RIETI Discussion Paper Series 14-E-053

Localization of Knowledge-creating Establishments

INOUE Hiroyasu
Osaka Sangyo University

NAKAJIMA Kentaro
Tohoku University

SAITO Yukiko Umeno
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 14-E-053
August 2014

Localization of Knowledge-creating Establishments[*]

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 establishment-level knowledge creation by using data from the Japanese patent database. Using distance-based methods, we obtained the following results. First, Japanese patent-creating establishments are significantly localized at the 5% level, with the range of localization at approximately 80 km. Second, localization was found for all patent technology classes, while the extent of localization differs among the classes. Third, the extent of localization is stronger in more creative establishments, in terms of both the number of patents created and the number of citations. These results suggest that geographical proximity is important for knowledge spillover regardless of the concerned technology and that creative establishments require external knowledge.

Keywords: Patent, K-density approach, Knowledge spillover
JEL classification: R11

RIETI Discussion Papers Series aims at widely disseminating research results in the form of professional papers, thereby stimulating lively discussion. The views expressed in the papers are solely those of the author(s), and neither represent those of the organization to which the author(s) belong(s) nor the Research Institute of Economy, Trade and Industry.

[*] This study was conducted as a part of the Inter-organizational and Inter-inventors Geographical Proximity and Networks project undertaken by the Research Institute of Economy, Trade and Industry (RIETI). We thank the RIETI for providing us the access to the micro database on Japanese establishments, and the Center for Spatial Information Science, University of Tokyo for providing us with the geocoding service. We gratefully acknowledge the financial support from the Japan Society for the Promotion of Science (Nos. 24530506, 25380275, and 25780181).

[†] 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

Knowledge creation is a major driver of the growth of modern economies, and knowledge spillovers and idea exchanges are crucial for knowledge creation. Since Marshall (1890), it has well been recognized that geographical proximity enhances knowledge spillovers and idea exchanges, which causes industrial agglomeration.

Several preceding studies have examined the localization of knowledge spillovers using patent citation (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Murata et al., 2014) and inter-organizational collaboration (Inoue et al., 2013) as a proxy for knowledge spillovers. These studies suggest that firms should promote the geographic concentration of their R&D laboratories in order to facilitate knowledge spillover.

On the other hand, localization of knowledge creation has been examined in previous studies. Carlino et al. (2012) found that R&D laboratories are geographically localized, relative to the overall geographic distribution of manufacturing employments. They used address information of R&D laboratories from Directory of American Research and Technology, and found that for most of the industries, R&D laboratories are significantly localized relative to the overall manufacturing activities. However, they only focus on the limited states in the US, and spatial distribution of manufacturing industry might not be appropriate reference for R&D activities because firms other than the manufacturing industries also conduct R&D.

With this background, this study investigates the localization of Japanese patent creations. We utilize the entire patent database of Japan, and capture the geographical features of all patent-creating establishments comprehensively in Japan. A convention in the Japanese patent system allows us to analyze establishment-level localizations in patent creation. In Japan, inventors register the address of the establishments to which they belong. This convention gives us information on establishment-level patent creation and comprehensive geographical features of patent-creating establishments.

To investigate the localization of patent-creating establishments, we conduct a distance-based analysis, as developed by Duranton and Overman (2005). This approach focuses on the distribution of bilateral distance between pairs of establishments and is, therefore, free from the problems of administrative boundaries. The critical idea is to compare the distribution of bilateral distances with the counterfactual distributions generated by a random allocation of establishment locations to all the potential sites. This method allows us to evaluate the extent of localizations based on the deviation from random location assignments. For the potential sites of patent-creating establishments, we use all establishments of all industries in Japan.

We obtain the following results. First, locations of patent-creating establishments are significantly localized at the 5% level, with the range of localization being approximately 80 km. Second, localization was found for all patent technology classes, while the extent of localization differs among the classes. Third, the extent of localization is stronger in more creative establishments, in terms of both the number of patents created and the number of citations. This implies that highly creative establishments require external knowledge from other establishments. These findings suggest that knowledge spillovers would be an important determinant of agglomeration of economic activities, especially in more creative establishments.

The rest of this paper is organized as follows. In the next section, we introduce the dataset and identification of patent-creating establishments. Section 3 describes the empirical strategy based on the micro geographic information of each establishment. Section 4 presents our baseline results and differences in technology. Section 5 focuses on the differences in creativity of establishments. Next, Section 6 investigates the robustness of our result. Finally, Section 7 concludes the paper.

2. Data

We utilize the Institute of Intellectual Property (IIP) patent database (Goto and Motohashi, 2007) that includes all the patent publications (the Patent Gazette) in Japan. This database includes basic patent information, like patent ID, publication date, names and addresses of applicants, and name and addresses of inventors. The database also includes citation information on each patent, such as the number of times the patent has been cited. From this database, we use all the patents published from 1993 to 2008.

This study focuses on the localization of patent-creating establishments. We identify the patent-creating establishments from the patent database, taking advantage of a convention of the Japanese patent system where inventors register the address of the establishments to which they belong as "inventors address" (Inoue, Nakajima, and Saito, 2013).

Here, we describe the algorithm how to identify the patent-creating establishments from our patent database following Inoue, Nakajima, and Saito (2013). First, firms are identified by name and address of applicants. Here, we define the firm as an applicant whose name includes the term "company limited," or "*kabushikigaisha*" in Japanese. This definition simultaneously excludes all relatively small firms, such as private limited companies. Second, the patent-creating establishments are identified as follows. We check whether the firm's name is included in the inventor's address. Then, we consider the inventor's address with the firm name as the address of the establishment owned by the firm.

Using this identification method, we get the following information. Table 1 gives the summary of the dataset. The dataset includes 1,967,361 patents and 1,189,262 are applied for by establishments. The number of firms as applicants is 56,592 and the number of patent-creating establishments is 74,452.

[Table 1 here]

Furthermore, our analysis requires potential sites where the patent-creating establishments can be located. We assume that patent-creating establishments can be located at every site where all establishments of all industries are located. To obtain information on the location of establishments of all industries, we use micro data of the Establishment and Enterprise Census. This database includes the address and number of employment information. Then, we convert the establishments' address into a latitude and longitude format.[1]

3. Empirical Strategy

In order to examine the localization of patent-creating establishments, we apply Duranton and Overman's (2005) distance-based approach. Intuitively, we first calculate the distribution of bilateral distances between every pair of patent-creating establishments, and then, compare the distribution with the counterfactual distributions generated from the random assignment of locations from potential sites.

3.1. K-density approach

We now describe how to measure the localization of patent-creating establishments by the K-density approach in detail. First, we estimate the distribution of bilateral distances between every pair of patent-creating establishments.

Let n be the number of establishments that have applied for at least one patent, and we have $n(n-1)/2$ number of unique bilateral distances in the patent-creating establishments. Next, let d_{ij} be the great circle distance between patent-creating establishments i and j. The estimator of the density of bilateral distances at any point d is

$$\widehat{K}(d) = \frac{1}{n(n-1)h}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} f\left(\frac{d-d_{ij}}{h}\right),$$

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

[1] We use geocoding service provided by the Center for Spatial Information Science, the University of Tokyo.

3.2. Counterfactual distribution and statistical testing

Overall economic activities (all establishments) have a tendency to agglomerate. To precisely detect the patent-creating establishments localization, we need to control for the localization of the overall economic activities. To do so, we generate a counterfactual distribution of locations for patent-creating establishments where establishments randomly choose their locations from all the potential sites as reference, and then, compare the actual localization of patent-creating establishments with the counterfactual localization. We consider the site of establishments of overall industries as potential sites for patent-creating establishments.

To test the localization of patent-creating establishments, we construct a two-sided confidence interval, with 95% of the K-densities based on randomly chosen counterfactual locations within the interval bands. To be more specific, we randomly choose n sites from the potential sites and estimated the K-density in the counterfactual situation. By iterating this trial 1,000 times, we can construct confidence bands. Then, we calculate the global confidence bands, i.e., an upper confidence band $K^U(d)$ and a lower confidence band $K^L(d)$, so that 95% of the 1,000 randomly drawn K-densities are within the bands over the entire distance range, which, in our case, is 0–180 km.

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

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

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

4. Baseline results

4.1. 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 within the 0–80 km range, the K-density is above the upper global confidence band. Thus, we consider the patent-creating establishments to be localized in the 0–80 km range[2]. We conclude that the patent-creating establishments are significantly localized relative to overall establishments. This implies the importance of knowledge spillovers as a driving force for agglomeration.

[Figure 1 here]

4.2. Differences in Patent Technology Classes

The baseline analysis pools all establishments that apply for patents regardless of their patent technology class. The extent of localization, however, might be different across patent technology classes. Establishments that create sophisticated, high technology patents would be more localized in order to pursue more knowledge transfers. To grasp this difference across patent technology classes, we repeat the analysis by patent class.

Our K-density estimator is modified for the technology-level analysis. Let S^A be a set of establishments that have applied for at least one patent in the patent technology class $A \in \mathfrak{A}$, where $\mathfrak{A}$ represents a set of patent technology classes. Let n_A be the number of establishments having patents in the patent technology class A. Similarly, let d_{ij} be the great circle distance between establishments i and j in the set S^A. The estimator of the density of bilateral distances at any point d for patent technology class A is

$$\widehat{K^A}(d) = \frac{1}{n_A(n_A - 1)h} \sum_{i=1}^{n_A - 1} \sum_{j=i+1}^{n_A} f\left(\frac{d - d_{ij}}{h}\right).$$

For the counterfactual distribution, similar to the baseline analysis, we consider the site of every establishment as a potential site for patent-creating establishments in the patent technology-class $A \in \mathfrak{A}$.

To denote patent-technology class, we use the first-three letters in the International Patent Classification (IPC). This classification includes a total of 120 patent-technology classes in our dataset.

Figure 2 shows the number of patent technology classes that are localized at each distance. In the range of 0-60 km, all 120 patent classes are localized. Then, after 60 km, the number of localized patent classes declines gradually. This pattern is quite similar to industrial localization in the manufacturing industry

---

[2] Range of localization is 40 km for firm-level industrial localization (Nakajima, Saito, and Uesugi, 2012a), 40km for inter-firm transaction localization (Nakajima, Saito, and Uesugi, 2012b), and 100 km for inter-establishment collaboration localization (Inoue, Nakajima, and Saito, 2013).

(Duranton and Overman, 2005; Nakajima et al., 2012)[3].

[Figure 2]

Next, we investigate the differences in the extent of localization among patent technology classes in detail. Table 2 shows the top 10 patent technology classes in terms of the extent of localization, Γ. Most of the patent technology classes in the table are high-tech industries, such as Aircraft; Aviation; Cosmonautics (IIP B64), etc. Table 3 shows the bottom 10 patent technology classes. In this table, the patent technology classes are low-tech industries, such as Butchering; Meat Treatment; Processing Poultry or Fish (IIP A22), etc. These tables suggest that the establishments in higher-technology industries may require more advanced knowledge transfers.

[Tables 2 and 3]

5. Differences in Establishments Creativity

We now consider the differences in establishments creativity. In the baseline analysis, we treat each establishment as homogeneous. However, establishments are heterogeneous in terms of their patent creativity. The number of patent created by each establishment extremely differ among establishments and each patent has different impact or quality. Importance of knowledge spillovers might be different depending on establishments creativity, resulting in different geographical pattern of establishments location. In order to examine this differences, we repeat the similar analysis after including establishments heterogeneity as in Duranton and Overman (2005).

Our new estimator of the K-density function is as follows,

$$\widehat{K}(d) = \frac{1}{h\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w(i)w(j)} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w(i)w(j)f\left(\frac{d-d_{ij}}{h}\right),$$

where $w(i)$ is the weight on creativity for establishment i. We consider the two measures for establishment creativity, the number of patent created and the number of total citations.. The former measure can be interpreted as creativity in terms of quantity and the later one in terms of quality

[3] Half of the industries are localized within 0-60 km, then the number of localized industries starts declining gradually. Note that localization of location is examined relatively within manufacturing industries, which leads to small ratio of localized industries compared with our analysis.

Figure 3 shows results in terms of quantitative creativity, i.e., the number of patents for the weight. The solid line in the figure represents the K-density weighted by the number of patents created, and dashed lines represent the global confidence bands. For every distance within the 0–80 km range, the K-density is above the upper global confidence band. Thus, we consider the location to be localized in the 0–80 km range, even if we weight each establishment by the number of patents created.

We also show the baseline K-density (without weighting) by the dotted line in the figure. Then, we clearly find that the weighted K-density is above the non-weighted K-density within a 0-50 km range.[4] That is, establishments that create more patents are more localized. This implies that establishments that require more knowledge transfers are more localized, or the greater concentration of establishments benefits the creativity of each establishment located in the area through larger knowledge transfers.

[Figure 3 here]

Next, we focus on qualitative creativity (Figure 4). The solid line in the figure represents the K-density weighted by the total number of patent citations, with dashed lines representing the global confidence bands, and the dotted line representing the baseline K-density. We obtain a similar result to previous results weighted by the number of patents created. In the close range (0-80 km), establishments are localized, and the weighted K-density is more localized than a non-weighted one. Even if we use patent quality as a measure of the establishments creativity, creative establishments are more localized.

[Figure 4 here]

6. Robustness

One may concern that our results on the localization of more creative establishments may come from the establishment size. That is, larger establishment might be more creative in terms of both quantity and quality, and lager establishments might be more geographically concentrated. Another concern is that patent creating establishments might be larger than the non-patent creating establishments. In these cases,

[4] The comparison between weighted and un-weighted distribution can be tested empirically. Under the null hypothesis that all of the patent creative establishments have the same tendency to localize, we can construct confidence interval bands by a Monte Carlo simulation similar to the baseline analysis.

stronger localization in patent-creating establishments may be caused by the greater concentration of larger establishments, and it does not necessarily represent the importance of knowledge transfers. To check this concern, we generate a counterfactual random distribution by weighting it by the establishment's size of employment.

Figure 5 shows the results. The solid lines in the figure represent the global confidence bands weighted by the employment size, and the dashed lines represent the global confidence bands without weighting. Indeed, weighted confidence bands are above the non-weighted confidence bands within the 0-60 km range. That means that larger establishments are more localized.

[Figure 5 here]

Next, we confirm the effect of this difference of localization in terms of establishment size on our results. Figure 6 shows the results that use employment-size weighted K-densities as the counterfactual distributions. Panel (a) in Figure 6 shows the baseline result. The actual K-density is not weighted and the same as the K-density in Figure 1. Qualitatively, the result is unchanged. Patent-creating establishments are localized in close proximities. That is, localization of patent-creating establishments is significant even if we control the size of establishments in reference. Note that the extent of localization in terms of Γ is smaller than the baseline result in Figure 1. That is, the extent of localization in patent-creating establishments partially includes establishments size-effects.

Panels (b) and (c) in Figure 6 show the results on the weighted K-density by the number of patents created and total number of patent citation, respectively. Results are similar to the baseline results, while the weighted results are less localized. Even after controlling for establishments size as counterfactual, the extent of localization is still stronger in more creative establishments as shown in differences of these panels (Panels (b) and (c)) from Panel (a).

[Figure 6 here]

Finally, we consider the differences between industries. Overall tendency of localization might be different between manufacturing and other industries. Then, we consider the possibility that manufacturing industries tend to create more patents and that localization of the patent-creating establishments comes from the localization of manufacturing industries. To control this possible tendency, we calculate counterfactual distributions by restricting samples to manufacturing establishments and weight them by the number of employees.

Figure 7 shows the results. The solid lines in the figure represent the global confidence bands after restricting the sample to the manufacturing establishments weighted by employment size, and dashed lines represent the global confidence bands by overall industries weighted by employment size. Indeed, confidence bands of manufacturing establishments are above confidence bands of other industries by a range of over 20 km, but they have almost similar shapes. Bias from industry difference would be small.

[Figure 7 here]

Next, we confirm the effect of this difference in localization in terms of industries on our results. Figure 8 shows the results that use K-densities by manufacturing establishments as compared to the counterfactual distributions. Panel (a) in Figure 8 shows the baseline result. The actual K-density is not weighted and is the same as the K-density in Figure 1. Panels (b) and (c) in Figure 8 show the results on the weighted K-density by the number of patent created and the total number of patent citation, respectively. Results are similar to the baseline results.[5]

[Figure 8 here]

7. Concluding Remarks

This study investigates the localization of patent-creating establishments in Japan. Using Duranton and Overman's (2005) K-density approach, we found the following results. First, Japanese patent-creating establishments are significantly localized with the range of 0-80 km. Second, localization was found for all patent technology classes, while the extent of localization differs among the classes. Third, the degree of localization is stronger in more creative establishments, in terms of both quantity and quality. Finally, these findings are robust when controlled for the size of the establishment and industry effects.

These results suggest that geographical proximity is important for knowledge spillover regardless of concerned technology and that creative establishments require external knowledge. Further, it is suggested that knowledge spillovers are an important determinant of agglomeration of economic activities, especially for creative establishments.

References

[5] These results are robust if we use plant-level data from manufacturing census.

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

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

Ellison, G., E. Glaeser, and W. Kerr (2010) "What Causes Industry Agglomeration?" *American Economic Review* 105, pp. 889-927.

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

Hall, B.H. and A.B. Jaffe (2001) The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools.

Inoue, H., K. Nakajima, and Y.U. Saito (2013) "Localization of Collaborations in Knowledge Creation," *RIETI Discussion Paper Series,* 13-E-70.

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

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.

Marshall, A. (1890) *Principles of Economics*, Macmillan, London.

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,* forthcoming.

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

Nakajima, K., Y.U. Saito, and I. Uesugi (2012b) "Localization of Interfirm Transaction Relationships and Industry Agglomeration," RIETI Discussion Paper Series, 12-E-23.

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

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

Wuchty, S., B.F. Jones, and B. Uzzi (2007) "The Increasing Dominance of Teams in Production of Knowledge," Science 316(5827), 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 applicants (Firms)	56,592
Number of establishments	74,452

Table 2: Top 10 patent-technology classes in localization

Rank	IPC	Technology-class	Gamma
1	B64	Aircraft; Aviation; Cosmonautics	0.348
2	G07	Checking-Devices	0.346
3	G04	Horology	0.329
4	G06	Computing; Calculating; Counting	0.315
5	H03	Basic Electronic Circuitry	0.313
6	G11	Information Storage	0.312
7	H04	Electric Communication Technique	0.306
8	G12	Instrument Details	0.295
9	B42	Bookbinding; Albums; Files; Special Printed Matter	0.290
10	B43	Writing or Drawing Implements; Bureau Accessories	0.287

Table 3: Bottom 10 patent-technology classes in localization

Rank	IPC	Technology-class	Gamma
1	A22	Butchering; Meat Treatment; Processing Poultry or Fish	0.000
2	C06	Explosives; Matches	0.031
3	B27	Working or Preserving Wood or Similar Material; Nailing or Stapling Machines In General	0.058
4	A24	Tobacco; Cigars; Cigarettes; Smokers' Requisites	0.083
5	C21	Metallurgy of Iron	0.086
6	F26	Drying	0.094
7	F22	Steam Generation	0.094
8	C05	Fertilisers; Manufacture Thereof	0.096
9	B22	Casting; Powder Metallurgy	0.096
10	B02	Crushing, Pulverising, or Disintegrating; Preparatory Treatment of Grain for Milling	0.105

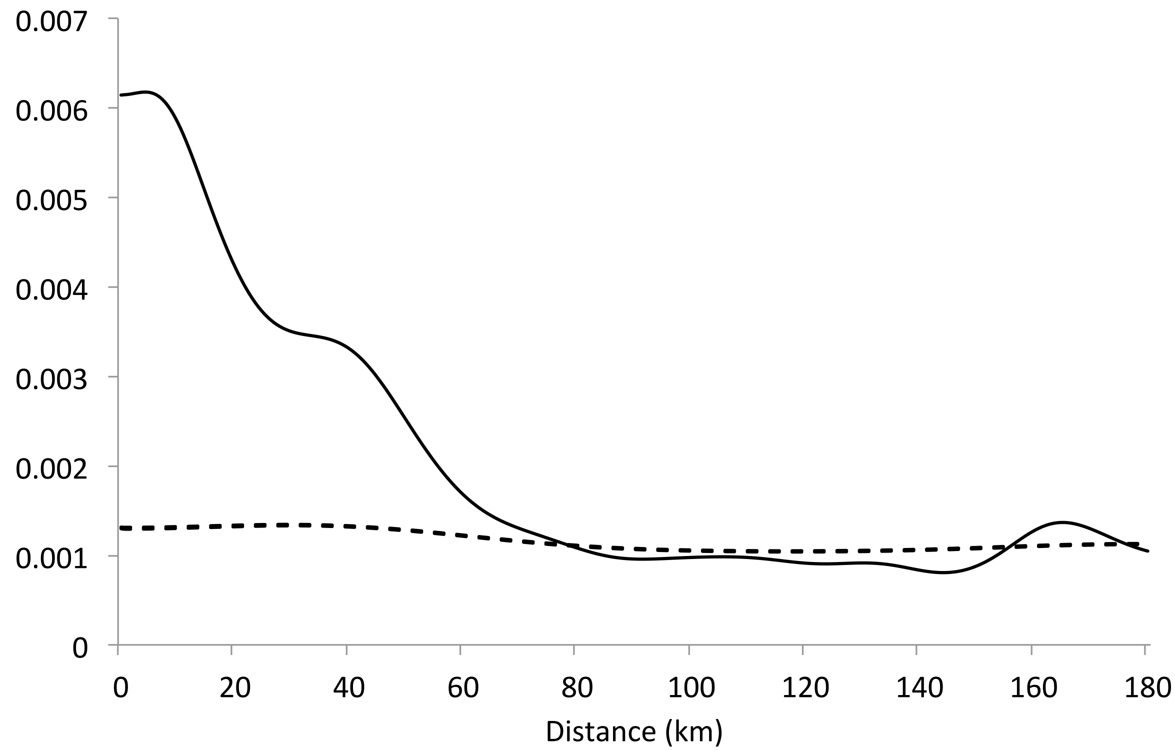


Figure 1: Result of the baseline analysis

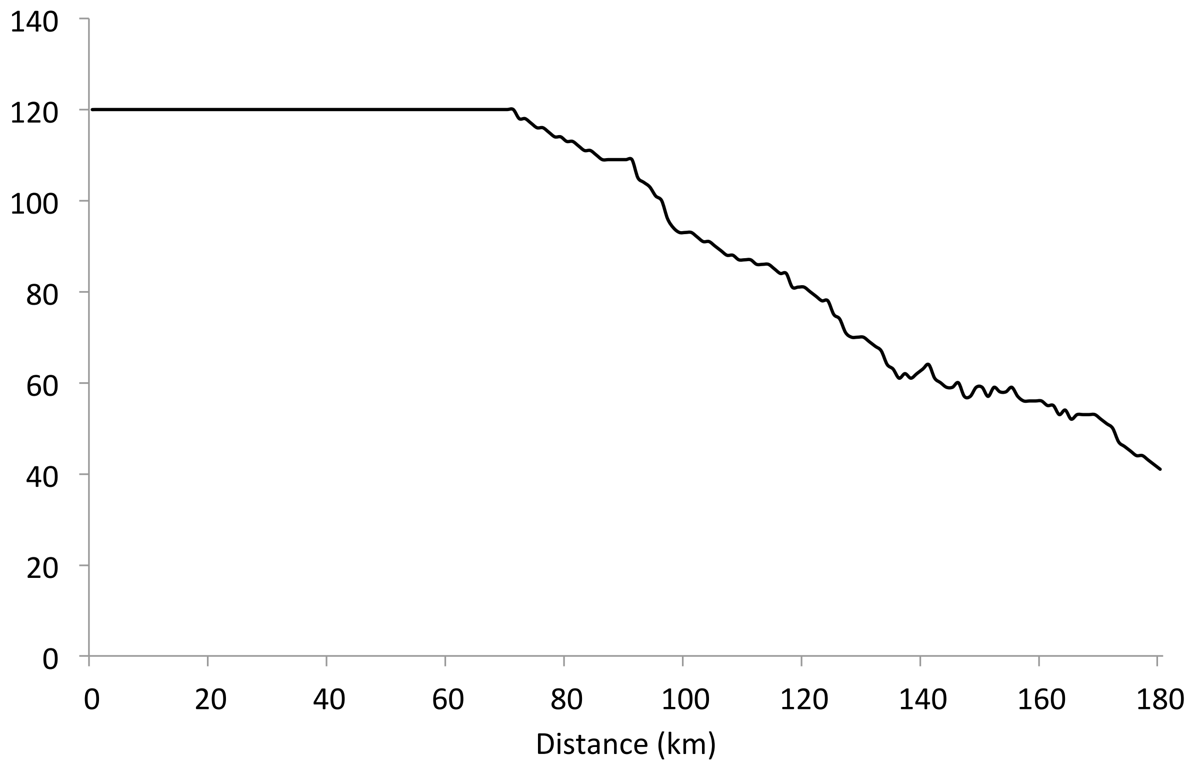


Figure 2: Number of localized patent classes in each distance

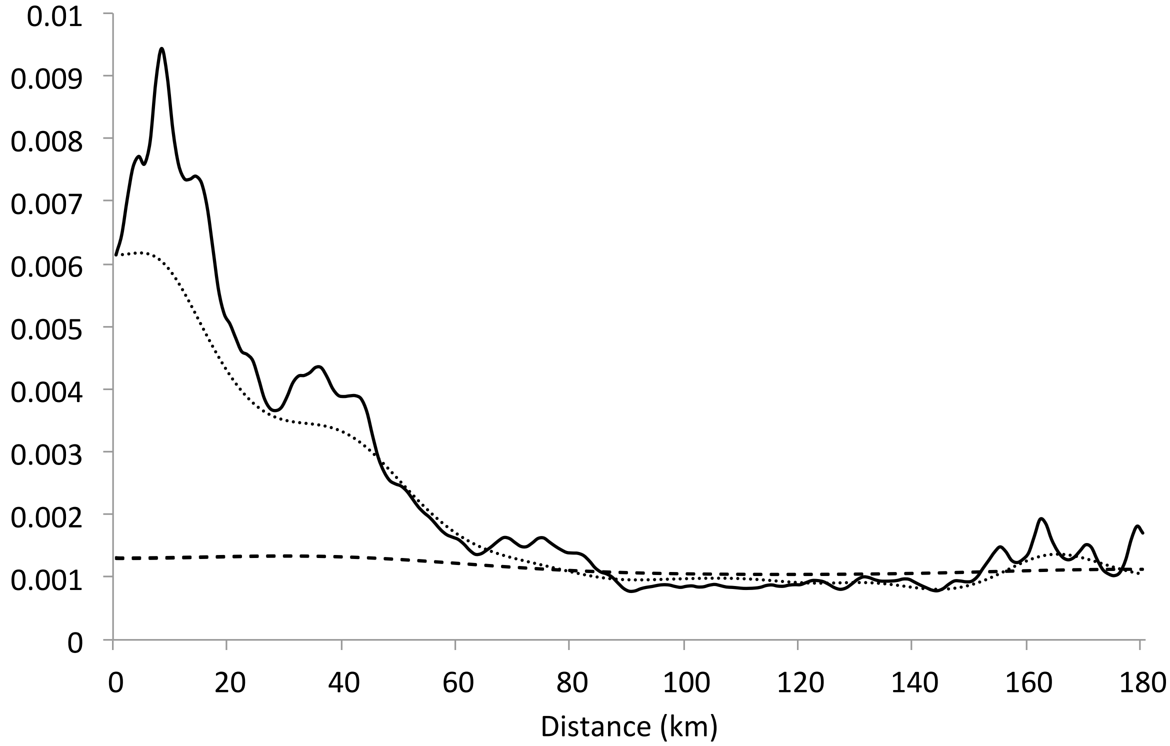


Figure 3: Results weighted by number of patent creations

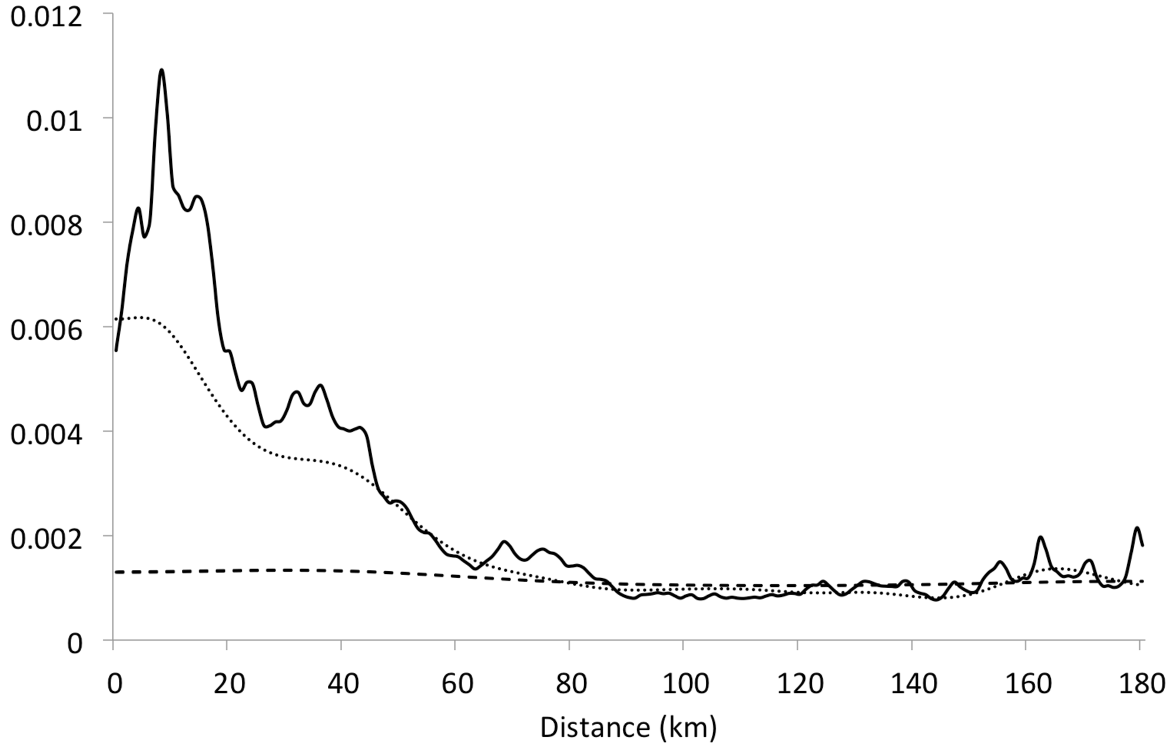


Figure 4: Results weighted by average number of patent citations

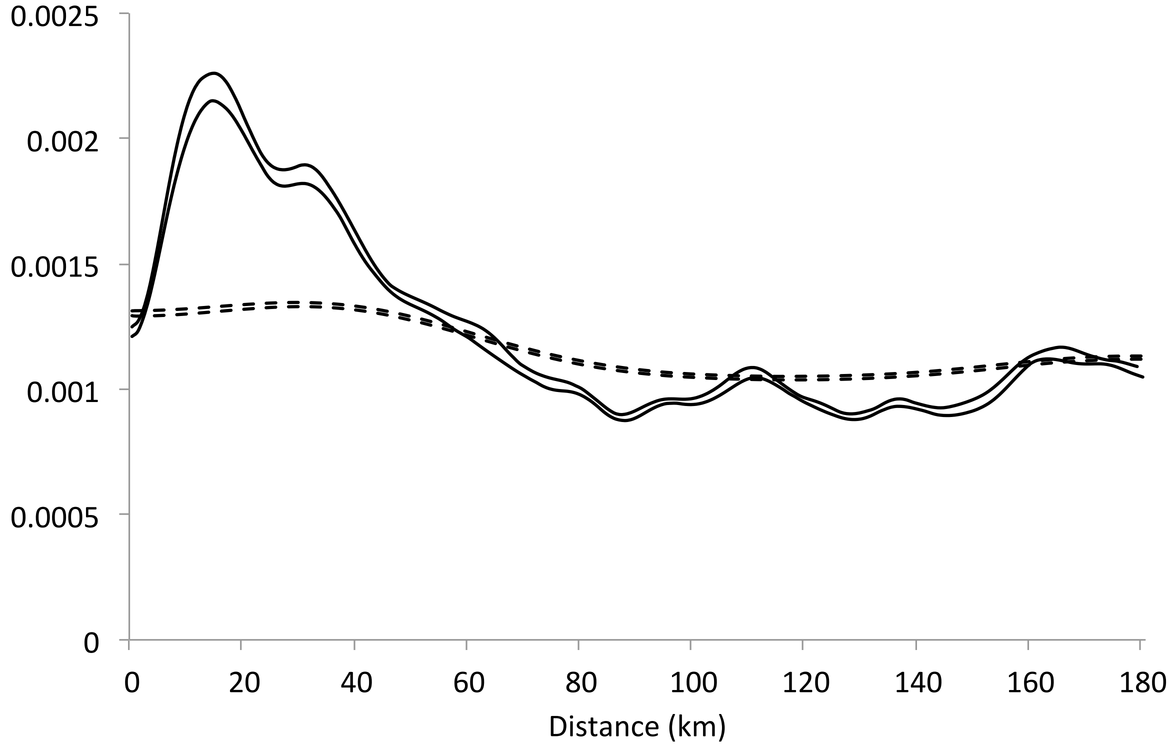
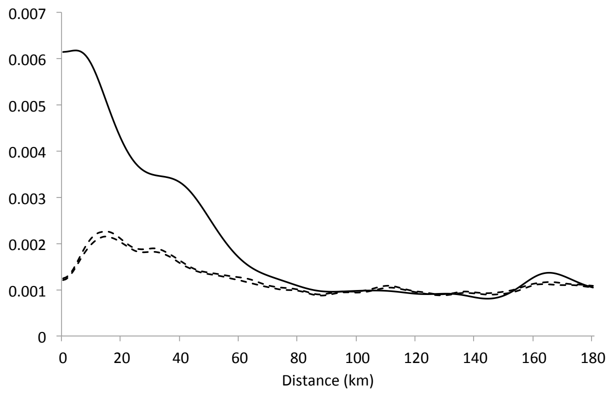
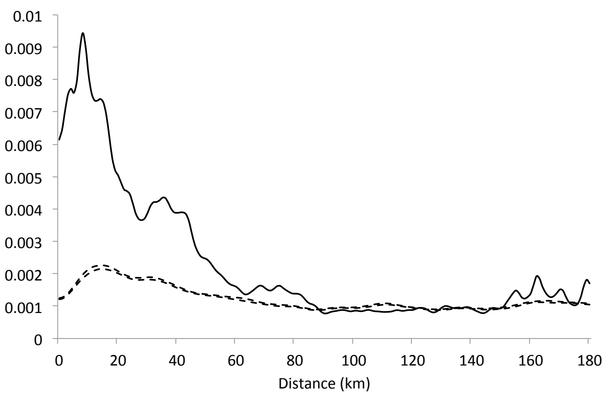


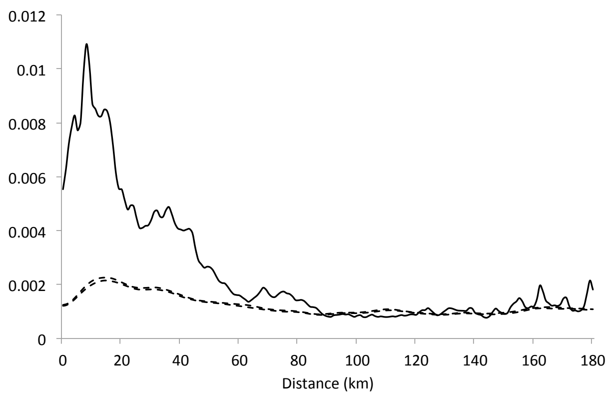
Figure 5: Confidence bands weighted by employment size



(a) Baseline



(b) Weighted by the number of patents



(c) Weighted by number of citations

Figure 6: Results with weighted confidence bands by establishment size

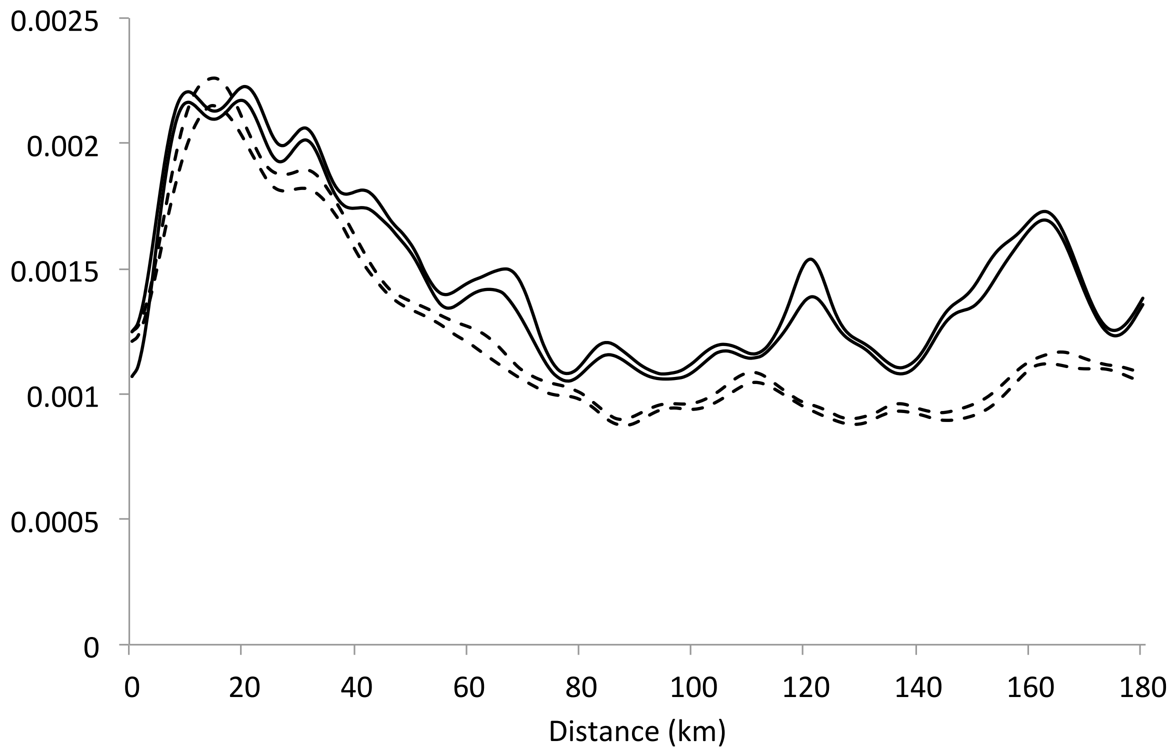
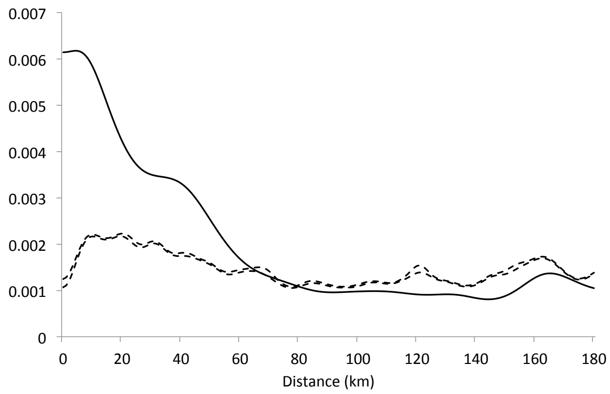
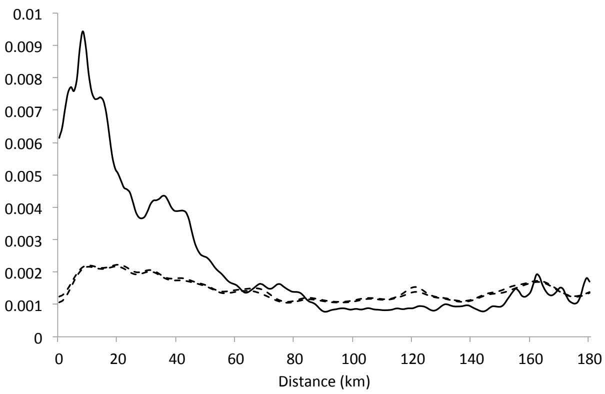


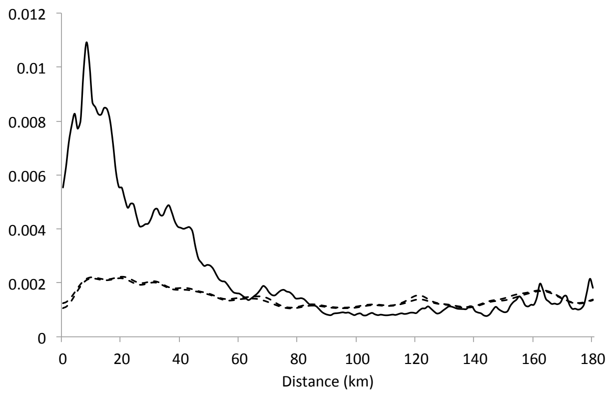
Figure 7: Confidence bands (Overall vs. manufacturing industry)



(a) Baseline



(b) Weighted by the number of patents



(c) Weighted by the number of citations

Figure 8: Results with confidence bands in manufacturing industry