Localization of Collaborations in Knowledge Creation

Hiroyasu INOUE
Kentaro NAKAJIMA
and
Yukiko Umeno SAITO

March, 2016

Localization of Collaborations in Knowledge Creation[*]

Hiroyasu INOUE[†], Kentaro NAKAJIMA[‡], and Yukiko Umeno SAITO[§]

March 2016

Abstract

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

Key words: Knowledge creation; Collaboration; Geographic frictions; Firm-border effects
JEL codes: R12; O31

[†]Graduate School of Simulation Studies, University of Hyogo, Computational Science Center Building, 7-1-28 Minatojima-minamimachi, Chuo-Ku Kobe 650-0047, Japan
[‡]Graduate School of Economics and Management, 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 as one of the determinants of agglomerations of economic activities (Marshall, 1920; Rosenthal and Strange, 2001; Ellison et al., 2010). Several studies have examined the localization of knowledge spillovers by using patent citation as a proxy for knowledge spillovers (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Murata et al., 2014, Kerr and Kominers, 2015).

This paper investigates the localization of knowledge spillovers by using the collaborative works in knowledge creation instead of patent citation. Workers in a collaborative work intensively exchange ideas and information. Thus, collaborative work can be considered an alternative proxy for knowledge spillover. Furthermore, collaborative works basically require face-to-face communications that demand researchers' actual presence. Thus, geographical distance will have a substantially important role to play in the knowledge spillovers that occur through collaboration.

The localization of collaborative works has been investigated by regionally aggregated data and a gravity equation approach (e.g., Ponds et al., 2007; Hoekman et al., 2009). However, several studies mention that knowledge spillovers significantly occur in quite short ranges (e.g., Rosenthal and Strange, 2003; Arzaghi and Henderson, 2008; Kerr and Kominers, 2015). Regionally aggregated data cannot capture these rather localized interactions in a region. To address this issue, this paper employs the micro-geographic information of collaborators without any regional aggregation.

Particularly, we use establishment-level collaboration information from Japanese patent data to precisely identify the place where the R&D activity actually occurred. The analyses of knowledge spillover by using disaggregated patent data have been mainly investigated in an inventor-level analysis (e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Murata et al., 2014; Kerr and Kominars, 2015). However, inventor-level analysis has a limitation when registered inventors' addresses are their residential addresses, not their working place addresses. When an inventor's residential address is used, the address does not represent the place where the invention and interaction actually occurred. As an alternative approach, firm-level analysis is also available by using firm headquarter address information that is designated on each patent, which appears as the applicant's address. However, this information is unsuitable for the investigation of the geographic features of collaboration. Because several firms can have multiple establishments, the address information of firms does not necessarily represent the actual location of the

---

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

invention. This paper addresses this problem by constructing an establishment-level dataset on collaborative works by using a convention of Japanese patents: the address that inventors register is the address of their establishment instead of the address of their residence.

Notably, establishment-level analysis enables us to investigate the 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 et al., 1994; Pittaway et al., 2004). By comparing inter- with intra-firm collaborations, we can quantitatively evaluate the 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. Because knowledge-creating establishments are concentrated in certain areas (e.g., Carlino et al., 2012; Inoue et al., 2014), a naive measurement of the geographic proximity between collaborating establishments inevitably reflects the proximity of the establishments.

To control for the geographic proximity of the establishments, we consider the counterfactual collaborations in which establishments choose collaborating partners regardless of distance. These counterfactual collaborations reflect the geographic proximity of the establishments. Then, we test the localization of collaboration by a comparison of actual with counterfactual collaborations. This concept is the same as the control patent that is often used in the literature on the localization of patent citations (e.g., Jaffe et al., 1993; Thompson and Fox-Kean, 2005). Specifically, we apply Duranton and Overman's (2005) K-density approach that is similar to Murata et al. (2014) to avoid the bias that comes from regional aggregation.

We obtain the following results. First, the collaborations between establishments are significantly localized at the 5% level with a localization range of approximately 100 km. Second, the localization of collaboration is observed in most technologies. Third, 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 stability suggests that geographic distance continues to play an important role in knowledge spillovers that require face-to-face communication. Fourth, the extent of localization is substantially larger in inter-firm collaborations than in intra-firm collaborations. These results show that geographic frictions in collaborations are different in inter-firm and intra-firm collaborations and imply 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. Overall, our results suggest that inter-establishment collaborations are

localized and stable and that geographic proximity mitigates the 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 the firm-border effects in the collaborations between establishments. Section 6 investigates the firm size effect on inter-firm collaborations. Finally, Section 7 concludes the paper.

2. Data

We constructed an establishment-level collaboration dataset that is based on the Institute of Intellectual Property (IIP) patent database (DB) (Goto and Motohashi, 2007). From this database, we use all the patents that were applied for from 1986 to 2005. We use the application date instead of the publication date. The publication date differs from the application date because of the reviewing time. To appropriately capture the timing when R&D activities actually occurred, we use the application date because it is closer to the timing when the research was actually conducted.

In this study, we focus on the collaborations between the establishments to which the inventors belong. Therefore, we identify the establishment-level collaboration information from each patent by using both applicant and inventor information. Similar to other countries, in Japan, both the firm and the individual can be registered as an applicant, whereas an inventor is limited only to individuals. Thus, if an invention is conducted in an establishment of a firm, the firm name and address is registered as the applicant, and the names and addresses of the inventors who actually performed the R&D activities are registered as the inventors. Basically, the firm registers its headquarter address as the applicant's address. Thus, the applicant's address does not necessarily represent the place where the R&D activities actually occurred when the firm has multiple establishments. Similarly, an inventor's address does not represent the place where the R&D activities were conducted when an inventor registers his or her residential address.

We address this problem by utilizing a convention in Japanese patent applications that the addresses of the establishments to which inventors belong instead of their residential addresses are registered as the address of the inventors. Accordingly, the inventor address information in the registered patent data provides information concerning the establishments where the inventions actually occurred. Further, the registered address normally includes the firm name; thus, we can identify the firm to which the

3

establishments belong. Additional details concerning the extraction of the establishments are shown in Appendix A.[2]

After identifying the establishments, we extract the collaboration between establishments as follows. If a patent includes inventors who belong to different establishments, we consider that there is a collaborative relation between these establishments.[3] In this paper, we focus on the bilateral collaboration relation. Thus, if a patent includes more than two establishments, we count all the bilateral relations. For example, in the case that a patent includes three establishments, A, B, and C, we identify three bilateral collaboration relations between A and B, B and C, and A and C.

Table 1 provides the summary of the dataset. Out of 1,967,361 patents, we can identify the establishments for 1,189,262 patents.[4] These patents include 56,592 firms and 74,452 establishments according to our procedures.

[Table 1 here]

Out of the 1,189,262 patents that were applied for by establishments, 7.9% of the patents (93,939 patents) involve establishment-level collaborations. In all, 59.7% of the patents (56,074 patents) out of the collaborative patents involve collaborations between different firms (i.e., inter-firm collaborations).

---

[2] For a brief survey of the Japanese patent system, there are annual reports from the Japan Patent Office (Japan Patent Office 2014). In addition, Maskus and McDaniel (1998) is a helpful overview of the Japanese patent system.

[3] One may concern the transfer of workers. Consider the case that an inventor transfers from an establishment during the process of collaboration before the collaborating patent is applied for, and she registers her new establishment in the patent application form. In this case, the observed distance between the two establishments that are registered in the patent does not necessarily represent the actual collaborating distance. To check this issue, we estimate from the previous literature the probability of the uncertain collaborations by using the frequency of the inventor's transfer and average duration of invention. Saito and Yamauchi (2015) found that 22.4 % of researchers transfer their establishments in 16 years. Suzuki (2011) found that it takes a median of 18 months from starting a project to applying for a patent in Japan. Thus, we roughly estimate that 2.3% (=1-(1-0.224)$^{18/12/16}$) of the workers transferred establishments within 18 months, and the magnitude of the effect is not very large.

[4] Our procedure identifies that 60.4% of all patents are applied for by firms. It is difficult to check the validity of our procedure because there is no data on the share of patents that are applied for by firms in Japan. Alternatively, we compare the share of patents that are applied for by firms in OECD countries, which is approximately 80% (OECD, 2008). If we assume that the proportion of the patent applications that are issued by firms in Japan is similar to the proportion of the patent applications that are issued by firms in OECD countries, our procedure successfully matches three-fourths of the total patents that are applied for by firms.

Furthermore, firms can be 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 the establishments that published no patents during the focus period when we counted the number of establishments for each firm.

3. Empirical Strategy

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

To control for this geographic proximity of the establishments, we consider counterfactual collaborations in which establishments choose partners regardless of distance. These counterfactual collaborations reflect the proximate locations of the establishments. Therefore, a comparison between the actual and counterfactual collaborations that reflect the geographic proximity of the establishments enables an appropriate assessment of the proximity of the collaborations. Specifically, we use Duranton and Overman's (2005) K-density approach.

Here, we describe in detail how to define 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 the International Patent Classification (IPC) to categorize the patents.[5] There are 120 patent technology classes in our data. Let P_A be the set of patent applications that is submitted by the collaborating establishments, let $p_{ij}^A \in P_A$ be a patent application that is submitted by collaborating establishments i and j in technology class A, and let n_A^p be the number of patent applications that is 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. Elements from the set of potential collaborating partners S^A can collaborate with one another. By using the definition of potential collaborating partners, we can generate counterfactual collaborations. For these counterfactual collaborations, we randomly choose n_A^p pairs of establishments from the set of potential collaborating

---

[5] A patent often has multiple IPCs. We use the primary IPC that is assigned to each patent.

partners S^A.[6]

These counterfactual collaborations represent a situation in which establishments randomly choose their partners regardless of the distance from potential collaborating partners. Thus, these counterfactual collaborations capture the geographic proximity of the establishments in each technology class.[7]

Then, 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 the collaborations. Our approach consists of three steps according to 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 from potential collaborating partners. 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 the bilateral distances between collaborating partners. There are n_A^p collaboration patents in the patent technology class A. Let $d\left(p_{ij}^A\right)$ 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) of the bilateral distances between the collaborating partners. The K-density estimator at distance d is

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

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

---

[6] For a robustness check, we conduct analyses by using a more conservative definition for potential collaborating partners that restricts the establishments that have at least one experience of collaboration. However, the results remained qualitatively unchanged. See Appendix B.

[7] One may argue that establishments can choose their locations by considering the expectations of future collaborations, and if this is true, these establishments choose their locations where many potential collaborators are located. In this case, the counterfactual collaborations will be more localized than the collaborations of the establishments that choose their location regardless of the existence of future collaborators, and the K-densities of the counterfactual collaborations become similar to the K-densities of the actual collaborations. Our approach identifies the localization of collaborations by the gap of the distribution between the actual and counterfactual collaborations, and this endogeneity may underestimate the localization of collaborations.

To statistically test the localization of collaborations, we construct two-sided confidence intervals. Specifically, we randomly choose n_A^p pairs of establishments from the set of potential collaborating partners S^A for every patent technology class $A \in \mathfrak{U}$ 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 the 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).[8]

By using the upper global confidence band $K^U(d)$ and the lower global confidence band $K^L(d)$, Duranton and Overman (2005) proposed an index of localization in distance *d* as follows:

$$\gamma_A(d) \equiv \max\{\widehat{K}(d) - K^U(d), 0\}.$$

If $\gamma_A(d) > 0$ for at least one $d \in [0, 180]$, the collaborations can be defined as globally localized at the 5% level. As Duranton and Overman (2005) mentioned, the index $\gamma_A(d)$ is used to reject the hypothesis that the actual density is the same as the counterfactual density. Furthermore, Duranton and Overman (2005) also discussed that $\gamma_A(d)$ represents the extent of localization in distance d. Thus, by focusing on $\gamma_A(d)$, we can measure the tendency of localization in each distance.

To discuss the overall extent of localization, Duranton and Overman (2005) defined the index of localization as the sum of $\gamma_A(d)$ in every focusing distance:

$$\Gamma = \sum_{d \in [0,180]} \gamma_A(d).$$

4. Baseline Results

Figure 1 shows the baseline result. The solid line in the figure represents the K-density, and the dashed lines represent the global confidence bands. For every distance in the 0–100 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 0–100 km range.

[Figure 1 here]

---

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

The choice of the upper bound does not qualitatively change our results.

We presume that the localization tendency varies across technology classes. Thus, strong localization tendencies in a small number of technology classes may cause this overall localization. To address this issue, we investigate the differences in the localization tendencies between patent technology classes. We conduct the same analysis for each technology class. 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 the 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 2 shows the number of technology classes that are localized for each distance. At shorter distances, collaborations are localized in 108 of the 120 technology classes. At distances of more than 100 km, the number of localized technology classes gradually declines.

[Figure 2 here]

Table 2 shows the frequency distributions of Γ that represent the extent of localization in each technology class., Collaborations are localized in most technology classes, while the value of Γ is heterogeneous. This finding implies that the overall result is not because of several strongly localized technology classes.

[Table 2 here]

Next, we focus on the dynamics of geographic friction in collaborations. 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 periodic changes of the localization tendency, we conduct the analysis for each five-year period. Potential collaborators are defined for each period to control for the change in location pattern.

The results are shown in Figure 3. Panel (a), (b), (c), and (d) shows the results of 1986-1990, 1991-1995, 1996-2000, and 2001-2005, respectively. Basically, the shapes of the K-densities are mostly stable. Furthermore, the estimated values of Γ are 0.205, 0.216, 0.243, and 0.249 for 1985-1990, 1991-1995 1996-2000,

and 2001-2005, respectively. The extent of localization has an increasing rather than a decreasing tendency through the periods. This tendency implies that the geographic frictions in collaborations are unchanged or increasing over time, whereas geographic frictions in patent citations decline (e.g., Griffith, Lee, and Van Reenen, 2011). This contrasting result concerning collaboration suggests the importance of geographical proximity in transferring tacit knowledge that requires face-to-face communications.

[Figure 3 here]

5. Firm-Border Effects

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

To capture these firm-border effects, we divide collaborations into two groups, namely, intra-firm collaborations and inter-firm collaborations. Intra-firm collaborations are collaborations between establishments that belong to the same firm, and inter-firm collaborations are collaborations between establishments that belong to different firms. We now separately examine the localization of the collaborations in the two groups and compare the results.

The geographic location pattern differs between inter- and intra-firm establishments. To control for this difference, we generate counterfactuals for inter- and intra-firm collaborations independently. When we generate counterfactual intra-firm collaborations, we determine the establishments in a firm to be potential collaborators. Similarly, when we generate counterfactual inter-firm collaborations, we determine the establishments that belong to different firms to be potential collaborators. Thus, we can control for location pattern differences between inter- and intra-firm establishments.

Figure 4 shows the results. Figure 4(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 in the 0–100 km range. Figure 4(b) shows the K-density and global confidence bands of the inter-firm collaborations. The K-density is above the upper global confidence band for every distance in the 0–77 km range.

[Figure 4 here]

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

intra-firm collaborations.

We also conduct a technology-by-technology analysis. Figure 5 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. While 99 of the 120 technology classes are localized in inter-firm collaborations, only approximately 62 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 5 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 findings suggest that inter-firm collaborations are localized in more technology classes than intra-firm collaborations.

These results reveal the difference in the geographic frictions in inter-firm and intra-firm collaborations, and imply that the firm border in collaborations can be complemented by geographic proximity. 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 et al., 1994; Pittaway et al., 2004). From our data, it is difficult to identify how geographic proximity mitigates these impediments to inter-firm collaborations. This mechanism, however, can be discussed according to Lerner (1995) and Agarwal and Hauswald (2010) who indicated that geographic proximities make it easier to monitor firms' private information in venture capital investment and credit lending relations between small firms and banks. A similar mechanism would work in a collaborative context.[9]

6. Firm Size Effect on Inter-firm Collaboration

The firm-border effects may differ according to firm size. To examine the firm size effects, we divide the

---

[9] To confirm the role of geographical proximity, we conducted interviews to firms that have experiences of collaborative works on March 2016 in a provincial city in Japan. Managers of firms mention that geographical proximity promotes trust between firms through monitoring and spillovers of each firm's information, and facilitates the collaboration between firms.

sample into multiple-establishment firms and single-establishment firms. Multiple-establishment firms can be regarded as large firms, and single-establishment firms can be regarded as small firms. Our multiple- and single-establishment firms are defined according to the number of establishments that have applied for patents in our database. Thus, we cannot capture the establishments that have not applied for patents in the analysis period from 1986–2005.

Figure 6 shows the results. Each panel in Figure 6 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 statistically localized in every case.

[Figure 6 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 finding implies that the estimated value of Γ is higher for the collaborations with small firms, which suggests that small firms are more sensitive to collaboration distances.

7. Concluding Remarks

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

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

Moreover, this study quantitatively finds the firm-border effects in collaborations, which has long been a

subject of debate (e.g., Häusler et al., 1994; Pittaway et al., 2004). The greater localization of inter-firm collaborations compared with intra-firm collaborations suggests that the collaboration between different firms requires additional management costs considering the fear of unwanted knowledge transfer and organizational secrets in inter-firm collaborations. This indication can mean that geographic proximity encourages trust between firms and reduces monitoring costs, which reduces the firm-border effects and facilitates collaboration.

Localized knowledge spillovers have been the theoretical background for cluster policy. However, distance may no longer matter 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 can facilitate research collaborations particularly between firms that may have different types of knowledge and can more effectively induce innovation.

References

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

Arzaghi, M. and V. Henderson (2008) "Networking off Madison Avenue," *Review of Economic Studies* 75(4), pp. 1011-1038.

Boschma, R. (2005) "Proximity and Innovation: A Critical Assessment," *Regional Studies* 39(1), pp. 61-74.

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.

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

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

Goto, A. and K. Motohashi (2007). "Construction of a Japanese patent database and a first look at Japanese patenting activities," *Research Policy*, 36(9), pp. 1431–1442.

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.

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

Hoekman, J., K. Frenken, and F. van Oort (2009) "The geography of collaborative knowledge production in Europe," *Annals of Regional Science* 43, pp. 721-738.

Inoue, H., K. Nakajima, and Y.U. Saito (2014) "Localization of Knowledge Creative Establishments," *RIETI*

Discussion Paper Series, 14-E-53.

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.

Japan Patent Office (2014), *Japan Patent Office Annual Report 2014*, Tokyo.

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

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

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

Maskus, Keith E., and Christine McDaniel. (1999) "Impacts of the Japanese patent system on productivity growth, " *Japan and the World Economy* 11.4: 557-574.

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.

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.

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

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.

Ponds, R., F. van Oort, and K Frenken (2007) "The geographical and institutional proximity of research collaboration," *Papers in Regional Science* 86(3), pp. 423-444.

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

Rosenthal, S. and W. Strange (2003) "Geography, Industrial Organization, and Agglomeration," *Review of Economics and Statistics* 85(2), pp. 377-393.

Saito, U.Y. and I. Yamauchi (2015) "Inventors' Mobility and Organizations' Productivity: Evidence from Japanese Rare Name Inventors," *RIETI Discussion Paper Series*, 15-E-128.

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

Suzuki, J. (2011) "The Structural Characteristics of Research and Development by Japanese Companies, and Issues for the Future," *RIETI Discussion Paper Series*, 11-J-002.

Thompson, P. and M. Fox-Kean (2005) "Patent citations and the geography of knowledge spillovers: a reassessment," *American Economic Review* 95(1), pp. 450–460.

Appendix A. Methodology of extracting establishments

We identify the establishments and their collaborations from the Japanese patent database. Here, we describe in detail the methodology.

For each patent, we conduct the following procedure. First, we identify whether the patent is applied for by firms. This identification is done by using an applicant's name. If the applicant's name includes the term "company limited" (in Japanese, *"kabushikigaisha"*), the applicant is defined as a firm. This definition simultaneously excludes all relatively small firms, such as private limited companies. Second, we identify the establishment information. We check whether the inventor's address includes the firm's name. If the inventor's address includes the firm name, we consider the inventor's address to be the address of the firm's establishment. In practice, there are many different written forms in addresses. Thus, we convert addresses to longitude and latitude information by using the geocoding service that is provided by the Center for Spatial Information Science, the University of Tokyo. This definition of establishment by latitude and longitude simultaneously aggregates two different establishments in the same building.

In some patents, the firm name is not included in the inventor's address, and there are three possible causes for this. First, the inventor registers their establishment address but does not include the firm name. In this case, we try to match the address to the establishment address that includes a firm name. Second, the inventor's residential address is registered although the inventor works in an establishment. Third, the inventor does not belong to the firm. Because there is no solution for these last two cases, we ignore them.

Appendix B. 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 whether the establishment has experience in collaborative work. However, our dataset includes many establishments that have never applied for collaborative patents with other establishments. These establishments may not be potential collaborators. To control for this possibility, we now adopt another definition of potential collaborators. We restrict the establishments to the establishments that have applied for at least one collaboration patent in technology

14

class $A \in \mathfrak{A}$. Below, we describe the results by using the definition of potential collaborating partners and counterfactual collaborations.

Figure B1 shows the baseline result. Similar to Figure 1, the collaborative relations are statistically localized with a localization range of approximately 85 km. Our baseline results remain unchanged although a more conservative definition of potential collaborators is applied.

[Figures B1 here]

Panels (a) and (b) of Figure B2 show the results. Similar to the baseline results that are shown in panels (a) and (b) of Figure 4, the localization of both collaborations are statistically significant, and the range of localization is shorter in inter-firm collaborations.

[Figure B2 here]

Finally, we check the results for firm size. Figure B3 shows the results. Similar to the baseline results that are shown in Figures 6, 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 B3 here]

Our baseline results remain completely unchanged although a more conservative definition of potential collaborators than the originally adopted definition is applied. The localization of the establishments that collaborate for innovative activities is statistically significant, and we observe firm-border effects, especially regarding small firms.

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
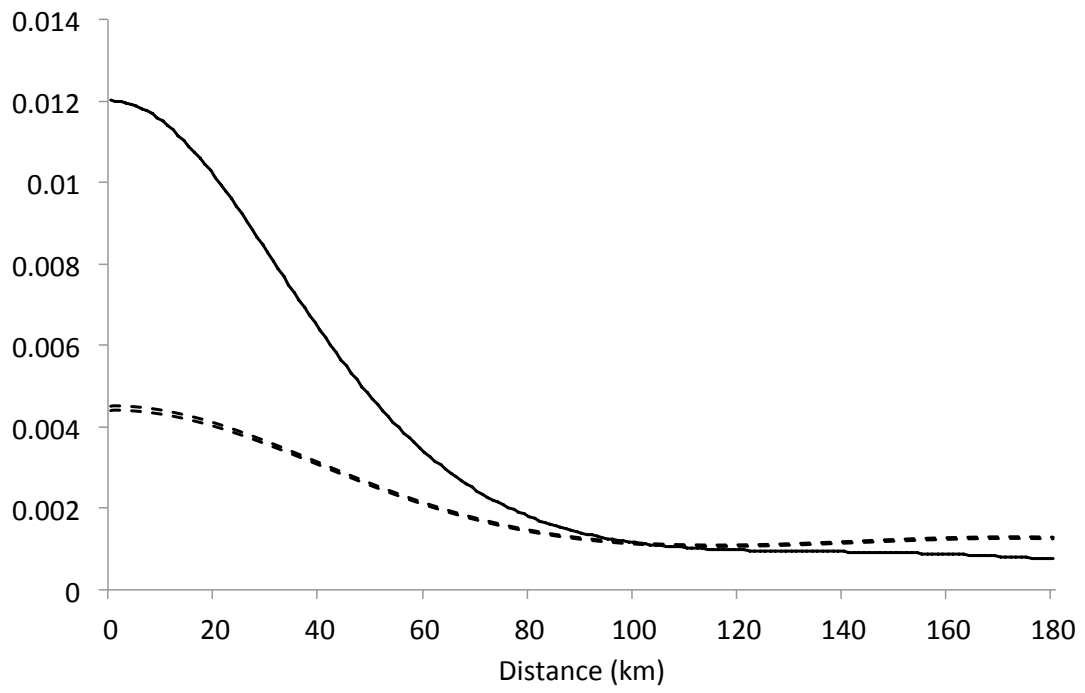
Figure 1: K-density of collaborations
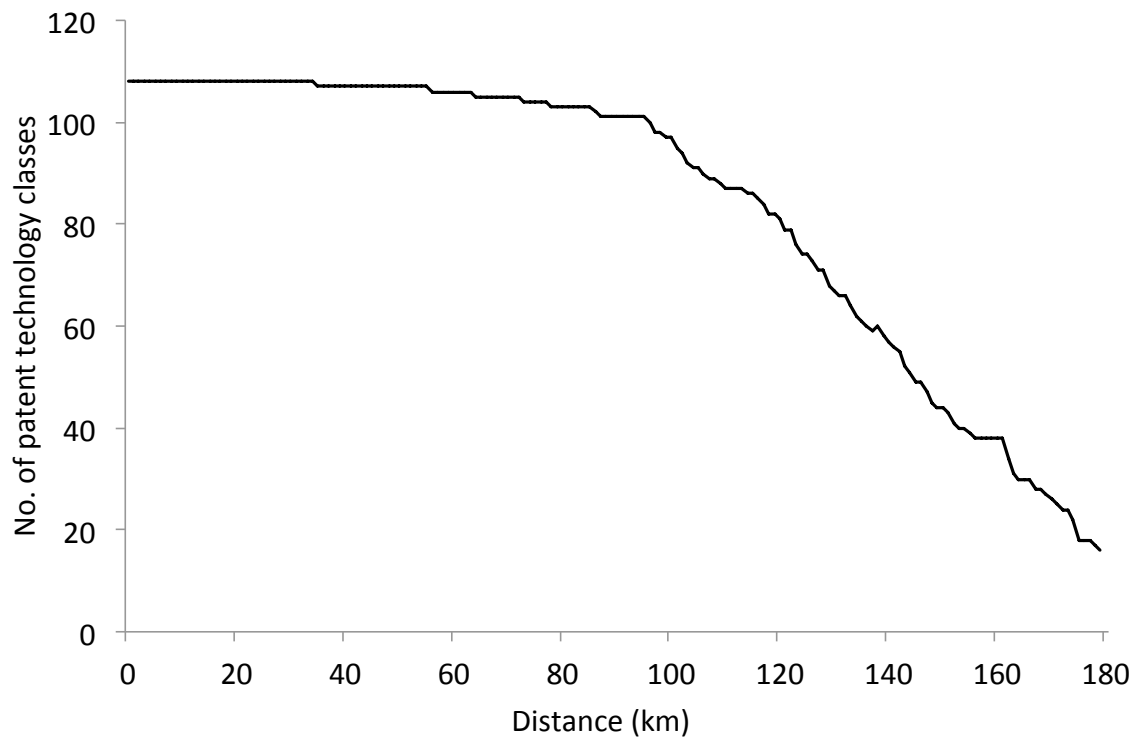
Figure 2: Number of localized technology classes for each distance
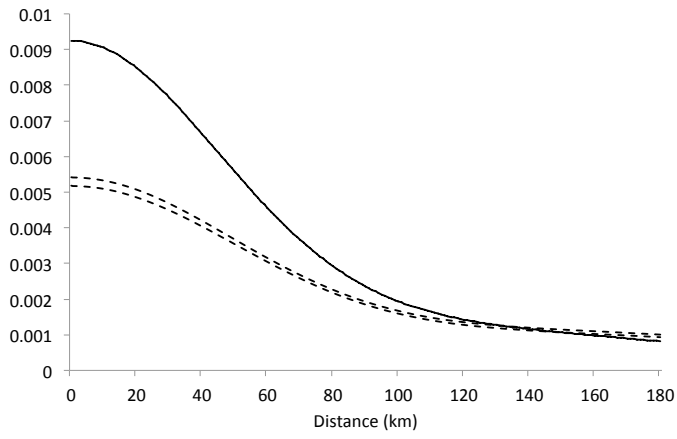
(a) 1986-1990
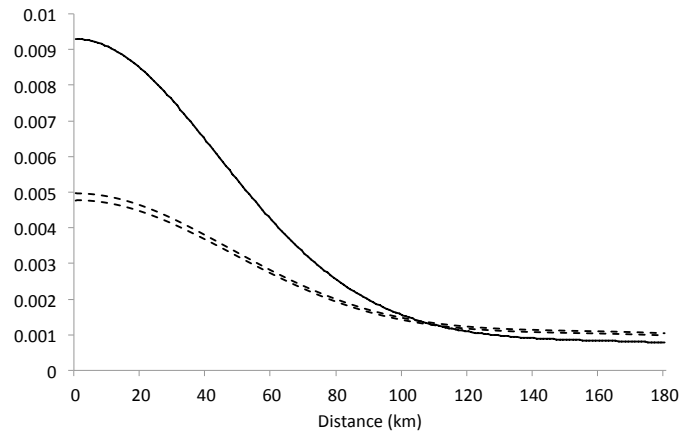
(b) 1991-1995
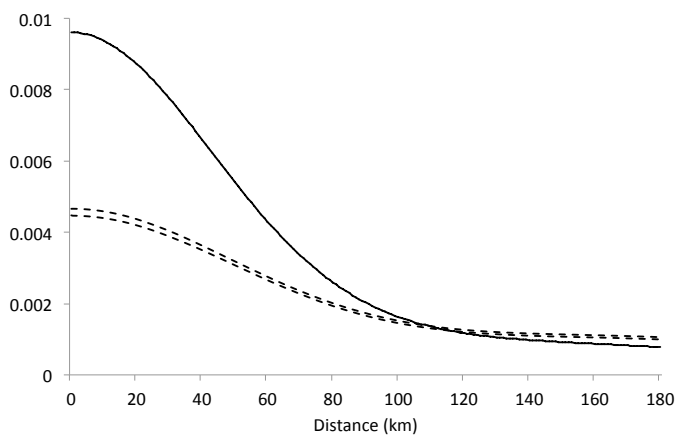
(c) 1996-2000
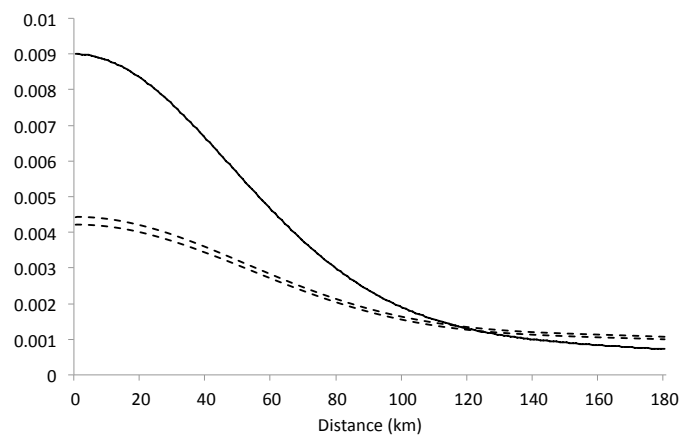
(d) 2001-2005

Figure 3: Relative densities for each five-year period
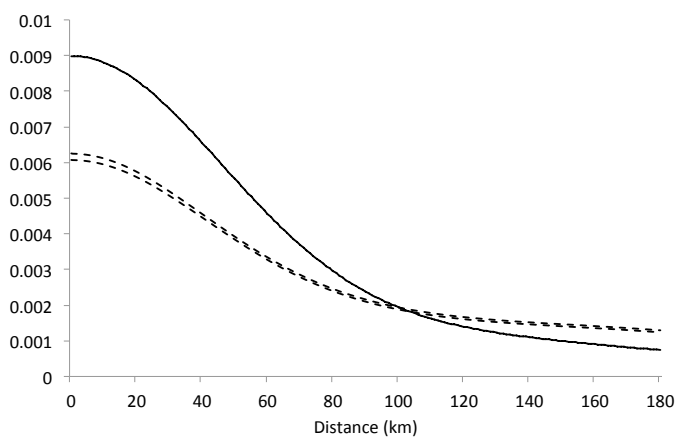
(a) Intra-firm collaborations
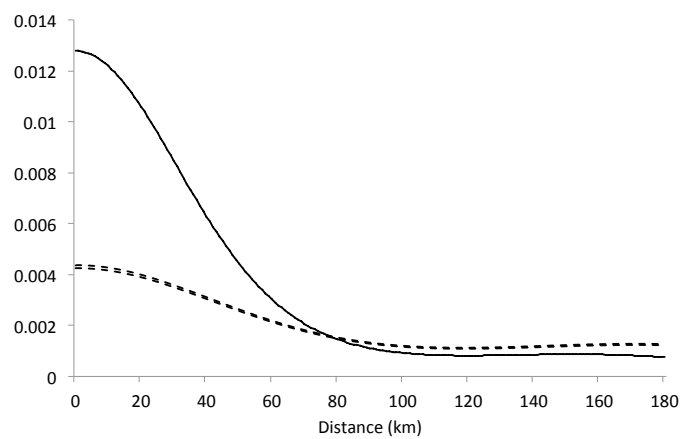
(b) Inter-firm collaborations

Figure 4: K-densities and $\gamma_A(d)$ of intra- and inter-firm collaborations
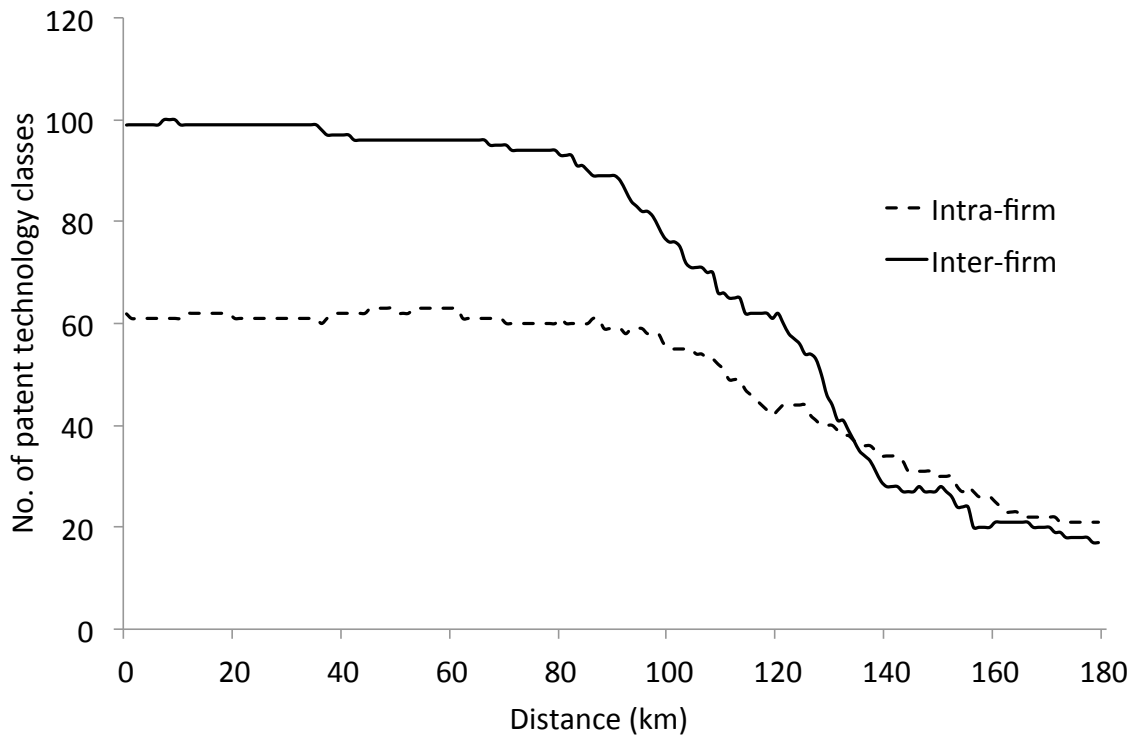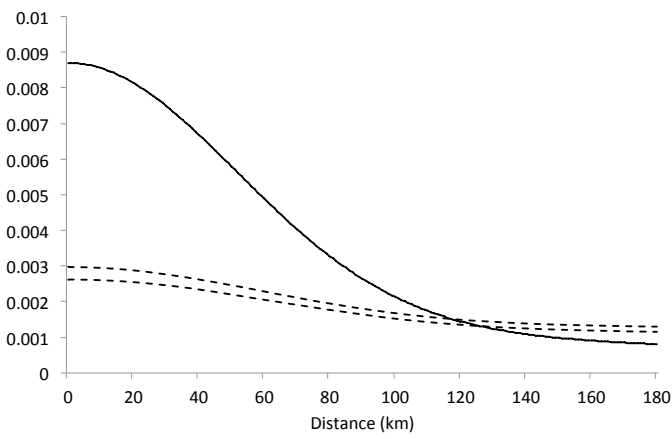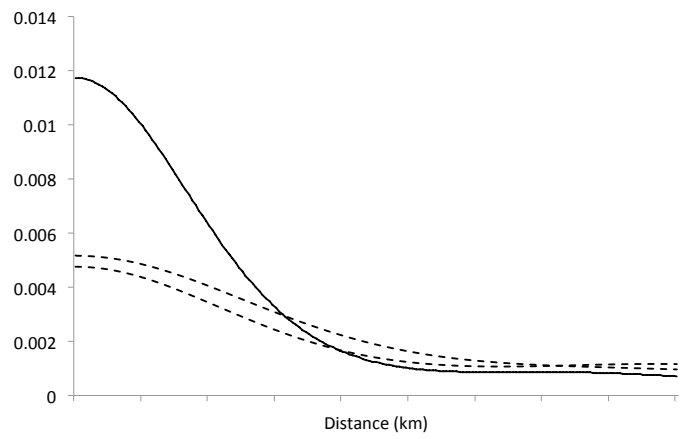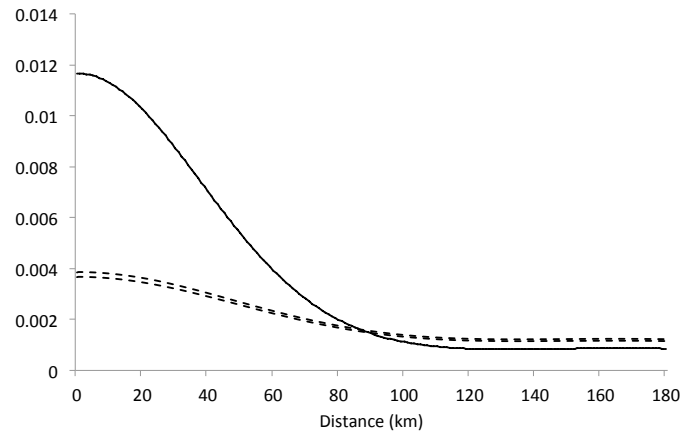
Figure 5: Number of localized technology classes in each distance (intra- compared with inter-firm)

(a) Single–single collaborations

(b) Multiple–multiple collaborations

(c) Multiple–single collaborations

Figure 6: K-densities of inter-firm collaborations
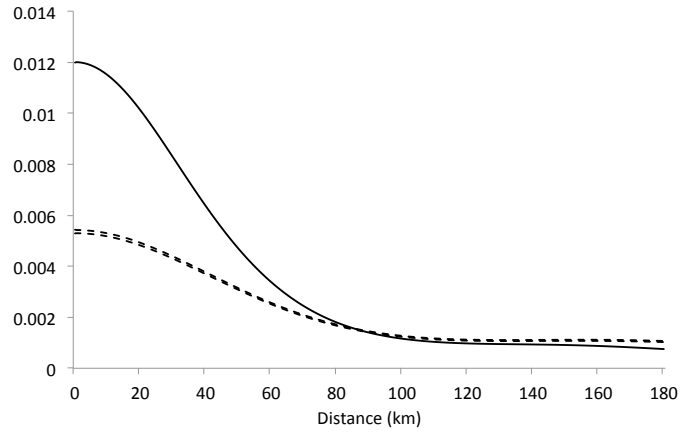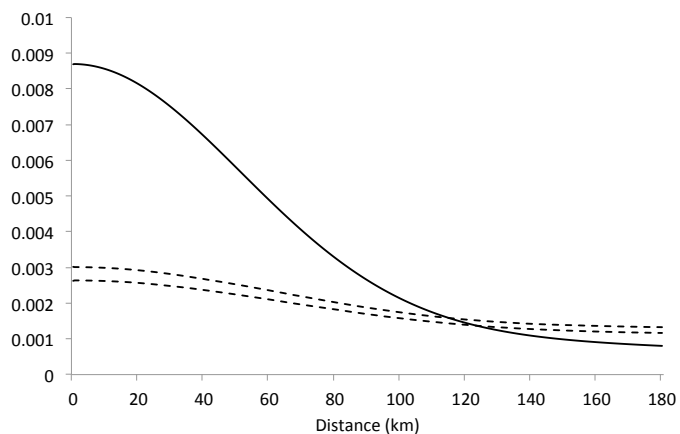
Figure B1: All collaborations
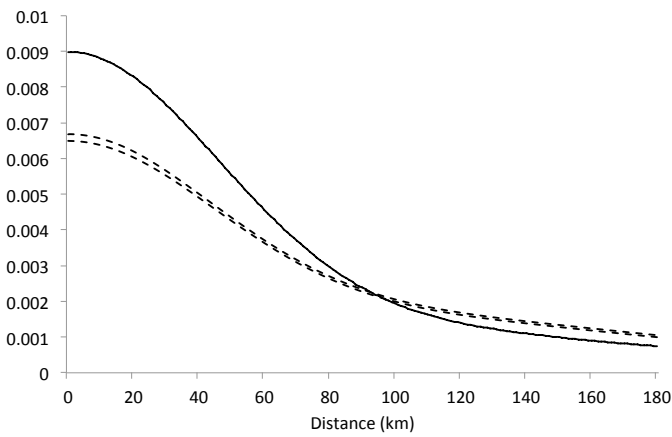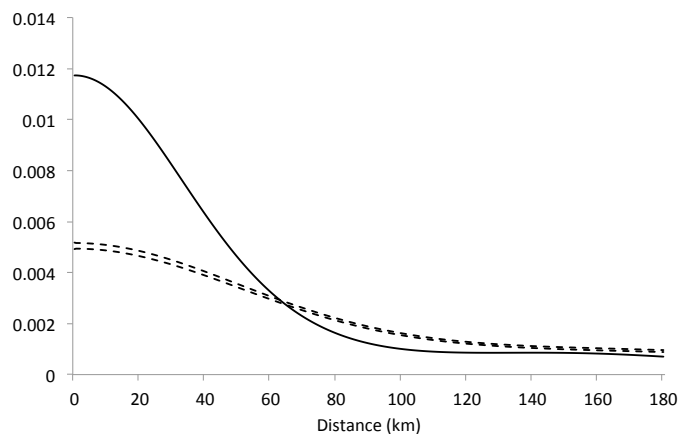

(a) Intra-firm collaborations
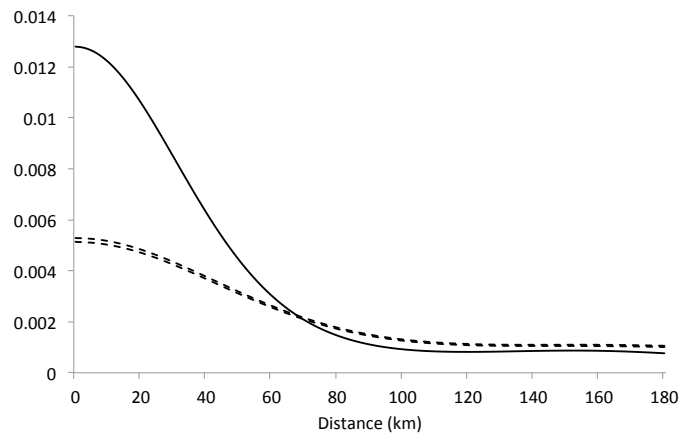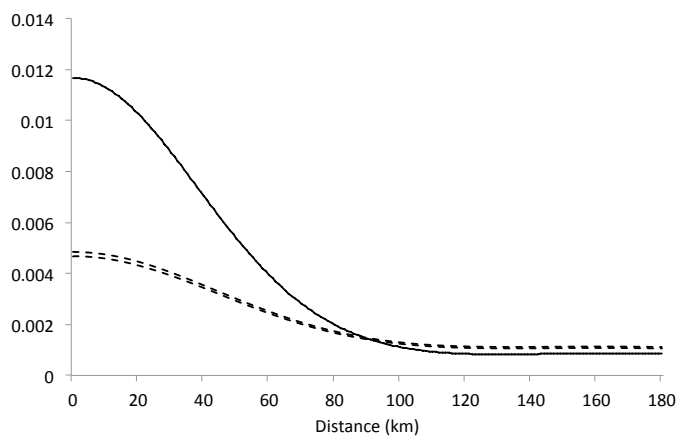

(b) Inter-firm collaborations

Figure B2: K-densities of collaborating relations

(a) Single–single collaborations

(b) Multiple–multiple collaborations

(c) Single–multiple collaborations

Figure B3: K-densities of inter-firm collaborations