



RIETI Discussion Paper Series 10-E-030

Measuring Economic Localization: Evidence from Japanese firm-level data

NAKAJIMA Kentaro

Institute of Economic Research, Hitotsubashi University

SAITO Yukiko

Fujitsu Research Institute

UESUGI Iichiro

RIETI



Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry

<http://www.rieti.go.jp/en/>

Measuring Economic Localization: Evidence from Japanese firm-level data

Kentaro NAKAJIMA

Institute of Economic Research, Hitotsubashi University

Yukiko SAITO

Fujitsu Research Institute

Iichiro UESUGI

Institute of Economic Research, Hitotsubashi University

Abstract

This paper examines the extent of localization in Japan's manufacturing sector using a unique firm-level dataset on the geographic location of firms. Following the point-pattern approach proposed by Duranton and Overman (2005), we find the following. First, approximately half of Japan's manufacturing industries can be classified as localized and the number of localized industries is largest for a distance of 40 km or less. Second, several industries in the textile mill products sector are among the most localized, which is similar to findings for the UK. This suggests that there exist common factors across countries that determine the concentration of industrial activities. Third, the distribution of distances between entrant (exiting) firms and remaining firms is, in most industries, not significantly different from a random distribution. The results by Duranton and Overman (2008) for the UK and our results for Japan suggest that most industries neither become more localized nor more dispersed over time.

Key words: Micro-geographic data; Economic geography

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 do not present those of the Research Institute of Economy, Trade and Industry.

We would like to thank Chihiro Shimizu for providing indispensable help geocoding firms' locations. We also would like to thank Tomoya Mori, Suminori Tokunaga and participants of the JEA meeting at Senshu University and seminars at the Bank of Japan, Hitotsubashi University, Kyoto University, RIETI, Tohoku University, the University of Tokyo, and Yokohama National University for helpful comments.

1 Introduction

Economic activities and industries are not uniformly distributed but tend to be localized (i.e., concentrated over and above overall economic activity) in certain areas. The agglomerations of automobile assemblers and suppliers in places such as Detroit in the United States and Toyota City in Japan are famous examples.

As laid out in uncountable theoretical models developed by economists, localization of industries arises through Marshallian externalities such as knowledge spillovers, labor pooling, and cost reductions. Along with the theoretical literature, many empirical studies have focused on how to accurately measure localization of industries. Procedures that have been proposed to measure spatial localization include the Gini, Isard, Herfindahl, and Theil indices. These “first generation” indices have been followed by a “second generation” of indices which seek to measure overall industry concentrations. The study by Ellison and Glaeser (1997), which is the pioneering work of this second generation, compares the degree of the spatial concentration of employment in a given sector with the degree of concentration that would arise if all plants in this sector were located randomly across locations. Maurel and Sedilló (1999), Devereux, Griffith, and Simpson (2004), and Mori, Nishikimi, and Smith (2005) have presented similar localization measures which control for the tendency of industrial activity to concentrate as a whole. There have been a number of applications of these indices using data from a variety of countries.¹

However, these indices are still susceptible to several problems. First, they limit the analysis only to one administrative spatial unit, be it city, prefecture, or region. This results in wildly different spatial scales after aggregation since administrative spatial scales are often very diverse in population and in size. For example, Japan’s largest city in terms of area (Takayama City) is 2179.35 square kilometers in size, while the smallest one (Warabi City) comprises merely 5.10 square kilometers. Second, even when spatial units with exactly the same geographic area are used for analysis, these indices are not robust in the way they define the shape and size of each spatial unit (Modifiable Areal Unit Problem). Finally, these indices disregard the distance between spatial units, that is, they treat adjacent spatial units as exactly the same as those in opposite ends of the country. This problem results in a downward bias to the detection of

¹Applications using Japanese data include Mori, Nishikimi, and Smith (2005), and Tokunaga and Akune (2005).

localization when spatial units with dense population of firms are located closely with each other.

Against this background, the aim of this study is to examine the extent of localization in Japan, using the spatial point-pattern approach proposed by Duranton and Overman (2005) to alleviate the above problems. Given that the empirical literature on localization utilizing firm-level location information is still limited, our results for Japan may help to understand patterns of localization more generally. Our findings can be summarized as follows.

First, we find that about half of the 561 four-digit manufacturing industries can be classified as localized and that the number of localized industries is largest for distances of 40 km or less. In addition, aggregating the degree of localization across all industries for each distance, we find that the aggregated localization index is highest at the shortest end of distances. Second, we aggregate the degree of localization across all distances for each industry and find that several four-digit industries within the textile mill products sector are among the most localized. These patterns of localization in Japan resemble those in the UK, which suggests that there exist common factors across countries determining the localization of industrial activities. Further, we repeat our calculations for a sub-sample of small, single-establishment firms in order to correct for any possible bias resulting from the use of firm-level data. The empirical regularities we obtain are qualitatively similar to those for the entire sample.

Finally, using observations on entrant and exiting firms, we examine the dynamics of industry localization by analyzing the distances between entrant (exiting) firms in an industry on the one hand and remaining firms on the other. In most of the industries, the distribution of distances between entrant (exiting) firms and remaining firms is not significantly different from a random distribution. This suggests that most industries neither become more localized nor more dispersed over time and the location distributions are stable.

The remainder of the paper is organized as follows. Section 2 describes our firm-level dataset and the methodology we employ to measure industry localization. Section 3 then provides the empirical results. Section 4 examines the dynamics of localization by focusing on entrant and exiting firms. Finally, Section 5 concludes.

2 Data and Empirical Approach

2.1 Data

We employ a unique and massive dataset of Japanese firms compiled by Tokyo Shoko Research (TSR). The TSR dataset covers 826,169 firms, which is equivalent to over half of all incorporated firms in Japan, and provides information on firms' location, two-, three- and four-digit industry classification code,² and number of employees. We geocode the firm location data using the CSV Address Matching Service provided by the Center for Spatial Information Science, University of Tokyo.³ Following previous studies on industry concentration, we focus on the manufacturing sector, which reduces the sample used for our analysis to 143,628 firms. The dataset was purchased from TSR only once, at the end of March 2006, so that we only have a cross-section and no longitudinal observations. However, the TSR dataset does contain information that allows inferences on the dynamics of industry concentration, namely a firm's establishment year and a dummy for firms which existed at the end of March 2006 but disappeared in the following year, 2007. We use these variables to identify entrant and exiting firms, respectively.

Two caveats are in order regarding the TSR dataset. First, the dataset is not a census and thus does not cover all manufacturing firms in Japan. The potential bias caused by the fact that the dataset is not a census depends on the methodology employed by TSR in selecting firms to be included in the database. If the TSR firm selecting strategy is regionally biased, (e.g., firms in urbanized area are more likely to be chosen) the localization indices calculated based on the dataset are also biased. Second, the dataset does not consist of establishment-level but of firm-level data. The potential bias resulting from the use of firm-level data could go in either direction. If non-headquarter establishments in an industry are concentrated in a particular location, then the use of firm-level data will fail to pick up such agglomeration and result in an underestimation of the degree of concentration in the industry. On the other hand, if headquarters in an industry are all located in a small number of highly confined areas, then the use of firm-level data will exaggerate the level of concentration in the industry. We examine this issue by aggregating the number of firms in each industry at the municipal level and calculate the Ellison and Glaeser (EG) and Mori, Nishikimi, and Smith (MNS) indices. We then compare the results using data from the *Census of Manufactures*. The correlation and ordered

²Industry classifications follow the Japanese Standard Industry Classification (JSIC).

³<http://newspat.csis.u-tokyo.ac.jp/geocode/>

correlation coefficients are quite high, but localization indices calculated using firm-level data indicate a higher degree of localization. In sum, using firm-level data results in an upward bias in the detection of localization, but the high correlation and ordered correlation indices suggest that our results are robust. See Appendix A for details. Moreover, as another robustness check of our results, we limit the sample to small firms that have only one establishment and compare the results with the full sample results.

To illustrate the different patterns found in different industries, Figures 1(a) to 1(d) show the geographical distribution across Japan of firms in four industry classifications, namely

Figures 1(a) to 1(d)

Gelatin and Adhesives (JSIC1794), Miscellaneous Seafood Products (JSIC0929), Fabricated Plate Work and Sheet Metal Work (JSIC2543), and Fabric Mills, Woven Woolen and Worsted (JSIC1143), with each dot representing the location of a firm in the industry. The maps show that the Gelatin and Adhesives industry (JSIC1794) appears to be concentrated in the Tokyo area, whereas the Miscellaneous Seafood Products industry (JSIC0929) is dispersed along the coast. The Fabricated Plate Work and Sheet Metal Work industry (JSIC2543) can be found mainly in the major MEAs, including Tokyo, Osaka, Nagoya, and Fukuoka, and the distribution coincides with the location distribution of the manufacturing sector as a whole. In contrast, the Fabric Mills, Woven Woolen and Worsted industry (JSIC1143) appears to be concentrated in two distinct locations.

2.2 Empirical Approach

This subsection provides an overview of our empirical approach.⁴ Our empirical analysis consists of three steps. First, we calculate the pairwise distances between all firms in an industry and estimate a kernel density function of the distance distribution. Second, in order to implement statistical tests, we consider the counterfactual that all firms in the industry randomly choose their location and simulate counterfactual location distributions. Third, based on the counterfactual random location distributions, we construct confidence interval bands and test whether an industry can be considered to be localized.

⁴For more details, see Duranton and Overman (2005, 2008).

2.2.1 Kernel densities

We begin by estimating the density distribution of pairwise distances. For each industry A, with n firms, we calculate the Euclidean distance between every pair of firms in the industry. Thus, we obtain $nC_2 = \frac{n(n-1)}{2}$ bilateral distances in industry A. We then estimate kernel-smoothed distributions (K -densities hereafter) of these pairwise distances. The estimator of the density at distance d is

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

where d_{ij} is the Euclidean distance between firm i and j , h is the bandwidth, and f is the kernel function.⁵

2.2.2 Counterfactuals

We construct counterfactual random location distributions as benchmarks for the statistical tests. For the counterfactuals, we assume that the set of all existing sites currently occupied by manufacturing firms in the TSR database constitutes the set of all possible locations for any manufacturing firm and that firms randomly choose their location from these potential sites. In each trial, we randomly draw sites of the same number as the number of firms in the industry, then calculate the pairwise distances of the sites and estimate the K -density. This procedure ensures that we control for the overall patterns of concentration in the manufacturing sector as a whole. Following Duranton and Overman (2005, 2008), we run 1,000 trials for each industry.

2.2.3 Identifying localization and dispersion

We examine whether an industry is localized or dispersed by comparing the actual K -density with that of the counterfactual distribution. Intuitively, if we observe a higher actual K -density at short distances than the density of randomly drawn distributions, we define the industry as localized. Similarly, if we observe a lower actual K -density at short distances than the density of randomly drawn distributions, we define the industry as dispersed. Specifically, we construct two-sided confidence intervals containing 95 % of the randomly drawn K -densities. Following

⁵Following Silverman (1986), we use a Gaussian kernel with optimal bandwidth.

Durantón and Overman (2005, 2008), we construct *global confidence bands*⁶ between 0–180 km⁷ such that 95 % of the randomly drawn K -densities lie above the lower band and another 95 % of the randomly drawn K -densities lie below the upper band.

Using this procedure, we obtain the upper global confidence band $\bar{K}_A(d)$ and the lower global confidence band $\underline{K}_A(d)$ of industry A. If $\hat{K}_A(d) > \bar{K}_A(d)$ for at least one $d \in [0, 180]$, industry A is defined as globally localized at the 5 % confidence level. On the other hand, if $\hat{K}_A(d) < \underline{K}_A(d)$ for at least one $d \in [0, 180]$, and industry A is not defined as localized, we define industry A as globally dispersed. We can also define an index of localization,

$$\Gamma(d) \equiv \max(\hat{K}_A(d) - \bar{K}_A(d), 0), \quad (2)$$

and an index of dispersion,

$$\Psi(d) \equiv \begin{cases} \max(\underline{K}_A(d) - \hat{K}_A(d), 0) & \text{if } \sum_{d=0}^{d=180} \Gamma(d) = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

For illustration, we examine the K -densities and corresponding two-sided confidence intervals of the four previously introduced industries in Figures 2(a) to 2(d). The solid lines in the figures represent the K -densities.

Figures 2(a) to 2(d)

The K -density of the Gelatin and Adhesives industry (JSIC1794) is higher at short distances, while that of the K -density of the Miscellaneous Seafood Products industry (JSIC0929) gets gradually higher at larger distances. Thus, the former industry seems to be more localized than the latter at short distances. The dashed lines in Figures 2(a) to 2(d) are the global confidence bands. Figure 2(a) for the Gelatin and Adhesives industry (JSIC1794) provides an example of a localized industry. For every distance within the range of 0–80 km, the K -density is above the upper global confidence bands, which provides evidence that this industry is localized. On the

⁶Durantón and Overman (2005) also define local confidence levels for each distance, so called *local confidence bands*. But as they pointed, the local confidence intervals only provide statements for given level of distance. Following Durantón and Overman (2008), to focus on the statements about deviations over the entire range of distances, we mainly use *global confidence bands* rather than *local confidence bands* for our analysis.

⁷The median of the pairwise distances of all manufacturing firms in Japan is about 400 km. However, in order to make our results comparable to those obtained by Durantón and Overman (2005, 2008) for the UK, where the median distance is 180km, we set our threshold value to 180 km. The results are similar to those we obtain when setting the threshold value to 400 km.

other hand, Figure 2(b) for the Miscellaneous Seafood Products industry (JSIC0929) provides an example of a dispersed industry. For every distance within the range of 0–180 km, the K -density is below the lower global confidence band and never above the upper global confidence band. Thus, this industry exhibits global dispersion within 180 km.

Figure 2(c) shows the K -density and the global confidence bands for the Fabricated Plate Work and Sheet Metal Work industry (JSIC2543). This industry is localized in the 50–80 km range, but the K -density falls between the confidence bands for most other distances. The location pattern of this industry with concentrations in the Tokyo, Osaka, and Aichi MEAs and along the Pacific industrial belt,⁸ as shown in Figure 1(c), is similar to the pattern for Japan’s manufacturing sector as a whole, and the K -density also resembles that of manufacturing overall. Next, Figure 2(d) shows the K -density and the global confidence bands for the Fabric Mills, Woven Woolen and Worsted industry (JSIC1143). The K -density of the industry has two peaks, with the second one located around a distance of 150 km. This kind of pattern suggests that there are two different areas of concentration at a distance of about 150 km, which is confirmed by a look at Figure 1(d): the largest concentration of firms in this industry can be found in the Bishu area of Aichi prefecture, with another concentration in the Senshu area of Osaka prefecture. The distance between Bishu and Senshu is around 150 km.

3 Results

3.1 Baseline results

This section presents the results. We use 561 four-digit JSIC code manufacturing industries, each of which has 10 or more firms. First, we examine how many industries are localized or dispersed. Figure 3 depicts the number of localized and dispersed industries for each distance d .

Figure 3

The solid line represents the number of localized industries. For short distances (0–40 km), between 267 and 276 industries, or about half of all industries, are localized. For medium-range distances (40–100 km), the number of localized industries falls rapidly, with a small bump around 110 km. Turning to the number of dispersed industries, which is represented by the dashed line, this is stable over the entire range of distances from 0–180 km. These patterns of the number

⁸As is well known, Japan’s manufacturing industries are mainly concentrated in these areas.

of localized and dispersed industries are quite similar to the results for the UK obtained by Duranton and Overman (2005).

Second, in addition to measuring the number of localized and dispersed industries, we construct an index which refers to the extent of localization across all industries for each distance, $\Gamma(d) \equiv \sum_A \Gamma_A(d)$, and an index of dispersion, $\Psi(d) \equiv \sum_A \Psi_A(d)$. Figure 4 depicts these two indices.

Figure 4

The solid line refers to the localization index, $\Gamma(d)$, and the dashed line refers to the dispersion index, $\Psi(d)$. Similar to the results for the UK, the extent of localization is much greater at small distances. From these two figures, we can infer that localization of manufacturing industries also takes place within small areas in Japan.

Third, we want to examine differences in the degree of localization across broad industry categories. We do so by focusing on two-digit industries and comparing the ratio of localized four-digit industries in the total number of four-digit industries in a particular two-digit industry. The results are shown in Table 1.

Table 1

As can be seen, most of the four-digit industries within the Printing and Allied Industries category (JSIC16), the Textile Mill Products industry (JSIC11), and the Electrical Machinery, Equipment and Supplies industry (JSIC27) are localized. On the other hand, the four-digit industries within the Petroleum and Coal Products industry (JSIC18), the Lumber and Wood Products industry (JSIC13), and the Food industry (JSIC09) are less frequently localized. This pattern of localization is similar to that in the UK, where the Textile and Publishing industries (SIC 17–19, 22) are localized and the Food and Drink industries (SIC15) and Wood and Petroleum industries (SIC 20 and 23) are not (see Duranton and Overman, 2005).

Finally, we look at indices that measure the degree of localization and dispersion among four-digit industries. Following Duranton and Overman (2005), we construct cross-distance measures of the localization and dispersion by summing up $\Gamma_A(d)$ and $\Psi_A(d)$ for all $d \in [0, 180]$, that is, $\Gamma_A = \sum_{d=0}^{180} \Gamma_A(d)$ and $\Psi_A = \sum_{d=0}^{180} \Psi_A(d)$. Figure 5 presents the rank-order distribution of localization and dispersion indices for each industry measured by Γ_A and Ψ_A .

Figure 5

The solid line represents the measure of localization, and the dashed line refers to the dispersion. Similar to the UK, only a limited number of industries are highly localized or dispersed, and most of the industries do not have extreme values in these measures. We also show the industries with the highest Γ_A and Ψ_A , that is, the most localized and the most dispersed industries (Tables 2 and 3, respectively)

Tables 2 and 3

The most localized industry is the Blankets industry (JSIC1292), with a concentration in the Senshu area of Osaka prefecture, where 98 % of total production in the industry in Japan takes place. In addition, similar to the case of the UK (see Duranton and Overman, 2005), we find a large Γ_A -value for the Tableware industry (JSIC2521), which reflects the large concentration of tableware factories in Tsubame (Niigata prefecture) that the city is renowned for.

Table 3 presents the most dispersed industries. It shows that the three most dispersed industries are all related to seafood products, with the Miscellaneous Seafood Products industry (JSIC0929) being at the top. This result is again similar to that for the UK, where the most dispersed industry is the Processing and Preserving of Fish and Fish Products industry (SIC1520).

3.2 Results for small firms

The use of firm-level data possibly results in a bias toward detecting localization. Large-sized firms tend to locate their headquarters in large MEAs, which increases the degree of localization. In order to correct for this potential bias, in this subsection we limit the sample to small firms that have only one establishment and repeat the exercise of Section 3.1. Figures 6 to 8 present the results for small firms and correspond to Figures 3 to 5 for the sample of all firms.

Figure 6 shows the number of localized (solid line) and dispersed (dashed line) industries for each distance d .

Figure 6

Similar to the baseline results, Figure 6 indicates that the number of localized industries is largest between 0–40 km. The number in this range is about 250, which is slightly smaller than the number in the baseline result. The number of dispersed industries is stable over the entire

range of distances between 0–180 km. Next, Figure 7 shows that $\Gamma(d)$ and $\Psi(d)$ are similar to the values presented in the baseline result in Figure 4.

Figure 7

The rank-order distribution of localized and dispersed industries defined by the size of Γ and Ψ and presented in Figure 8 is quite similar to the baseline results.

Figure 8

Furthermore, we examine the relationship between Γ for the entire sample of firms and Γ for small firms. As shown in Figure 9, the correlation is very high. The correlation coefficient is 0.98.

Figure 9

Similarly, Figure 10 depicts the relationship between the Ψ for all firms and that for small firms.

Figure 10

Again, the correlation is very high, with a correlation coefficient of 0.98. In sum, we find no significant difference in the location pattern from the baseline results when restricting the sample to small firms.

4 Dynamics of industry localization

In this section, we look at the localization dynamics of industries to examine whether they are becoming more or less localized. We employ the method developed by Duranton and Overman (2008) and focus on entrant and exiting firms. Let industry A consist of n firms, which are composed of n_1 entrant firms, n_2 exiting firms, and n_3 staying firms. We define firms that are less than 10 years old as entrant firms and those which exit the market in the year after the sample year 2007 as exiting firms. We classify all other firms in the sample as staying firms. In order to examine the behavior of entrant firms in the industry A, we first measure all the bilateral distances between entrant firms and other firms, including staying and exiting firms, and calculate the K -densities for each distance. Second, to construct the confidence bands, we randomly draw sites of the same number as the number of entrant firms in the industry from the location of all the firms in the industry A consisting of n_1 entrant firms and $n_2 + n_3$

other firms. Third, we calculate all the bilateral distances between randomly chosen entrant firms and other firms and estimate the K -densities. We run 1,000 trials and construct two-sided confidence bands. Then, we investigate whether the actual entrants are located more closely to staying and exiting firms than the entrants in the counterfactual simulations. We follow the same procedure when examining the pattern of firm exits. Note that this procedure focuses on how closely entrant (exiting) firms are located to staying and exiting (staying and entrant) firms in the industry rather than how closely firms are located to each other. If the location pattern of entrant (exiting) firms resembles that of staying and exiting (staying and entrant) firms, this indicates that the industry is becoming neither more concentrated nor more dispersed and that the location pattern is stable over time. To examine this, we measure the bilateral distances between n_1 (n_2) firms and $n_2 + n_3$ ($n_1 + n_3$) firms rather than all pairwise distances among n firms and modify the K -density estimator as follows:

$$\hat{K}_A(d) = \frac{1}{n_1(n_2 + n_3)h} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2+n_3} f\left(\frac{d - d_{ij}}{h}\right), \quad (4)$$

for the case of the examination of entrant firms.⁹ Note that the numbers of entrant and exiting firms are extremely small in the four-digit categories. Therefore, in order to ensure that we have a sufficient number of observations, we focus on three-digit industries in this section.

Figure 11 shows the number of industries in which entrant firms are localized (solid-line) and dispersed (dashed-line).

Figure 11

Localization and dispersion of entrant (exiting) firms relative to staying and exiting (entrant and staying) firms are most frequently observed for very short distances of 0–40 km. However, both when we focus on firm entry and when we focus on firm exit, the number of localized and dispersed industries each does not exceed ten, which is very small in comparison with the total number of three-digit industries. Next, Figure 12 shows the number of industries in which exiting firms are localized (solid-line) or dispersed (dashed-line).

Figure 12

Again, localization is most common at short distances. However, overall, the number of localized and of dispersed industries is very small.

⁹When examining exiting firms, we replace n_1 and $n_2 + n_3$ with n_2 and $n_1 + n_3$.

To summarize, we find that in most industries, the location patterns of both entrant and exiting firms are not significantly different from those of existing firms, which suggests that in most manufacturing industries in Japan the location distribution is actually stable.

5 Concluding remarks

This paper examined the pattern of industry localization in Japan's manufacturing sector following the methodology employed by Duranton and Overman (2005, 2008) and utilizing firm-level micro-geographic data. We arrived at the following empirical findings. First, about half of the 561 four-digit manufacturing industries can be classified as localized, and the number of localized industries is largest for distances of up to 40 km. We also calculated the extent of localization by summing up the gaps between a kernel density and the upper bound of the confidence band, across all industries at each distance, and found that the localization tended to take place in quite small areas. Second, we found that most of the four-digit industries within several two-digit industries including Printing and Allied Industries category and the Textile Mill Products industry. We also calculated a measure of the extent of localization in each industry and found that several four-digit industries related to textile mills products are among the most localized. Overall, the patterns of localization in Japan resemble those in the UK, suggesting that there exist common factors across countries determining the concentration of industrial activities. Moreover, even when restricting the sample to small firms with a single establishment, we observed qualitatively similar empirical regularities to those using the entire sample. Finally, focusing on entrant and exiting firms, we examined industry location dynamics by analyzing the distribution of distances between entrant (exiting) and staying and exiting (staying and entrant) firms. In most of the industries, the location pattern of entrant (exiting) firms resembles that of staying and exiting (staying and entrant) firms. This indicates that these industries are becoming neither more concentrated nor more dispersed and the location patterns are stable over time.

One intriguing research question emerging from our results concerns the stability of industry localization. Since the degree of localization varies over time as documented by Kim (1995) and Dumais, Ellison, and Glaeser (2002), how industry localization develops over time is an interesting research question. Our results suggest that, at least in Japan in recent years, entrant firms and exiting firms are neither more localized nor more dispersed than other firms, meaning

that industry localization appears to be stable over time in most industries. Note, however, that there exist significant differences in the degree of localization among industries. How these differences emerged, with firm entries and exits neither strengthening or weakening industry localization, is an issue that deserves further scrutiny.

Another intriguing research issue is to identify the factors that contribute to industry localization, something that previous studies such as Rosenthal and Strange (2001) and Ellison, Glaeser, and Kerr (2010) have sought to address. In our analysis, we detect localization not only among high-tech industries such as information equipment and precision instruments industries which benefit from inter-firm knowledge spillovers but also among low-tech industries such as Blanket industry which depend less on such spillovers. This indicates that knowledge spillovers, contrary to what much of the extant theoretical literature suggests, may not necessarily be the major driving force underlying industry localization. Since our firm-level dataset contains several unique variables, including information on the suppliers and customers of each firm, we should be able to examine if there are other factors, such as transactions in intermediate goods and services, that may explain localization. Examining these issues using the unique set of variables provided by our dataset is a task we hope to address in future research.

A The implications of using the TSR data rather than census data

This appendix examines the potential bias caused by the use of firm-level non-census data. We use two concentration indices, the G -index developed by Ellison and Glaeser (1997) and the D -index developed by Mori, Nishikimi, and Smith (2005), which we calculate using two different datasets, the non-census, firm-level TSR dataset and the *Census of Manufactures*, an establishment-level census, and then compare the results. We aggregate the data in the TSR dataset by municipality at the two-digit industry level, the only format in which the general public has an access to the data from the Census of Manufactures. The results are shown in Figures A1 and A2.

Figures A1 and A2

In the figures, the horizontal axes represent the G -index or D -index values calculated using the TSR dataset, while the vertical axes represent the index values calculated using the Census of

Manufactures data.

Both the G -index and the D -index are higher for the TSR data than the Census data since most of the scatter points are below the 45° degree-line. This means that we arrive at a higher degree of localization when using the TSR data than when using the Census data. This suggests that headquarters tend to be more concentrated than establishments. However, note that the correlation between the index values using the two different databases is high. The correlation coefficient and the ordered correlation coefficient for the G -index are 0.82 and 0.67, respectively, while for the D -index, they are 0.77 and 0.81, respectively. To summarize, we tend to observe higher degrees of concentration when using firm-level non-census data than when using establishment-level census data, although the correlation of the degree of concentration measured using the two different kinds of data is high. These characteristics of our TSR data need to be taken into account when interpreting the results of our main analysis.

References

- Devereux, G., R. Griffith, and H. Simpson (2004), The Geographic Distribution of Production Activity in the UK, *Regional Science and Urban Economics* 35, 533-564.
- Dumais, G., G. Ellison, and E. Glaeser (2002), Geographic Concentration as a Dynamic Process, *Review of Economics and Statistics* 84, 193-204.
- Duranton, G. and H. Overman (2005), Testing for Localization Using Micro-geographic Data, *Review of Economic Studies* 72, 1077-1106.
- Duranton, G. and H. Overman (2008), Exploring the Detailed Location Patterns of U.K. Manufacturing Industries Using Microgeographic Data, *Journal of Regional Science* 48, 213-243.
- Duranton, G. and D. Puga (2004), Micro-foundations of Urban Agglomeration Economies, in V. Henderson and J. Thisse (eds.), *Handbook of Regional and Urban Economics* 4, 2063-2117, North-Holland, Amsterdam.
- Ellison, G. and E. Glaeser (1997), Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach, *Journal of Political Economy* 105, 889-927.
- Ellison, G., E. Glaeser, and W. Kerr (2010), What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *American Economic Review*, forthcoming.

- Fujita, M. and J.-F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge.
- Kim, S. (1995), Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860–1987, *The Quarterly Journal of Economics* 110, 881–908.
- Maurel, F. and B. Sedilló (1999), A Measure of the Geographic Concentration of French Manufacturing Industries, *Regional Science and Urban Economics* 29, 575-604.
- Marshall, A. (1890), *Principles of Economics*, Macmillan, London.
- Mori, T., K., Nishikimi, and T. Smith (2005), A Divergence Statistic for Industrial Localization, *Review of Economics and Statistics* 87, 635-651.
- Rosenthal, S. and W. Strange (2001), The Determinants of Agglomeration, *Journal of Urban Economics* 50, 191-229.
- Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.
- Tokunaga, S. and Y. Akune (2005) Wagakuni Seizogyo no Shuseki no Dotai Bunseki: Ellison-Glaser no Shuseki Sisu ni yoru Approach [A Measure of the Agglomeration in Japanese Manufacturing Industries: Using an Index of Agglomeration by Ellison and Glaeser (sic)], *Chiikigaku Kenkyu* [*Studies in Regional Science*] 35, 155–175 (in Japanese).

Table 1: Number of localized four-digit industries in each two-digit industry

Two-digit industry		Number of four-digit industries	No. of localized industries ≤ 60 km	No. of localized industries > 60 km
9	Food	40	3	3
10	Beverages, tobacco and feed	11	2	3
11	Textile mill products	36	28	32
12	Apparel	32	21	14
13	Lumber and wood products	18	1	2
14	Furniture and fixtures	10	4	5
15	Pulp, paper and paper products	20	11	12
16	Printing and allied industries	5	5	3
17	Chemical and allied products	39	27	28
18	Petroleum and coal products	5	1	1
19	Plastic products	23	15	16
20	Rubber products	12	7	6
21	Leather tanning, leather products and fur skins	10	8	8
22	Ceramic, stone and clay products	46	21	24
23	Iron and steel	24	7	6
24	Non-ferrous metals and products	17	10	13
25	Fabricated metal products	30	23	18
26	General machinery	47	33	37
27	Electrical machinery, equipment and supplies	24	22	22
28	Information and communication equipment	11	9	9
29	Electronic parts and devices	9	6	6
30	Transportation equipment	14	9	9
31	Precision instruments and machinery	21	17	16
32	Miscellaneous manufacturing industries	36	25	23

Table 2: The most localized industries (industries with the highest Γ_A)

Rank	Industry	# of firms	Γ
1	Blankets	29	0.72
2	Tableware (occidental type)	75	0.522
3	Fabric mills, woven woolen and worsted	186	0.5
4	Manufacture of towels	135	0.475
5	Ophthalmic goods, including frames	236	0.418
6	Jewelry products of precious metal and precious stone	562	0.385
7	Manufacture of textile mill products at cotton spinning mills	51	0.357
8	Tiles and mosaics, except quarry tiles	98	0.349
9	Microscopes and telescopes	162	0.348
10	Hull blocks	55	0.343

Table 3: The most dispersed four-digit industries (industries with the highest Ψ_A)

Rank	Industry	# of firms	Ψ
1	Miscellaneous seafood products	1537	0.216
2	Canned seafood and seaweed	148	0.164
3	Frozen seafood products (processed and packaged)	187	0.157
4	Crushed stones	570	0.155
5	Tatami mats (straw mats)	480	0.147
6	Wood chip mills	144	0.145
7	General sawing and planing mills (lumber)	2786	0.143
8	Fresh concrete	2213	0.128
9	Frozen seafood products (unprocessed and packaged)	209	0.127
10	Elemental feeds (animal foods)	91	0.123

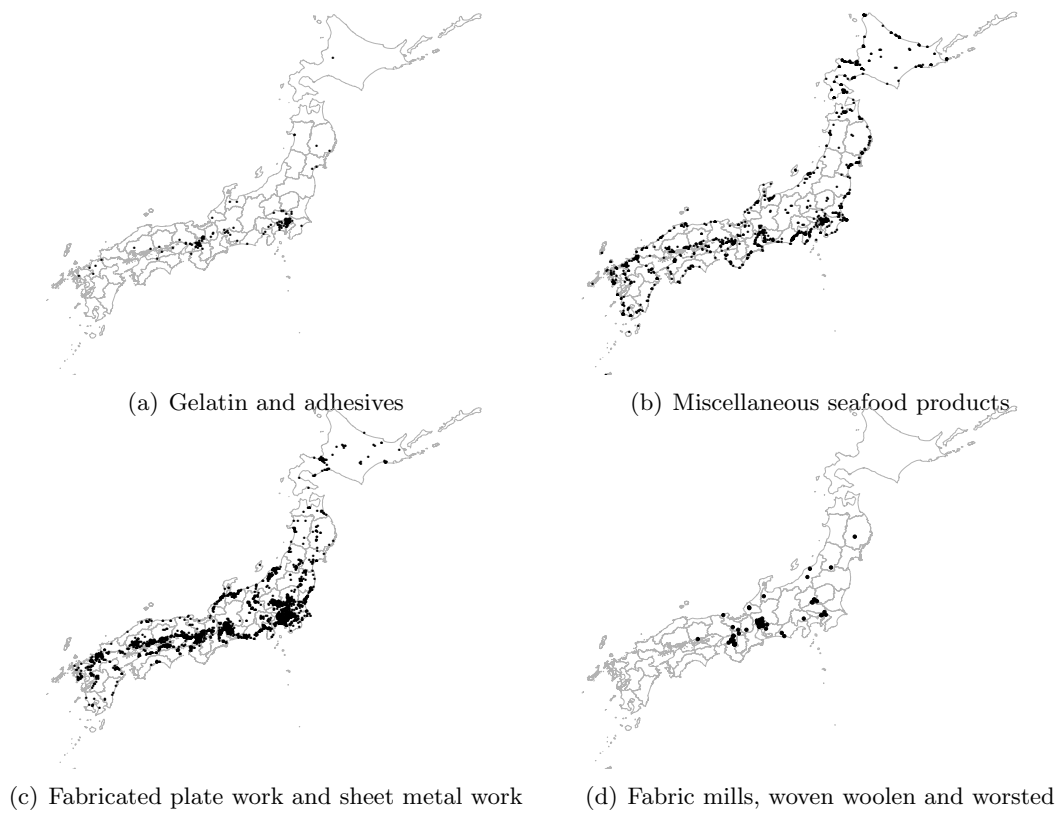
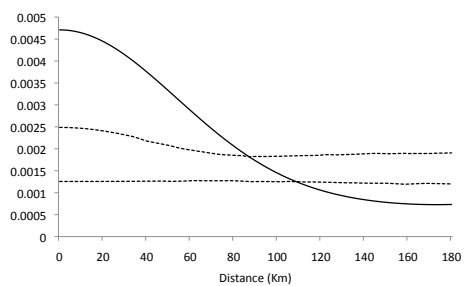
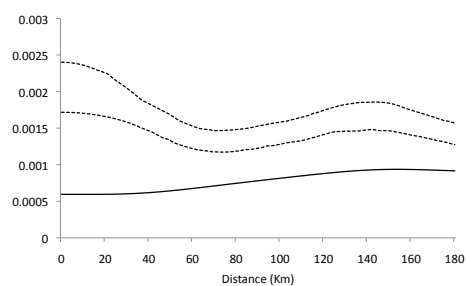


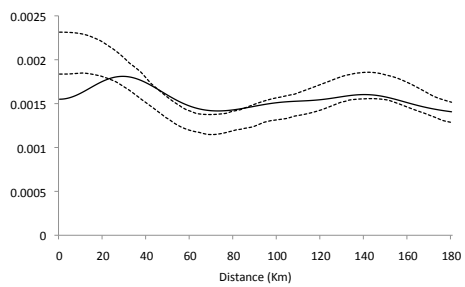
Figure 1: Geographical distribution of firms by industry



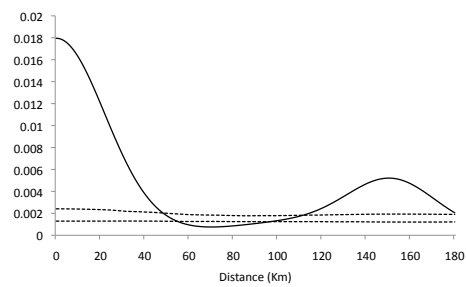
(a) Gelatin and Adhesives



(b) Miscellaneous Seafood Products



(c) Fabricated Plate Work and Sheet Metal Work



(d) Fabric Mills, Woven Woolen and Worsted

Figure 2: K -densities

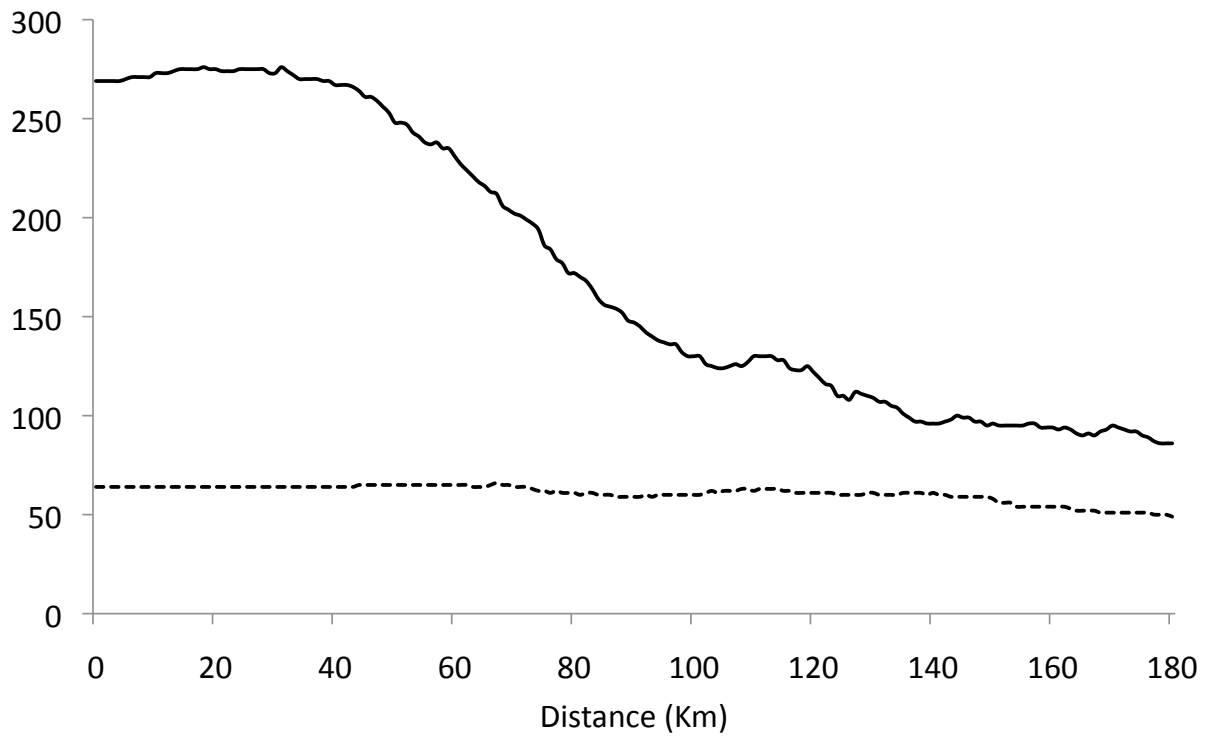


Figure 3: Number of localized and dispersed industries

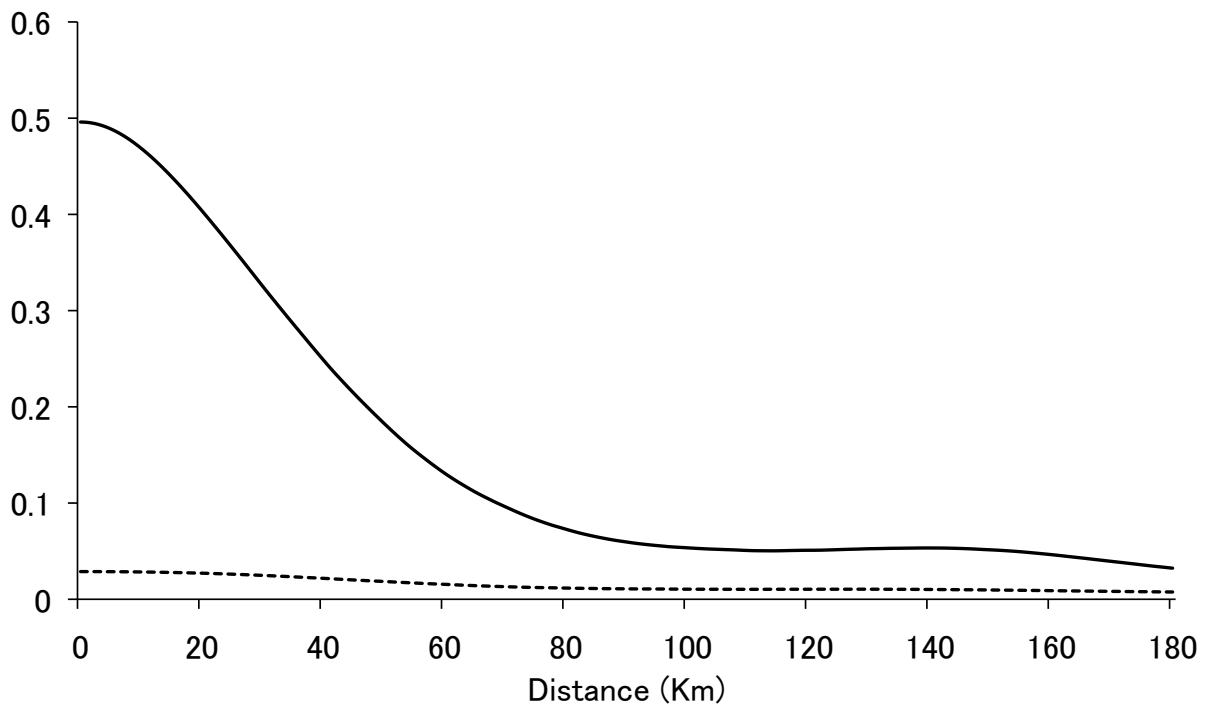


Figure 4: Localization and dispersion indices at each distance: $\Gamma(d)$ and $\Psi(d)$

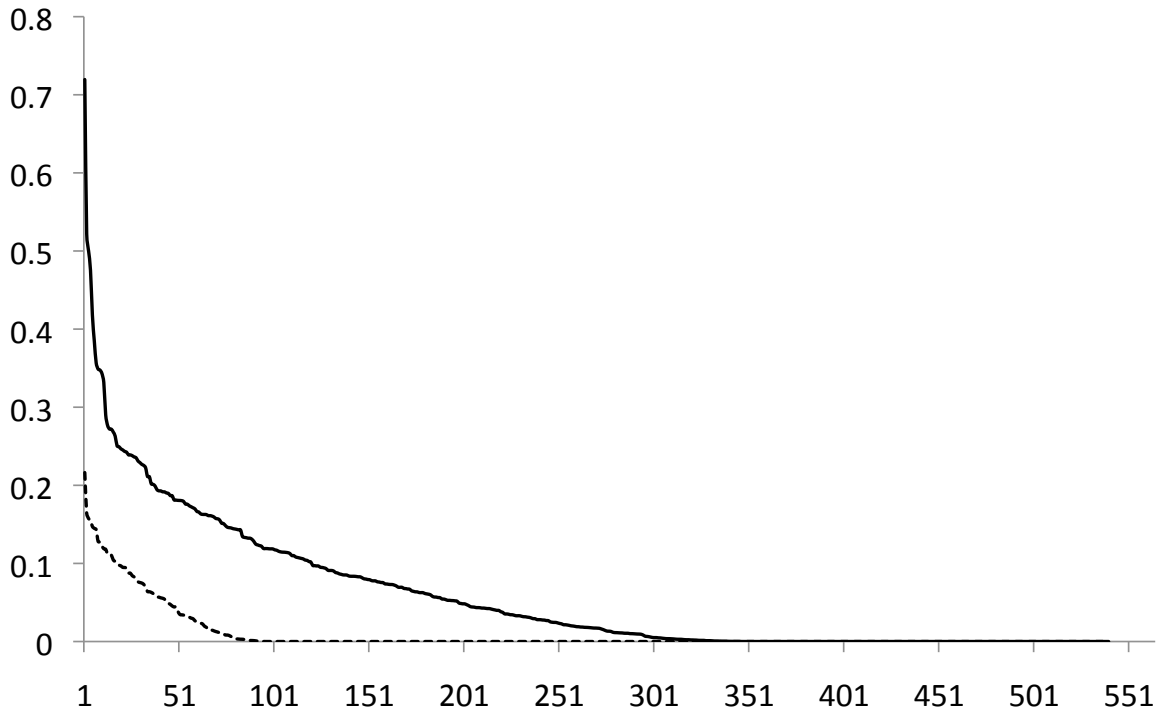


Figure 5: Rank-order distribution of localization and dispersion indices for each industry: Γ_A and Ψ_A

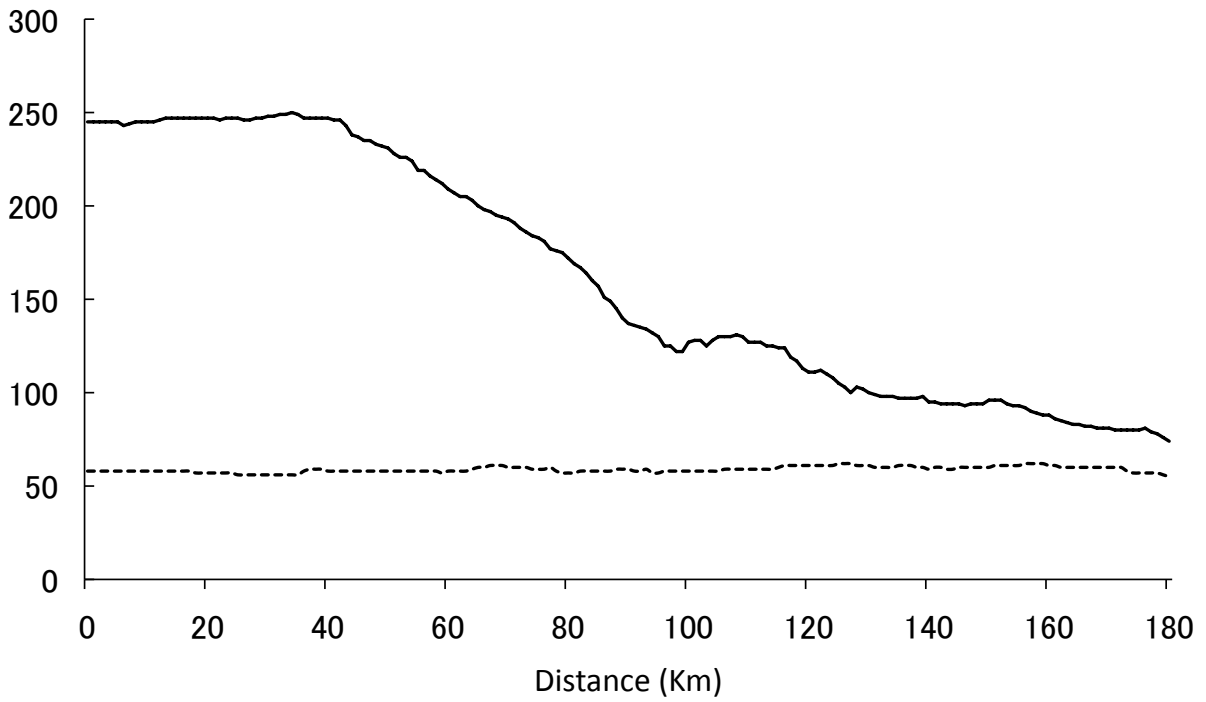


Figure 6: Number of localized and dispersed industries for small firms

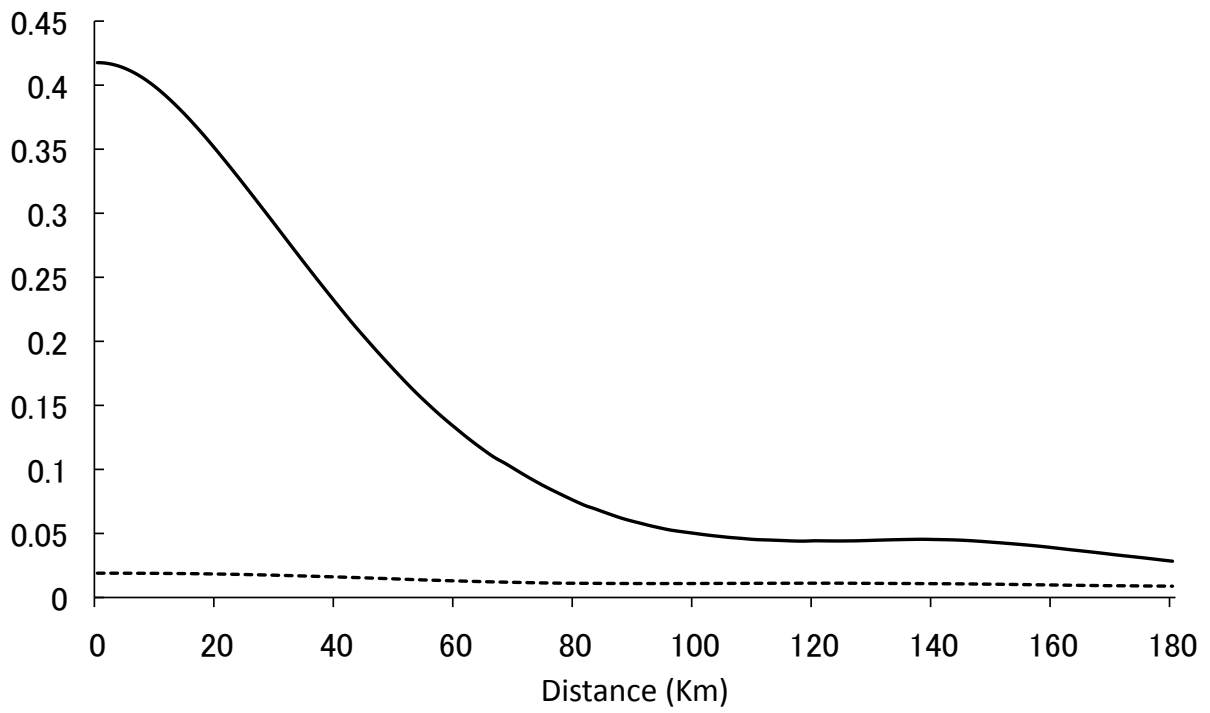


Figure 7: Localization and dispersion indices for each distance: $\Gamma(d)$ and $\Psi(d)$ for small firms

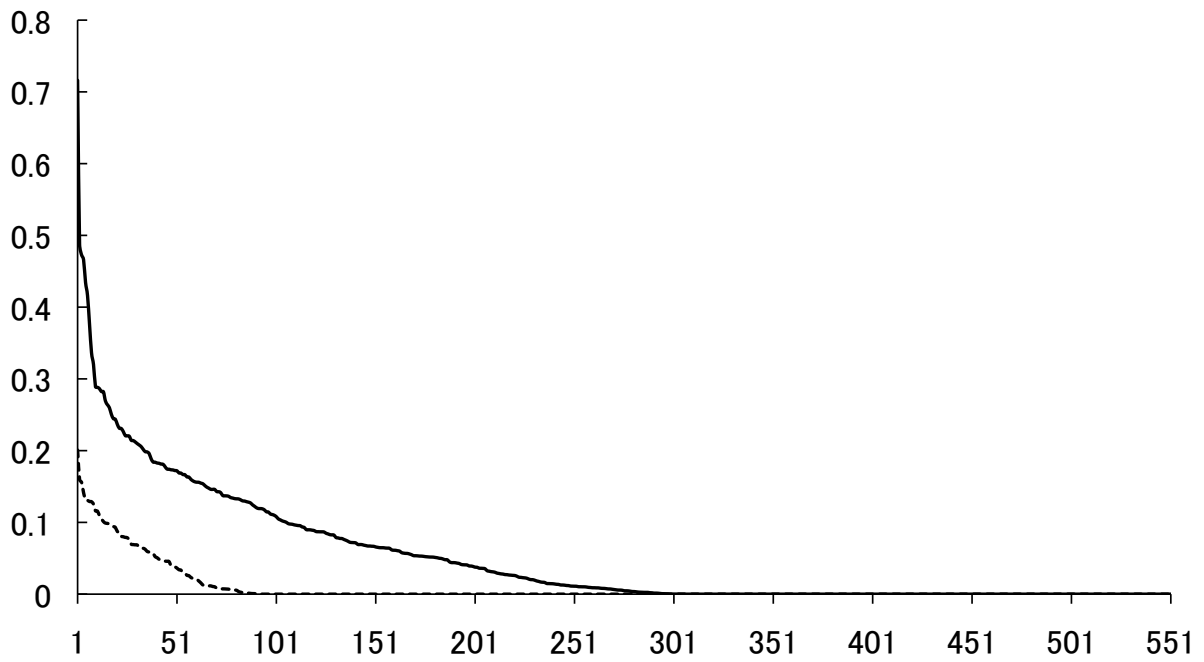


Figure 8: Rank-order distribution of localization and dispersion indices for each industry: Γ_A and Ψ_A for small firms

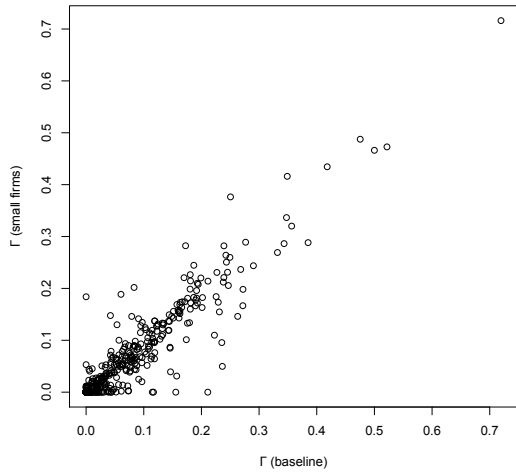


Figure 9: Γ_A for all firms and for small firms

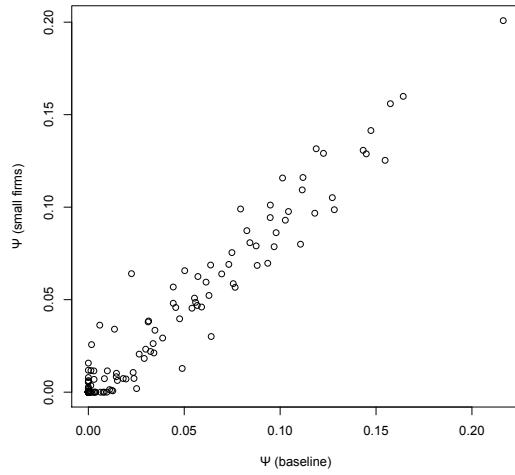


Figure 10: Ψ_A for all firms and for small firms

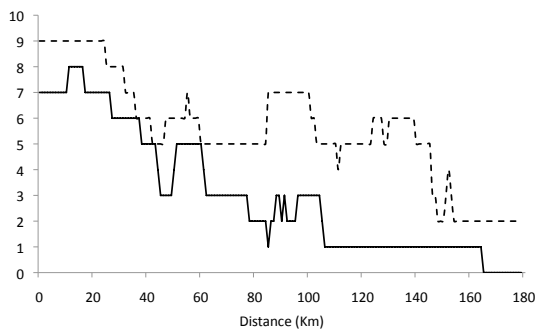


Figure 11: Number of localized and dispersed industries for entrant firms

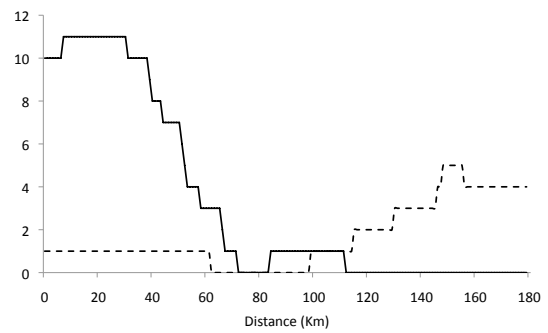


Figure 12: Number of localized and dispersed industries for exiting firms

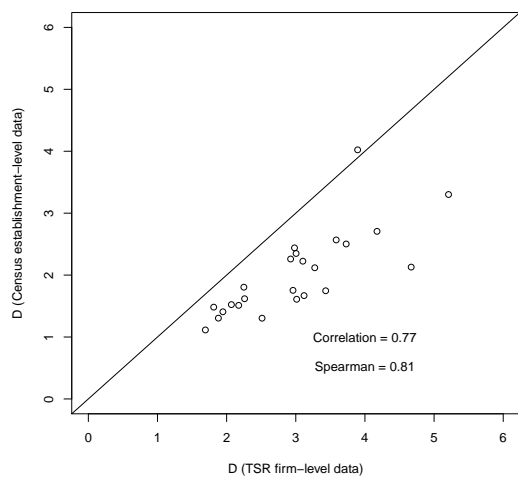
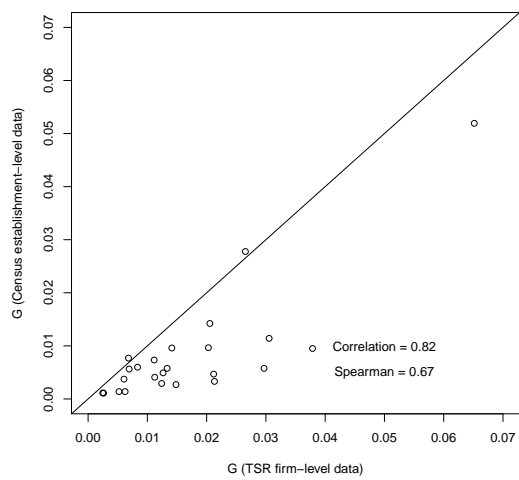


Figure A 1: *G*-index based on Census establishment-level data vs. TSR firm-level data
 Figure A 2: *D*-index based on Census establishment-level data vs. TSR firm-level data