Grant-in-Aid for Scientific Research（S）

Real Estate Markets, Financial Crisis, and Economic Growth

: An Integrated Economic Approach

Working Paper Series No.47

Localization of Knowledge-creating Establishments

Hiroyasu INOUE

Kentaro NAKAJIMA

and

Yukiko Umeno SAITO

March, 2016

HIT-REFINED PROJECT

Institute of Economic Research, Hitotsubashi University

Naka 2-1, Kunitachi-city, Tokyo 186-8603, JAPAN

Tel: +81-42-580-9145

E-mail: hit-refined-sec@ier.hit-u.ac.jp

http://www.ier.hit-u.ac.jp/hit-refined/

Localization of Knowledge-creating Establishments*

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

March 2016

Abstract

This study investigates the localization of establishment-level knowledge creation using data from a Japanese patent database. Using distance-based methods, we obtain the following results. First, Japanese patent-creating establishments are significantly localized at the 5% level, with a localization range of approximately 80 km. Second, localization is observed for all patent technology classes, and the extent of localization has a positive relationship with the level of technology measured by R&D investment. Finally, the extent of localization is stronger for establishments that are more productive in terms of both the number of patents and the number of citations received, i.e., quantitatively and qualitatively. These results indicate that geographical proximity is important for knowledge spillover, particularly for knowledge-demanding establishments.

Keywords: Knowledge spillover; Agglomeration; Micro-geographic data
JEL codes: R11

---

* 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 access to the micro-database on Japanese establishments and the Center for Spatial Information Science, University of Tokyo for providing the geocoding service. We gratefully acknowledge financial support from the Japan Society for the Promotion of Science (Nos. 24530506, 25220502, 25380275, 25780181, 26245037, 26285051, 15K01217).
†Graduate School of Simulation Studies, University of Hyogo, Computational Science Center Building, 7-1-28 Minatojima-minamimachi, Chuo-Ku Kobe 650-0047, 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 spillovers are crucial for innovation and productivity growth. Beginning with Marshall (1890), it has been widely recognized that geographical proximity enhances knowledge spillovers, which cause industrial agglomeration. Several prior 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 proxies for knowledge spillovers. This implies that knowledge-demanding establishments agglomerate more than do other types of establishments.

The role of knowledge spillover as an agglomeration force has long been empirically examined. Rosenthal and Strange (2001) find that the intensity of R&D investment is positively related to the extent of industrial agglomeration. Ellison, Glaeser, and Kerr (2010) find that the intensity of citation relationships between industries positively relates to the extent of co-agglomeration between pair industries. These studies, however, identify agglomeration determinants with industry-level estimations. Thus, establishment-level heterogeneity within an industry disappears due to aggregation. The role of knowledge spillovers, however, may differ across establishments within an industry, depending on the extent of their demand for knowledge spillovers. Thus, the location pattern may also differ by establishments within an industry.

A seminal paper by Carlino et al. (2012) examines the localization of knowledge-demanding establishments. They use the address information of R&D laboratories from the Directory of American Research and Technology and find that R&D laboratories are significantly localized for most industries. They also identify core clusters of R&D laboratories in the U.S. and localized spillovers within the identified clusters. However, their focus is revealing the local structure of R&D laboratories' localization rather than the overall location pattern of knowledge-creating activities across a country.

With this background, this study investigates the localization of Japanese knowledge-creating activities by constructing an establishment-level database drawn from the entire patent database in Japan. A convention in the Japanese patent application allows us to construct an establishment-level database. That is, inventors in Japan register the address of the establishments to which they belong as the "inventor's address". We can detect 74,452 patent-creating establishments, which covers establishments in all regions. This enables us to capture the pattern of localization of knowledge-creating activities across an entire country.

In addition to the global pattern of localization across Japan, our sample encompasses all industries. The industrial agglomeration literature mainly focuses on manufacturing industries. As we will show, establishments of non-manufacturing industries also intensively create knowledge. However, the role of knowledge spillovers in industrial agglomeration in non-manufacturing industries has been scarcely investigated. This study covers both manufacturing and non-manufacturing industries in the analysis.

Furthermore, using a patent database enables us to address differences in localization patterns within patent-creating establishments, depending on the heterogeneity of their demand for knowledge spillovers. If knowledge spillovers are determinant of agglomeration, establishments that demand more knowledge spillovers should be more localized. The technology level of patents that establishments apply for represents the extent of their demand for knowledge spillovers. High-technology invention ought to require more knowledge spillovers. From another perspective, establishments that create more patents or higher-quality patents require more knowledge spillovers. We examine differences in localization patterns depending on the above-mentioned heterogeneity.

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 all pairs of patent-creating 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 distribution generated by a random allocation of patent-creating establishments' locations to all potential sites. For the potential sites of patent-creating establishments, we use all establishments of all industries in Japan from micro-data in the Establishment and Enterprise Census.

We obtain the following results. First, the locations of patent-creating establishments are significantly localized at the 5% level, with a localization range of approximately 80 km. Furthermore, patent-creating establishments are more localized within an industry. Second, localization is found for all patent technology classes, and the extent of localization has a positive relationship with the level of technology, as measured by R&D investment. Finally, the extent of localization is stronger for more productive establishments in terms of both the number of patents created and the number of citations, i.e., quantitatively and qualitatively. This implies that productive establishments require more external knowledge from other establishments. These findings suggest that knowledge spillovers are important determinants of economic agglomeration, particularly for knowledge-demanding establishments.

The remainder of this paper is organized as follows. In the next section, we introduce the dataset and the 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 robustness checks. Section 5 focuses on the differences in the extent of demand for knowledge spillovers across patent-creating establishments. Finally, Section 6 concludes.

2. Data

We utilize the Institute of Intellectual Property (IIP) patent database (Goto and Motohashi, 2007), which includes Japanese patent publications (the Patent Gazette) over more than two decades. This database includes basic patent information, such as patent IDs, publication dates, names and addresses of applicants, and names 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 construct an establishment-level database of all 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 in the Japanese patent application where inventors register the address of the establishments to which they belong as the "inventor's address" (Inoue, Nakajima, and Saito, 2013).

Here, we describe the algorithm used to identify the patent-creating establishments from our patent database, following Inoue, Nakajima, and Saito (2013). First, firms are identified by the name and address of the applicants. Here, we define the firm as an applicant whose name includes the term "company limited," or "*kabushikigaisha*" in Japanese. This definition simultaneously excludes 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 obtain the following information. Table 1 provides the summary of the dataset, which includes 1,967,361 patents. A total of 1,189,262 patents are applied for by the firms with identified patent-creating establishments. The number of firm applicants is 56,592, and the total number of patent-creating establishments is 74,452.

[Table 1 here]

Figure 1 shows the map of patent-creating establishments identified by our methodology. As the map shows, the identified patent-creating establishments span Japan.

[Figure 1 here]

Furthermore, our analysis requires the potential sites of patent-creating establishments. We assume that patent-creating establishments can be located at any site where the establishments of all industries are located. To obtain information on the locations of establishments of all industries, we use micro-data from the Establishment and Enterprise Census. This database includes the address, the number of employees, and the industry code information of each establishment. Then, we convert the establishments' address into a latitude-and-longitude format.[1] The number of establishments in the data is 5,722,559.

3. Empirical strategy

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 all pairs of patent-creating establishments; then, we compare the distribution with counterfactual distributions generated by the random assignment of locations from potential sites.

3.1. K-density approach

We now describe in detail the procedure for measuring the localization of patent-creating establishments using the K-density approach. First, we estimate the distribution of bilateral distances between all pairs of patent-creating establishments.

Let n be the number of establishments that have applied for at least one patent, and we have $n(n-1)/2$ pairs of the patent-creating establishments. Next, let d_{ij} be the great circle distance between the pair of 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.

3.2. Counterfactual distribution and statistical testing

Overall economic activities (i.e., all establishments) have a tendency to agglomerate. To precisely detect the localization of patent-creating establishments, we need to control for the localization of overall economic activity. To do so, we generate a counterfactual distribution of locations for patent-creating establishments where establishments randomly choose their locations from all potential sites as a reference. Then, we

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

compare the actual localization of patent-creating establishments with the counterfactual localization. We consider all the sites of the establishments of all industries as potential sites for patent-creating establishments.

To test the localization of patent-creating establishments, we construct a two-sided confidence interval. Specifically, we randomly choose n sites from the potential sites and estimate the K-density in the counterfactual situation. By iterating this trial 1,000 times, we can construct the "global confidence bands" introduced by Duranton and Overman (2005), i.e., an upper confidence band $K^U(d)$ and a lower confidence band $K^L(d)$. Of the 1,000 randomly drawn K-densities, 95% lie below the upper band $K^U(d)$, and the other 95% lie above the lower band $K^L(d)$ over the entire distance range, which, in our case, is 0–180 km.

If $\widehat{K}(d) > K^U(d)$ for at least one $d \in [0, 180]$, patent-creating establishments are defined as globally localized at the 5% level. To discuss the extent of localization, following Duranton and Overman (2005), we define the extent of localization as follows:

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

4. Baseline results and robustness check

4.1. Baseline results

Figure 2 provides heat maps of overall economic activities (panel (a)) and knowledge-creating activities (panel (b)). In panel (a), the color of each parcel shows its share of overall establishments in Japan. Blue shows a smaller share of establishments, and red shows a larger share of establishments. This figure shows that economic activities are broadly distributed across Japan, but there are variations in the share. In panel (b), the color of each parcel shows its share of knowledge-creating establishments. Comparing panel (b) to panel (a), we see that knowledge-creating activities are heavily concentrated in narrower areas.

[Figure 2 here]

To measure the above-mentioned concentration, Figure 3 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 patent-creating establishments to be significantly localized in the 0–80 km range[2] relative to the overall establishment locations.

[Figure 3 here]

4.2. Robustness: Controlling for industrial localization

Our baseline analysis uses the overall establishments of all industries as potential sites for knowledge-creating establishments. One may be concerned that differences in the location distributions across industries may affect the results. For example, if there are many knowledge-creating establishments in specific industries that have a strong tendency to localize, our results capture the localization not of knowledge-creating establishments but of specific industries. To control for this industry heterogeneity in localization tendencies, we conduct a within-industry analysis. In this analysis, we restrict the sample by industry. Then, we test the localization of the knowledge-creating establishments in the industry via a comparison to the overall establishments in the same industry.

To do so, we need to identify the industry to which each patent-creating establishment belongs. We use industry information from a large-scale, firm-level database provided by Tokyo Shoko Research (TSR). The TSR data cover 826,169 Japanese firms, which is over half of the total firms in Japan. This database includes the industry code for each firm. Merging the TSR data with the patent database by the firm's name and address information, we obtain the applicant firms' industry information. Note that there are significant numbers of patent-creating establishments that belong to the non-manufacturing sector. In our dataset, 33% of knowledge-creating establishments are in non-manufacturing industries.

Our K-density estimator is modified for the within-industry analysis. Let S^I be a set of establishments that have applied for at least one patent and belong to industry $I \in \mathfrak{J}$, where $\mathfrak{J}$ represents a set of industries. Let n_I be the number of patent-creating establishments in industry I. Similarly, let d_{ij} be the great circle distance between establishments i and j in set S^I. The estimator of the density of bilateral distances at any point d for industry I is

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

---

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

For the counterfactual distribution, we consider the site of every establishment in industry $I \in \mathfrak{I}$ as a potential site for patent-creating establishments in the industry.

In the estimation, we use a two-digit industry code to ensure a sufficient sample size, and we restrict the sample to industries that have more than ten patent-creating establishments. As a result, we find in 83.6% industries (61 of 73 industries) that patent-creating establishments are significantly localized at the 5% level. The fact that knowledge-creating establishments are more localized within the same industry indicates that our baseline result is not caused by specific industries with strong agglomeration.

Regarding the difference between manufacturing and non-manufacturing industries, 91% of industries are localized in manufacturing, and 79% of industries are localized in non-manufacturing sector. As Nakajima, Saito, and Uesugi (2012) show, service industries tend to be more localized in Japan. Therefore, the counterfactual distributions in our analysis are also more localized in service industries. The difference in the counterfactual distribution may partly cause the lower share of localized non-manufacturing industries. Therefore, the lower share of localized non-manufacturing industries does not necessarily imply a lower role of knowledge spillover as an agglomeration determinant in the non-manufacturing sector than in the manufacturing sector.

5. Heterogeneity within patent-creating establishments

We now consider differences in the demand for knowledge spillovers within patent-creating establishments. In the baseline analysis, we treat each patent-creating establishment as homogeneous. However, the establishments are heterogeneous in terms of the extent of their demand for knowledge spillovers. Thus, the extent of localization may vary across establishments depending on their demand for knowledge spillovers. In this section, we analyze differences in the tendency for localization across establishments with varying demand for knowledge spillovers.

5.1. Differences by patent technology classes

The demand for knowledge spillovers may vary with the patent technology class of published patents. Inventions in higher technology are thought to require many knowledge spillovers. Then, establishments that publish higher technology classes should be more localized to pursue more knowledge transfers. To grasp this difference across patent technology classes, we conduct 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 patent-creating establishments in the patent technology class A. Similarly, let d_{ij} be the great circle distance between establishments i and j in 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 sites of all establishments as potential sites for patent-creating establishments in the patent technology-class $A \in \mathfrak{A}$.

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

Figure 4 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 similar to industrial localization in the manufacturing industry (Duranton and Overman, 2005; Nakajima et al., 2012).[3]

[Figure 4]

Next, we investigate in detail the differences in the extent of localization among patent technology classes. 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, and cosmonautics (IIP B64). 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, and poultry and fish processing (IIP A22). These tables suggest that establishments in higher-technology industries may require more advanced knowledge transfers.

[Tables 2 and 3]

To examine the relationship between localization and the demand for knowledge spillovers more precisely, we define level of the technology level by R&D investments. In general, a higher technology class requires more investments in invention. Therefore, we define the level of technology as the portion of R&D

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

investment over total sales. That is, the technology classes that require more investment for invention can be defined as high-class technologies. Using the Basic Survey of Japanese Business Structure and Activities, we calculate the R&D investment share over total sales in each patent technology class. For the detailed process of data construction, see Appendix A.

Figure 5 shows the relationship between the degree of localization and extent of R&D investments. The horizontal line refers to R&D investments, and the vertical line refers to the degree of localization on each technology class. The solid line represents the linear fitted line. We can see a clear positive relationship between them. With the increase in the R&D investment share, the degree of localization is increased.

[Figure 5]

These results imply that the establishments publishing patents in higher technology classes measured by R&D investment are more localized to acquire knowledge spillovers.

5.2. Differences by establishment productivity

Next, we consider other measures of demand for knowledge spillovers. In this subsection, we consider the heterogeneity of establishments in terms of their productivity. More productive establishments may require external knowledge from other establishments. We consider two measures of productivity: the number of patent publications and the number of citations received. The number of patent publications measures establishments' patent productivity in terms of quantity, and the number of patent citations received measures it in terms of quality. We modify the baseline analysis by weighting by these productivity measures for each establishment.

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 productivity for establishment i. We consider the two measures for establishment productivity: the number of patents created and the number of total citations received.

Figure 6 (a) shows the results of quantitative productivity weighted by the number of patents. The solid line in the figure represents the K-density weighted by the number of patents created, and the dashed lines represent the global confidence bands. For every distance within the 0–85 km range, the K-density is above the upper global confidence band. Thus, we consider the location to be localized in the 0–85 km range, even if we weight each establishment by the number of patents created.

We also show the baseline K-density (without weighting) as the dotted line in the figure. Then, we clearly find that the weighted K-density is above the unweighted K-density within a 0-50 km range.[4] We can compare the extent of localization between weighted and unweighted results by Γ. The weighted result ($\Gamma = 0.221$) is larger than the baseline result ($\Gamma = 0.163$). These results show that establishments that publish more patents are more localized, implying that establishments that require more knowledge transfers are more localized or that the greater concentration of establishments benefits the productivity of each establishment located in the area through larger knowledge transfers.

[Figure 6 here]

Next, we focus on productivity in terms of quality (Figure 6 (b)). The solid line in the figure represents the K-density weighted by the total number of patent citations, with the dashed lines representing the global confidence bands and the dotted line representing the baseline K-density. We obtain a result similar 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 an unweighted one. The estimated Γ in the weighted result ($\Gamma = 0.245$) is larger than that in baseline result ($\Gamma = 0.163$). Even if we use patent quality as a measure of establishment productivity, productive establishments are more localized.

6. 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 within the range of 0-80 km. Second, even within an industry, knowledge-creating establishments are significantly more localized than overall establishments in the industry. Third, localization was found for all patent technology classes, and the extent of localization has a positive relationship with the level of technology. Finally, the degree of localization is stronger in more productive establishments in terms of both quantity and quality. These results indicate that geographical proximity is important for all knowledge-creating establishments, particularly for more knowledge-demanding establishments. These findings suggest that knowledge spillovers are an important determinant of the agglomeration of economic activities, particularly for knowledge-demanding establishments.

---

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

Appendix A. Classification of technology by R&D investments

We define high- and low-technology classes by their R&D investment share over total sales. In this appendix, we explain how to calculate the R&D investment share of total sales in each technology-class.

R&D investment information can be obtained from the Basic Survey of Japanese Business Structure and Activities. It covers firms with more than 50 employees and capital stock of over 30 million yen. This database includes each firm's sales, R&D investment, and industry code (JSIC). From this dataset, we create R&D intensity by dividing the aggregated R&D investment and the aggregated sales for each three-digit level industry code.

Next, we identify the correspondence between three-digit level industry codes and patent technology classes. To do so, we use patent data and a large-scale firm database. We merge firms in the patent dataset with firms in the Tokyo Shoko Research (TSR) firm database. The TSR covers 826,169 Japanese firms, which is more than half of the total number of firms in Japan. This database includes the industry code (JSIC) for each firm. Merging the TSR data with the patent database by the firm's name and address information, we obtain the applicant firm's industry information for each patent.

Using the merged patent database, we calculate the composite of the industry code for each technology class. Focusing on the patents of one technology class, we calculate the number of patents per industry to which each applicant firm belongs.

Finally, by using information on industry-level R&D investment shares and calculating the weighted average of the R&D investment shares based on the abovementioned industry composite, we obtain the patent-technology class levels using R&D investment shares. The results are available upon request.

References

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 knowledge-creating establishments	74,452
Number of overall establishments	5,722,559

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	Fertilizers; Manufacture Thereof	0.096
9	B22	Casting; Powder Metallurgy	0.096
10	B02	Crushing, Pulverizing, or Disintegrating; Preparatory Treatment of Grain for Milling	0.105

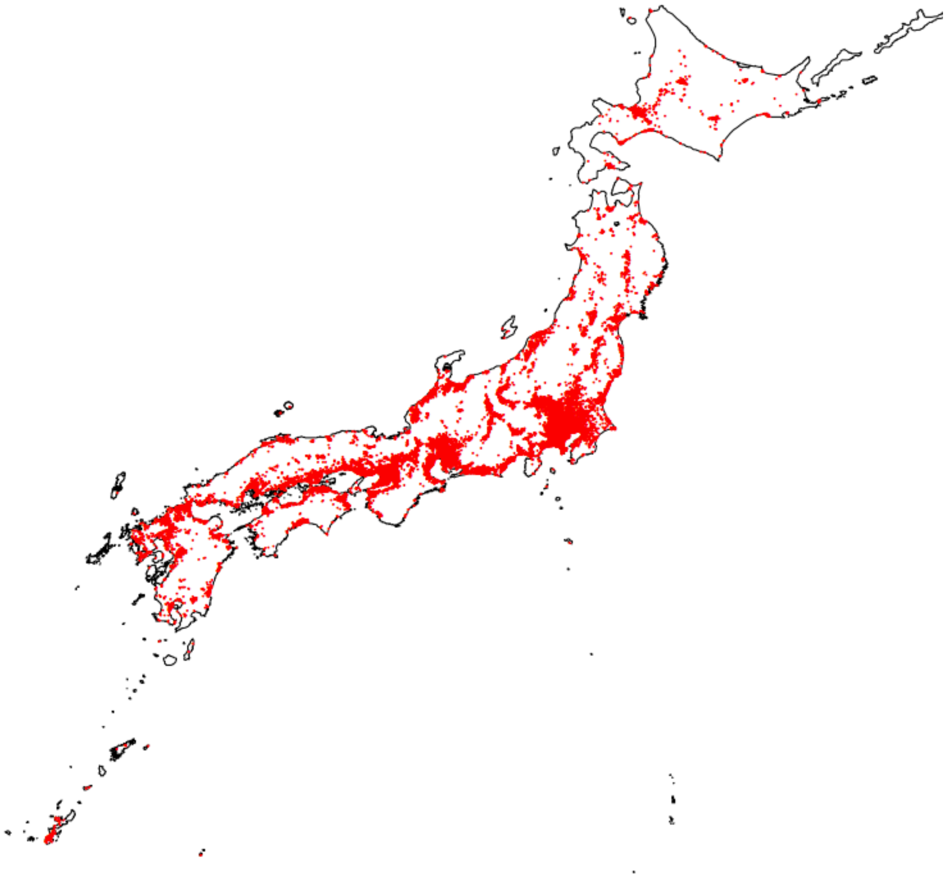
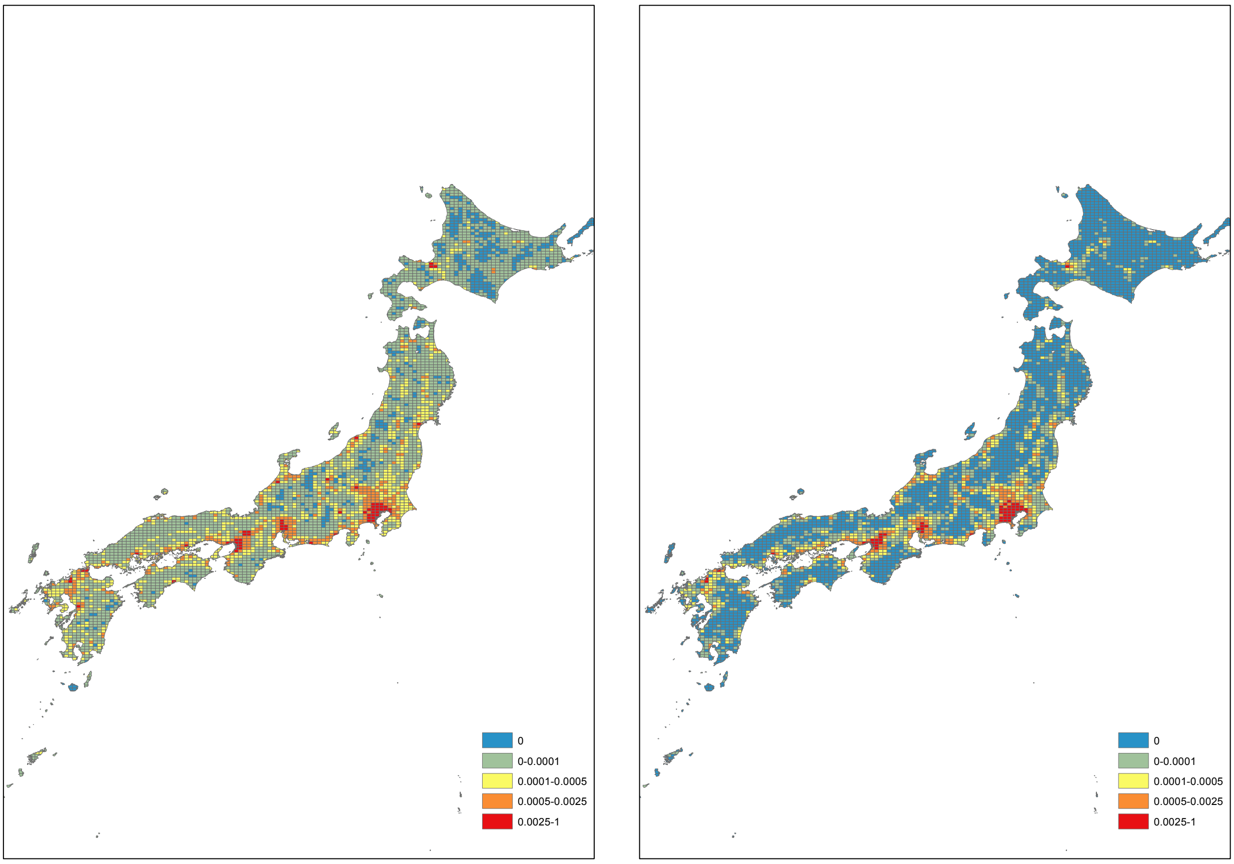


Figure 1: Map of patent-creating establishments



Panel (a): Overall economic activities
Panel (b): Knowledge-creating activities

Figure 2: Map of overall economic and knowledge-creating activities

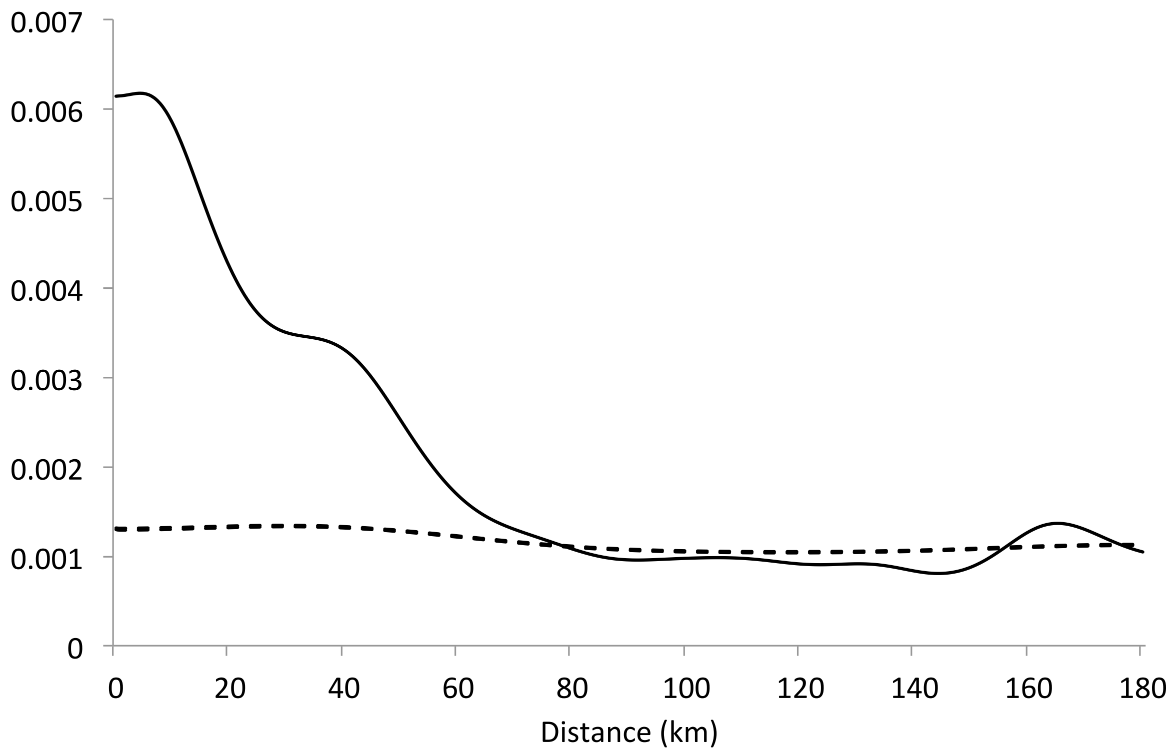


Figure 3: Result of the baseline analysis

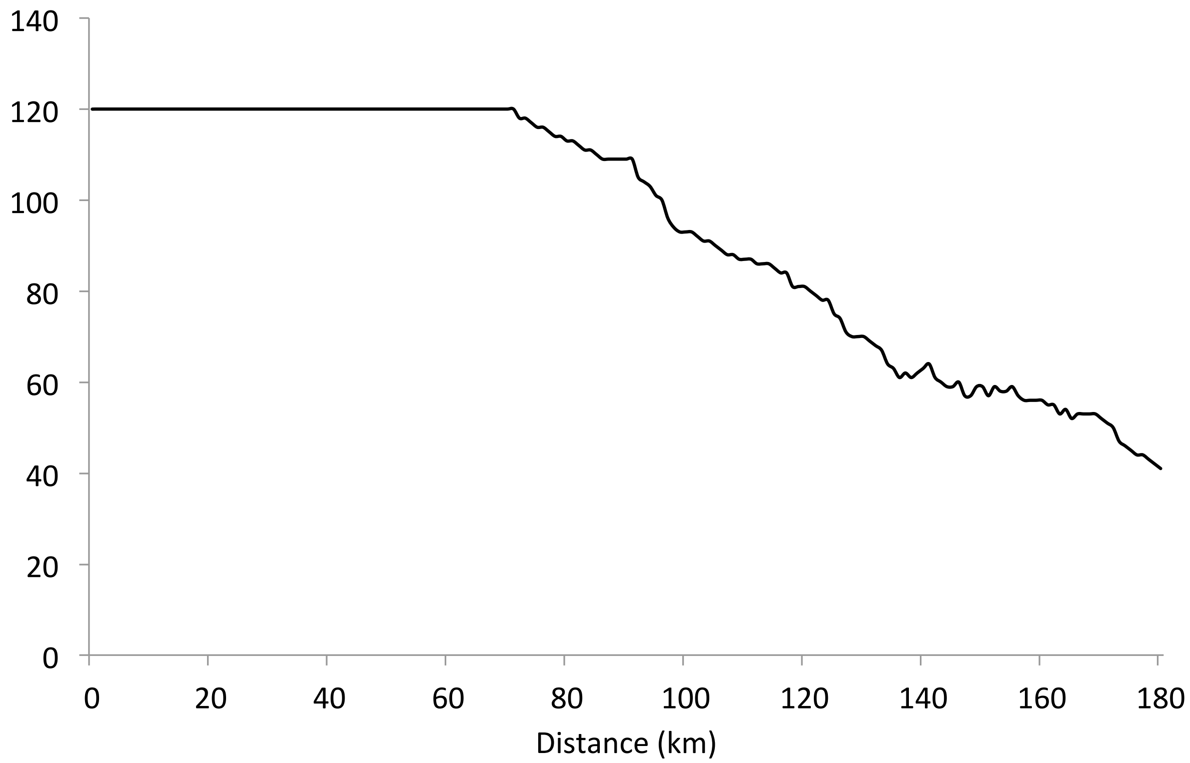


Figure 4: Number of localized patent classes by distance

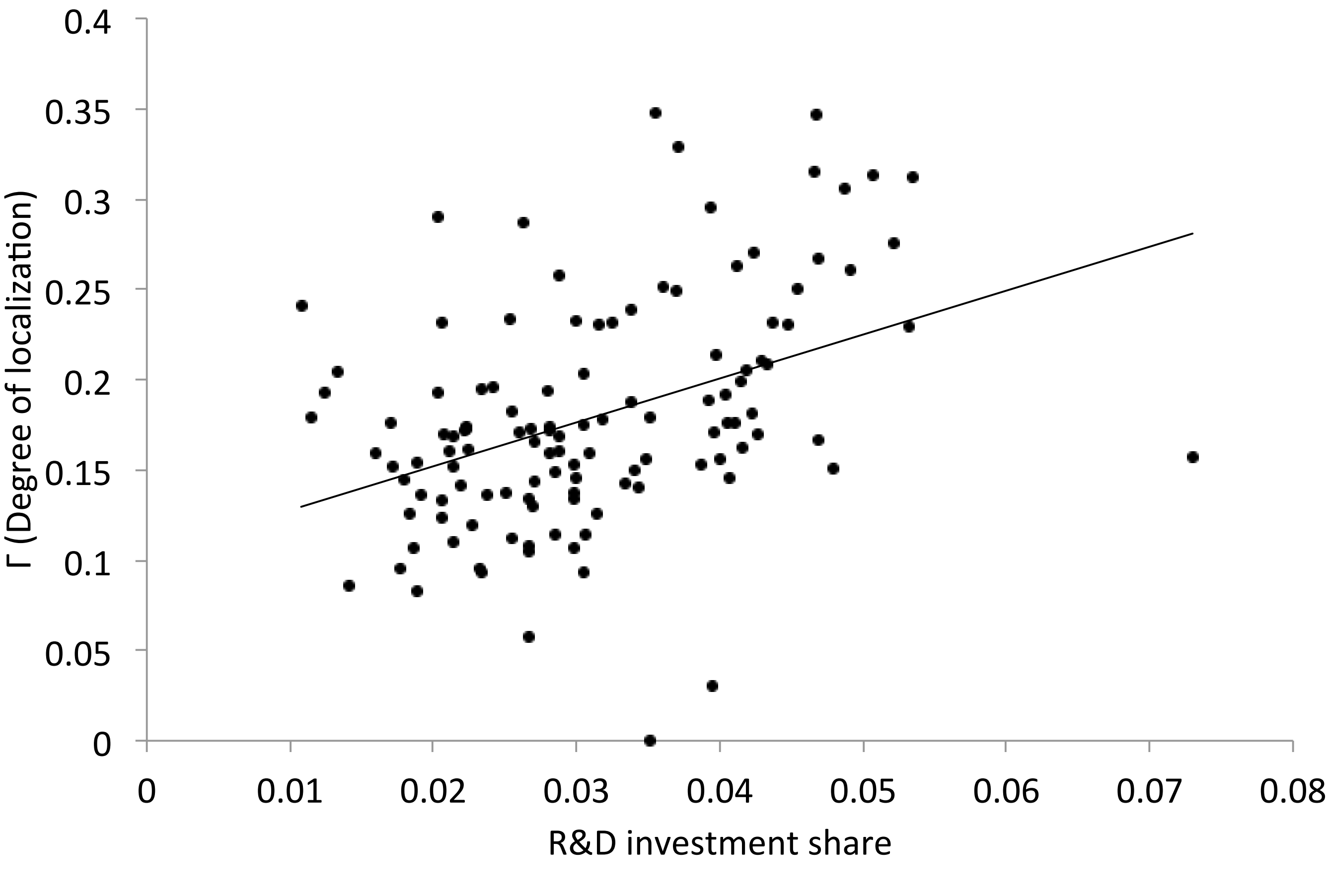
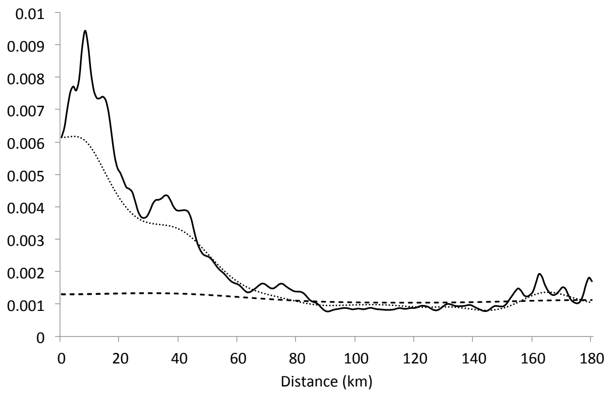
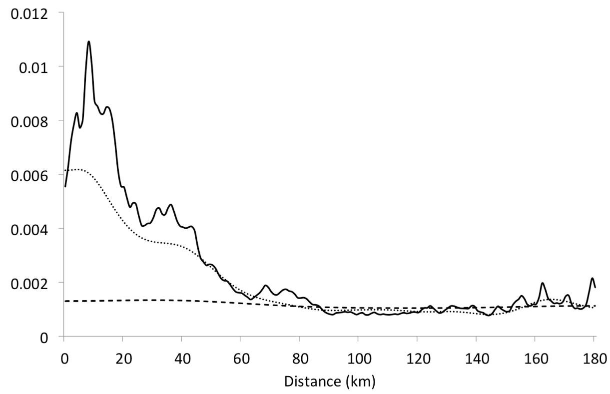


Figure 5: Relationship between R&D investment share and degree of localization



(a): Weighted by number of patents



(b): Weighted by number of patent citations

Figure 6: Results weighted by establishment creativity