L type, and 0.3832 for the mixed type. In both cases, we find supporting evidence on PAM.

large measure of innovators yields high expected profits from solo innovations. The setup captures the existence of knowledge spillovers in that more innovators contribute to a higher overall productivity and the parametric form embeds the property that the knowledge spillover decays with distance.

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

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

where H is an aggregator which combines the expected profits from collaborative and solo innovations. If the two types of innovations were independent, H would be a simple summation of the two. But H does not have to be a linear aggregator as collaborative and solo innovations may affect each other (complements or substitutes). η_m is an observed feature of location m (e.g. disamenities) and we assume it to be exogenous. ξ captures the elasticity of η_m.[4]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Under the three assumptions, we obtain a closed-form solution of the model. Assumption 1 allows collaborative and solo innovations to be substitutable and regulates the elasticity of substitution to be one. Assumption 3 allows us to obtain the level of solo innovations conveniently and can be rationalized by considering that the value of each solo innovation is similar. Thus total innovation value is proportional to the counts of solo inventions.[5]

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

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

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

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

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

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

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

Using Equations (OA7), (OA9) and (OA10), we further decompose the distribution effect, con-

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

sisting of the $\ln g_m + \ln g_n$ terms, as follows:

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

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

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

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

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

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

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

References

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: sorting, selection, and agglomeration. *Journal of Political Economy 122*(3), 507–553.

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.

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

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

Duranton, G. and M. A. Turner (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review 101*(6), 2616–2652.

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.

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.

Melitz, M. J. and S. J. Redding (2021). Trade and Innovation. *NBER Working Paper*, 1–41.

Perlman, E. R. (2016). Connecting the periphery: Three papers on the developments caused by spreading transportation and information networks in the nineteenth century United States.

C. Assignment of Applicants to Grids

For each patent in the CNIPA database, address for the first-listed applicant is reported. For example, for a collaborated patent with three applicants, we have the address for only the first-listed applicant but not the other 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 containing exact addresses in the data, using information on applicants of solo patents and first-listed applicants of collaborated patents. 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. If there is no patent by the same collaborator in the same year, we instead search for other patents by the same collaborator across other years. 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: 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 398,121 observations of applicants for 175,232 collaborated patents between 2000 and 2018. 175,232 (44.01%) observations are first-listed applicants 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.

There are some caveats to note. First, there may be different applicants with the same name. Among 398,121 observations of applicants, 92.82% are non-individuals (e.g. firms, universities, research institutes, hospitals, etc.) Institutions are less likely to share the same name, but there may be ambiguous cases among individual applicants. For example, if there are two observations of the same applicant name in two different locations in a year, it is unclear whether a single applicant relocated or there are two applicants in different locations. For simplicity, we assume that as long as the two addresses are within 2 *km*, we treat them as the same applicant and randomly select an address to use in that year. Second, the accuracy of imputation is likely to be lower for applicants who appear less in our search database. For example, suppose we have the precise address of an applicant 2018 and we are trying to impute the address back in 2010. Whereas we assume that the applicant's location is the same in 2010 as it was in 2018, an applicant may have moved during this time. This measurement error is likely to be severe if there is a large share of applicants who produce many patents but rarely appear as either solo or first-listed applicants. However, we find that this is not the case. 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, rather than being to only a small number of applicants.

D. Assigning Movers vs. Non-movers

In Panel B of Tables 5 and 6, we divide applicant pairs into "Nonmover Pairs" and "Mover Pairs." To determine relocation status, we conduct the following data work. We place pairs of applicants who were for sure both non-movers in that year in the "Nonmover Pairs" category, while placing all other remaining pairs in the "Mover Pairs" category. For this aim, we compile the entire list of all first-listed applicants using both solo and collaborated patents in our data. We use these reported addresses of these applicants to identify whether an applicant moved across years. If an applicant's location remained the same, then we assign that applicant to be a "non-mover" during that time period. For example, if an applicant was observed at the same address in 2000, 2005, and 2007, then that applicant is a non-mover from 2000 to 2007.[6]

[6] Although it is possible that the applicant moved to another location and returned to the original location between 2001 and 2004, the likelihood is not high.

E. Robustness Analyses

E.1. Analysis under Alternative Sample Selections

One identification concern with the IV approach discussed in Section 3 is that the exposure measure based on the Euclidean distance may be non-random. Evidence in Figure OA4 shows that the grid pairs with very large Euclidean distance tend to be more concentrated in peripheral areas. Table OA9 shows the estimates when we either drop the top 10% (columns (1) and (3)) or the top 20% (columns (2) and (4)) of observations with very large Euclidean distance from our sample. We find that dropping such samples does not change our baseline estimates much but in fact increase the magnitudes slightly.

E.2. Analysis Using Control Pair Approach

As an alternative research design, we examine whether collaboration between innovators is more likely to be formed when connectivity between their locations improves due to subway expansion. The idea is to match *actual* collaboration pairs to *control* collaboration pairs based on observable characteristics, as proposed by Jaffe et al. (1993). If potential endogeneity concern resides on the selection on observables, the matching procedure allows us to address such concern. We use application year and the first-listed IPC at the 3-digit level to match patents. More specifically, we first identify all pairwise combinations of applicants listed on a collaborated patent.[7] This gives us 77,749 applicant pairs whose respective locations are identified in the data. Then we extract the first-listed IPC code at the 3-digit level and application year from all collaborated patents.[8] We create control pairs by taking one applicant from the original patent and randomly drawing one applicant from a control patent, for which the two patents share the same 3-digit first-listed IPC code and application year.[9]

Using matched *control* collaboration pairs, we estimate a linear probability model:

$$\mathbb{1}(CollabPatents)_{ijpt} = \beta \log(Length)_{ijpt} + \theta_{ij} + \lambda_t + \xi_{ijpt}, \tag{OA17}$$

where $\mathbb{1}(CollabPatents)_{ijpt}$ is a binary indicator that equals 1 if an applicant pair p is an *actual* pair and equals 0 otherwise; $Length_{ijpt}$ is a proxy for the extent of subway build-up between grids i and j in which applicants are located in; θ_{ij} represents grid pair fixed effects; λ_t represents year fixed effects; ξ_{ijpt} is the idiosyncratic error. 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 grid pair and year fixed effects, the conditional independence assumption is

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

[8] A patent typically has multiple IPC codes, but the first-listed IPC code represents the primary classification of the patent's content.

[9] For example, if the actual pair of collaborators is (A, B), we create a matched control applicant pair (B, C), where C is an applicant of another collaborated patent that shares the same first-listed IPC code and application year as the patent invented by (A, B).

satisfied (i.e. $\mathrm{cov}(Connect_{ijpt}, \xi_{ijpt}) = 0$).

In Table OA10, we either create one control pair for every actual pair (i.e. "One-to-One Match") or two control pairs for every actual pair (i.e. "One-to-Two Match"). The results are quite similar. Without controlling for grid pair fixed effects in Columns (1) and (3), applicants in a pair are less likely to collaborate as $Length$ between their locations increases, compared with control pairs. However, once we further control for grid pair fixed effects in Columns (2) and (4), enhanced connectivity captured by a longer $Length$ improves the collaboration probability. Such findings are consistent with our main results.

E.3. Analysis Using Planned but Unbuilt Stations as Counterfactuals

To further alleviate the concern on endogenous selection into the treatment group, we follow another commonly adopted identification strategy used in the literature. The idea is to compare the impact of stations that were planned and built to that of the stations that were planned but unbuilt during the sample period (Duranton and Turner 2011; Donaldson 2018). Namely, we exploit the unforeseen fate of Line 11 in the planning of the Beijing subway network. Line 11 was originally proposed in the 1993 Beijing subway network plan as shown in Panel (a) of Figure OA5. The initial plan was to start its construction from 2000 onwards, after completing Lines 3 and 5. However, unexpected challenges arose in constructing Line 3 as standards in protecting cultural relics were significantly raised and this made previous plans infeasible.[10] Thus, Line 3 was later partially replaced by Line 13, which was competing for resources with the proposed Line 11 post the 1998 economic recession. As a result, Line 11 never ended up being constructed according to its original plan. Some stations were built but became part of other lines, and construction of other stations never initiated during our sample period. To the extent that the original plan reflects the selection criteria of stations and construction was postponed due to unforeseen reasons, these planned but unbuilt stations serve as a valid control set in identifying the impact of improved connectivity on collaborative innovations.

To this end, we create a control grid set which consists of the grids that are within one kilometer of the five stations (i.e. Dachengezhuang, Banjieta, Nanduzi, Dongbaxi, and Dongba) that were planned but not built during our sample period. These are the five stations that lie at the east end of Line 11 in Figure Panel (a) of Figure OA5, according to the original plan.[11] The geographical location of these five stations in a map is also marked in stars in Figure OA6. To further exclude any potential effect arising from other nearby stations, we further drop from this control grid set any grids that are within one kilometer from other existing nearby stations.

We consider four different ways of defining a treatment grid set. In the first approach, we include the rest of all other grids in the treatment group. In other three approaches, we focus on a narrower group of grids that are within one kilometer from a selected set of stations. More specifically, we

[10] Detailed discussions can be found in Beijing Metro Operation Co., Ltd (2011).

[11] The full set of stations that were actually built by 2020 is shown in Panel (b) of Figure OA5.

select stations that are 1) actually built during our sample period; 2) geographically close to the control stations; and 3) roughly of the same distance to the center of the city (Tiananmen Square) as the control stations. Under such criteria, one treated group (i.e. "treat 1") consists of the grids within one kilometer from the following stations: Futong, Wangjing, Donghuqu, Laiguangying, and Shan'gezhuang. These five stations, marked in triangles in Figure OA6, were originally part of Line 3 but were eventually built along Line 14 in 2014. In another treated group (i.e. "treat 2"), we use the grids that are within one kilometer from the following stations: Shilipu, Qingnianlu, Dalianpo, Huangqu, and Changying. These stations, marked in diamonds in Figure OA6, were built along Line 6 in 2012. Lastly, we also consider using the combined treated group, which consists of all grids that are in "treat 1" and "treat 2."

In Figure OA6, we first illustrate the magnitude of increase in collaborative patents between 2000 and 2018 using a thematic map. A darker shade implies a larger increase in collaborative patents. Based on the contrast of the shades in the thematic map, it is evident that the areas near the planned but unbuilt stations (marked in stars) experienced a much smaller increase in collaborative patents, compared to that in areas near the stations considered in the treatment groups (marked in either stars or diamonds). The contrast is also evident if one were to compare the control set to the rest of all other grids.

Table OA11 shows our findings based on this identification strategy. The dependent variable, Δ IHS(Collab. patents), is defined as the change in IHS(Collab. patents) from its value in 2000 to its cumulative value between 2001 and 2018. In both Panels, Column (1) includes the full sample, Column (2) contains the grids that are near the "treat 1" stations and control stations; Column (3) contains the grids that are near the "treat 2" stations and control stations; Column (4) contains the grids that are near the combinations of "treat 1" stations, "treat 2" stations, and control stations. Panel A shows analysis conducted at the grid level. We find that relative to the control group, the treated grids experience a significantly higher increase in collaborative patents and such evidence is robust to different ways of selecting the treated stations. Panel B shows analysis at the grid pair level, for which a grid pair is defined to be treated if both grids are from the treatment group. The results at the grid pair level are also consistent.

E.4. Analysis Using Alternative Distance Cutoffs

In our grid pair analysis, we construct pairs using all grids ever located within 2 *km* from a subway station during 2000-2018. The 2 *km* distance cutoff was based on the spatial decay patterns as shown in Figure 5 and Table OA2. In Table OA12, we alternatively apply different cutoffs (i.e. 1.5 *km*, 2.5 *km*, and 3 *km*, respectively) and confirm that our results remain robust.

E.5. Analysis Using an Alternative Definition of Mobility

We apply an alternative definition of movers. As in Section D, we first we assign an applicant to be a mover when the applicant's location changes. Then we identify the pairs which contain at least one applicant who moved for sure in that particular year as "Mover Pairs," while all remaining pairs are classified as "Nonmover Pairs" in Table OA13. We find that our estimates based on this alternative mobility definition remain comparable.

E.6. Analysis Using an Alternative Definition of Productivity

Instead of using patent counts in the previous year to determine productivity status of the current year, we apply an alternative measure of productivity. We use the overall patent approval rate between 2000 and 2015 for each applicant to determine productivity status. Whereas our productivity measure in the main analysis focuses on the quantity aspect of production in the past, this alternative measure is informative of how successful an applicant is gaining an exclusive legal right. We drop years from 2016 to 2018 in our analysis, given that the number of approved patents in the data from 2016 falls. This is because applications take time to be approved and their approval status is not properly captured by the end of our sample period. Tables OA14 and OA15 summarize our findings based on this alternative productivity measure, using the sample from 2001 to 2015.[12] An applicant with larger than 65% patent approval rate is defined to be the H-type, for which the cutoff is approximately the median value in the distribution. The findings on the dominant mechanism still remain consistent and robust.

E.7. Analysis Using Approved Patents

Instead of using data on patent applications, we alternatively focus on the sample of patent applications that were eventually approved in Tables OA16 and OA17. Given such the time lag in patent approval process, we use the sample of collaborated patents in 2000-2015 that were approved. Tables OA16 and OA17 show that our findings still remain valid.

E.8. Analysis Controlling for the Impact of Internet Penetration and Increased Car Ownership

Potential factors that may weaken the identification in the IV strategy include expansion of the citywide internet access and increase in car ownership rate during our sample period. We address each of these concerns respectively by controlling for internet penetration rate and per capita car ownership rate in our baseline analysis. Specifically, we obtain data on the internet penetration rate in Beijing for each year between 2000 and 2017. We interact this rate with the initial travel time between a grid

[12] Year 2000 still needs to be dropped because incumbent and mobility status is defined based on the previous year's information.

pair to be used as an additional control variable.[13] This is to address the concern that the internet allows easier communication between locations and such effect is likely to be more pronounced for grid pairs that initially faced higher communication barriers. Similarly, we also obtain data on per capita car ownership rate in Beijing in each year between 2001 and 2010 and interact this measure with the initial travel time to form another control variable.[14] Table OA18 reports findings when these two control variables are added to our baseline specifications. We find that the results are still similar to our baseline outcomes.

E.9. Analysis Instrumenting the Discrete Treatment Variable

To alleviate concerns on the potential endogeneity of the discrete treatment variable, we use the same shift-share instrument as in our main analysis but to use this to instrument for the discrete treatment variable. Results are reported in Table OA19. Evidence is consistent with our baseline findings: a 30-minute reduction in travel time between a pair of grids significantly increases the probability of having collaborated patents and the number of collaborated patents.

E.10. Analysis Using More Regular Shaped Grids

We demonstrate in Section 3 that the Euclidean distance serves as a reasonable exposure measure, because of the "very orderly and grid-like" layout of Beijing subway lines and economic activities in general. To make better correspondence with this argument, we conduct a robustness check based on a sub-sample of the grids that forms a more regular shape overall. Figure OA7 displays the selection of the sub-sample. Corresponding results are reported in Table OA20. In this sub-sample analysis, we find similar results on the impact of improved connectivity on patent collaborations for both the extensive margin and the intensive margin. The estimated coefficients become larger likely because of a stronger instrument that better mitigates attenuation bias arising from potential errors in measuring connectivity.

[13] We obtain the internet penetration rate in Beijing between 2006 and 2017 from the CEIC (`https://www.ceicdata.com/en/china/internet-number-of-internet-user-and-penetration-rate`). For years earlier than 2006, we extrapolated the rate based on the change in the national internet penetration rate in China (`https://chinapower.csis.org/web-connectedness/`).

[14] Data are obtained from the Beijing Traffic Management Bureau (2012).

F. Appendix Tables

Table OA1: Treatment Effect on Patents—Grid Level Analysis

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

Notes: The sample contains all grids during 2000-2018. In Panel B, we use the never-treated as the control group. Treatment equals one for the first and the following years when a subway station in a grid opens; and zero otherwise. In Panel A, 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 OA2: Impact of Station Opening on Patents —Grid-Level Analysis (Spatial Decay)

Dep. variable	IHS(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 the count of all patents in grid i in year t, in its IHS transformation. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA3: Impact of Subway Expansion on Travel Time —Grid Pair-Level Analysis

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

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

Table OA4: Treatment Effect on the Size of Inventors — An Applicant-Level Analysis

	IHS(Number of Inventors)		
	(1)	(2)	(3)
$\mathbb{1}$(Located in a grid with a subway station)	0.0276		
	(0.0655)		
$\mathbb{1}$(Located within 1km from a subway station)		-0.0158	
		(0.0272)	
$\mathbb{1}$(Located within 2km from a subway station)			-0.0133
			(0.0293)
Observations	41,867	41,867	41,867
Method	TWFE	TWFE	TWFE
Year FE	Yes	Yes	Yes
Grid FE	Yes	Yes	Yes
Applicant FE	Yes	Yes	Yes

Notes: The sample contains all non-individual applicants that produced solo patents during 2000-2018. The dependent variable is constructed using the unique list of inventor names in solo patents for an applicant in a given year. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the applicant level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA5: Including Grid by Treatment Fixed Effects — A Grid Pair Level Analysis

	$\mathbb{1}$(Collab. patents)		IHS(Collab. patents)	
	(1)	(2)	(3)	(4)
log(length)	0.0053	0.0065	0.0303**	0.0302***
	(0.0055)	(0.0049)	(0.0131)	(0.0117)
Observations	188,081	188,081	188,081	188,081
Method	IV	IV(LOO)	IV	IV(LOO)
Year FE	Yes	Yes	Yes	Yes
Grid Pair FE	Yes	Yes	Yes	Yes
Origin Grid by Treated FE	Yes	Yes	Yes	Yes
Dest. Grid by Treated FE	Yes	Yes	Yes	Yes

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

Table OA6: Impact of Subway Expansion on Collaborators of Different Productivity and Mobility—Grid Pair-Level Analysis
Local Average Causal Response - Full Sample
Dependent Variable: IHS(Collab. patents)

	Method: IV				Method: IV(LOO)			
	(1) by H-H non-Movers	(2) by Other non-Movers	(3) by H-H Movers	(4) by Other Movers	(5) by H-H non-Movers	(6) by Other non-Movers	(7) by H-H Movers	(8) by Other Movers
log(Length)	0.0356**	0.0064**	0.0002	0.0002	0.0331**	0.0053**	0.0002	0.0001
	(0.0155)	(0.0030)	(0.0002)	(0.0002)	(0.0130)	(0.0025)	(0.0001)	(0.0002)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-0.6781	-0.1219	-0.0038	-0.0038	-0.6452	-0.1033	-0.0039	-0.0019
Offline = 10 km/h; Subway = 36 km/h	-0.4331	-0.0779	-0.0024	-0.0024	-0.4086	-0.0654	-0.0025	-0.0012
Offline = 8 km/h; Subway = 36 km/h	-0.2878	-0.0517	-0.0016	-0.0016	-0.2693	-0.0431	-0.0016	-0.0008
Kleibergen-Paap rk Wald F statistic	249.244	249.244	249.244	249.244	249.244	249.244	249.244	249.244
Observations	178,182	178,182	178,182	178,182	178,182	178,182	178,182	178,182
Root MSE	0.395	0.161	0.011	0.010	0.395	0.161	0.011	0.010
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Dest. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We first identify all pairs that include at least one applicant who collaborated after moving as "Movers," and the rest as "non-Movers." If each applicant'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. In columns (1) - (4), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year interacted with the Euclidean distance between those two locations. In columns (5) - (8), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year subtracting own pair length before interacted with the Euclidean distance between those two locations (Leave-one-out IV). Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA7: Impact of Subway Expansion on Collaborators of Different Productivity and Mobility—Grid Pair-Level Analysis
Local Average Causal Response - Robustness with Sample Cut at 90th Percentile
Dependent Variable: IHS(Collab. patents)

	Method: IV				Method: IV(LOO)			
	(1) by H-H non-Movers	(2) by Other non-Movers	(3) by H-H Movers	(4) by Other Movers	(5) by H-H non-Movers	(6) by Other non-Movers	(7) by H-H Movers	(8) by Other Movers
log(Length)	0.0415***	0.0059**	0.0002	0.0002	0.0376***	0.0049**	0.0001	0.0002
	(0.0143)	(0.0027)	(0.0002)	(0.0002)	(0.0120)	(0.0023)	(0.0001)	(0.0002)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-0.7905	-0.1124	-0.0038	-0.0038	-0.7329	-0.0955	-0.0019	-0.0039
Offline = 10 km/h; Subway = 36 km/h	-0.5049	-0.0718	-0.0024	-0.0024	-0.4642	-0.0605	-0.0012	-0.0025
Offline = 8 km/h; Subway = 36 km/h	-0.3355	-0.0477	-0.0016	-0.0016	-0.3059	-0.0399	-0.0008	-0.0016
Kleibergen-Paap rk Wald F statistic	382.878	382.878	382.878	382.878	382.878	382.878	382.878	382.878
Observations	172,206	172,206	172,206	172,206	172,206	172,206	172,206	172,206
Root MSE	0.396	0.161	0.012	0.010	0.395	0.161	0.012	0.010
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Dest. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We first identify all pairs that include at least one applicant who collaborated after moving as "Movers," and the rest as "non-Movers." If each applicant'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. In columns (1) - (4), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year interacted with the Euclidean distance between those two locations. In columns (5) - (8), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year subtracting own pair length before interacted with the Euclidean distance between those two locations (Leave-one-out IV). Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA8: Impact of Subway Expansion on Collaborators of Different Productivity and Mobility—Grid Pair-Level Analysis
Local Average Causal Response - Robustness with Sample Cut at 80th Percentile
Dependent Variable: IHS(Collab. patents)

	Method: IV				Method: IV(LOO)			
	(1) by H-H non-Movers	(2) by Other non-Movers	(3) by H-H Movers	(4) by Other Movers	(5) by H-H non-Movers	(6) by Other non-Movers	(7) by H-H Movers	(8) by Other Movers
log(Length)	0.0509***	0.0050**	0.0001	0.0002	0.0460***	0.0042**	0.0001	0.0002
	(0.0115)	(0.0023)	(0.0002)	(0.0002)	(0.0100)	(0.0020)	(0.0001)	(0.0002)
Wald Estimators Based on Travel Speed Assumptions of ...								
Offline = 12 km/h; Subway = 36 km/h	-0.9695	-0.0952	-0.0019	-0.0038	-0.8967	-0.0819	-0.0019	-0.0039
Offline = 10 km/h; Subway = 36 km/h	-0.6192	-0.0608	-0.0012	-0.0024	-0.5679	-0.0519	-0.0012	-0.0025
Offline = 8 km/h; Subway = 36 km/h	-0.4115	-0.0404	-0.0008	-0.0016	-0.3743	-0.0342	-0.0008	-0.0016
Kleibergen-Paap rk Wald F statistic	468.084	468.084	468.084	468.084	468.084	468.084	468.084	468.084
Observations	158,004	158,004	158,004	158,004	158,004	158,004	158,004	158,004
Root MSE	0.398	0.162	0.012	0.011	0.397	0.162	0.012	0.011
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Dest. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We first identify all pairs that include at least one applicant who collaborated after moving as "Movers," and the rest as "non-Movers." If each applicant'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. In columns (1) - (4), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year interacted with the Euclidean distance between those two locations. In columns (5) - (8), we instrument for the *Length* variable between two locations in a year using the total cumulative length of the subway lines in Beijing in the same year subtracting own pair length before interacted with the Euclidean distance between those two locations (Leave-one-out IV). Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA9: Robustness Checks – Alternative Sample Selection

	(1)	(2)	(3)	(4)
Sample	90%	80%	90%	80%
Panel A.				
	𝟙(Collab. patents)		IHS(Collab. patents)	
log(Length)	0.0141***	0.0149***	0.0411***	0.0474***
	(0.004)	(0.0038)	(0.0113)	(0.009)
Observations	181,773	166,782	181,773	166,782
Panel B.				
	log(Collab. patents) [asinh transf] by Incumbent Pairs		IHS(Collab. patents) by Non-Incumbent Pairs	
log(Length)	0.0363***	0.0460***	0.0084***	0.0063***
	(0.0118)	(0.0098)	(0.0024)	(0.0021)
Observations	172,206	158,004	172,206	158,004
Panel C.				
	log(Collab. patents) [asinh transf] by Nonmover Pairs		IHS(Collab. patents) by Mover Pairs	
log(Length)	0.0400***	0.0479***	0.0003	0.0003
	(0.0121)	(0.0101)	(0.0002)	(0.0002)
Observations	172,206	158,004	172,206	158,004
Panel D.				
	log(Collab. patents) [asinh transf] by H-H Nonmover Pairs		IHS(Collab. patents) by Other Nonmover Pairs	
log(Length)	0.0376***	0.0460***	0.0049**	0.0042**
	(0.0120)	(0.0010)	(0.0023)	(0.0020)
Observations	172,206	158,004	172,206	158,004
Method	IV (LOO)	IV (LOO)	IV (LOO)	IV (LOO)
Year FE	YES	YES	YES	YES
Grid Pair FE	YES	YES	YES	YES

Notes: In Panel A, we use the sample from 2000 to 2018. In Panels B to D, we drop the initial year of 2000 since we use the previous year's information to determine incumbent, mobility, and productivity status. The *Length* variable measures the length of the Beijing subway lines that connect a station near the centroid of one grid to that of another grid in a given year along the fastest travel route. We instrument for the *Length* variable for a grid pair in a year using the leave-one-out aggregate length (total cumulative length of the subway lines in Beijing in that year minus own pair length) interacted with the Euclidean distance between those two grids. Estimates for a full expansion of the detailed breakdowns are provided in Tables OA7 and OA8. 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 OA10: Impact of Subway Expansion on Probability of Collaboration
—Collaborator Pair-Level Analysis

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

Notes: Using the empirical methodology developed by Jaffe et al. (1993), we create the control group of non-collaborators by matching the first listed IPC code at the three digits and application year. For "One-to-One Match" ("One-to-Two Match") sample, we use one (two) randomly selected control pair(s) 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 OA11: Counterfactual Design using Planned but Unbuilt Stations

	Δ IHS(Collab. patents)			
Panel A. Grid Level	(1)	(2)	(3)	(4)
𝟙(Treated)	2.3733***	4.0128***	3.4207***	3.6864***
	(0.1573)	(0.8273)	(0.8734)	(0.6136)
Observations	7,010	73	81	116
Sample	Full Sample	Treat 1 & Control	Treat 2 & Control	Treat 1, 2 & Control
Adjusted R^2	0.0006	0.2637	0.1471	0.1380

	Δ IHS(Collab. patents)			
Panel B. Grid-pair Level	(1)	(1)	(3)	(4)
𝟙(Treated)	1.4011***	1.5695***	1.8253***	1.6791***
	(0.0096)	(0.2446)	(0.4412)	(0.2374)
Observations	102,844	835	1,017	1,830
Sample	Full Sample	Treat 1 & Control	Treat 2 & Control	Treat 1, 2 & Control
Adjusted R^2	0.6258	0.8363	0.7397	0.7802

Notes: The dependent variable Δ IHS(Collab. patents) is defined as the change in IHS(Collab. patents) from its value in 2000 to its cumulative value between 2001 and 2018. In Panel A, 𝟙(Treated) equals to one for the grids within one kilometer of the treated stations; and zero otherwise. In Panel B, 𝟙(Treated) equals to one for the grid-pairs that contain both treated grids; and zero otherwise. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA12: Impact of Subway Expansion on Patent Collaboration—Alternative Distance Thresholds for Sample Selection

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

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

Table OA13: Impact on Different Types of Collaborators using IV Approach
—Grid Pair-Level Analysis
Robustness with Alternative Definition of Movers

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

Notes: We use the sample from 2001 to 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. In Columns (1) and (3), we instrument for the *Length* variable for a grid pair 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 grids. In Columns (2) and (4), the instrument is similarly constructed except that we now subtract own pair length from the total cumulative length of the subway lines in Beijing in a year (i.e. Leave-one-out). 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 OA14: Impact on Different Types of Collaborators using TWFE and DiD Approaches —Grid Pair-Level Analysis
Robustness with Alternative Definition of Productivity

	(1)	(2)	(3)	(4)
Dep. variable	IHS(Collab. patents)		IHS(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
𝟙(Treated)	0.0289***	0.0203***	-0.0003	0.0016
	(0.0083)	(0.0048)	(0.0022)	(0.0023)
Observations	148,485	148,485	148,485	148,485
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
𝟙(Treated)	0.0283***	0.0217***	-0.00002	-0.0002
	(0.0088)	(0.0054)	(0.0001)	(0.0002)
Observations	148,485	148,485	148,485	148,485
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
𝟙(Treated)	0.0067***	0.0047**	-0.0013	-0.0019
	(0.0026)	(0.0020)	(0.0018)	(0.0012)
Observations	148,485	148,485	148,485	148,485
Method	TWFE	DiD	TWFE	DiD
Year FE	YES		YES	
Grid Pair FE	YES		YES	

Notes: We use the sample from 2001 to 2015. We drop the initial year of 2000, since we use the previous year's information to determine incumbent and mobility status. We drop year from 2016 onward because we determine productivity status based on patent approval rate until 2015. An applicant is considered to be of H type if it has patent approval rate of 0.65, which is approximately the median in the data. 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 OA15: Impact on Different Types of Collaborators using IV Approach
—Grid Pair-Level Analysis
Robustness with Alternative Definition of Productivity

	(1)	(2)	(3)	(4)
Dep. variable	IHS(Collab. patents)		IHS(Collab. patents)	
Panel A.				
	By Incumbent Pairs		By Non-Incumbent Pairs	
log(Length)	0.0577***	0.0495***	0.0105***	0.0084***
	(0.0156)	(0.0126)	(0.0036)	(0.0030)
Observations	148,485	148,485	148,485	148,485
Panel B.				
	By Nonmover Pairs		By Mover Pairs	
log(Length)	0.0636***	0.0543***	0.00008	0.00006
	(0.0164)	(0.0132)	(0.0003)	(0.0002)
Observations	148,485	148,485	148,485	148,485
Panel C.				
	By H-H Nonmover Pairs		By Other Nonmover Pairs	
log(Length)	0.0175***	0.0144***	0.0145***	0.0116***
	(0.0051)	(0.0040)	(0.0036)	(0.0029)
Observations	148,485	148,485	148,485	148,485
Method	IV	IV (LOO)	IV	IV (LOO)
Year FE	YES	YES	YES	YES
Grid Pair FE	YES	YES	YES	YES

Notes: We use the sample from 2001 to 2015. We drop the initial year of 2000, since we use the previous year's information to determine incumbent and mobility status. We drop year from 2016 onward because we determine productivity status based on patent approval rate until 2015. An applicant is considered to be of H type if it has patent approval rate of 0.65, which is approximately the median in the data. 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. In Columns (1) and (3), we instrument for the *Length* variable for a grid pair 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 grids. In Columns (2) and (4), the instrument is similarly constructed except that we now subtract own pair length from the total cumulative length of the subway lines in Beijing in a year (i.e. Leave-one-out). 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 OA16: Treatment Effect on Patent Collaboration—Grid Pair-Level Analysis Using Approved Collaborated Patents Only (2000 to 2015)

	$\mathbb{1}$(Collab. approved patents)		IHS(Collab. approved patents)	
Panel A.	(1)	(2)	(3)	(4)
$\mathbb{1}$(Treated)	0.013***	0.010***	0.026***	0.024***
	(0.002)	(0.003)	(0.005)	(0.006)
Observations	158,384	158,384	158,384	158,384
Method	TWFE	TWFE	TWFE	TWFE
Adjusted R^2	0.047	0.075	0.066	0.171
Year FE	Yes	Yes	Yes	Yes
Grid FE	Yes	No	Yes	No
Grid Pair FE	No	Yes	No	Yes

Panel B.		
$\mathbb{1}$(Treated)	0.0102***	0.0213***
	(0.0024)	(0.0047)
Observations	158,384	158,384
Method	DiD	DiD

Notes: We focus on all collaborated patents in 2000-2015 which were approved. In Panel B, the control group consists of the pairs of grids that were never treated, but were located within 2*km* from a subway station during our sample period. For a grid pair, $\mathbb{1}$(Treated) equals one for the first and the following years when the travel time between the grids gets reduced by more than 30 minutes compared to the previous year; and zero otherwise. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA17: Impact of Subway Expansion on Patent Collaboration—Grid Pair-Level Analysis Using Approved Collaborated Patents Only (2000 to 2015)

Dep. variable	$\mathbb{1}$(Collab. patents)		IHS(Collab. patents)	
	(1)	(2)	(3)	(4)
log(Length)	-0.0019***	0.0020***	-0.0031***	0.0033***
	(0.0003)	(0.0005)	(0.0007)	(0.0012)
Wald Estimators Based on Travel Speed Assumptions of ...				
Offline = 12 km/h; Subway = 36 km/h	-0.1418	-0.1538	-0.2313	-0.2538
Offline = 10 km/h; Subway = 36 km/h	-0.1881	-0.0952	-0.3069	-0.1571
Offline = 8 km/h; Subway = 36 km/h	-0.3725	-0.0606	-0.6078	-0.1000
Year FE	YES	YES	YES	YES
Grid FE	YES	NO	YES	NO
Grid Pair FE	NO	YES	NO	YES
Observations	158,384	158,384	158,384	158,384
R-squared	0.0676	0.1329	0.0869	0.2231

Notes: We focus on all collaborated patents in 2000-2015 which were approved. 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 OA18: Impact of Subway Expansion on Patent Collaboration
Controlling for Internet Penetration Rate and/or Car Ownership Rate

	(1)	(2)	(3)
Dep. variable		IHS(Collab. patents)	
Panel A			
$\mathbb{1}$(Treated)	0.0212***	0.0142***	0.0143**
	(0.0077)	(0.0047)	(0.0046)
Panel B			
log(Length)	0.0041***	0.0035***	0.0035***
	(0.0015)	(0.0012)	(0.0012)
Panel C (IV)			
log(Length)	0.0544***	0.0596***	0.0616***
	(0.0087)	(0.0099)	(0.0103)
Panel D (IV-LOO)			
log(Length)	0.0484***	0.0448***	0.0461***
	(0.0074)	(0.0072)	(0.0074)
Internet Penetration Rate × Initial Travel Time	YES	NO	YES
Car Ownership Rate × Initial Travel Time	NO	YES	YES
Year FE	YES	YES	YES
Grid Pair FE	YES	YES	YES
Observations	178,182	98,990	98,990

Notes: Due to data availability on Beijing Internet Penetration Rate and Car Ownership Rate, results reported in Column (1) are based on sample between 2000 and 2017 and results reported in Columns (2) and (3) are based on sample between 2001 and 2010. For a grid pair, $\mathbb{1}$(Treated) equals one for the first and the following years when the travel time between the grids gets reduced by more than 30 minutes compared to the previous year; and zero otherwise. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA19: Impact of Subway Expansion on Patent Collaboration
Controlling for Internet Penetration Rate and/or Car Ownership Rate

	(1)	(2)	(3)
Dep. variable		𝟙(Collab. patents)	
Panel A (IV)			
𝟙(Treated)	0.3130***	0.2949***	0.3042**
	(0.0491)	(0.0533)	(0.0550)
Kleibergen-Paap rk Wald F statistic	158.435	168.247	158.183
Panel B (IV-LOO)			
𝟙(Treated)	0.2924***	0.2395***	0.2462**
	(0.0437)	(0.0424)	(0.0435)
Kleibergen-Paap rk Wald F statistic	188.568	244.483	231.578
Dep. variable		IHS(Collab. patents)	
Panel C (IV)			
𝟙(Treated)	0.6949***	0.5734***	0.5942***
	(0.1170)	(0.0985)	(0.1023)
Kleibergen-Paap rk Wald F statistic	158.435	168.247	158.183
Panel D (IV-LOO)			
𝟙(Treated)	0.6431***	0.4678***	0.4836***
	(0.1037)	(0.0774)	(0.0801)
Kleibergen-Paap rk Wald F statistic	188.568	244.483	231.578
Internet Penetration Rate × Initial Travel Time	YES	NO	YES
Car Ownership Rate × Initial Travel Time	NO	YES	YES
Year FE	YES	YES	YES
Grid Pair FE	YES	YES	YES
Observations	178,182	98,990	98,990

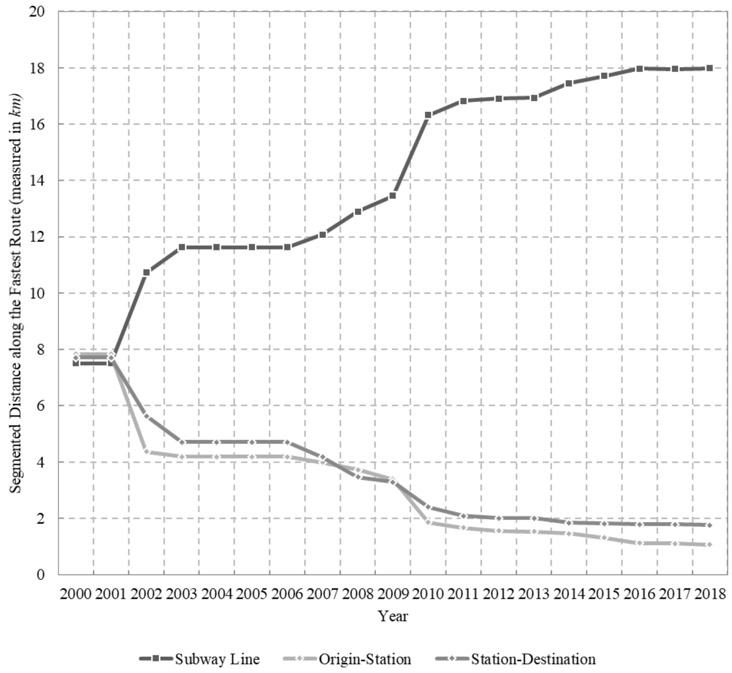
Notes: Due to data availability on Beijing Internet Penetration Rate and Car Ownership Rate, results reported in Column (1) are based on sample between 2000 and 2017 and results reported in Columns (2) and (3) are based on sample between 2001 and 2010. For a grid pair, 𝟙(Treated) equals one for the first and the following years when the travel time between the grids gets reduced by more than 30 minutes compared to the previous year; and zero otherwise. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the grid pair level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1 respectively.

Table OA20: Using More Regular-Shaped Set of Grids — A Grid Pair Level Analysis

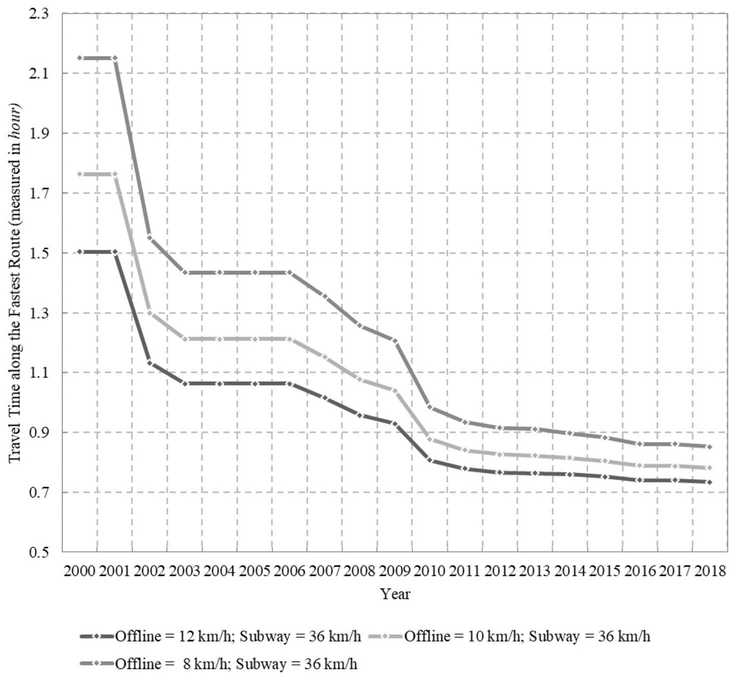
	$\mathbb{1}$(Collab. patents)		IHS(Collab. patents)	
	(1)	(2)	(3)	(4)
log(length)	0.0227***	0.0216***	0.0637***	0.0606***
	(0.0039)	(0.0036)	(0.0088)	(0.0082)
Observations	105,450	105,450	105,450	105,450
Method	IV	IV(LOO)	IV	IV(LOO)
Year FE	Yes	Yes	Yes	Yes
Grid Pair FE	Yes	Yes	Yes	Yes

Notes: The sample contains all grids ever located within 2 *km* from a subway station during 2000-2018 that form a more regular shape overall. Figure OA7 displays the selection of the sub-sample. 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.

G. Appendix Figures



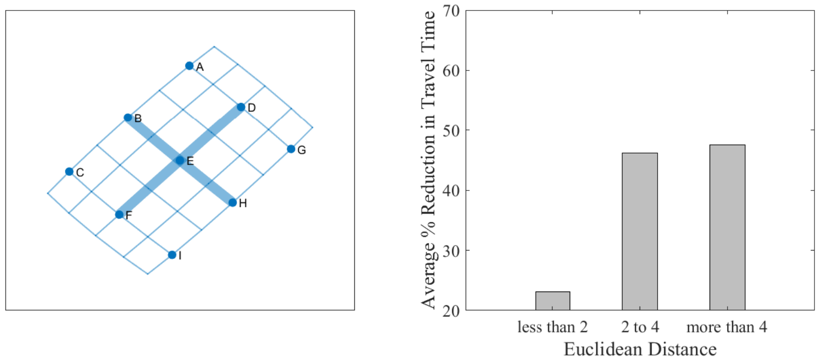
(a) Segmented Distance along the Fastest Route



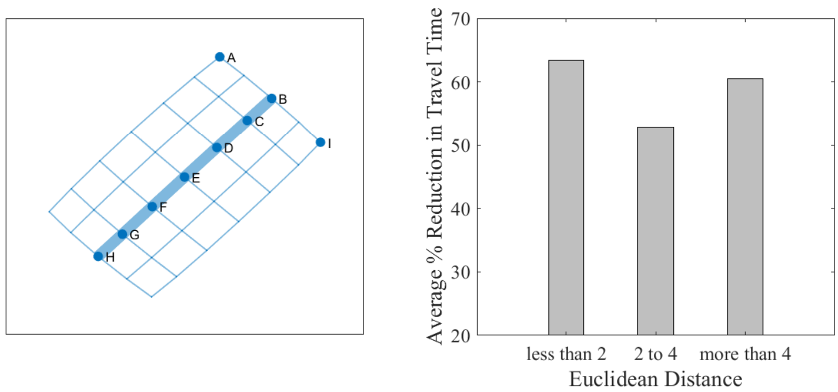
(b) Travel Time along the Fastest Route

Figure OA1: Subway Expansion and Travel Time

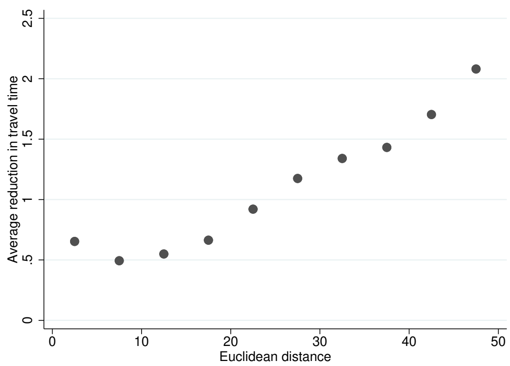
Notes: Panel (a) shows the segmented distance measures (*km*) along the fastest travel route. Panel (b) shows the travel time measures (*hour*) along the fastest travel route.



(a) Example 1



(b) Example 2



(c) Euclidean Distance and Average Reduction in Travel Time in the Data

Figure OA2: Euclidean Distance and Layout of Subway Stations

Notes: In Panel (a), Example 1 assumes a symmetric and orderly layout of 9 locations (Points A-I) and the subway lines (thick blue lines) that connect some of those locations. In Panel (b), Example 2 assumes an alternative layout which is less grid-like. We assume that one can travel between location points along the grid lines and using the subway increases the travel speed by a factor of four. Each bar graph summarizes the average percentage reduction in travel time given the subway construction for corresponding example. In Panel (c), we summarize the relationship between Euclidean distance and average reduction in travel time (unit: hour) between 2000 and 2018 using all grid pairs in our data.

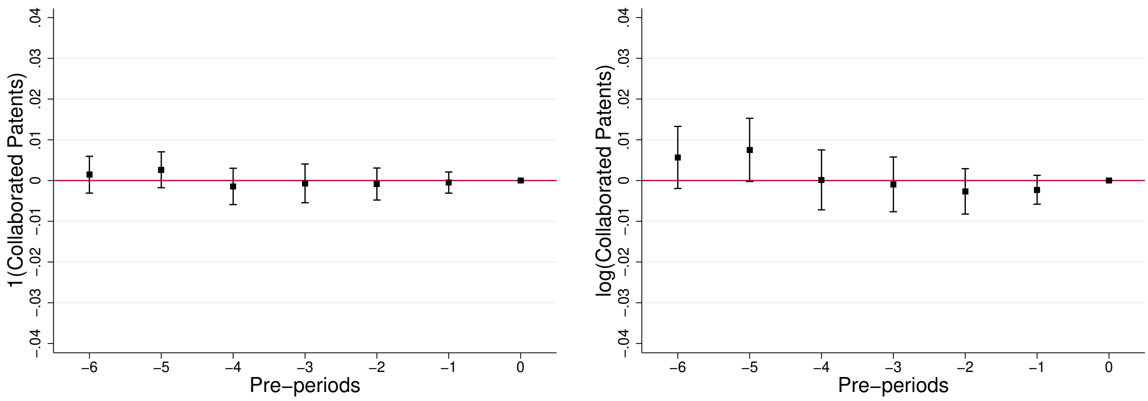


Figure OA3: Pre-trends

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

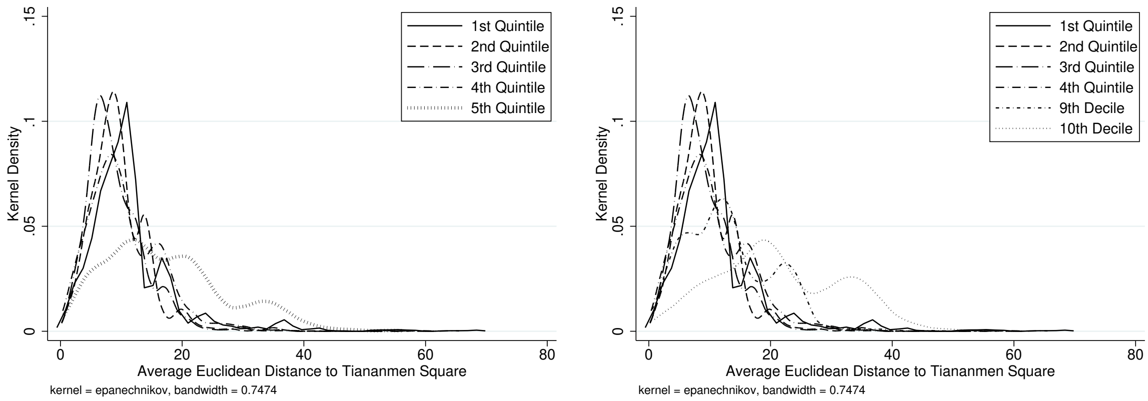
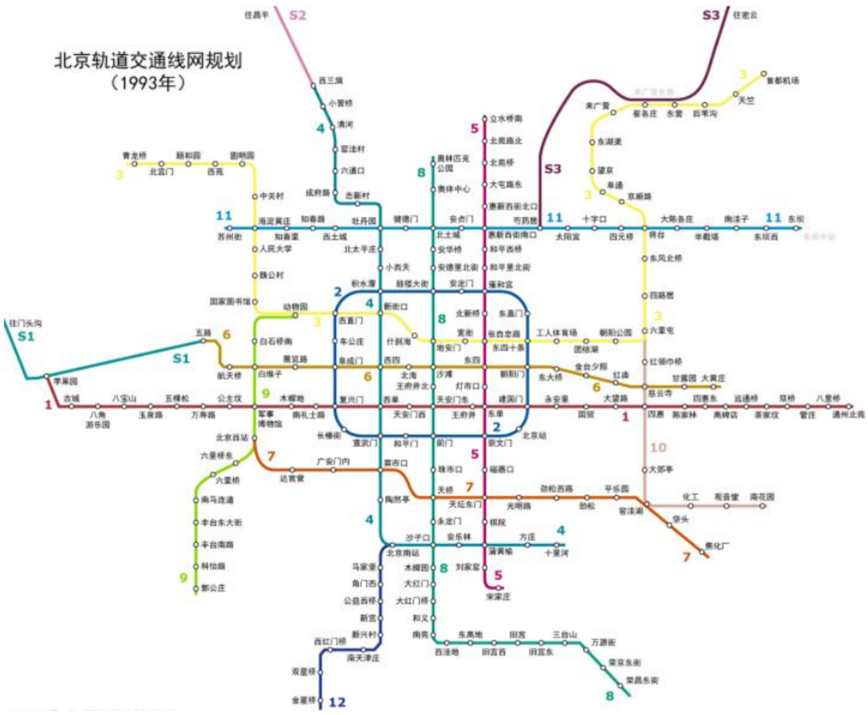
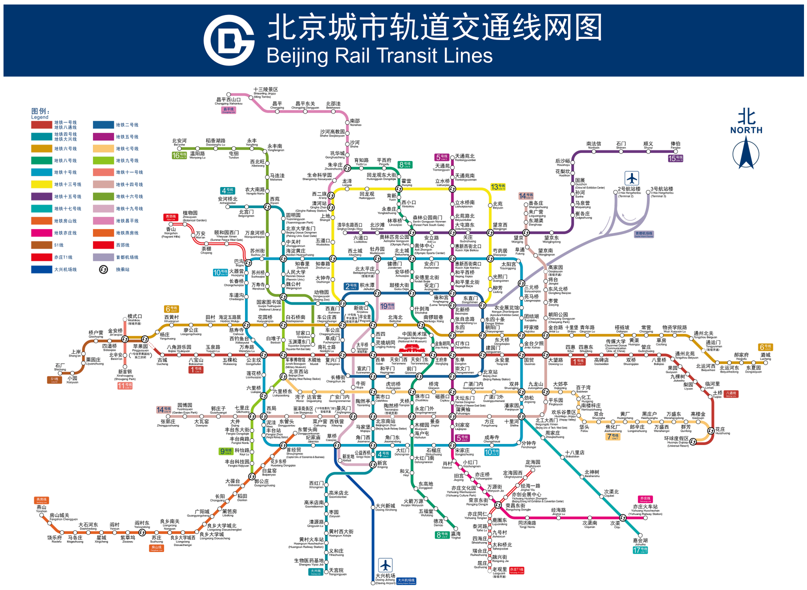


Figure OA4: Kernel Density

Notes: For a grid pair, we calculate the Euclidean distance between the centroid of a grid and Tiananmen Square, and the Euclidean distance between centroid of the other grid and Tianamen Square. Then we compute their average value and derive this average distance measure for every grid pair used in the empirical analysis. The kernel density is plotted based on these average distance values.



(a) Beijing Subway Network Planned in 1993



(b) Actual Beijing Subway Network in 2020

Figure OA5: Planned Subway Network in 1993 and Actual Subway Network in 2020

Notes: In Panel (a), the source is: Compilation of the Chronicle of Beijing Urban Construction Design and Research Institute (1958-2008), Beijing Urban Construction Design and Research Institute Co., Ltd., 2008. In Panel (b), the source is: `https://www.bjsubway.com/station/xltzs/`

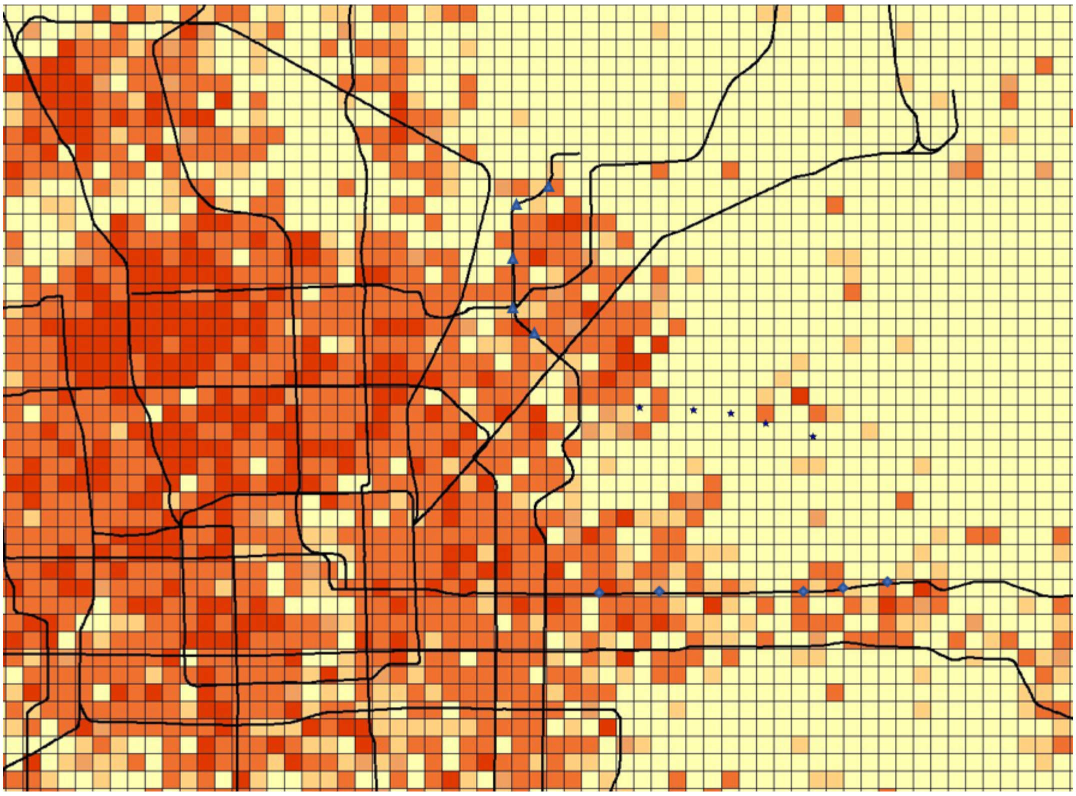


Figure OA6: Counterfactual Stations and Two Selected Groups of Treated Stations

Notes: Counterfactual stations refer to stations that were planned in 1993 but not built during our sample period. The location of those stations are marked in stars. One set of treated stations, "treat 1" are chosen to be along Line 14 and are marked in triangles. The other set of treated stations, "treat 2", are chosen to be along line 6 and are marked in diamonds. The background of the map shows the grids on which our empirically analysis is based. Darker shades indicate a higher increase in collaborative patents between 2000 and 2018.

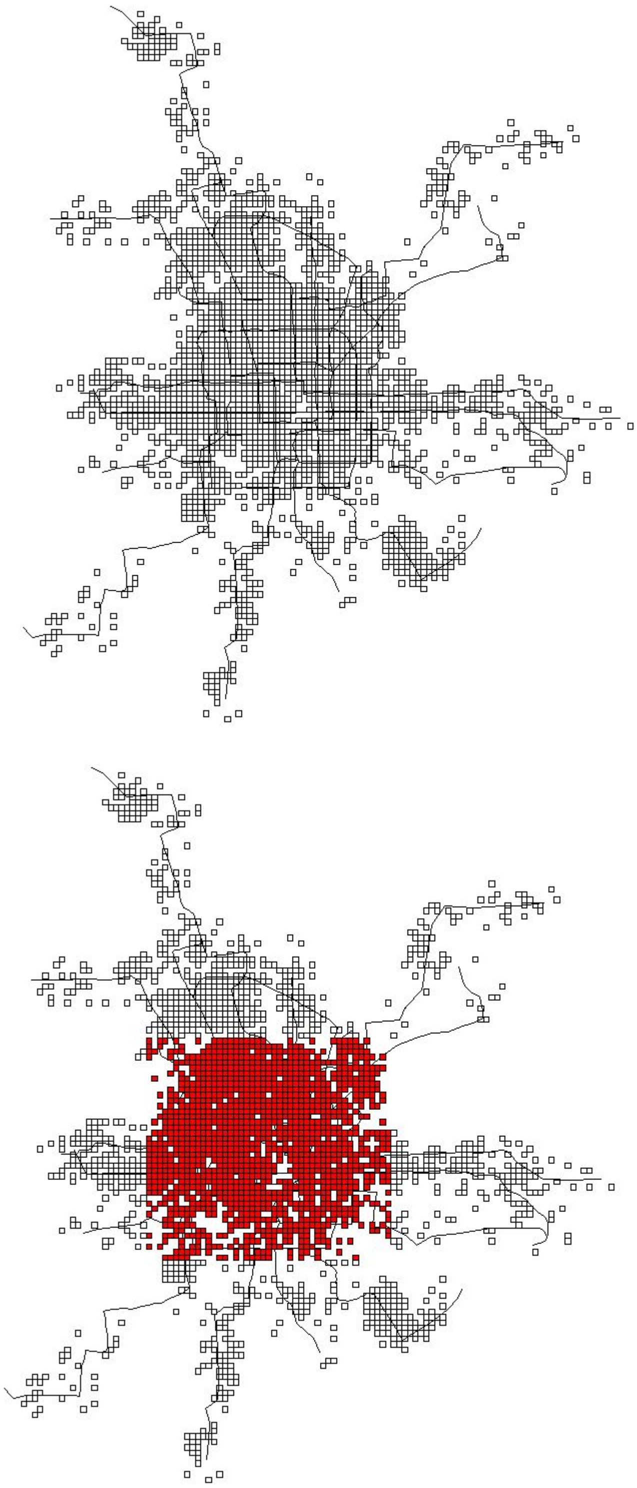


Figure OA7: Sub-sample of Grids

Notes: The upper panel displays the full set of grids used for our baseline grid pair-level analysis. The grids in red displayed in the lower panel are the selected more regular-shaped set of grids used to produce results in Table OA20.