

Program for Promoting Social Science Research

Aimed at Solutions of Near-Future Problems

Design of Interfirm Network to Achieve Sustainable Economic Growth

Working Paper Series No.10

Measuring Economic Localization: Evidence from Japanese Firm-level Data

Kentaro Nakajima
Yukiko Umeno Saito
and
Iichiro Uesugi

September 8, 2011

Research Center for Interfirm Network
Institute of Economic Research, Hitotsubashi University
Naka 2-1, Kunitachi-city, Tokyo 186-8603, JAPAN
Tel: +81-42-580-9145
E-mail: hit-tdb-sec@ier.hit-u.ac.jp
<http://www.ier.hit-u.ac.jp/ifn/>

Measuring Economic Localization: Evidence from Japanese Firm-level Data *

Kentaro Nakajima[†] Yukiko Umeno Saito[‡] Ichiro Uesugi[§]

August 23, 2011

Abstract

This paper examines location patterns of Japan's manufacturing industries 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, about half of Japan's manufacturing industries can be classified as localized and the number of localized industries is largest for a distance level 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, suggesting that there exist common factors across countries determining 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. These results suggest that most industries in Japan neither become more localized nor more dispersed over time and are in line with similar findings by Duranton and Overman (2008) for the UK. Fourth, a comparison with the service sector indicates that the share of localized industries is higher in manufacturing than in services, although the extent of localization among the most localized manufacturing industries is smaller than that among the most localized service industries, including financial service industries.

Keywords: Micro-geographic data; Economic geography

JEL classification: R11

*We would like to Tomoya Mori, Suminori Tokunaga, and participants of the JEA meeting at Senshu University and of seminars at the Bank of Japan, Hitotsubashi University, Kyoto University, RIETI, Tohoku University, the University of Tokyo, and Yokohama National University for helpful comments.

[†]Faculty of Economics, Tohoku University.

[‡]Fujitsu Research Institute and Organization for Asian Studies, Waseda University.

[§]Institute of Economic Research, Hitotsubashi University and RIETI.

1 Introduction

Economic activities and industries are not uniformly distributed but tend to cluster in certain areas. The automobile industry clusters in places such as Toyota City in Japan and Detroit in the United States are famous examples.

As laid out in uncountable theoretical models developed by economists, clustering 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 the localization (i.e., concentration over and above overall economic activity) of industries. Procedures that have been proposed to measure localization include the Gini, Isard, Herfindahl, and Theil indices. These “first generation” indices have been followed by a “second generation” of indices which adopt a so-called dartboard approach. The pioneering study for this second generation of indices is that by Ellison and Glaeser (1997), which compares the degree of 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 (like darts thrown randomly at a map). Other studies that have developed alternative measures, but follow a similar approach, include Maurel and Sedilló (1999), Devereux, Griffith, and Simpson (2004), and Mori, Nishikimi, and Smith (2005).¹

However, these indices are still susceptible to several problems due to aggregation at a given level of geographical units. First, the aggregation limits 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 at opposite ends of the country. This problem results in a downward bias to the detection of localization when spatial units with dense populations of firms are located close to each other.

¹Empirical application of some of these measures in a Japanese context can be found in Mori, Nishikimi, and Smith (2005) and Tokunaga and Akune (2005).

Against this background, the aim of this study is to examine the extent of localization in Japan, employing the distance-based approach proposed by Duranton and Overman (2005).² Their approach measures localization by utilizing the address data of establishments in continuous space and thus alleviates the problem incurred by aggregation and zoning. Given that the empirical literature on localization utilizing firm-level location information is still limited, our results may help to compare localization patterns of manufacturing industries across countries.³

Another shortcoming of the existing empirical literature on localization using firm-level data is that very few studies examine the situation in the service sector—despite the fact that, in most developed economies, including Japan, services account for a larger share of economic activity than manufacturing. Hence, another contribution of the present study is that, in addition to the manufacturing sector, we also examine the extent of localization in the service sector. Comparing the two, we find interesting differences in localization patterns, so that our results contribute to a richer understanding of 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. Third, we repeat our calculations for a sub-sample of small, single-establishment firms in order to correct

²Another strand of studies that is similar to the study by Duranton and Overman (2005) in its focus on a distance-based approach includes the studies by Marcon and Puech (2003) and Marcon and Puech (2010). The most salient difference between the two strands concerns the counterfactuals. The former takes the industry-average tendency to localize as the counterfactual, while the latter takes as the counterfactual the assumption that establishments are uniformly distributed. Since the aim of our study is to examine heterogeneity in location patterns across industries rather than a comparison with unrealistic random spatial location patterns, we follow Duranton and Overman's (2005) approach to detect industrial localization. Note, however, that in a later study, Marcon and Puech (2010) propose another distance-based approach with counterfactuals that are similar to Duranton and Overman's.

³Admittedly, the distance-based approach of Duranton and Overman (2005) has its own limitations. Marcon and Puech (2010) point this out with specific examples. One such example is an industry that has a number of regularly distributed clusters of establishments on a squared grid. That is, the distance between the neighboring clusters is approximately the same and the size of each cluster is also the same. This means that even though there obviously is localization of establishments in the industry, the distance-based approach of Duranton and Overman cannot precisely detect such localization.

for any possible biases resulting from the use of firm-level data. The empirical regularities we obtain are qualitatively similar to those for the entire sample.

Furthermore, we apply the distance-based approach to service industries. We find that 35% of the 469 four-digit service industries can be classified as localized for distances of 40 km or less. The share of localized industries is smaller in the service sector than in the manufacturing sector, although the range of localization is similar in both sectors. However, looking at the most localized service industries, their degree of localization is higher than that of the most localized manufacturing industries. Moreover, the degree of localization at the shortest end of the distance scale in services is much higher than the degree of such localization in manufacturing. These results are consistent with the fact that in some service industries, such as financial industries, headquarters are heavily agglomerated in small areas.

Finally, using observations on entrant and exiting manufacturing 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. Note, however, that in some industries, such as metal works, the distribution of distances between exiting and remaining firms is significantly skewed toward localization. This suggests that the significance of well-known industrial clusters in Tokyo's Ota Ward and in Higashi-Osaka, which have attracted many small but high-quality metal work manufacturers in the past, is waning.

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,⁴ 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.⁵ Following previous studies on industry concentration, the main focus of our analysis is the manufacturing sector, which reduces the sample to 143,628 firms. However, in addition, we later compare the results for the manufacturing sector with those for the service sector. The sample size for the service sector is 396,856 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 information on 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 will also be 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 calculating 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 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

⁴Industry classifications follow the Japanese Standard Industry Classification (JSIC).

⁵<http://newspat.csis.u-tokyo.ac.jp/geocode/>

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. It should be noted, however, that the extent of bias caused by the use of firm-level data is difficult to assess precisely and that our results may not be directly comparable to those based on establishment-level data.

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 Gelatin and Adhesives (JSIC 1794), Miscellaneous Seafood Products (JSIC 0929), Fabricated Plate Work and Sheet Metal Work (JSIC 2543), and Fabric Mills, Woven Woolen and Worsted (JSIC 1143), with each dot representing the location of a firm in the industry. The maps show that the Gelatin and Adhesives industry (JSIC 1794) appears to be concentrated in the Tokyo area, whereas the Miscellaneous Seafood Products industry (JSIC 0929) is dispersed along the coast. The Fabricated Plate Work and Sheet Metal Work industry (JSIC 2543) can be found mainly in the major metropolitan economic areas (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 (JSIC 1143) appears to be concentrated in two distinct locations.

Figures 1(a) to 1(d)

2.2 Empirical Approach

This subsection provides an overview of our empirical approach.⁶ 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.

⁶For 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.⁷

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 (service) firms in the TSR database constitutes the set of all possible locations for any manufacturing (service) firm and that firms randomly choose their location from these potential sites.⁸

In each trial, we randomly draw sites of the same number as the number of firms in the industry and then calculate the pairwise distances of the sites and estimate the K -density. This procedure ensures that we control for the overall pattern 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

⁷Following Silverman (1986), we use a Gaussian kernel with optimal bandwidth.

⁸It should be noted that other assumptions for the counterfactuals are also possible. Marcon and Puech (2003) in their study, for example, assume complete spatial randomness—that is, they assume that firms can locate anywhere in a certain area—and based on this calculate the L-function. Comparing the counterfactual based on Marcon and Puech’s (2003) assumption of complete spatial randomness with that employed here following Duranton and Overman (2005), the latter is likely to be skewed toward the shorter end of the distance scale.

localized.⁹ 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 Duranton and Overman (2005, 2008), we construct *global confidence bands*¹⁰ between 0–180 km¹¹ 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,

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

and an index of dispersion,

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

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 (JSIC 1794) is higher at short distances, while that of the Miscellaneous Seafood Products industry (JSIC 0929) gets

⁹Note that kernel smoothing may result in a possible bias in detecting localized industries. Since K -densities fluctuate less than actual densities, industries that are classified as not localized with kernel smoothing may actually be localized when we use the actual densities. We examine this possibility by using actual densities and find that the number of such industries is limited.

¹⁰Duranton and Overman (2005) also define local confidence levels for each distance, so called *local confidence bands*. However, as they point out, the local confidence intervals only allow statements with regard to a given distance. Following Duranton and Overman (2008), to be able to make statements about deviations over the entire range of distances, we mainly use *global confidence bands* rather than *local confidence bands* for our analysis.

¹¹The 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 Duranton 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.

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 (JSIC 1794) 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 other hand, Figure 2(b) for the Miscellaneous Seafood Products industry (JSIC 0929) 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 (JSIC 2543). 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,¹² 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 (JSIC 1143). 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 share of localized and dispersed industries for each distance d .

Figure 3

¹²As is well known, Japan’s manufacturing industries are mainly concentrated in these areas.

The solid line represents the share of localized industries. For short distances (0–40 km), about half of all the manufacturing industries (between 267 and 276 industries out of 561) are localized. For medium-range distances (40–100 km), the share of localized industries falls rapidly, with a small bump around 110 km. Turning to the share 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 share of localized and dispersed industries are quite similar to the results for the UK obtained by Duranton and Overman (2005).¹³

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 in Japan also takes place within small areas.

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 (JSIC 16), the Textile Mill Products industry (JSIC 11), and the Electrical Machinery, Equipment and Supplies industry (JSIC 27) are localized. On the other hand, the four-digit industries within the Petroleum and Coal Products industry (JSIC 18), the Lumber and Wood Products industry (JSIC 13), and the Food industry (JSIC 09) 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 industry (SIC 15) and Wood

¹³Note that the range of localization depends on the smoothing parameters of the K -densities. The actual distributions tend to fluctuate more than the K -density distributions and the range of localization may become smaller.

and Petroleum industries (SIC 20 and 23) are not (see Duranton and Overman, 2005). The pattern is also similar to that observed for France by Marcon and Puech (2003), although their empirical approach is somewhat different from ours; that is, industries that are localized are the Clothing and Leather industry (C1), the Printing and Publishing industry (C2), and the Textile industry (F2).

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 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 $\Psi_A = \sum_{d=0}^{180} \Psi_A(d)$.¹⁴ Figure 5 presents the distributions of Γ_A and Ψ_A by decreasing order of these measures.

Figure 5

The solid line represents the measure of localization, while 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 (JSIC 1292), 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 (JSIC 2521), 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 (JSIC 0929) 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 (SIC 1520).

¹⁴We follow Duranton and Overman (2005) in calculating Γ_A and Ψ_A . There may be alternative metrics that measure the extent of localization and dispersion in a more appropriate manner. We can calculate distance-weighted Γ_A and Ψ_A , in which we give higher weights to $\Gamma_A(d)$ and $\Psi_A(d)$ for smaller d . We can also calculate Γ_A and Ψ_A using actual densities rather than the densities with kernel-smoothing. However, the rank order distributions we see in Figure 5 do not change significantly with these alternative metrics.

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 share of localized (solid line) and dispersed (dashed line) industries for each distance d .

Figure 6

Figure 6 indicates that, similar to the baseline result, the share of localized industries is largest between 0–40 km. The share in this range is about 45%, which is slightly smaller than the share in the baseline result. The share 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_A and Ψ_A and presented in Figure 8 is quite similar to the baseline result.

Figure 8

Next, we examine the relationship between Γ_A for the entire sample of firms and Γ_A for small firms. The correlation is very high and statistically significant. The correlation coefficient is 0.98, and the rank correlation coefficient is 0.91. The null hypothesis that there is no rank correlation between Γ_A for the entire sample of firms and Γ for small firms is rejected with a p -value of 0.000. Similarly, we analyse the relationship between Ψ_A for all firms and Ψ_A for small firms. Again, the correlation is very high, with a correlation coefficient of 0.98 and a rank correlation coefficient of 0.87. The rank correlation is again statistically significant with a p -value of 0.000. In sum, we find no significant difference in the location pattern from the baseline result when restricting the sample to small firms.

3.3 Results for firms in the service sector

While there are many studies that have examined localization in the manufacturing sector, few studies have analyzed localization in the service sector, mainly due to a lack of data. Yet, in most developed countries, including Japan, the service sector accounts for a much larger share of economic activity than the manufacturing sector. This means that if we really want to understand the spatial features of economic activity, it is important to also examine locational patterns in the service sector. Fortunately, our TSR dataset includes not only data on manufacturing firms but also data on service firms. Therefore, in this subsection, we examine location patterns in the service sector.

As in the analysis focusing on the manufacturing sector, we need to construct counterfactual random location distributions for the service sector. For the counterfactuals, we therefore again assume that the set of all existing sites currently occupied by service firms in the TSR database constitutes the set of all possible locations and that firms randomly choose their location from these potential sites.

In the analysis, we use 469 four-digit JSIC code service industries, each of which has 10 or more firms. Note that we focus on service industries only and therefore do not consider omit industries such as agriculture, forestry, fishery, mining, and construction. Thus, our dataset consists of the following service industries: information and telecommunication, transportation, wholesale and retail, financial services, real estate, restaurants and accommodations, medical and health care services, education, and other services industries. The total number of firms in the dataset is 396,856. Figures 9 to 11 present the results for service firms and correspond to Figures 3 to 5 for the sample of manufacturing firms.

First, Figure 9 shows the share of localized (solid line) and dispersed (dashed line) industries for each distance d .

Figure 9

Similar to the results for the manufacturing sector, the share of localized industries is largest between 0–40 km, but the share itself is only 35% and therefore lower than that in the manufacturing sector. The share of dispersed industries is largest between 0–40 km at slightly above 20% and then gradually decreases for medium and long distances.

Second, Figure 10 shows the distributions of $\Gamma(d)$ and $\Psi(d)$, which measure the extent of

localization across all service industries for each distance. Figure 10 resembles Figure 4 in that $\Gamma(d)$ decreases over the range of distances between 0–80 km and that $\Psi(d)$ is small and steady over the entire range of distances.

Figure 10

On the other hand, there exists a significant difference between these two figures. For the shortest-range of distances (0–5 km), the extent of localization $\Gamma(d)$ is much higher than that in the manufacturing sector. This suggests that firms in service industries are more heavily agglomerated at the shortest distances than those in manufacturing industries. At distances greater than 5 km, $\Gamma(d)$ rapidly decreases.

Third, Figure 11 presents the distributions of Γ_A and Ψ_A by decreasing order of these measures, where the solid line represents the measure of localization and the dashed line refers to the dispersion. The shape of these lines is similar to that in Figure 5 for the manufacturing sector.

Figure 11

Note, however, that the highest Γ_A in the service sector at around 0.8 is higher than the corresponding Γ_A in the manufacturing sector, which is around 0.7. This can be confirmed by comparing Table 5, which shows the most localized industries in the service sector, with Table 2, which shows the same for the manufacturing sector. The top 10 industries in the service sector have higher values of Γ_A than their top 10 counterparts in the manufacturing sector. The service industry with the highest Γ_A is Foreign Banks in Japan (JSIC 6124). Moreover, among the top 10, there are four other financial services industries (Trust Banks, JSIC 6122; Life Insurance Stock Companies, JSIC 6711; Investment Businesses, JSIC 6491; and Investment Advisory Businesses, JSIC 6522). The headquarters of firms in these industries are heavily concentrated in the financial district of Tokyo. Three of the other top 10 industries fall under the heading of information and communication. Overall, Figure 11 and Table 5 suggest that certain types of industries in the service sector have a strong tendency to localize.

Next, Table 6 shows the industries with the highest Ψ_A in the service sector. Many of these industries are related to agricultural and fishery products (Fisheries Cooperative Associations, n.e.c., JSIC 7912; Agriculture Cooperative Associations, n.e.c., JSIC 7911; Fish Processors Product Processing Cooperative Associations, n.e.c., JSIC 7913; and Fresh Fish Stores, JSIC

5741). Interestingly, many of the manufacturing industries with the highest Ψ_A are also related to fishery products, as seen in Table 3, which shows that four of the top 10 manufacturing industries are related to fishery products. These results suggest that even though firms in each four-digit fishery-related industry are dispersed, many firms from different four-digit fishery-related industries may be collocated in a certain area, say around a fishing port.

In sum, the share of localized industries in the service sector is smaller than that in the manufacturing sector. On the other hand, however, when we compare industries with the highest Γ_A , which possibly represents the extent of localization, the value tends to be higher in the service sector than in the manufacturing sector.

Tables 5 and 6

4 Dynamics of industry localization

In this section, we again concentrate on the manufacturing sector and 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 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 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.¹⁵ 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 12 shows the number of industries in which entrant firms are localized (solid line) and dispersed (dashed line).

Figure 12

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 13 shows the number of industries in which exiting firms are localized (solid line) or dispersed (dashed line).

Figure 13

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

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. One interesting result worth mentioning, however, is that while in most industries, no localization in firm entry or exit can be observed, the Metal Machine Parts and Tooling Products (JSIC 255) and the Metal Working Machinery (JSIC 264) industries are notable exceptions in terms of firm

¹⁵When examining exiting firms, we replace n_1 and $n_2 + n_3$ with n_2 and $n_1 + n_3$.

exit. Since the end of World War II, small manufacturers in these industries have traditionally clustered in certain urban districts such as Tokyo's Ota Ward and Higashi-Osaka. These clusters are well-known for their concentration of firms supplying high-quality parts and components to Japan's electronics and automobile assemblers, thus underpinning the country's competitiveness in these industries. However, in recent years, many of these firms have gone out of business. Our results indicate that these exiting firms tend to be located close to remaining firms, meaning that these exits significantly reduce the degree of industrial agglomeration in these industries.

5 Concluding remarks

This paper examined the location patterns of Japan's manufacturing sector following the methodology employed by Duranton and Overman (2005, 2008) and utilizing firm-level micro-geographic data. In addition, it was examined how localization differs between the manufacturing and the service sector. 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 for each manufacturing industry at each distance, summed it up across industries at each distance, and found that 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, are localized. We also calculated a measure of the extent of localization in each industry and found that several four-digit industries related to textile mill products are among the most localized. Overall, the patterns of localization in Japan resemble those in the UK and France, 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. Third, we examined location patterns of the service sector and found that about 35 % of industries are localized, which is lower than the figure for the manufacturing sector. On the other hand, however, the most localized service industries, including financial services businesses, were found to be more heavily concentrated in a small area than the most localized manufacturing industries. 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 issue concerns 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 detected localization not among high-tech industries such as information equipment and precision instruments industries which benefit from inter-firm knowledge spillovers, but among low-tech industries, such as the 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 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* 100, 1195-1213.

- Fujita, M. and J.-F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press, Cambridge.
- Marcon, E. and F. Puech (2003), Evaluating the Geographic Concentration of Industries Using Distance-based Methods, *Journal of Economic Geography* 3, 409–428.
- Marcon, E. and F. Puech (2010), Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods, *Journal of Economic Geography* 3, 409–428.
- 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-Glaeser 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 in the manufacturing sector

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 in the manufacturing sector (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 industries in the manufacturing sector (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

Table 4: Number of localized four-digit industries in each two-digit industry in the service sector

Two-digit industry	Number of four-digit industries	No. of localized industries ≤ 60 km	No. of localized industries > 60 km
33 Production, transmission and distribution of electricity	2	0	0
34 Manufacture of gas	3	0	0
35 Heat supply	1	1	1
36 Collection, purification and distribution of water, and sewage collection processing and disposal	3	0	0
37 Communications	4	2	2
38 Broadcasting	5	1	2
39 Information services	5	5	5
40 Internet based services	1	1	31
41 Video picture, sound information, character information production and distribution	8	8	7
42 Railway transport	3	0	1
43 Road passenger transport	5	0	0
44 Road freight transport	6	5	5
45 Water transport	8	4	5
46 Air transport	2	1	0
47 Warehousing	2	1	1
48 Services incidental to transport	12	6	6
49 Wholesale trade, general merchandise	2	2	2
50 Wholesale trade (textile and apparel)	12	11	11
51 Wholesale trade (food and beverages)	16	1	6
52 Wholesale trade (building materials, minerals and metals, etc.)	17	8	16
53 Wholesale trade (machinery and equipment)	13	8	10
54 Miscellaneous wholesale trade	18	11	13
55 Retail trade, general merchandise	2	0	0
56 Retail trade (dry goods, apparel and apparel accessories)	9	0	2
57 Retail trade (food and beverages)	20	0	5
58 Retail trade (motor vehicles and bicycles)	5	0	3
59 Retail trade (furniture, household utensil and household appliances)	11	0	4
60 Miscellaneous retail trade	24	0	9
61 Banking	3	2	0
62 Financial institutions for cooperative organizations	7	0	0
63 Institutions dealing with postal savings, government-related financial institutions	0	0	0
64 Non-deposit money corporations engaged in the provision of finance, credit and investment	9	5	5
65 Securities and futures commodity dealing activities	7	7	6
66 Financial auxiliaries	3	1	1
67 Insurance institutions, including insurance agents, brokers and services	9	6	6
68 Real estate agencies	3	3	3
69 Real estate lessors and managers	7	4	5
70 General eating and drinking places	9	3	2
71 Spree eating and drinking places	3	0	0
72 Accommodations	4	0	1
73 Medical and other health services	8	2	3
74 Public health and hygiene	4	0	0
75 Social insurance and social welfare	14	5	5
76 School education	7	4	4
77 Miscellaneous education, learning support	11	3	2
78 Postal services, except otherwise classified	0	0	0
79 Cooperative associations, n.e.c.	5	0	1
80 Professional services, n.e.c.	21	14	13
81 Scientific and development research institutes	5	4	4
82 Laundry, beauty and bath services	10	3	4
83 Miscellaneous living-related and personal services	13	3	5
84 Services for amusement and hobbies	25	7	9
85 Waste disposal business	11	1	2
86 Automobile maintenance services	2	0	2
87 Machine, etc. repair services, except otherwise classified	8	1	1
88 Goods rental and leasing	12	4	6
89 Advertising	3	3	2
90 Miscellaneous business services	15	10	9
91 Political, business and cultural organizations	6	5	5
92 Religion	3	3	3
93 Miscellaneous services	3	0	0

Table 5: The most localized industries in the service sector (industries with the highest Γ_A)

Rank	Industry	# of firms	Γ
1	Foreign banks in Japan	27	0.827
2	Trust banks	26	0.823
3	Life insurance stock companies	31	0.731
4	Satellite broadcasting	28	0.594
5	Dramatic companies	229	0.537
6	Recording and disk production	51	0.455
7	Miscellaneous data processing and information services	584	0.432
8	Investment businesses	772	0.406
9	Miscellaneous wholesale trade, general merchandise	4523	0.406
10	Investment advisory businesses	158	0.405

Table 6: The most dispersed industries in the service sector (industries with the highest Ψ_A)

Rank	Industry	# of firms	Ψ
1	Fisheries cooperative associations, n.e.c.	372	0.137
2	Surveying services	1913	0.110
3	Fuel stores, except gasoline service stations	3244	0.099
4	Convenience stores, primarily for sale of staple food and beverages	1431	0.095
5	Agriculture cooperative associations, n.e.c.	938	0.084
6	Fish processors product processing cooperative associations, n.e.c.	86	0.072
7	Stores selling rice, barley and other cereals	1086	0.072
8	Grocery stores	1995	0.069
9	Funeral services	839	0.065
10	Fresh fish stores	505	0.063

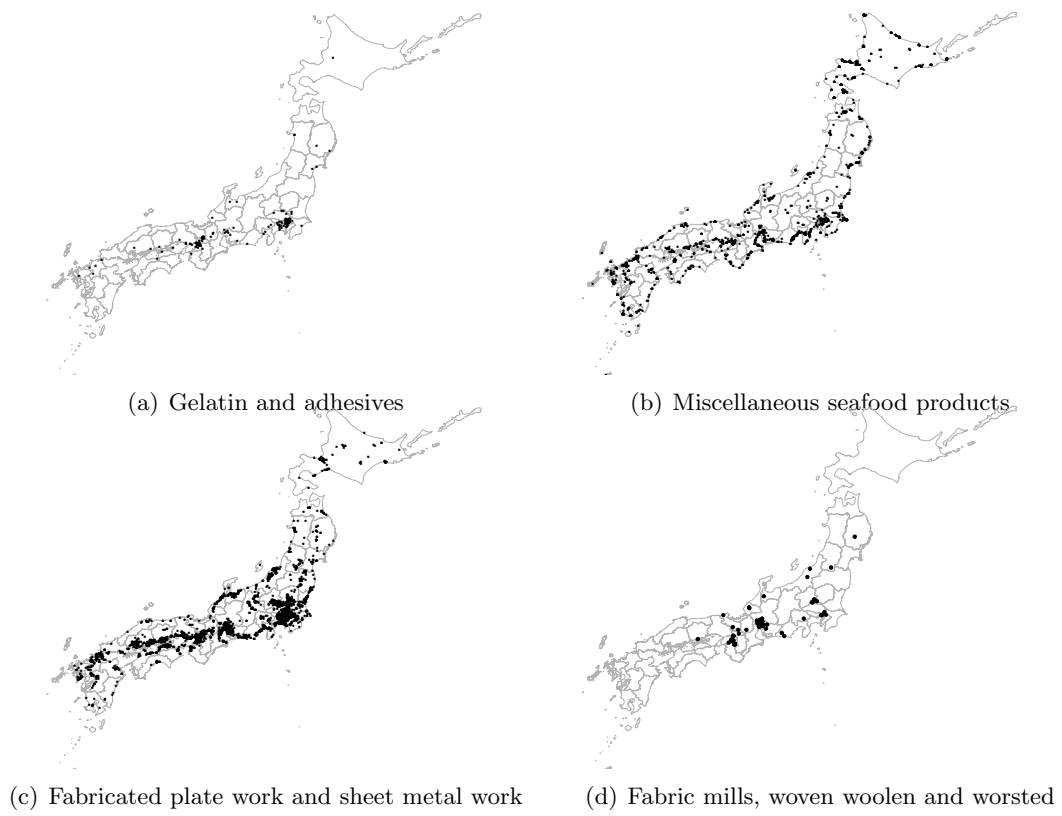
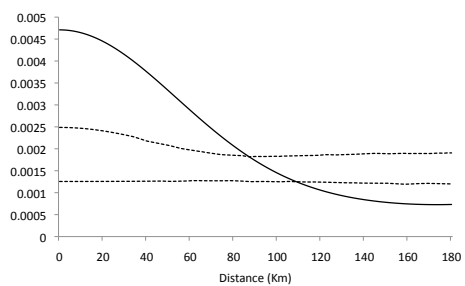
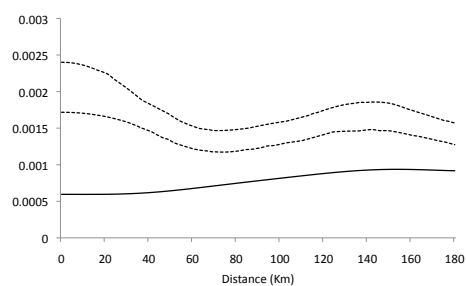


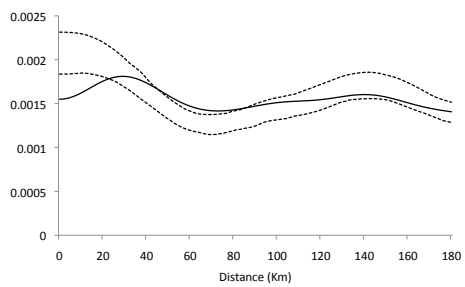
Figure 1: Geographical distribution of firms by industry



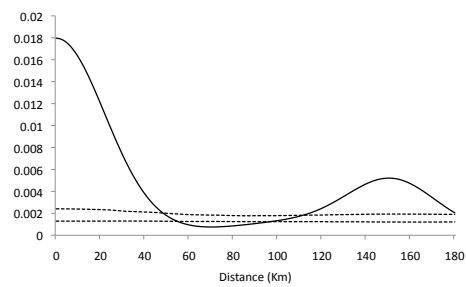
(a) Gelatin and Adhesives



(b) Miscellaneous eafood Products



(c) Fabricated Plate Work and Sheet Metal Work



(d) Fabric Mills, Woven Woolen and Worsted

Figure 2: K -densities

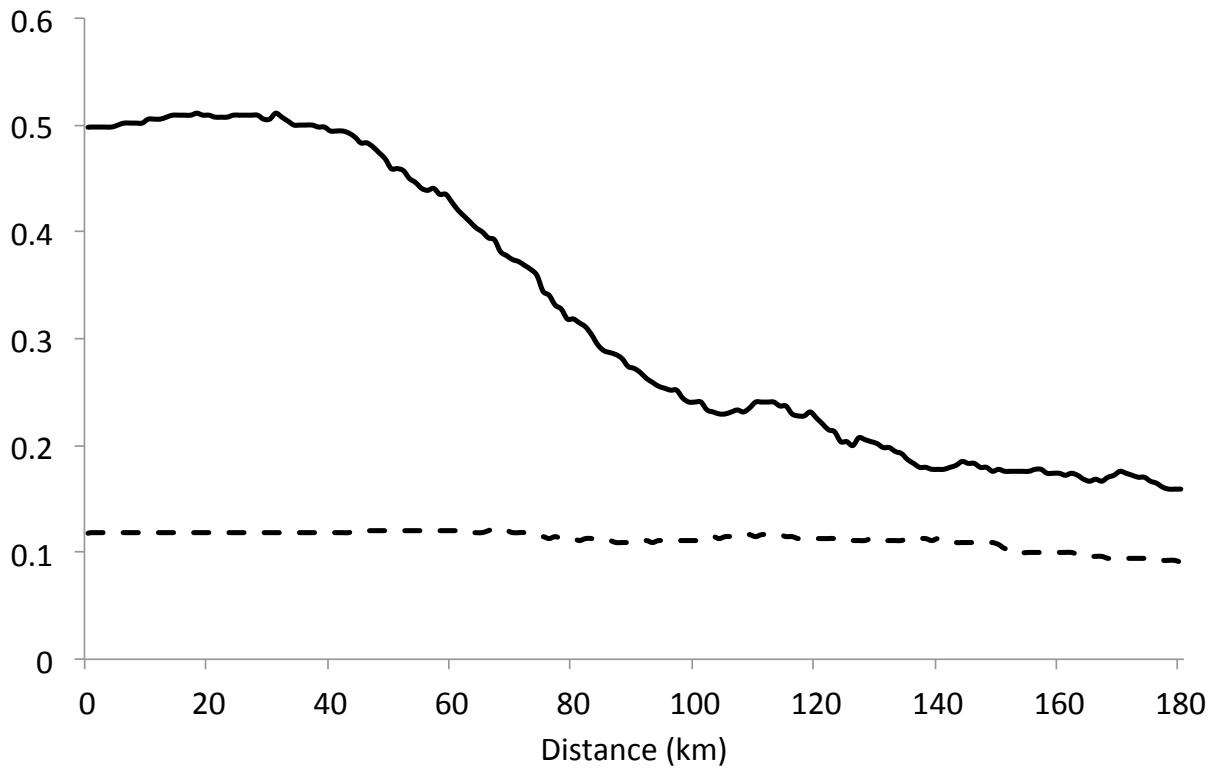


Figure 3: Share of localized and dispersed industries

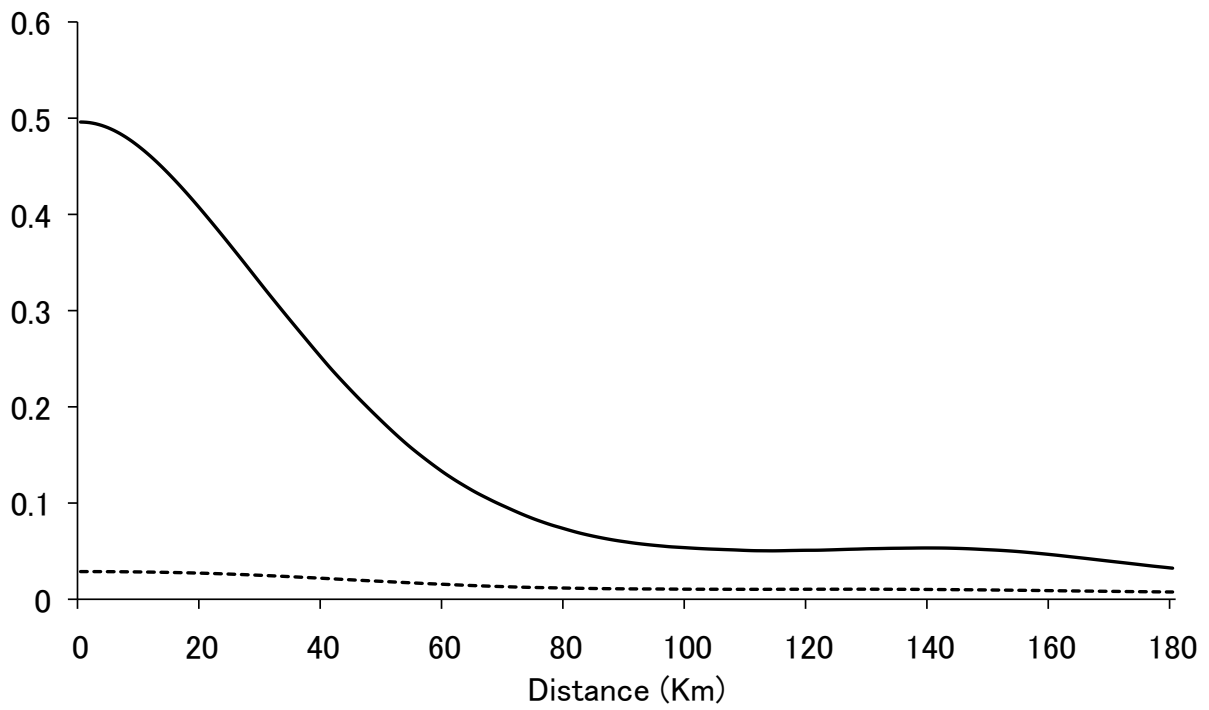


Figure 4: Localization and dispersion indices for each distance: $\Gamma(d)$ and $\Psi(d)$

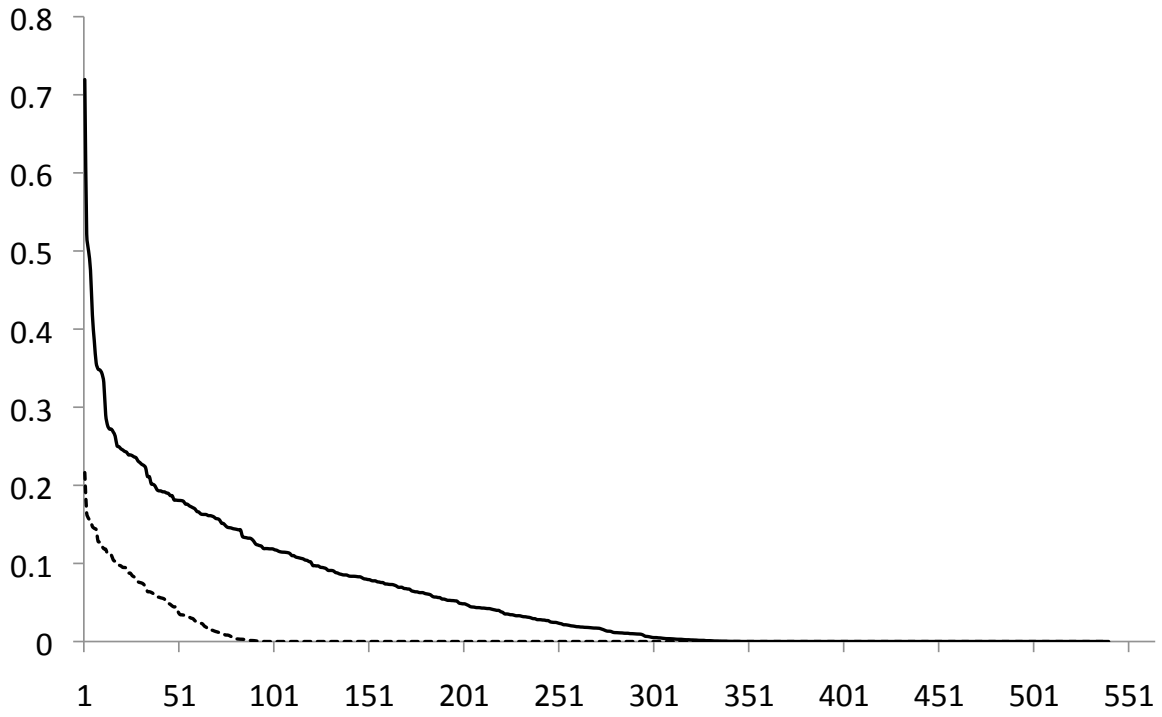


Figure 5: Rank-order distribution of localization and dispersion indices for each industry: Γ_A and Ψ_A

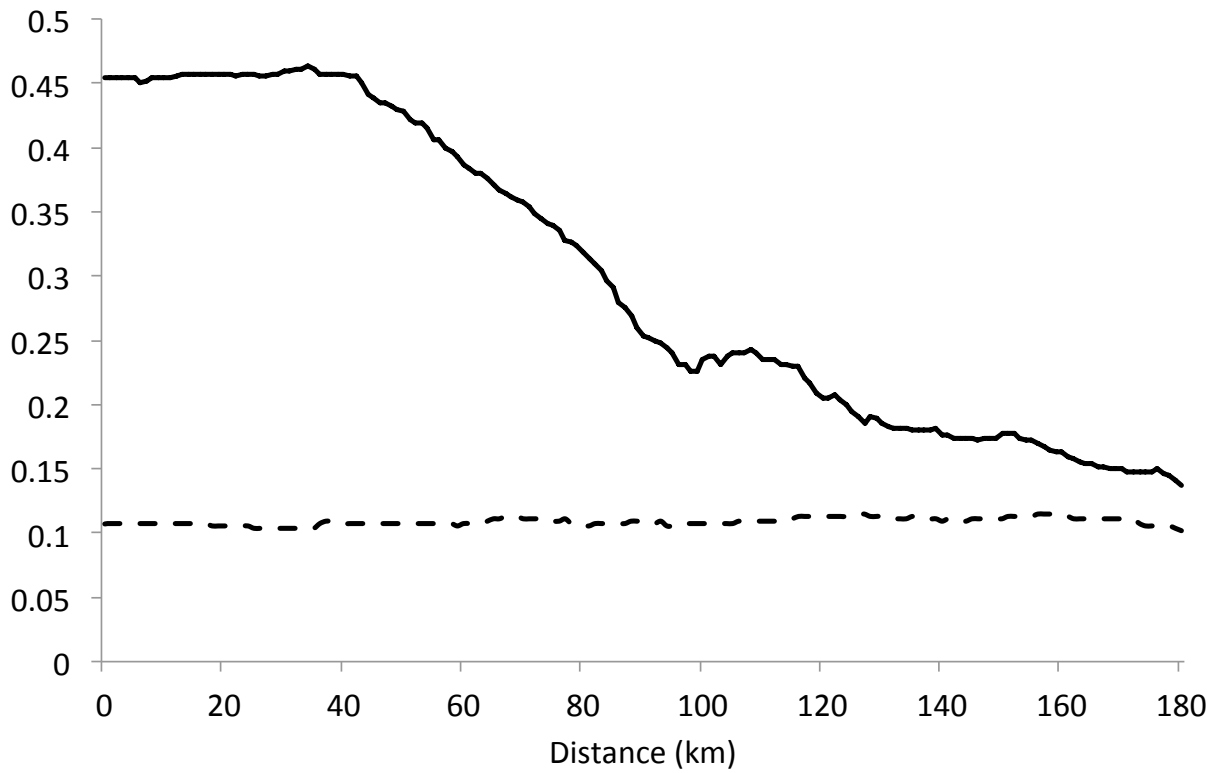


Figure 6: Share of localized and dispersed industries for small firms

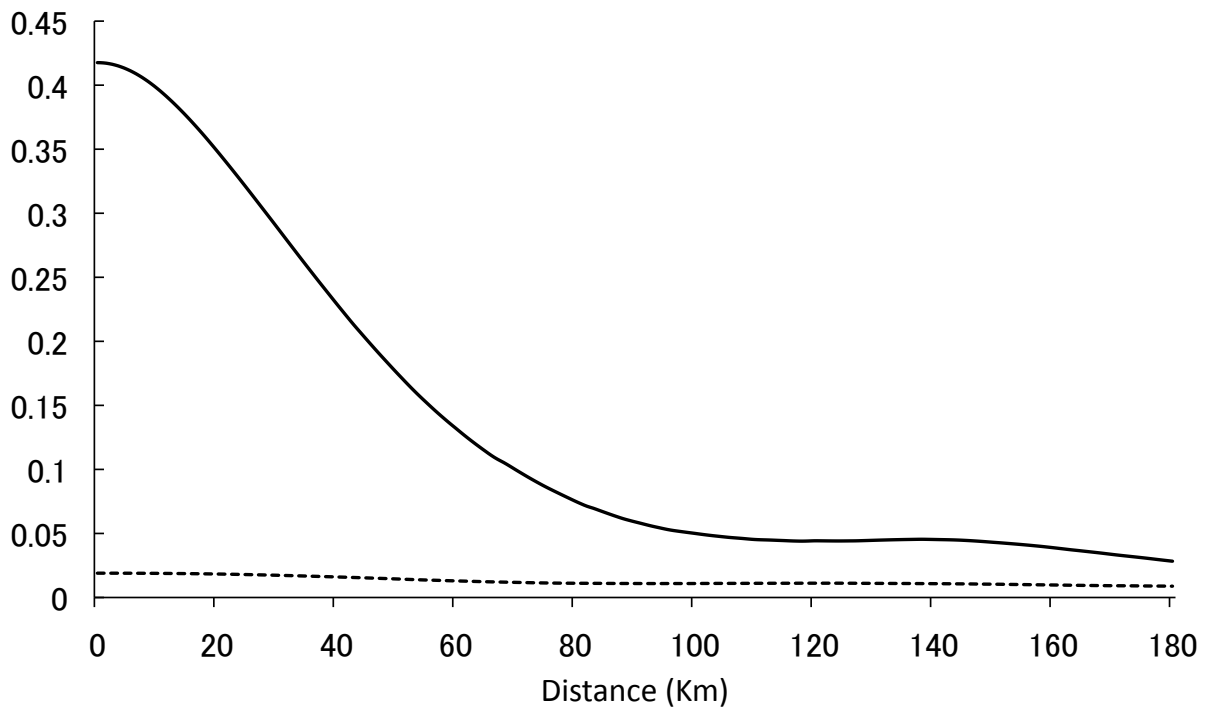


Figure 7: Localization and dispersion indices for each distance: $\Gamma(d)$ and $\Psi(d)$ for small firms

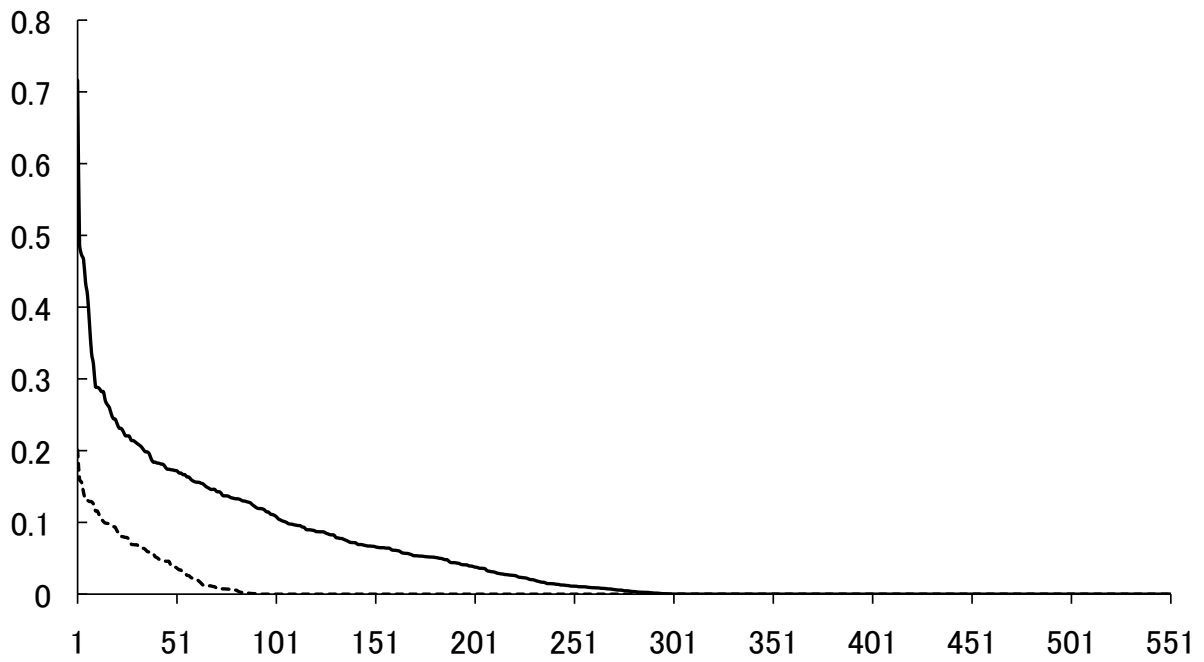


Figure 8: Rank-order distribution of localization and dispersion indices for each industry: Γ_A and Ψ_A for small firms

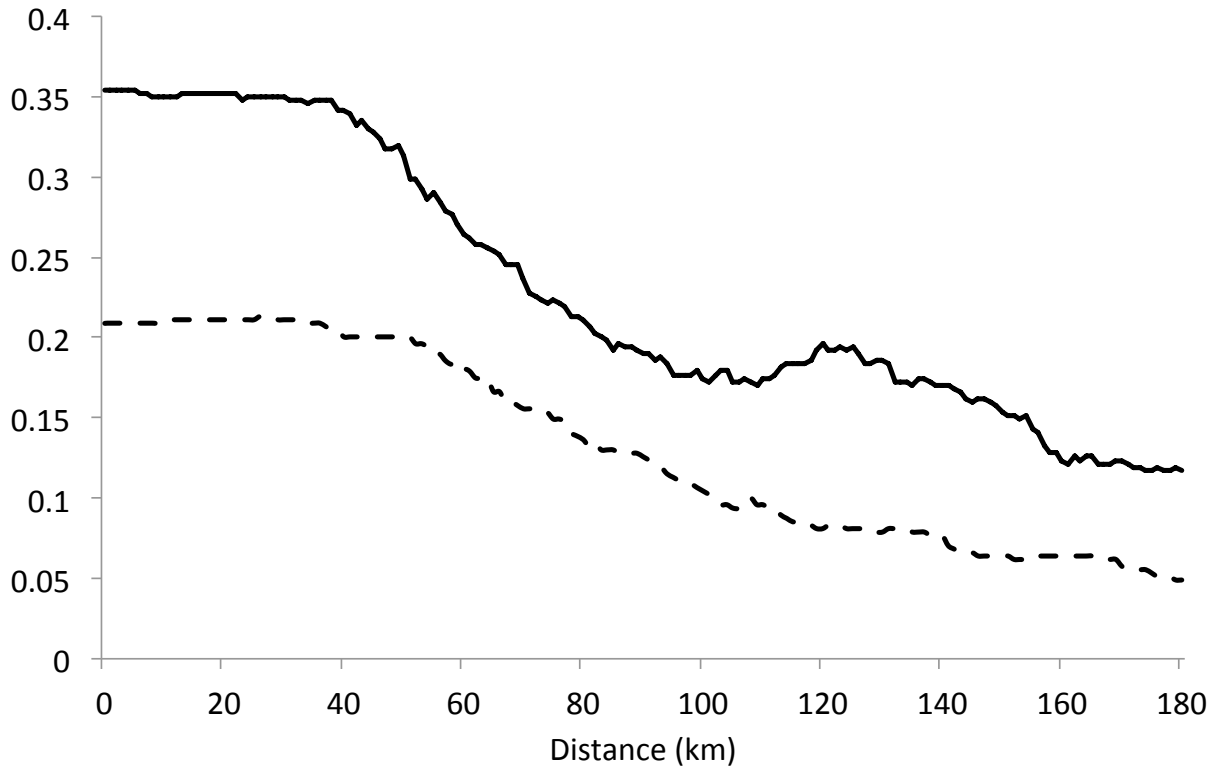


Figure 9: Share of localized and dispersed industries in the service sector

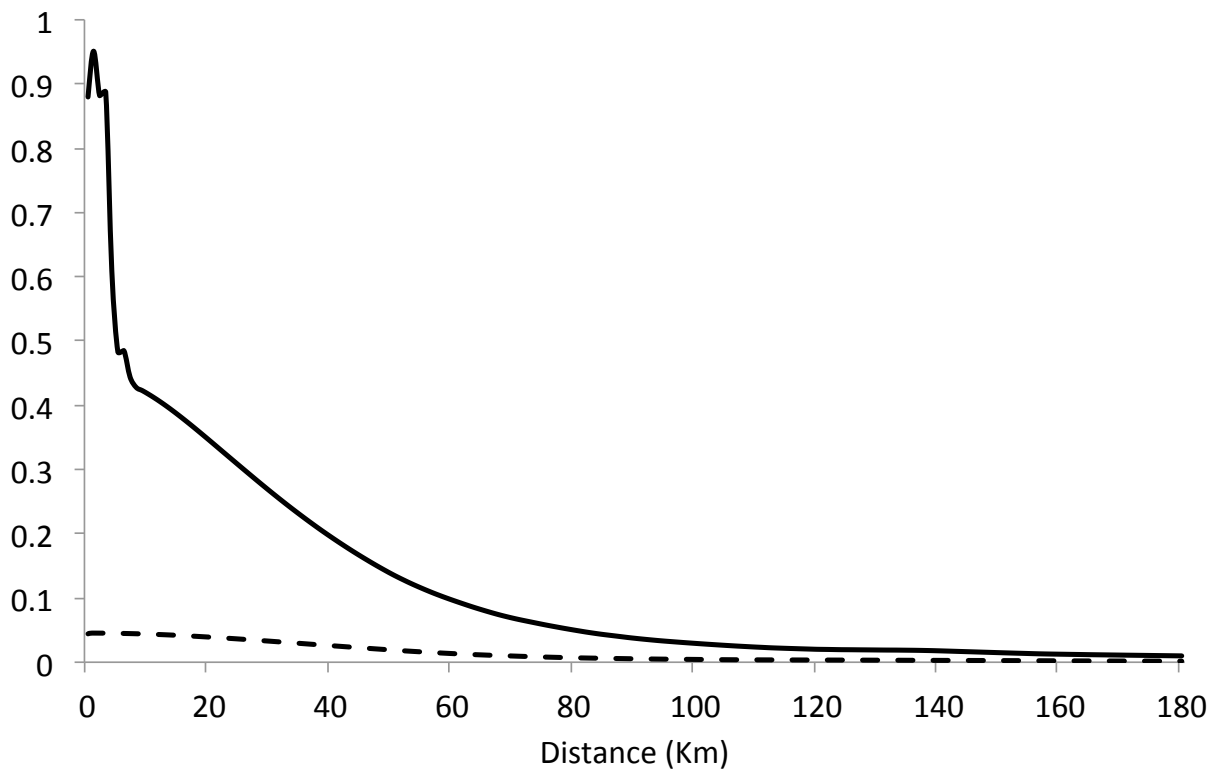


Figure 10: Localization and dispersion indices for each distance: $\Gamma(d)$ and $\Psi(d)$ in the service sector

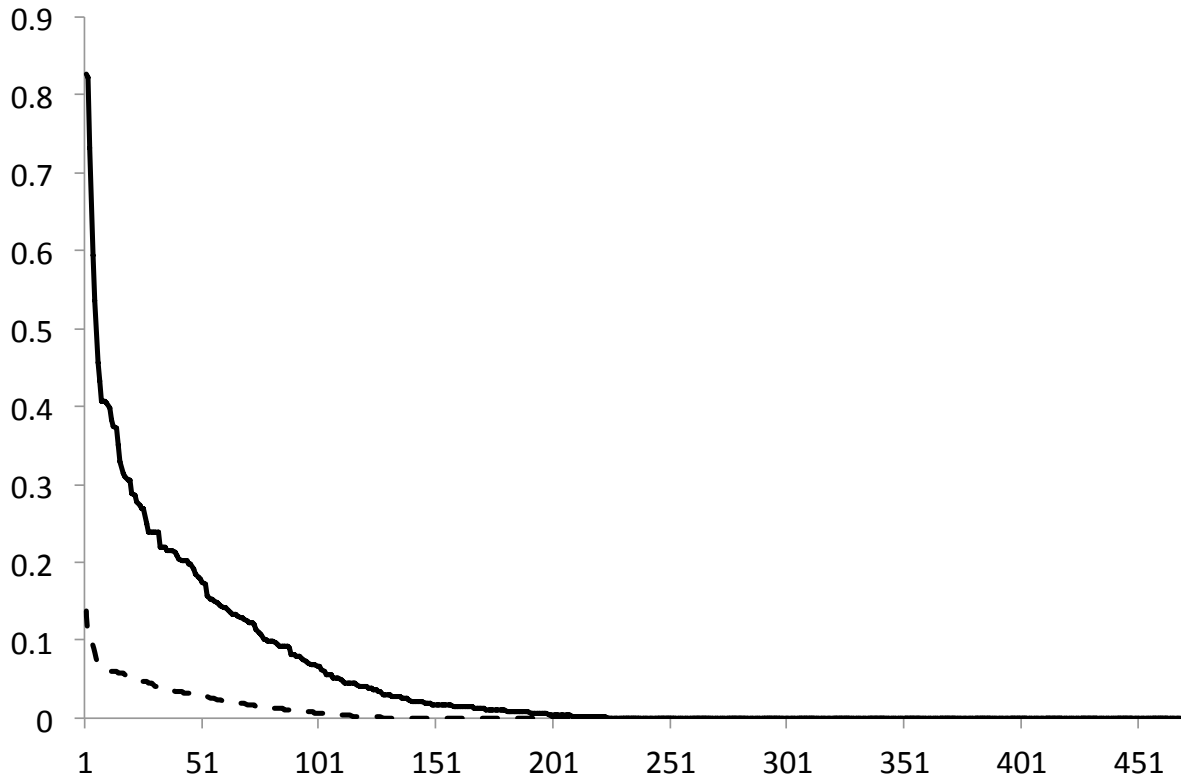


Figure 11: Rank-order distribution of localization and dispersion indices for each industry: Γ_A and Ψ_A in the service sector

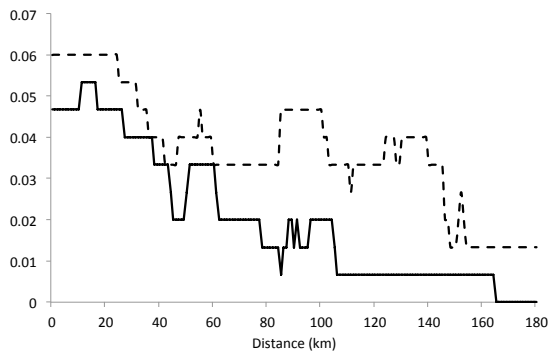


Figure 12: Share of localized and dispersed industries for entrant firms

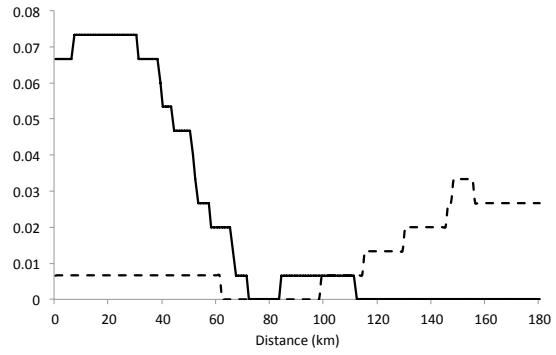


Figure 13: share of localized and dispersed industries for exiting firms

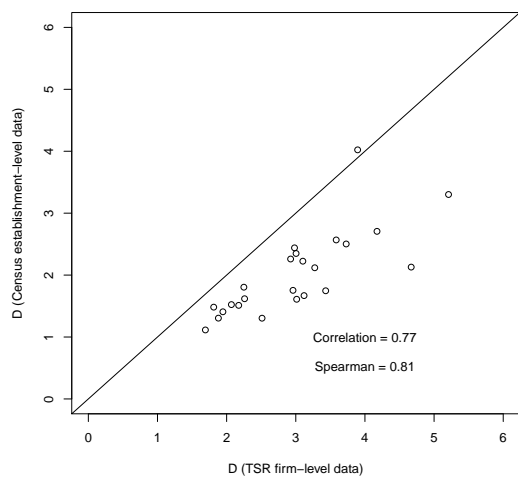
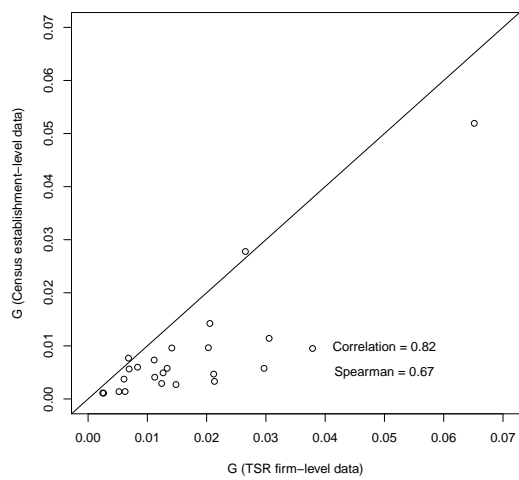


Figure A 1: G -index based on census establishment-level data vs. TSR firm-level data
 Figure A 2: D -index based on census establishment-level data vs. TSR firm-level data