RIETI Discussion Paper Series 13-E-070

Localization of Collaborations in Knowledge Creation (Revised)

INOUE Hiroyasu
Osaka Sangyo University

NAKAJIMA Kentaro
Tohoku University

SAITO Yukiko Umeno
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 13-E-070
First draft: August 2013
Revised: March 2015

Localization of Collaborations in Knowledge Creation[*]

INOUE Hiroyasu[†]
Osaka Sangyo University
NAKAJIMA Kentaro[‡]
Tohoku University
SAITO Yukiko Umeno[§]
Research Institute of Economy, Trade and Industry

Abstract

This study investigates the localization of collaborative works in knowledge creation, using data on Japanese patent applications. Applying distance-based methods, we obtained the following results. First, collaborations are significantly localized at the 5% level, within a localization range of approximately 100 km. Second, the extent of localization was stable during 1986–2005 despite extensive developments in information and communications technology that facilitate communication between remote organizations. Third, the extent of localization is substantially greater in inter-firm collaborations than in intra-firm collaborations. Furthermore, in inter-firm collaborations, the extent of localization is greater in collaborations with small firms. This suggests that geographic proximity mitigates firm-border effects in collaborations, especially for small firms.

Keywords: Knowledge creation, Collaboration, Geographic frictions, Firm-border effects
JEL codes: R12; O31

[*]This study is conducted as a part of the Project "Inter-organizational and Inter-inventors Geographical Proximity and Networks" undertaken at Research Institute of Economy, Trade and Industry (RIETI). The authors thank various supports from the institute. We thank Victor Couture, Gilles Duranton, Tatsuaki Kuroda, Yasusada Murata, Sadao Nagaoka, Ryosuke Okamoto, Yukako Ono, and the participants at the UEA in Ottawa, ARSC at Toyama University, and seminars at Fukuoka University, Hitotsubashi University, Kobe University, Nihon University, RIETI, Temple University, and the University of Tokyo for their helpful comments. We also thank the Center for Spatial Information Science, University of Tokyo, for providing us with geocoding services. We gratefully acknowledge financial support from the Japan Society for the Promotion of Science (Nos. 22330073, 23730228, 23330076, 24530506, 25220502, 25380275, 25780181, and 26245037).

[†] Faculty of Business Administration, Osaka Sangyo University, 3-1-1, Nakagaito, Daitoshi, Osaka, 574-0013 Japan
[‡] Faculty of Economics, Tohoku University, 27-1 Kawauchi Aoba-ku, Sendai, 980-8576 Japan
[§] Research Institute of Economy, Trade and Industry, 1–3-1, Kasumigaseki Chiyoda-ku, Tokyo, 100-8901 Japan

1. Introduction

Economic activities are concentrated in certain areas. The agglomeration of information technology (IT) firms in the Silicon Valley is a well-known example of industry agglomeration.[1] Knowledge spillovers are known to constitute one of the determinants of agglomerations of economic activities (Marshall, 1920; Rosenthal and Strange, 2001; Ellison, Glaeser, and Kerr, 2010). Several studies have examined the localization of knowledge spillovers (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Murata et al., 2011, Kerr and Kominers, 2014).

These studies use patent citation as a proxy for knowledge spillovers. However, patent citations may not fully capture the geographic features of knowledge spillovers. As Polanyi indicates in his seminal works (1958, 1966), knowledge can be classified into two types: explicit (codified) knowledge and tacit knowledge. Published patents are typically codified knowledge, as many articles note (e.g., Hansen, 1999; Hansen et al. 2005; Nonaka and Takeuchi, 1995). Therefore, knowledge transferred by the patent citation is generally limited to explicit knowledge, and only a negligible amount of tacit knowledge transfer is captured in the patent citation approach.

To capture tacit knowledge spillovers, this paper focuses on collaborative works in knowledge creation. Tacit knowledge is shown to be transferred by close-range observation, demonstration, and hands-on experience (Nonaka and Takeuchi, 1995). These processes are involved in collaborative works. Thus, collaborative works can be considered as a proxy for tacit knowledge transfers. Wuchty, Jones, and Uzzi (2007) noted that collaborative works in knowledge creation are rapidly increasing and that the quality of the created knowledge is higher in collaborative works than in solo works. This suggests that the exchange of tacit knowledge is crucial to knowledge creation.

This paper investigates the geographic features of tacit knowledge spillovers. Tacit knowledge is generally transferred through face-to-face contact between researchers. Since face-to-face contact requires researchers' trips, geographical distance has a substantially more important role to play in tacit knowledge transfers than explicit knowledge transfers. Thus, tacit knowledge transfers have a substantially stronger agglomeration force than explicit knowledge transfers in regard to economic activities.

Collaborative works have been investigated in a firm-level analysis (e.g., Hagedoorn, 2003). However, firm-level analysis is not suitable for the investigation of geographic features of collaboration. Several firms have multiple establishments, and the address information of firms does not necessarily represent the

[1] Duranton and Overman (2005) and Nakajima, Saito, and Uesugi (2012) show that nearly half of the manufacturing industries in the United Kingdom and Japan, respectively, tend to be localized using distance-based methods.

actual location of invention. To precisely investigate the geographic features of collaboration, we use establishment-level collaboration information from Japanese patent data.

Notably, establishment-level analysis enables us to investigate firm-border effects in collaborations. Collaborations between firms are considered difficult, because their interests conflict and unwanted transfers of knowledge and organizational secrets significantly harm firm competitiveness (e.g., Häusler, 1994; Pittaway et al., 2004). By comparing between inter- and intra-firm collaborations, we can quantitatively evaluate firm-border effects in collaborations.

To assess the geographic proximity of establishment-level collaborations, we should consider the geographic proximity of the overall research establishments. Knowledge-creating establishments are concentrated in certain areas (e.g., Carlino, Hunt, Carr, and Smith, 2012; Inoue, Nakajima, and Saito, 2014). Thus, a naive measurement of the geographic proximity between collaborating establishments inevitably reflects the proximity of establishments. Thus, to appropriately assess the proximity of collaborating establishments, we should control for the proximity of establishments.

To control for the geographic proximity of establishments, we consider counterfactual collaborations in which establishments choose collaborating partners regardless of distance. These counterfactual collaborations reflect the geographic proximity of establishments. Next, we compare actual and counterfactual collaborations. Specifically, we first assess the proximity of actual collaborations as normalized by counterfactual collaborations, which we call "relative density". Second, we provide a formal statistical inference for the localization of collaborations by applying Duranton and Overman's (2005) K-density approach.

We obtain the following results. First, collaborations between establishments are significantly localized at the 5% level with a localization range of approximately 100 km. Second, the extent of localization remains stable during the two decades (1986–2005) under consideration despite extensive developments in information and communications technology (ICT) that facilitate communication between remote researchers. This suggests that geographic distance continues to play an important role in knowledge spillovers that require face-to-face communication. Third, the extent of localization is substantially larger in inter-firm than in intra-firm collaborations. These results suggest that geographic proximity mitigates the firm-border effects in collaborations. Finally, in inter-firm collaborations, the extent of localization is larger in the collaborations with small firms. As a whole, our results suggest that inter-establishment collaborations are localized and stable and that geographic proximity mitigates firm-border effects, especially for small firms.

The remainder of this paper is organized as follows. In the next section, we introduce the dataset and identify both the establishments and the firms that own them. Section 3 describes our empirical strategy, through which we utilize the micro-geographic information for each establishment. Section 4 presents our baseline results. Section 5 focuses on firm-border effects in collaborations between establishments. Section 6 investigates the firm size effect on inter-firm collaborations. Finally, Section 7 concludes the paper.

2. Data

We constructed a patent dataset that includes all of the patents published in Japan from 1993 to 2008. We extracted the patent ID, application date, names and addresses of applicants, and addresses of inventors from the Institute of Intellectual Property (IIP) patent database (DB) (Goto and Motohashi, 2007)..

Note that the publication date differs from the application date because of the reviewing time. To appropriately capture the timing of collaboration activities, we used the application date because it is closer to the timing of collaboration. We then restricted our sample to patents applied for from 1986 to 2005.

In this study, we focus on collaborations between the establishments to which the inventors belong. Following a convention in the Japanese patent system, the address that inventors use when they register is that of their establishment. Typically, this establishment address includes its firm's name. Accordingly, registration data provide information on establishment-level collaborations in patent creation.

Here, we describe the algorithm that is used to identify the establishments and their collaborations from our patent database. For each patent, we conduct the following procedure. First, firms are identified using the applicants' names and addresses. Here, a firm is defined as an applicant if the name includes the term "company limited" (in Japanese, "*kabushikigaisha*"). This definition simultaneously excludes all relatively small firms, such as private limited companies. Second, the establishments are identified as follows. We check whether the firm's name is included in the inventor's address. We then consider the inventor's address with the firm name as the address of the firm's establishment.[2] Finally, we identify the collaborations as follows. When a patent includes information about more than one establishment, we define these establishments as collaborating partners.

In some patents, the firm name is not included in the inventor's address. There are three possible reasons

[2] In practice, there are many different written forms in addresses. Thus, we convert addresses to latitude and longitude information using the geocoding service provided by the Center for Spatial Information Science at the University of Tokyo.

for this. First, the inventor resisters their establishment address, but it does not include firm name. Second, the inventor's residence address is registered even though that person is part of an establishment. Third, the inventor does not belong to any establishment or firm. In the first case, we apply another algorithm to obtain a firm name using establishment addresses.[3]

One may raise the following question: How many of the establishments can we identify considering that our methodology is based on a Japan-specific convention? Table 1 provides the summary of the dataset. Out of 1,967,361 patents, we can identify establishments for 1,189,262 patents, i.e., 60.4% of the total. Since approximately 80% of patents are issued by firms in OECD countries (OECD, 2008), we have successfully matched three-fourths of the total patents registered by firms. Table 1 provides additional information. We identified 56,592 firms and 74,452 establishments using our algorithm.

[Table 1 here]

We focus on establishment-level collaborations. From 1,189,262 patents applied for by establishments, 8.1% patents (i.e., 93,939 patents) involve establishment-level collaborations. A total of 59.7% of the patents that involve establishment-level collaborations (56,074 patents) involve collaborations between different firms (i.e., inter-firm collaborations).

Firms can be further classified into multi-establishment and single-establishment firms. Our dataset includes 9,688 multi-establishment firms and 46,904 single-establishment firms. Here, we do not consider establishments that published no patents during the focus period when we count number of establishments.

3. Empirical Strategy

Knowledge-creating establishments are concentrated in certain areas (e.g., Carlino, Hunt, Carr, and Smith, 2012; Inoue, Nakajima, Saito, 2014). Therefore, even if establishments choose collaborating partners regardless of distance (e.g., they make a random choice), collaborating establishments are located within a short distance. To appropriately assess the proximity of collaborating establishments, we should control for the geographic proximity of the overall research establishments.

To control for this geographic proximity, we consider counterfactual collaborations in which establishments

[3] If the inventor's address does not include the firm's name, we attempt to match the address information (latitude and longitude) to an establishment address from other patents that includes firm name in inventor's address.

choose partners regardless of distance. These counterfactual collaborations reflect the proximate locations of establishments. Therefore, a comparison between actual and counterfactual collaborations that controls for the geographic proximity of establishments enables an appropriate assessment of the proximity of collaborations.

Specifically, we first assess the proximity of collaborating partners using actual distributions normalized by counterfactual distributions. We then provide a formal statistical inference for localization using Duranton and Overman's (2005) K-density approach.

3.1. Potential collaborating partners and counterfactual collaborations

Here, we describe in detail how to consider potential collaborating partners and construct counterfactual collaborations. Let S^A be a set of establishments that have applied for at least one patent in patent technology class $A \in \mathfrak{A}$ over the entire analysis period (1986–2005), where $\mathfrak{A}$ represents a set of patent technology classes. In the analysis, we use the first three letters of International Patent Classification (IPC) to categorize patents.[4] There are 120 patent technology classes in our data. Let n_A be the number of establishments publishing patents in technology class A, let P_A be the set of patent applications submitted by the collaborating establishments, let $p_{ij}^A \in P_A$ be a patent application submitted by collaborating establishments i and j in technology class A, and let n_A^p be the number of patent applications submitted by the collaborating establishment in technology class A. We set S^A as the set of potential collaborating partners in patent technology class A. In other words, elements from the set of potential collaborating partners S^A can collaborate with one another. Using the definition of potential collaborating partners, we can generate counterfactual collaborations. For counterfactual collaborations, we randomly choose n_A^p pairs of establishments from the set of potential collaborating partners S^A.[5]

These counterfactual collaborations can be considered to represent a situation in which establishments randomly choose their partners regardless of distance. Thus, these counterfactual captures the geographic proximity of the establishments.

One may argue that establishment location is determined by the likelihood of future collaborations. If an establishment chooses a region in which many potential collaborators are located, counterfactual collaborations will capture this location proximity due to the likelihood of future collaborations. Thus, our counterfactual approach controls for such endogenous location choices.

---

[4] A patent often has multiple IPCs. We use the first IPC assigned in each patent.

[5] For robustness, we use another definition for potential collaborating partners. See Appendix.

3.2. Methodology 1: Relative density

First, we depict the extent of collaboration localization by simply investigating the bilateral distance distributions of the collaborating establishments, which is normalized by those of all of the potential collaborating partners to control for the existing localization of establishments.

Specifically, we first calculate all of the bilateral distances between the collaborating establishments to obtain the density in each distance bracket, which we call the actual collaboration density. Next, to control for the localization of establishments, we calculate all of the bilateral distances between potential collaborating establishments to obtain their density in each distance bracket, which we call potential density. Finally, we define the relative density as the metric of collaboration intensity by taking the ratio of densities between the actual and potential collaborations in each distance bin as follows:

$$RelativeDensity(d) = \frac{ActualCollaborationDensity(d)}{PotentialDensity(d)}.$$

If $RelativeDensity\,(d)$ is greater than 1, the collaboration is concentrated in distance d. This measure captures the extent of localization of collaboration and controls for the localization of establishments.

3.3. Methodology 2: Duranton and Overman's method

Using the relative density, we can estimate the extent of collaboration localization. However, this method does not lead to a formal statistical test for localizations. We employ Duranton and Overman's (2005) distance-based approach to examine the statistical significance. This method was originally developed to test the localization of the manufacturing establishments' location. We adopt this method to assess the localization of collaborations. Our approach consists of three steps à la Duranton and Overman (2005). First, we calculate all of the bilateral distances between the collaborating establishments. We then estimate the kernel density function of the distance distribution. Second, to statistically test the localization, we construct a counterfactual in which each establishment randomly chooses its collaborating partner. Third, from the counterfactual distance distributions, we construct a confidence interval band and test whether the collaborations are localized.

Kernel Densities

We begin by estimating the density distribution of bilateral distances between collaborating partners. As stated above, there are n_A^p collaboration patents in the patent technology class A. Let $d\,(p_{ij}^A)$ be the great-circle distance between establishments i and j, which apply for patent p_{ij}^A. We then estimate the kernel-smoothed densities (K-densities hereafter) of bilateral distances between the collaborating partners. The K-density estimator at distance d is

$$\widehat{K}(d) = \frac{1}{h\sum_{A\in\mathfrak{A}} n_A^p} \sum_{A\in\mathfrak{A}} \sum_{p_{ij}^A \in P^A} f\left(\frac{d - d\,(p_{ij}^A)}{h}\right)$$

where h is the bandwidth, set as the optimal bandwidth proposed by Silverman (1986), and f is the Gaussian kernel function.

Identifying the Localization of Collaborations Based on Counterfactuals

To statistically test the localization of collaborations, we construct two-sided confidence intervals. More specifically, we randomly choose n_A^p pairs of establishments from the set of potential collaborating partners S^A for every patent technology class $A \in \mathfrak{A}$ and estimate the K-density in the counterfactual situation. By iterating this trial 1,000 times, we construct confidence bands. Following Duranton and Overman (2005), we calculate global confidence bands, that is, the upper confidence band $K^U(d)$ and lower confidence band $K^L(d)$, so that 95% of the 1,000 randomly drawn K-densities are within the confidence bands, over the entire distance range (in our case, 0–180 km).[6]

Thus, we obtain the upper global confidence band $K^U(d)$ and the lower global confidence band $K^L(d)$. If $\widehat{K}(d) > K^U(d)$ for at least one $d \in [0, 180]$, the collaborations can be defined as globally localized at the 5% level.

To discuss the strength of localization, we follow Duranton and Overman (2005), and define the index of localization as follows:

$$\Gamma = \sum_{d\in[0,180]} \max\{\widehat{K}(d) - K^U(d), 0\}.$$

4. Baseline Results

Figure 1 shows the relative density results. Collaborations occur more frequently within short distances.

---

[6] Following Duranton and Overman (2005) and Nakajima, Saito, and Uesugi (2012), we set 180 km as the upper bound of our focus distance.

The $RelativeDensity\ (d)$ has the largest value, 2.8, in the 0–20 km range. This means that the density of actual collaboration is 2.8-fold larger than that of potential collaborations in that range. The value of $RelativeDensity\ (d)$ is then shown to gradually decrease. In the 80–160 km range, the $RelativeDensity\ (d)$ is approximately 1.0. These results mean that collaborations are localized within short distance ranges.

[Figure 1 here]

We next follow Duranton and Overman's (2005) method to statistically test and identify the range of localization. Figure 2 shows the results. The solid line in the figure represents the K-density, and the dashed lines represent the global confidence bands. For every distance within the 0–105 km range, the K-density is above the upper global confidence band. Thus, we consider the collaborations to be statistically localized at the 5% level in the 0–105 km range.

[Figure 2 here]

We presume that localization tendency varies across technology classes. Thus, strong localization tendencies in a small number of technology classes may cause this overall localization. To respond to this argument, we investigate the differences in the localization tendencies between patent technology classes. We conduct the same analysis for each technology classes. In this case, the K-density of patent technology class A in distance d, $\widehat{K_A}(d)$, is as follows:

$$\widehat{K_A}(d) = \frac{1}{hn_A^p} \sum_{p_{ij}^A \in P^A} f\left(\frac{d - d\ (p_{ij}^A)}{h}\right).$$

In this analysis, we construct counterfactual distributions for each technology class and construct an upper global confidence band, $K_A^U(d)$, and a lower global confidence band, $K_A^L(d)$, for patent technology class A.

We can assess localization for each distance d. If $\widehat{K_A}(d) > K_A^U(d)$ at distance d, patent technology class A is localized at distance d. Figure 3 shows the number of technology classes that are localized for each distance. At shorter distances, collaborations are localized in 108 out of the 120 technology classes. At distances of more than 100 km, the number of localized technology classes gradually declines.

[Figure 3 here]

Table 2 shows the frequency distributions of Γ that represent the extent of localization in each technology class. The value of Γ is heterogeneous, while collaborations are localized in most technology classes.

[Table 2 here]

In summary, our overall finding of significantly localized collaborations is not due to a few strongly localized technology classes. As can be observed in Figure 3 and Table 2, although values of Γ are heterogeneous, collaborations are localized in most technology classes; thus, our overall results do not represent the localization tendencies of specific technology classes.

Geographic friction in knowledge spillovers as measured by patent citations declines over time (e.g., Griffith, Lee, and Van Reenen, 2011). Our focus period (1986-2005) was characterized by vast developments in ICT, and a dramatic decrease in communication costs. These developments may also reduce collaboration costs between geographically remote establishments. To investigate time period differences, we calculate the $RelativeDensity\,(d)$ for each five-year period. The results are shown in Figure 4. Essentially, a similar $RelativeDensity\,(d)$ tendency is observed for every five-year period. Collaborations are concentrated in the shortest distance ranges. The shape of $RelativeDensity\,(d)$ does not drastically change over time. Whereas geographic frictions in knowledge spillovers measured by patent citations decline over time, those in collaborations are unchanged. This suggests the importance of geographical proximity in transferring tacit knowledge that requires face-to-face communications.

[Figure 4 here]

5. Firm-Border Effects

We now consider firm-border effects. As reviewed in Pittaway et al. (2004), firm-border effects may significantly impede collaborations because the fear of leakage in inter-firm collaborations has necessitated additional management costs for collaboration.

To capture such firm-border effects, we split collaborations into two groups: intra-firm collaborations, that is, collaborations between establishments belonging to the same firm, and inter-firm collaborations, that is, collaborations between establishments belonging to different firms. We now separately examine the localization of collaborations in the two groups and compare the results.

Note that the geographic location pattern differs between inter- and intra-firm establishments. To control for this difference, we generate separate counterfactuals for inter- and intra- firm collaborations. When we generate counterfactual intra-firm collaborations, we assign establishments within a firm. Similarly, when we generate counterfactual inter-firm collaborations, we assign establishments belonging to different firms. Thus, we can control for location pattern differences between inter- and intra-firm establishments.

First, we present the relative densities. Figure 5 shows the results. The solid line in the figure represents the $RelativeDensity\ (d)$ for intra-firm collaboration and the dashed line for inter-firm collaboration. The figure clearly shows that in the shortest distance range (0–20 km), $RelativeDensity\ (d)$ is substantially larger for inter-firm than intra-firm collaboration. Inter-firm collaborations are substantially more concentrated in the shortest distance ranges than in intra-firm collaborations.

[Figure 5 here]

Figure 6 shows the results of our analysis via Duranton and Overman's (2005) approach. Figure 6(a) shows the K-density and global confidence bands of intra-firm collaborations. The K-density is above the upper global confidence band for every distance within the 0–102 km range. Figure 6(b) shows the K-density and global confidence bands of inter-firm collaborations. The K-density is above the upper global confidence band for every distance within the 0–77 km range.

[Figure 6 here]

With respect to localization strength, the estimated value of Γ is 0.155 and 0.300 for intra-firm and inter-firm collaborations, respectively. This means that inter-firm collaborations are more localized than intra-firm collaborations.

To capture patent-technology differences, we also conduct a technology-by-technology analysis. Figure 7 shows the number of technology classes that are localized in each distance. The dashed line represents intra-firm collaborations and the solid line represents inter-firm collaborations. We observe a clear difference between intra- and inter-firm collaborations. Although approximately 100 of the 120 technology classes are localized in inter-firm collaborations, only approximately 60 technology classes are localized in intra-firm collaborations at shorter distances. Furthermore, the number of localized industries in inter-firm collaborations has declined at shorter distances (approximately 80 km).

[Figure 7 here]

Table 3 shows the frequency distributions of Γ in intra- and inter-firm collaborations. Of the 120 technology classes, only 19 are not localized in inter-firm collaborations (Γ equals zero) compared with 42 in intra-firm collaborations. These suggest that inter-firm collaborations are localized in more technology classes than intra-firm collaborations.

The difference between intra- and inter-firm collaborations implies that the firm border in collaboration can be complemented by geographic proximity. It is difficult to identify the driving force of this mechanism from our data. This mechanism can be discussed in line with Lerner (1995) and Agarwal and Hauswald (2010). They discussed that geographic proximities make it easier to monitor firms' the private information in venture capital investment and credit lending relationships between small firms and banks. A similar mechanism would work in a collaboration context.

6. Firm Size Effect on Inter-firm Collaboration

The firm-border effects may differ according to firm size. To examine firm size effects, we split the sample into multiple-establishment firms, and single-establishment firms. Multiple-establishment firms can be regarded as large firms and single-establishment firms as small firms can be regarded. Note that our multiple- and single- establishment firms are defined according to the number of establishments that publish patents in our database. We cannot capture the establishments that do not publish patents in the analysis period 1986–2005.

Figure 8 shows the relative densities of inter-firm collaborations. The solid line shows the relationships between single-establishment firms, the dashed line between single- and multiple-establishment firms, and the dotted line between multiple-establishment firms.

[Figure 8 here]

With respect to the firm size effect on collaboration, we found that collaborations with single-establishment firms (single–single and single–multiple) have stronger localization tendencies.

As shown in Figure 9, we use the Duranton and Overman (2005) approach to statistically test localization. Each panel in Figure 9 represents the K-density and global confidence bands in each collaboration (single–single, multiple–multiple, and multiple–single). In every panel from (a) to (c), the K-densities are above the upper confidence bands at shorter distances; therefore, the collaborations are thus localized statistically in every case.

[Figure 9 here]

The estimated values of Γ are 0.334, 0.253, and 0.330 for single–single, multiple–multiple, and multiple–single collaborations, respectively. This implies that the estimated value of Γ is higher for collaborations with small firms. This suggests that small firms are more sensitive to collaboration distances.

7. Concluding Remarks

This study investigated the localization of collaborations in knowledge creation. Using data on establishment-level collaborations in patent applications, we arrive at the following results. First, collaborations between establishments are significantly localized at the 5% level with a localization range of approximately 100 km. Second, the extent of collaboration localization was stable during the two decades studied (1986–2005) despite extensive ICT developments that facilitate communication between remote researchers. Third, the extent of localization is substantially larger in inter-firm than in intra-firm collaborations. Finally, within inter-firm collaborations, the extent of localization is larger in collaborations with small firms.

The finding that collaboration localization was stable during the two decades provides a new perspective on geographic friction differences between collaborations and knowledge spillovers. Griffith, Lee, and Van Reenen (2011) found that frictions in knowledge spillovers observed in patent citations decline over time. This difference between collaborations and patent citations may indicate that ICT may not fully compensate for geographic frictions and that collaboration would require more face-to-face communication than patent citations.

Moreover, this study quantitatively finds the firm-border effects in collaborations, which have long been a subject of debate (e.g., Häusler, 1994; Pittaway et al., 2004). The greater localization of inter-firm collaborations compared with intra-firm collaborations suggests that collaboration between different firms would require additional management costs in light of the fear of unwanted knowledge transfer and organizational secrets in inter-firm collaborations. This can mean that geographic proximity fosters trust between firms and reduces monitoring costs, thereby reducing firm-border effects and facilitating collaboration.

Localized knowledge spillovers have been the theoretical background for cluster policy. On the other hand, it is pointed out that distance does not matter anymore because of the vast development of ICT. This paper’s findings suggest that geographically localized knowledge spillovers between establishments are still crucial for innovation. Thus, promoting the location of businesses within a proximate distance through an industrial cluster policy could facilitate research collaborations particularly between firms having different types of knowledge, and induce innovation more effectively.

Appendix A. A more conservative definition of potential partners

The key idea of our analysis is to control for potentially collaborative establishment locations. Therefore, our results may depend on the definition of potential collaborators. In this appendix, we assess the robustness of our results by adopting a more conservative definition of potential partners.

In our main analysis, we define a potential collaborating partner as an establishment that has applied for at least one patent in technology class A. It does not matter that the establishment has an experience of collaborative work. However, our dataset includes many establishments that have never applied for collaborative patents with other establishments. Such establishments may not be potential collaborators. To control for this, we now adopt another definition of potential collaborators. We restrict the establishments under analysis to those that have applied for at least one collaboration patent in technology class $A \in \mathfrak{A}$. Below, we describe the results using the definition of potential collaborating partners and counterfactual collaborations.

Figure A1 shows the baseline results of relative density. In the bins from 0 to 80 km, the relative densities exceed 1, which indicates collaboration localization. Figure A2 shows the results obtained using the Duranton and Overman (2005) method. Similar to Figure 2, collaborative relationships are statistically localized with a localization range of approximately 90 km. Our baseline results remain unchanged despite a more conservative definition of potential collaborators.

[Figures A1 and A2 here]

Figure A3 shows the relative density in each period. The shape of the graph does not change drastically in any period.

[Figure A3 here]

We next show the results of firm-border effects. Figure A4 shows the relative densities of intra-firm and inter-firm collaborations. Similar to the baseline results shown in Figure 5, in the shortest-distance bin, inter-firm collaborations show a stronger relative density compared with intra-firm collaborations.

[Figure A4 here]

Panels (a) and (b) of Figure A5 show the Duranton and Overman (2005) analysis results. Similar to the baseline results shown in panels (a) and (b) of Figure 6, the localization of both collaborations are statistically significant, and the range of localization is shorter in inter-firm collaborations.

[Figure A5 here]

Finally, we check the results for firm size. Figure A6 shows the relative density results, and Figure A7 shows the Duranton and Overman (2005) analysis results. Similar to the baseline results shown in Figures 8 and 9, in every collaboration pattern (single–single, multiple–multiple, and single–multiple), stronger collaboration localizations are found for small firms, and all collaboration localizations are statistically significant at the 5% level. Furthermore, the range of localization is shorter for small firms.

[Figures A6 and A7 here]

To summarize, our baseline results remain completely unchanged despite adopting a more conservative definition of potential collaborators than that originally adopted. The localization of establishments collaborating for innovative activities is statistically significant, and we observe firm-border effects for collaborations, especially regarding small firms.

References

[1] Agarwal, S. and R. Hauswald (2010) "Distance and Private Information in Lending," *Review of Financial Studies* 23(7), pp. 2757-2788.

[2] Carlino, G., R. Hunt, J. Carr, and T. Smith (2012) "The Agglomeration of R&D Labs," Working Paper Series 12-22, Federal Reserve Bank of Philadelphia.

[3] Duranton, G. and H. Overman (2005) "Testing for Localization Using Micro-geographic Data," *Review of Economic Studies* 72(4), pp. 1077-1106.

[4] Ellison, G., E. Glaeser, and W. Kerr (2010) "What Causes Industry Agglomeration?" *American Economic Review* 100(3), pp. 889-927.

[5] Griffith, R., S. Lee, and J. Van Reenen (2011) "Is Distance Dying at Last? Falling Home Bias in Fixed-Effects Models of Patent Citations," *Quantitative Economics* 2(2), pp. 211-249.

[6] Hagedoorn, J. (2003) "Sharing Intellectual Property Rights—An Exploratory Study of Joint Patenting Amongst Companies," *Industrial and Corporate Change* 12(5), pp. 1035-1050.

[7] Hansen, T. (1999) "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge Across Organization Subunits," *Administrative Science Quarterly* 44(1), pp. 82-111.

[8] Hansen, T., L. Mors, and B. Løvås (2005) "Knowledge Sharing in Organizations: Multiple Networks, Multiple Phases," *Academy of Management Journal* 48(5), pp. 776-793.

[9] Häusler, J., H.W. Horn, and S. Lütz (1994) "Contingencies of Innovative Networks: A Case Study of

Successful Interfirm R&D Collaboration," *Research Policy* 23(1), pp. 47-66

[10] Inoue, H., K. Nakajima, and Y.U. Saito (2014) "Localization of Knowledge Creative Establishments," RIETI Discussion Paper Series, 14-E-53.

[11] Jaffe, A., M. Trajtenberg, and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108(3), pp. 577-598.

[12] Kerr, W. and S. Kominers (2015) "Agglomerative Forces and Cluster Shapes," *Review of Economic Studies*, forthcoming.

[13] Lerner, J. (1995) "Venture Capitalists and the Oversight of Private Firms," *Journal of Finance* 50(1), pp. 301-318.

[14] Marshall, A. (1920) *Principles of Economics*, Macmillan, London.

[15] Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014) "Localized Knowledge Spillovers and Patent Citations: A Distance-based Approach," *Review of Economics and Statistics* 96(5), pp. 967-985.

[16] Nakajima, K., Y.U. Saito, and I. Uesugi (2012) "Measuring Economic Localization: Evidence from Japanese Firm-level Data," *Journal of the Japanese and International Economies* 26(2), pp. 201-220.

[17] Nonaka, I. and H. Takeuchi (1995) *The Knowledge Creating Company*, Oxford University Press, New York.

[18] OECD (2008), *Compendium of Patent Statistics*, OECD, Paris.

[19] Pittaway, L., M. Robertson, K. Munir, D. Denyer, and A. Neely (2004) "Networking and Innovation: A Systematic Review of the Evidence," *International Journal of Management Reviews* 5/6(3&4), pp. 137-168.

[20] Polanyi, M. (1958) *Personal Knowledge*, Routledge & Kegan Paul, London.

[21] Polanyi, M. (1966) *The Tacit Dimension*, The University of Chicago Press, London.

[22] Rosenthal, S. and W. Strange (2001) "The Determinants of Agglomeration," *Journal of Urban Economics* 50(2), pp. 191-229.

[23] Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

[24] Wuchty, S., B. Jones, and B. Uzzi (2007) "The Increasing Dominance of Teams in Production of Knowledge," *Science* 5827(316), pp. 1036-1039.

Table 1: Data summary

Number of patents	1,967,361
Number of patents applied for by establishments	1,189,262
Number of patents applied for by collaborating establishments	93,939
Number of patents (Intra-firm collaborations)	37,865
Number of patents (Inter-firm collaborations)	56,074
Number of establishments	74,452
Number of firms	56,592
Number of single-establishment firms	46,904
Number of multi-establishment firms	9,688

Table 2: Frequency distributions of Γ

Range of Γ	Number of technology classes
0	12
0–0.1	21
0.1–0.2	48
0.2–0.3	30
0.3–0.4	8
0.4–0.5	1

Table 3: Frequency distributions of Γ, intra- and inter-firm collaborations

Range of Γ	Number of technology classes	
	Intra-firm	Inter-firm
0	42	19
0–0.1	45	32
0.1–0.2	23	39
0.2–0,3	7	19
0.3–0.4	2	9
0.4–0.5	0	2
0.5–0.6	0	0
0.6–0.7	1	0

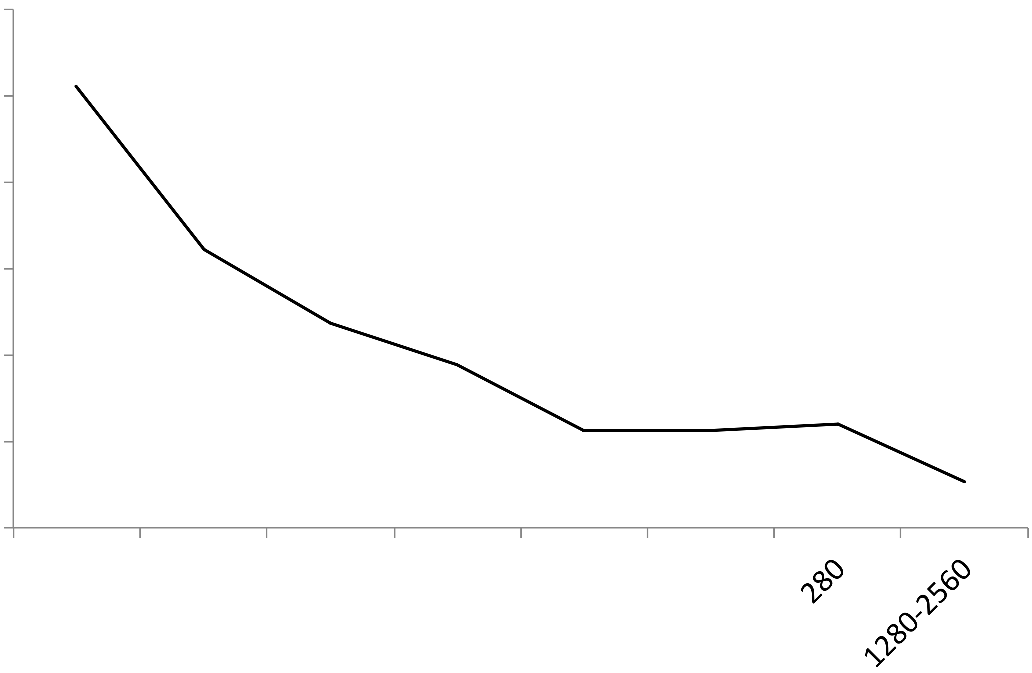


Figure 1: Result of relative density

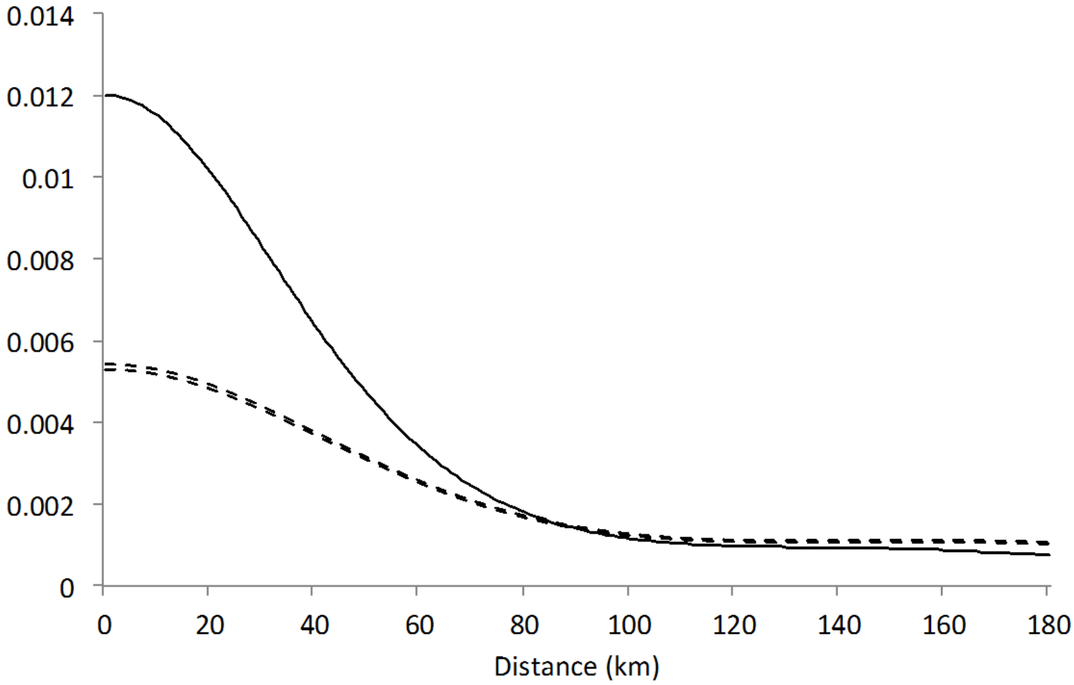


Figure 2: K-density of collaborations

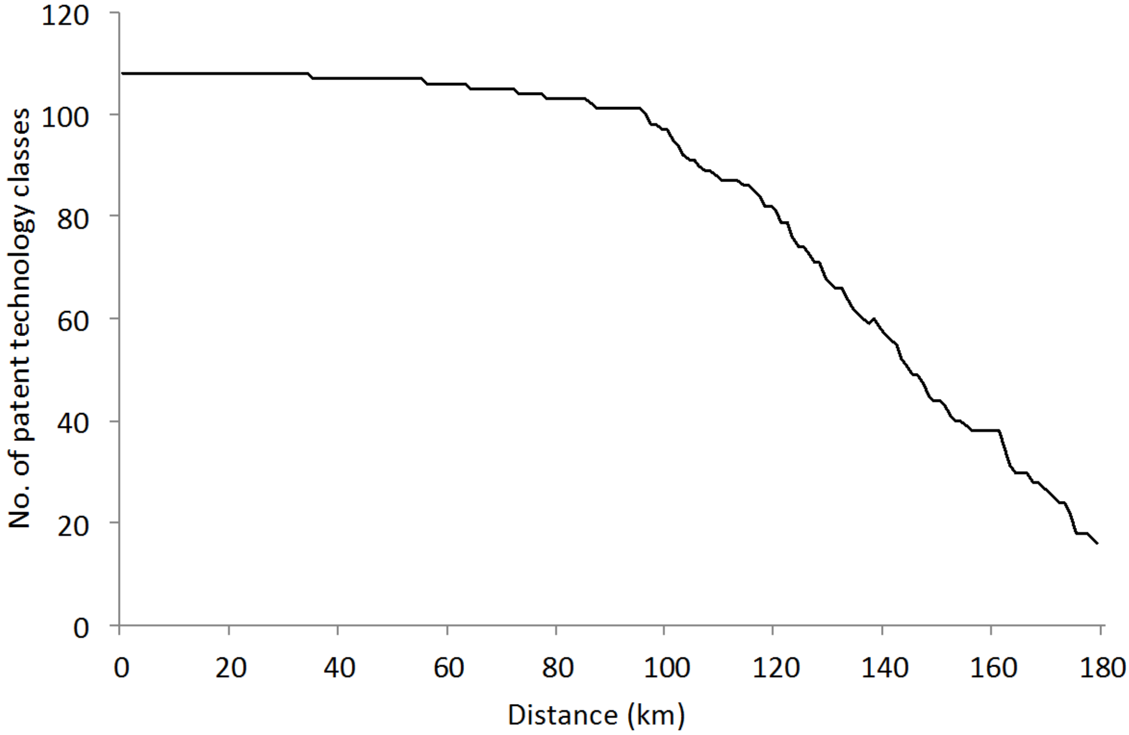


Figure 3: Number of localized technology classes for each distance

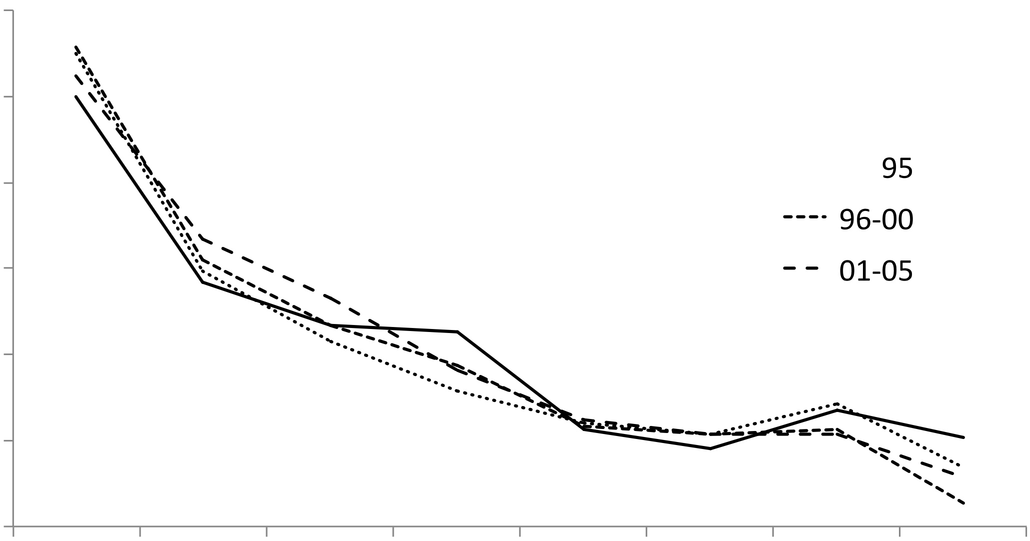


Figure 4: Relative densities for each five-year period

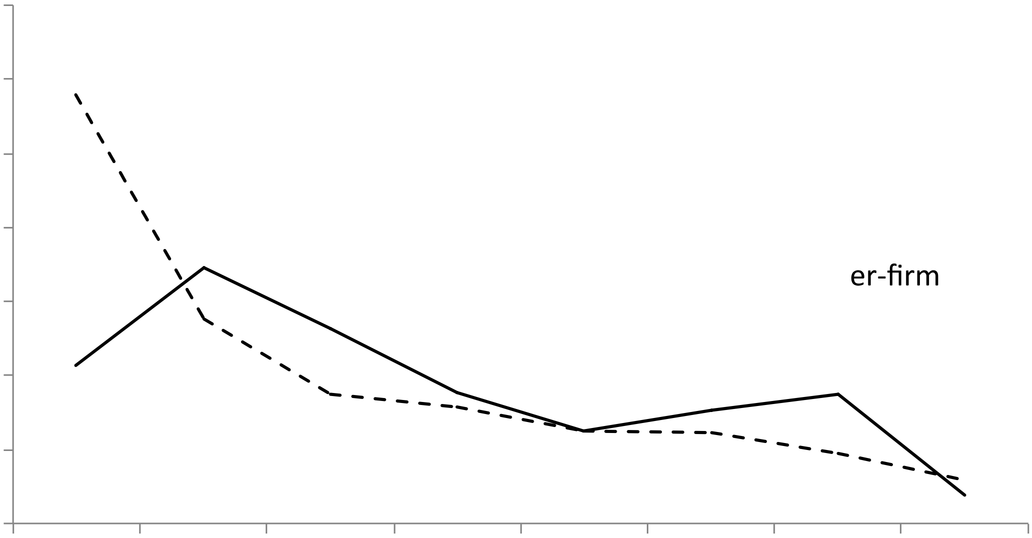
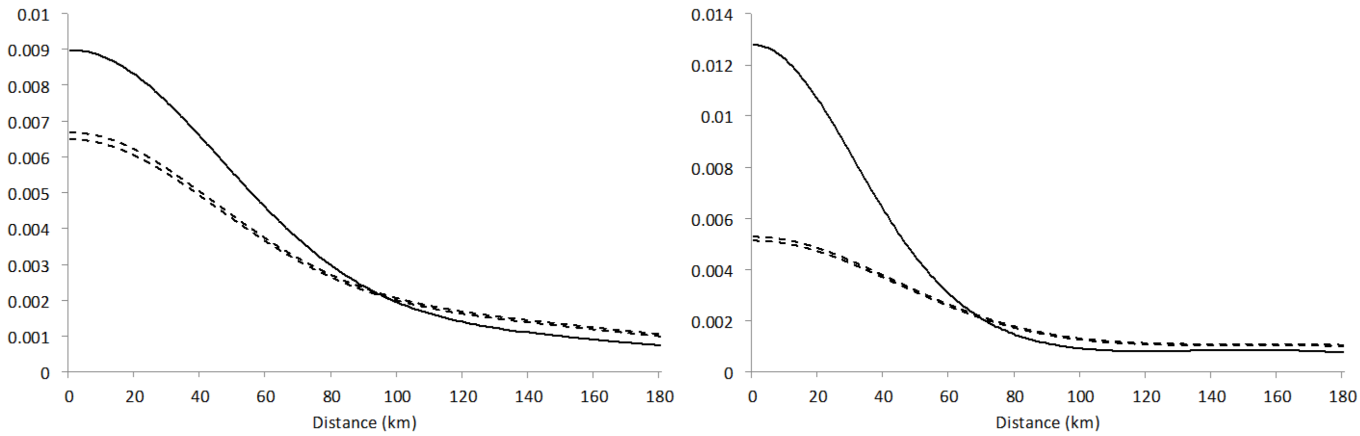


Figure 5: Relative densities of intra-firm and inter-firm collaborations



(a) Intra-firm collaborations

(b) Inter-firm collaborations

Figure 6: K-densities of collaborations

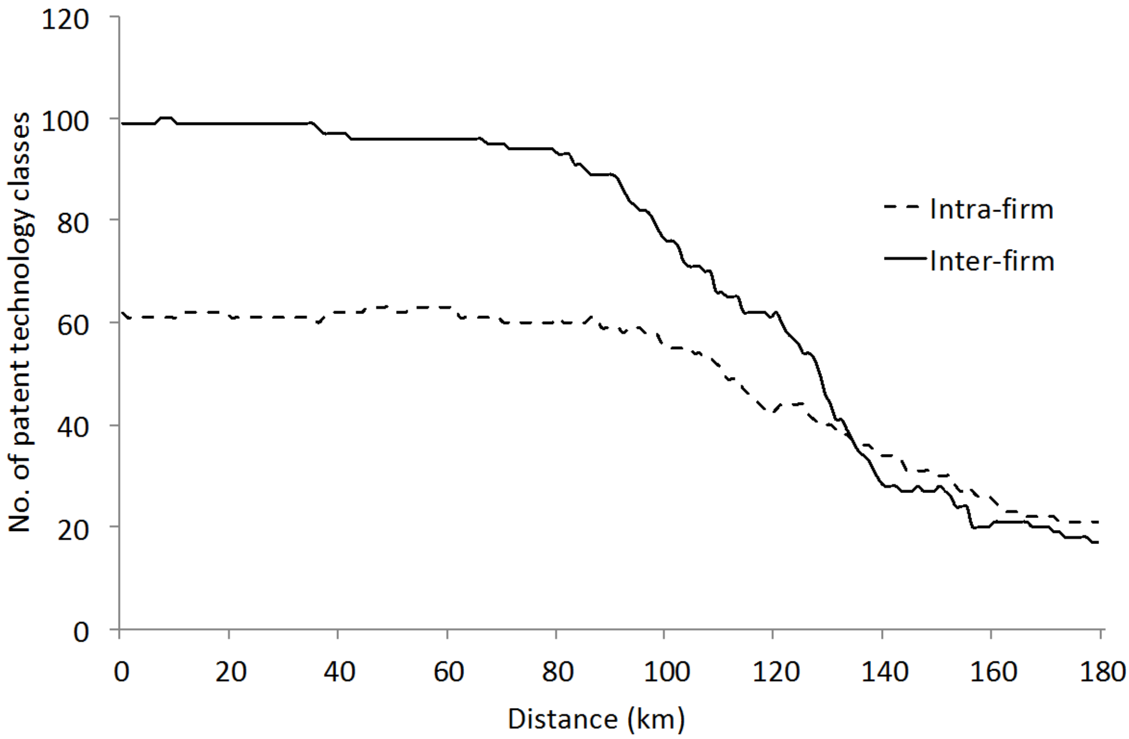


Figure 7: Number of localized technology classes in each distance (intra- versus inter-firm)

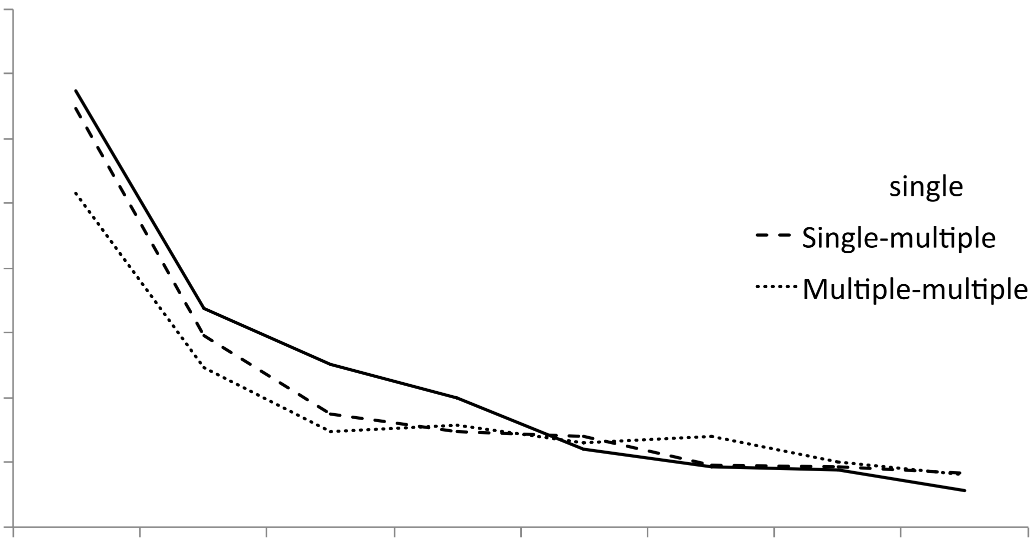
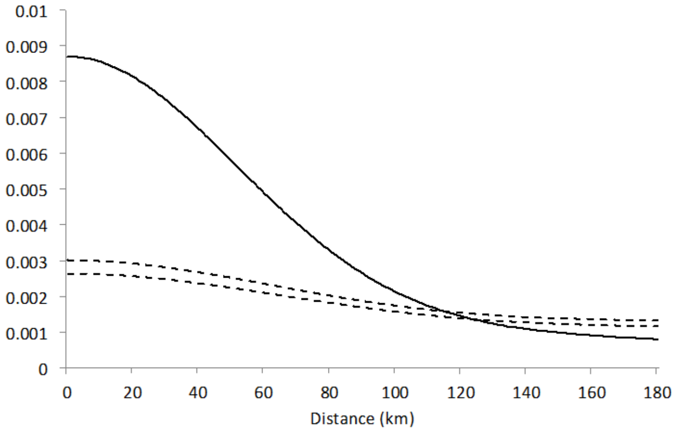
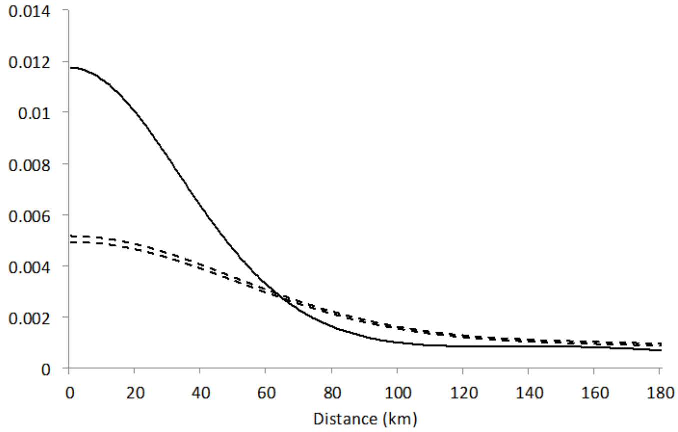


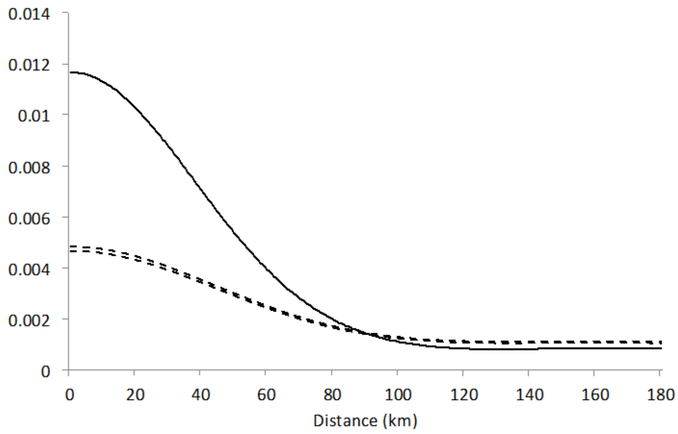
Figure 8: Relative densities of inter-firm collaborations



(a) Single–single collaborations



(b) Multiple–multiple collaborations



(c) Multiple–single collaborations

Figure 9: K-densities of inter-firm collaborations

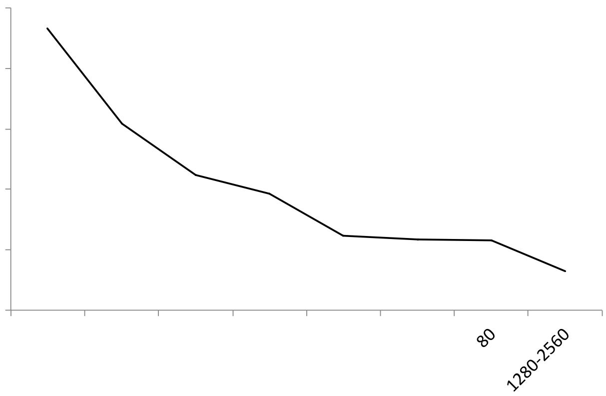


Figure A1: Relative density

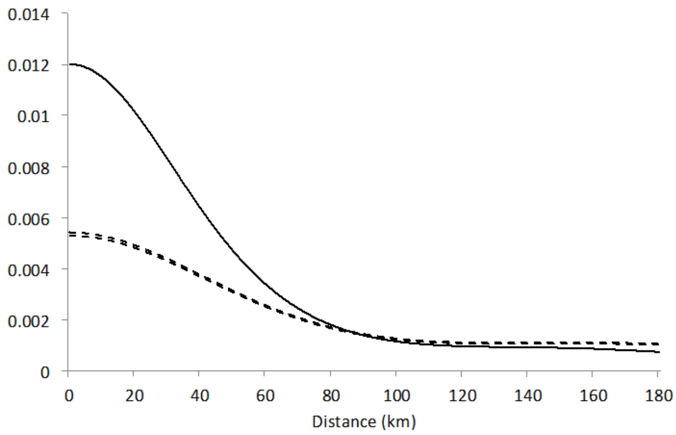


Figure A2: Duranton and Overman's results

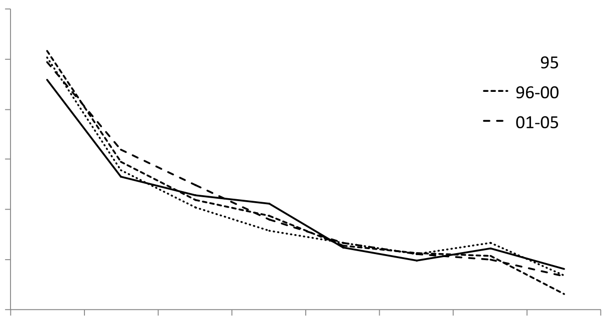


Figure A3: Relative densities for each five-year period

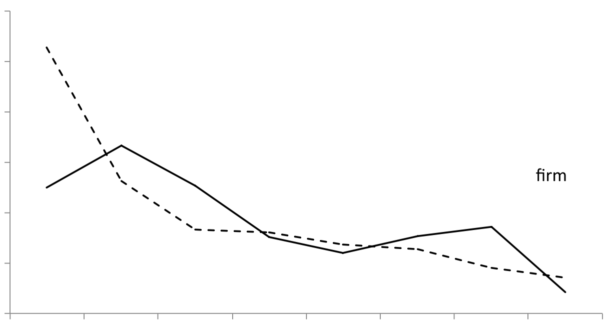
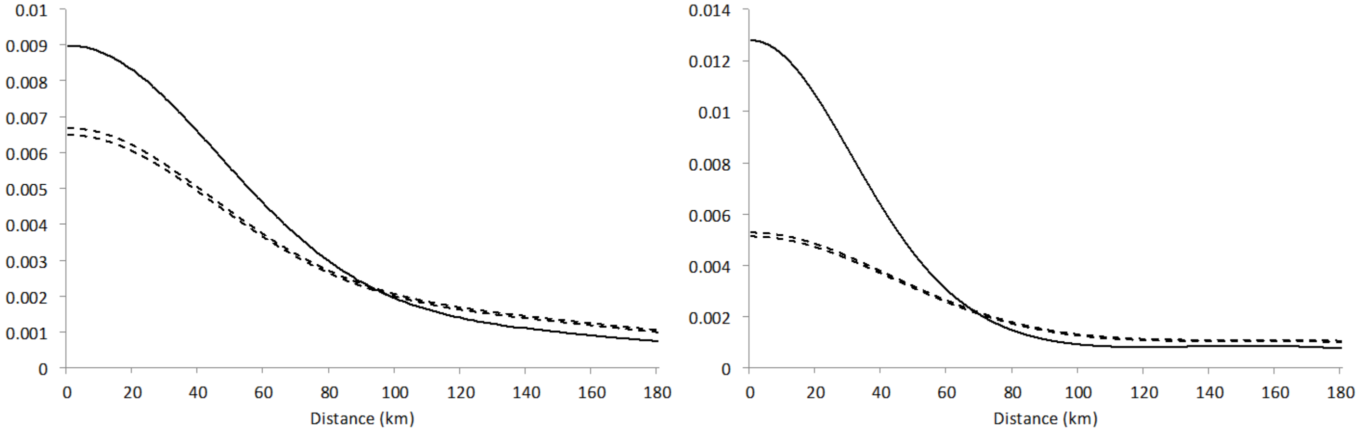


Figure A4: Relative densities of intra-firm and inter-firm collaborations



(a) Intra-firm collaborations (b) Inter-firm collaborations

Figure A5: K-densities of collaborating relationships

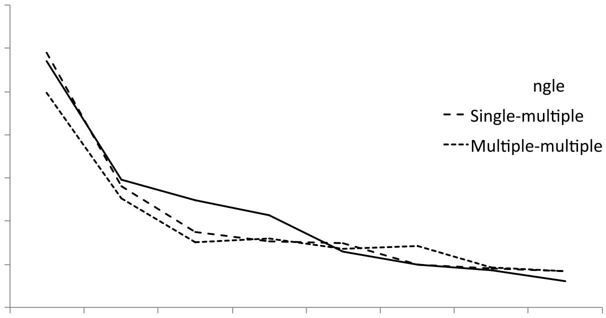
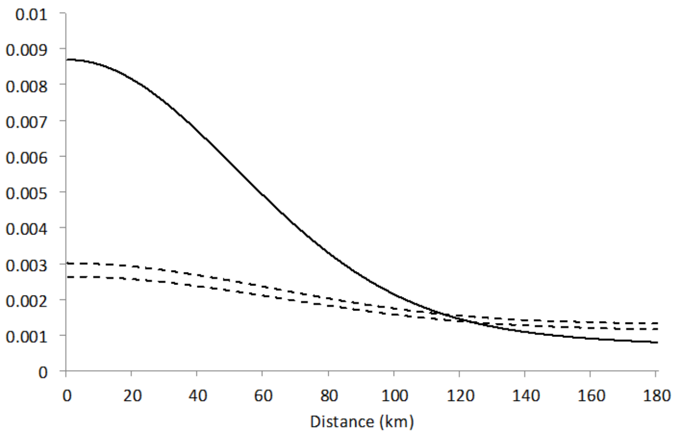
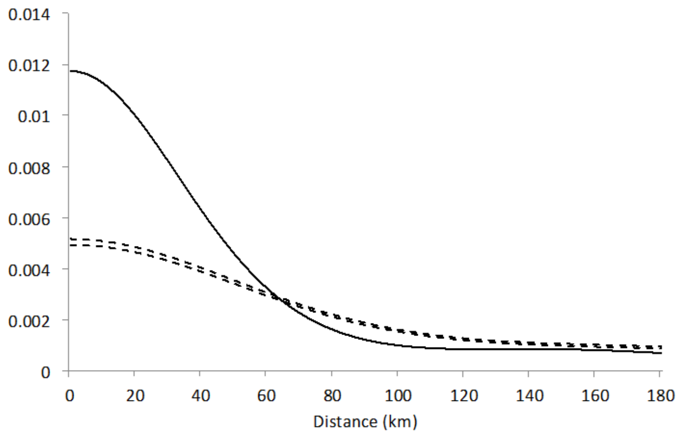


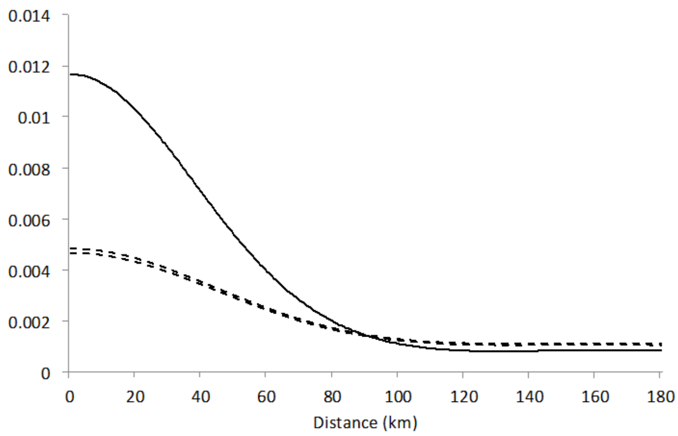
Figure A6: Relative densities of inter-firm collaborations



(a) Single–single collaborations



(b) Multiple–multiple collaborations



(c) Single–multiple collaborations

Figure A7: K-densities of inter-firm collaborations