Do More Productive Firms Locate New Factories in More Productive Locations? An Empirical Analysis Based on Panel Data of Japan's *Census of Manufactures*

Kyoji Fukao

(Hitotsubashi University, RIETI)

Kenta Ikeuchi

(NISTEP)

YoungGak Kim

(Senshu University)

Hyeog Ug Kwon

(Nihon University, RIETI)

Abstract

Using a Melitz-style model of heterogeneous firms, Baldwin and Okubo (2006) recently presented a theoretical model in which self-sorting occurs and more productive factories choose to locate in more productive areas. The model suggests that firm-specific factors and regional factors affect each other through the endogeneity of location decisions. However, to date there are few studies empirically testing this issue. Against this background, our aim is to examine the relationship between firm and location-specific factors in location decisions using factory-level panel data from Japan's *Census of Manufactures*. We begin by estimating how much of the differences in factories' TFP levels can be explained by firm and by location effects. The estimation results show that both effects have a significant impact on the productivity level of a factory, and that the firm effects are more important than the location effects. We also find a statistically significant negative correlation between firm effects and location effects, and investigate what causes this relationship. One potential explanation is that more productive firms may tend to set up new factories in less productive locations such as rural areas, where factor prices such as land prices and wage rates are usually low, in order to benefit from low factor prices. To examine this, we estimate a mixed logit model of location choice. The results indicate that more productive firms indeed tend to set up new factories in low-productivity locations, which is consistent with our hypothesis.

Key words: TFP, firm effects, location effects

JEL classification: R30, D24, O53

1. Introduction

The literature on productivity shows that there are large differences in the productivity of factories even in a narrowly defined, highly homogeneous industry (see Bartelsman and Doms, 2000). Many researchers on productivity have looked for firm- or factory-specific factors which may be responsible for such productivity differences, such as human capital, capital vintage, and the characteristics of the firm (in terms of R&D, IT, FDI, exports, etc.) to which the factory belongs. On the other hand, many researchers on regional economics have looked for regional factors which may explain the differences in productivity among different areas, such as agglomeration effects due to local industry-specific knowledge spillovers and natural cost advantages (e.g., Ciccone and Hall, 1996, and Henderson, 2003). Both groups of researchers, however, usually fail to incorporate the perspective of the other group. The former does not take into account location factors, while the latter usually does not fully take account of the characteristics of the firm to which a factory belongs.

Using a Melitz-style model of heterogeneous firms, Baldwin and Okubo (2006) recently presented a theoretical model in which self-sorting occurs and more productive factories choose to locate in more productive areas. Their result suggests that firm-specific factors and regional factors affect each other through the endogeneity of location decisions.

Despite the importance of this issue, there are few empirical studies on this topic, probably because of a lack of appropriate data.[1] Against this background, our aim is to examine this issue using factory-level panel data of Japan's *Census of Manufactures*.[2]

First, we decompose factories' TFP levels into firm effects, location effects, and factory-specific characteristics, such as size and age. Next, we investigate the characteristics of our estimated firm effects and location effects by calculating the coefficients of correlation between these effects and several firm- and location-specific characteristics. Based on the estimated firm and location effects, we examine how much of the total variation in TFP levels across factories can be explained by firm effects and by location effects and test whether more productive firms tend to have factories in more productive locations. We also estimate location choice models and test whether more productive firms tend to open up new factories in more productive locations.

The structure of the paper is as follows. The next section introduces our methodology for estimating firm and location effects on TFP, explains data sources and the construction of variables, and presents our estimation results of firm and location effects. In Section 3, we examine the characteristics of the estimated firm and location effects by calculating the coefficients of correlation

[1] Using start-up data of foreign-owned factories in the United States, Shaver and Flyer (2000) have shown that productive foreign-owned firms tend to locate their activities in less productive regions. Shaver and Flyer suggest that the reason is that they receive fewer net benefits from agglomeration and technology spillovers from other firms.

[2] We were able to gain access to the micro data of the *Census of Manufactures* and the *Basic Survey of Japanese Business Structure and Activities* as part of our research project on the "Firm and Industry Level Analysis of Productivity" at the Research Institute of Economy, Trade and Industry (RIETI).

between these effects and various firm and location characteristics. In addition, we conduct an analysis of variance in order to examine the relative importance of the two effects. In Section 4, we then test whether firm and location effects are positively correlated. We also estimate models of firms' location choices. Section 5 summarizes our results and discusses remaining issues to be investigated in future research.

2. Estimation of Firm and Location Effects on Factories' TFP

We begin our examination of the role of location and firm effects on factory-level TFP by presenting our methodology for measuring TFP and isolating the two effects from other, factory-specific effects. This is followed by an explanation of our data sources and the variables used. Finally, this section presents our estimation results of firm and location effects.

Let us start with our methodology for estimating how much of the differences in factories' TFP levels can be explained by firm effects and how much by location effects. The TFP level in year t of factory i, which belongs to firm f and is located in location l, is assumed to be determined by the following factors: factory age and size, firm effects (measured by a dummy for the firm to which the factory belongs), location effects (measured by a location dummy at the city/ward/town/village level), industry specific effects (measured by a dummy for the industry to which the factory belongs), and year effects:

$$TFP_{i,t} = G\left(Age_{i,t}, Scale_{i,t}, f_{i,t}, l_{i,t}, t\right) \tag{1}$$

Specifically, we estimate the following econometric model:

$$\begin{aligned} \ln TFP_{i,t} &= \beta_{a1,t} Age_{i,t} + \beta_{s1,t} Scale_{i,t} \\ &+ \sum_{l'} \beta_{l',t} DL_{l'}\left(l_{i,t}\right) + \sum_{f'} \beta_{f',t} DF_{f'}\left(f_{i,t}\right) + \sum_{j'} \beta_{j',t} DI_{j'}\left(j_{i,t}\right) + \sum_{t'} \beta_{t'} DY_{t'}(t) + R_{i,t} \end{aligned} \tag{2}$$

where $Age_{i,t}$ is the age of factory i in year t, $Scale_{i,t}$ is the number of employees of this factory in year t, $l_{i,t}$ is the location of factory i, and $f_{i,t}$ denotes the firm to which factory i belongs in year t. $DL_{l'}$ is a dummy variable for location l' which takes value one if the location of factory i, l_i, is equal to l', otherwise this dummy variable takes value zero. Similarly, $DF_{f'}$ is a dummy variable for firm f' which takes value one if the firm to which factory i belongs, $f_{i,t}$, is equal to f', otherwise this dummy variable takes value zero, $DI_{j'}$ is a dummy variable for industry j' which takes value one if the industry to which factory i belongs, $j_{i,t}$ is equal to j', otherwise this dummy variable takes value zero,[3] and $DY_{t'}$ is a dummy variable for year *t'* which takes value one if the observation year *t* is equal to *t'*, otherwise this dummy variable takes value zero. $R_{i,t}$ is the residual term. The coefficient $\beta_{l',t}$ shows the location effect of location *l*' in year *t* on factories' TFP level, while the coefficient $\beta_{f',t}$ shows the firm effect of firm *f*' in year *t* on factories' TFP level.

When we estimate firm and location effects, we ran three-year rolling panel regression. For example, we estimate values of year *t* using data of *t*-2, *t*-1, and *t*. We take this approach in order to avoid endogeneity problems in our estimation of a location choice model. We do not include productivity data of newly set-up factories in our estimation of the firm and the location effects, both of which are used in the estimation of a location choice model.

We estimate firm and location effects using equation (2). In order to take account of the possibility that firm and location effects may differ across industries, we estimated equation (2) separately for the following six manufacturing subsectors, as shown Table 1: materials, chemicals, general machinery, electric machinery, transportation machinery, and miscellaneous products.

Insert Table 1

We calculate the relative TFP level of each factory vis-à-vis the industry average TFP level. Following Good, Nadiri, and Sickles (1997), we measure the TFP level of factory *i* in year *t* in a certain industry in comparison with the TFP level of a hypothetical representative factory in year *t* in that industry using the following equation:

$$\ln TFP_{i,t} = (\ln Q_{i,t} - \overline{\ln Q_t}) - \sum_{n=1}^{N} \frac{1}{2}(s_{n,i,t} + \overline{s_{n,t}})(\ln X_{n,i,t} - \overline{\ln X_{n,t}}) \tag{3}$$

where $Q_{i,t}$ is the gross output of factory *i* in year *t*, $s_{n,i,t}$ is the cost share of the *n*-th input, and $X_{n,i,t}$ is the amount of the *n*-th input at factory *i* in year *t*. Variables with upper bars denote the arithmetic mean of each variable over all factories in that industry in year *t*. Three inputs, labor, capital, and intermediate input, are taken into account in our analysis.

The main data source for this paper is the longitudinal data of the *Census of Manufactures* conducted by the Ministry of Economy, Trade and Industry (METI). This census covers all manufacturing factories with four or more employees.[4] Since 1997, the data include information on factories' affiliation, so that we can group factories according to their parent firms, although the data

[3] Industry dummy variables are based on JIP industry classification which divides the whole manufacturing sector into 52 subsectors.

[4] Factories with three or fewer employees are included in specific years, starting with the 1981 survey and then in years ending with 0, 3, 5, and 8.

do not include detailed information on parent firms. We used data for the period of 1997-2007.

Gross output is measured as the sum of shipments, revenues from repairing and fixing services, and revenues from performing subcontracted work. Intermediate inputs are defined as the sum of raw materials, fuel, electricity and subcontracting expenses for consigned production used by the plant. Using industry level price deflators taken from the Japan Industrial Productivity Database (JIP) 2010, Gross output and intermediate inputs are converted to values in constant prices of 2000.

Following preceding studies using micro data of the *Census of Manufactures*, we measure labor input in terms of man-hours, calculated as the product of the number of employees and the industry average of annual working hours. The underlying assumption is that the labor service provided by one hour of work is the same for all workers in a particular industry.[5]

Capital input is measured as real capital stock, which is defined as the product of the nominal book value of tangible fixed assets (taken from the *Census of Manufactures*) and the book-to-market value ratio for each industry, which is calculated using industry level investment data and the book value of industry-level capital stock from the *Census of Manufactures*.

Labor costs are defined as total salaries and intermediate costs are defined as the sum of raw materials, fuel, electricity and subcontracting expenses for consigned production provided in the *Census of Manufactures*, respectively. Capital costs are calculated by multiplying the real net capital stock with the user cost of capital. The latter is estimated as follows:

$$c_k = \frac{1-z}{1-u} p_k \{(1-u)i + \delta_j - (\frac{\dot{p}_k}{p_k})\}$$

where p_k, i, δ, u and z are the price of investment goods, the interest rate, the depreciation rate, the corporate tax rate, and the present value of depreciation deduction on a unit of nominal investment, respectively. Data on investment goods prices, interest rates, and corporate tax rates are taken from the JIP 2010, the *Bank of Japan* and the *Ministry of Finance Statistics Monthly*. The depreciation rate for each sector is taken from the JIP 2010. We measure the cost share of each factor by dividing the costs of each factor by total costs, which is the sum of labor costs, intermediate input costs, and capital costs. Figure 1 plots sectoral average values of TFP level over the 11-year period 1997-2007. It shows that the largest increases in relative TFP level occurred in electrical machinery.

Insert Figure 1

[5] We should note that if the labor service provided by one hour of work differs across regions, our estimates of location effects will be biased. For example, factories in a certain ward of the Tokyo metropolitan area might employ more skilled workers than factories in other areas. If we do not take account of this difference and measure labor input by man-hours, we will overestimate the TFP level of factories in this ward.

In the case of independent factories, we cannot decompose the TFP level into the firm effect and the location effect. Similarly, in the case of locations with just one factory for the whole observation period, we cannot decompose the TFP of this factory into the firm effect and the location effect. In the case of locations with only a small number of factories, or in the case of firms that only have a small number of factories, our estimates of firm and location effects are likely to be unreliable. To obtain reliable results, we excluded observations of factories belonging to firms with fewer than three observations or observations of factories located in cities/wards/towns/villages with fewer than three observations. We should note that our approach is not free from sample selection bias problems.

Table 2 shows total number of factories, number of factories, of which TFP were measured, and number of factories, of which TFP data used for measuring firm effects and location effects for each year. The original data set for this paper consists of 3,786,975 (factory times year) observations. We can calculate the TFP level for 909,005 observations. We cannot derive the TFP level for the other 2,877,970 observations (most of them are for small factories) mainly because of the absence of information on capital stocks.[6] The number of observations used for the estimation is 189,270. Table 3 shows total number of cities/wards/towns/villages and number of cities/wards/towns/villages, of which location effects are estimated.

Insert Tables 2 and 3

Table 4 shows the descriptive statistics of this estimation. 51,130 location effects and 119,636 firm effects were estimated, and the standard deviations of firm effects and of location effects are almost of the same size.

Insert Table 4

3. The Characteristics of Firm and Location Effects and Their Relative Importance

In order to examine the characteristics of the estimated firm and location effects, we calculated the coefficients of correlation between these effects and several firm- and location-specific characteristics. The results are shown in Tables 5 and 6. In Table 5, correlation coefficients are calculated across firms, while in Table 6, correlation coefficients are calculated across locations.

[6] This problem is more serious for years after 2001. The reason is that from 2001, the *Census of Manufactures* stopped collecting capital stock data for factories with 29 or fewer employees in non-benchmark years.

Insert Tables 5 and 6

Starting with the correlation between firm effects and other firm characteristics (Table 5), we find a positive correlation between firm effects and the average TFP level of the firm.[7] Firm effects are also positively correlated with factories' gross output as well as the number of factories a firm owns. That is, larger firms tend to generate larger positive firm effects. These results are consistent with Adams and Jaffe (1996) which found a strong correlation between plant-level productivity and the number of plants of a firm.

One caveat with regard to these results, however, is that by relying on the *Census of Manufactures*, which only provides information on manufacturing establishments, the calculation of firms' TFP level does not fully cover firms' headquarter activities and non-manufacturing activities. Thus, in order to examine the correlation with the TFP level of firms' total activities, we also calculated firms' TFP level using micro-data from the *Basic Survey of Japanese Business Structure and Activities*. Again we find a positive correlation between these two variables.

Turning to the correlations between location effects and other location characteristics (Table 6), we find that location effects are positively correlated with weighted average of TFP level of all the factories in the same location.[8]

In order to measure congestion effects (that is, negative effects of economic agglomeration), we prepared data of two variables; wage premiums and land prices. We get land prices from the *Chiiki-keizai-deta CD-ROM* (*Regional Economy Data CD-ROM*) published by Toyo Keizai. We use regional wage premiums obtained by Daiji Kawaguchi and Ryo Kambayashi based on micro data of the *Basic Survey on Wage Structure* as part of their background analysis for a recent study (Kawaguchi and Kambayashi, 2009).[9]

We find that location effects are positively correlated with the regional wage premium, and the

---

[7] We calculate the average log value of the TFP of firm *f'* in year *t*, $lnTFP_{f',t}$, as a weighted average of the log value of the TFP level of all the establishments which belong to this firm:

$$\ln TFP_{f',t} = \sum_{f(i)=f'} \frac{sales_{i,t}}{\sum_{f(i')=f'} sales_{i',t}} \ln TFP_{i,t}$$

[8] We calculate the average TFP of location *l'*, $lnTFP_{l',t}$, as a weighted average of the log values of the TFP level of all the factories located in this location:

$$\ln TFP_{l',t} = \sum_{l(i)=l'} \frac{sales_{i,t}}{\sum_{l(i')=l'} sales_{i',t}} \ln TFP_{i,t}$$

where $sales_{i,t}$ denotes the total sales of factory *i* in year *t*.

[9] Kawaguchi and Kambayashi estimated regional wage premiums by estimating a Mincer-type wage function with each worker's educational attainment, work experience (defined as age minus years of education minus 6), tenure, quadratic terms of work experience and tenure, factory size, city dummies, and industry dummies as explanatory variables, using all the survey data for full-time male workers across all industries and for all of Japan.

average land price of that location. That is, location effects tend to be greater in areas with higher wage rates and higher land prices.

Next, in order to examine the relationship between location effects and industry agglomeration, we prepared three sets of indices of industry agglomeration for each subsector of each location; total shipment of subsector *j* in location *l*, total number of factories in subsector *j* of location *l*, and density of industry agglomeration ((total shipment of subsector *j* in location *l*)/(square kilometers of location *l*)).

We find a positive correlation between location effects and the three indices of industry agglomeration. It seems that industry agglomeration has a positive effect on factories' productivity.

To sum up our results, both the estimated firm effects and the location effects have plausible characteristics. For example, larger firms tend to generate larger positive firm effects, and firm effects are positively correlated with the average TFP level of all the factories of this firm. Location effects are positively correlated with the average TFP level of all factories in a particular location, and location effects tend to be greater for locations with higher industry agglomeration level.

The next question we address is how much of the total variation in TFP levels across factories can be explained by firm effects and how much by location effects. In order to answer this question, we conduct an analysis of variance (ANOVA). ANOVA provides a measure of the fit of the regression of the contribution of each variable by measuring how well the variation in each independent variable predicts the variation in the dependent variable.

The ANOVA results are shown in Table 7. In the table, "partial sum of squared deviations" denotes how much the variation of each variable contributes to the total variation (total sum of squared deviations) of the dependent variable, that is, each factory's TFP level. The results show that, in all six manufacturing subsectors, both location and firm effects are important in explaining factories' productivity level. About 40-50 percent of the total variation can be explained by these two effects. The table also shows that in all the manufacturing subsectors, the partial sum of squared deviations of the firm effects is greater than the partial sum of squared deviations of the location effects. Thus, to which firm a factory belongs is a more important determinant of this factory's TFP level than in which location this factory is located.

Insert Table 7

4. Do More Productive Firms Set Up Their Factories in More Productive Locations?

In this section, we examine the relationship between firm and location effects. Above, it was suggested that firm effects on factory-level productivity have a positive correlation with firm size and the average TFP level of all the factories of this firm. Similarly, location effects have a positive correlation with the average TFP level of all the factories in a particular location and the location's industry agglomeration. This raises the question whether "self-sorting" occurs in that more productive firms set up their factories in more productive locations.

However, before conducting our analysis of location decisions, we examine the static correlation between firm and location effects across factories. The results are shown in Table 8 and in Figure 2, indicating that there is a statistically significant negative correlation between firm and location effects for all six manufacturing subsectors.

Insert Table 8 and Figure 2

What causes this negative relationship? One potential explanation is that more productive firms may tend to set up new factories in less productive locations such as rural areas, where factor prices such as land prices and wage rates are usually low, in order to benefit from low factor costs. On the other hand, less productive firms may be unable to locate new factories in rural areas because of their inability to solve logistical problems, which are common in rural areas.

To determine the cause of the negative relationship, we calculate correlation coefficients across factories between firm effects and location characteristics. We use two variables for location characteristics: wage premiums, and land prices.[10] Table 9 shows the results, which indicate that there is no clear pattern suggesting that more productive firms tend to place their factories in locations with low wage premiums or low land prices. While a negative and significant correlation between firm effects and land price can be observed in the general machinery industry, the correlations between firm effects and wage premiums or land prices in all other subsectors are either insignificant or actually significantly positive.

Insert Table 9

So far, our analysis has focused on factories of all ages. However, many of the factories in our sample were set up a long time ago and the regional characteristics on which location decisions were originally based may have changed since the establishment of these factories. In order to take this into account, we also examine the correlation between firm effects for newly opened factories and regional characteristics in the year they were established.

To do so, we estimate the following mixed logit model of location choice:

[10] Examining the correlation between these variables across locations, we find a high positive correlation between wage premiums and land prices.

$$\Pr(y_{f,t} = l) = \frac{\exp(\beta' z_{f,l,t})}{\sum_{l'=1}^{L} \exp(\beta' z_{f,l',t})} \qquad for \quad l = 1...L \; and \; t = 1....T$$

where the left-hand side denotes the probability of observing the establishment of a factory in year *t* by firm *f* in location *l*, $z_{f,l,t}$ denotes a vector including characteristics of location *l* and cross terms between characteristics of firm *f* and characteristics of location *l* as its elements, and β is the coefficient vector.

The number of newly opened factories, of which it is possible to calculate firm and location effects, and their distribution over time are shown in Table 10. The numbers are not very large, partly because Japan's manufacturing sector has shrunk rapidly during this period and partly because we can calculate both firm and location effects only for a fraction of newly established factories.

Insert Table 10

Correlation coefficients between the location variables used in the estimation are reported in Table 11. High correlation coefficients suggest that there is a risk of multi-collinearly problems. Because of this risk, we do not use all the location variables within one equation.

Insert Table 11

The results of our estimation of the mixed logit model of location decision are presented in Table 12. Specification 1 represents the baseline estimation and as independent variables only includes the location effect, one of the location variables and the number of factories of the same firm in the same location.[11]. The estimated coefficients on the location effect are positive in most cases and significant. This result provides a strong evidence suggesting that firms tend to prefer more productive locations. Table 12.2 shows the same estimation results with the sample of the machinery industries (including General Machinery, Electric Machinery, and Transportation Machinery), which is almost same to that of Table 12.1.

Insert Table 12.1 and Table 12.2

In the estimation in Table 13, we added a cross term of location effects and firm effects. The estimated coefficients on the cross term of firm and location effects are negative and significant. This

[11] In all the estimation, we observe that the presence of factories of the same firm in the same prefecture has very strong effects on the probability of the selection of that location.

result implies that more productive firms tend to be less attracted to more productive locations. This finding is consistent with the results of our analysis of the correlation between firm and location effects across factories. In Table 13 we also estimated a model with industry agglomeration and its cross term with firm effect. The results show that industry agglomeration has positive impact on the location choice, and that it is more important for the productive firms in machinery industries.

Insert Table 13.1 and Table 13.2

As already mentioned, one possible explanation of the negative correlation between firm and location effects is that more productive firms tend to set up their factories in less productive locations, such as rural areas, so as to exploit the lower factor prices there. Moreover, they can do so because they are able to overcome the logistical problems that may be associated with locating production in rural areas – something that typically smaller less productive firms cannot do. If this hypothesis is correct, high factor prices should have a negative effect on firms' location decision. To test this, we added factor price variables such as the regional wage premium and land prices to our specification of Table 13.

The results are reported in Table 14. Against expectation, we do not find any statistically significant negative effects of factor prices on location decisions in specifications 1 and 2 of Table 14. One possible explanation for this result is that, because the equation does not include sufficient variables to control for the positive effects of agglomeration, the estimate for the local wage premium picks up these effects.

Another question of considerable interest is what causes location effects. As many studies in the field of economic geography have argued, one potential factor is the industry agglomeration. To test the importance of agglomeration effects for location decisions, we replaced the location effect variable with industry agglomeration in our analysis of location choice. If more productive firms are less attracted by industry agglomeration, a negative coefficient for the cross term of firm effects and industrial agglomeration is expected. The results are reported as specifications 3 and 4 in Table 14. We find that industry agglomeration has a positive and statistically significant effect on location decision in all industries. But we do not find negative and significant coefficient for the cross term of firm effects and industry agglomeration.

In specifications 5 and 6, we test another possibility that location effect and industry agglomeration captures pretty different aspects of the attraction of locations by adding the industry agglomeration to the location effects and its cross term with firm effects as an explanatory variable. The table shows that controlling such regional merits as location effects estimated above, industry agglomeration, and presence of factories of the same firm, factor prices such as wage premium and land price have negative effect on the choice. This result is consistent with our hypothesis.

Insert Table 14.1 and 14.2

5. Conclusion

Using micro data of Japan’s *Census of Manufactures*, we decomposed the TFP level of each factory into firm effects, location effects, and other, factory specific factors such as size and age. Both the estimated firm effects and the estimated location effects have plausible characteristics. That is, larger firms tend to generate larger positive firm effects. and firm effects are positively correlated with the average TFP level of all the factories of this firm. Location effects are positively correlated with the average TFP level of all factories in a particular location, and location effects tend to be positively correlated with industrial agglomeration.

Based on the estimated firm and location effects, we also conducted an analysis of variance and found that both location and firm effects play a role in explaining factories’ TFP levels. In addition, comparing the contribution of the two effects, it was found that both location and firm effects have a statistically significant and large influence on factories’ productivity, and firm effects were the more important of the two.

Next, our attention turned to the correlation between firm and location effects, which was negative. That is, more productive firms tended to have factories in cities with weaker location effects. We hypothesized that the reason for this may be that more productive firms set up their factories in less productive locations such as rural areas to benefit from low factor prices there. To test this hypothesis, we estimated a location choice model. Consistent with our hypothesis, the estimation results showed that location effects had a positive effect on location decisions and the cross term of firm and location effects a significant negative effect. However, against our expectation, we did not find a negative and significant effect of factor prices, such as wage premiums and land prices, on location choices.

These findings mean that we need to examine our estimates of firm and location effects further. We found a positive and significant correlation between location effects and industry agglomeration, but the correlation coefficient is not very large. When we replaced the location effect variable in our location choice analysis with industry agglomeration, we found a positive and statistically significant effect on location decisions. However, we did not find a negative and significant coefficient on the cross term, that is, firm effects times industrial agglomeration. These results indicate that we still do not understand well what local characteristics cause positive location effects.

We also tested another possibility that location effect and industry agglomeration captures pretty different aspects of the attraction of locations by adding the industry agglomeration to the location effects and its cross term with firm effects as an explanatory variable. We found that controlling such regional merits as location effects estimated above, industry agglomeration, and presence of factories of the same firm, factor prices such as wage premium and land price have negative effect on the choice. This result is consistent with our hypothesis.

In the case of firm effects, we found a positive and significant correlation between firm size and firm effects, but we were unable to examine how other firm specific factors, such as research and development expenditures, information technology investment, or the accumulation of intangible assets, affect firm effects, because of a lack of such data. We hope that we can investigate these issues in the future by matching our micro data of the *Census of Manufactures* with other firm level data.

There are also several additional issues that still need to be tackled. First, as we have already highlighted, our approach is not free from sample selection bias problems. And second, it is likely that firms make decisions on the location of new factories in the context of their networks already in place. These are factors that a more refined location choice model should address and we hope to develop and test such a model in the future.

References

Adam, J. D. and A. B. Jaffe (1996) "Bounding the Effects of R&D: An Investigation Using Matched Establish-Firm Data," *RAND Journal of Economics*, 27, pp.700-721.

Baldwin, R. E. and T. Okubo (2006) "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting," *Journal of Economic Geography*, 6(3), pp. 323-346.

Bartelsman, E. J. and M. Doms (2000) "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38, pp.569-594.

Ciccone, A. and R. E. Hall (1996) "Productivity and the Density of Economic Activity," *American Economic Review*, 86, pp.54-70.

Henderson, J. V. (2003) "Marshall's Scale Economies," *Journal of Urban Economics*, 53, pp.1-28.

Good, D. H., M. I. Nadiri and R. C. Sickles (1997) "Index Number and Factor Demand Approaches to the Estimation of Productivity," in M.H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics: Vol. 2. Microeconometrics*, Oxford, England: Basil Blackwell, pp. 14-80.

Kawaguchi, D. and R. Kambayashi (2009) "Seifu Tokei no Setsugo Data no Sakusei to Riyo: Kogyo Tokei Chosa to Chingin Kozo Kihon Chosa no Rei [Compilation and Use of Merged Micro-Data of Government Surveys: The Case of the *Census of Manufactures* and the *Basic Survey on Wage Structure*]," forthcoming in Yukinobu Kitamura (ed.), *Ohyo Mikuro Keiryo Keizai-gaku* [*Applied Micro-Econometrics*], Tokyo, Japan: Nihon Hyoron-sha.

Shaver, J. M. and F. Flyer (2000) "Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States," *Strategic Management Journal*, 21, pp.1175-1193.

Table 1. Industry Classification

JIP industry classification	Industry classification of this paper
8 Livestock products	6 Miscellaneous products
9 Seafood products	6 Miscellaneous products
10 Flour and grain mill products	6 Miscellaneous products
11 Miscellaneous foods and related products	6 Miscellaneous products
12 Prepared animal foods and organic fertilizers	6 Miscellaneous products
13 Beverages	6 Miscellaneous products
14 Tobacco	6 Miscellaneous products
15 Textile products	6 Miscellaneous products
16 Lumber and wood products	1 Material products
17 Furniture and fixtures	1 Material products
18 Pulp, paper, and coated and glazed paper	1 Material products
19 Paper products	1 Material products
20 Printing, plate making for printing and bookbinding	6 Miscellaneous products
21 Leather and leather products	1 Material products
22 Rubber products	1 Material products
23 Chemical fertilizers	2 Chemicals
24 Basic inorganic chemicals	2 Chemicals
25 Basic organic chemicals	2 Chemicals
26 Organic chemicals	2 Chemicals
27 Chemical fibers	2 Chemicals
28 Miscellaneous chemical products	2 Chemicals
29 Pharmaceutical products	2 Chemicals
30 Petroleum products	2 Chemicals
31 Coal products	2 Chemicals
32 Glass and its products	1 Material products
33 Cement and its products	1 Material products
34 Pottery	1 Material products
35 Miscellaneous ceramic, stone and clay products	1 Material products
36 Pig iron and crude steel	1 Material products
37 Miscellaneous iron and steel	1 Material products
38 Smelting and refining of non-ferrous metals	1 Material products
39 Non-ferrous metal products	1 Material products
40 Fabricated constructional and architectural metal products	1 Material products
41 Miscellaneous fabricated metal products	1 Material products
42 General industry machinery	3 General machinery
43 Special industry machinery	3 General machinery
44 Miscellaneous machinery	3 General machinery
45 Office and service industry machines	3 General machinery
46 Electrical generating, transmission, distribution and industrial apparatus	4 Electric machinery
47 Household electric appliances	4 Electric machinery
48 Electronic data processing machines, digital and analog computer equipment and accessories	4 Electric machinery
49 Communication equipment	4 Electric machinery
50 Electronic equipment and electric measuring instruments	4 Electric machinery
51 Semiconductor devices and integrated circuits	4 Electric machinery
52 Electronic parts	4 Electric machinery
53 Miscellaneous electrical machinery equipment	4 Electric machinery
54 Motor vehicles	5 Transportation machinery
55 Motor vehicle parts and accessories	5 Transportation machinery
56 Other transportation equipment	5 Transportation machinery
57 Precision machinery & equipment	3 General machinery
58 Plastic products	6 Miscellaneous products
59 Miscellaneous manufacturing industries	6 Miscellaneous products

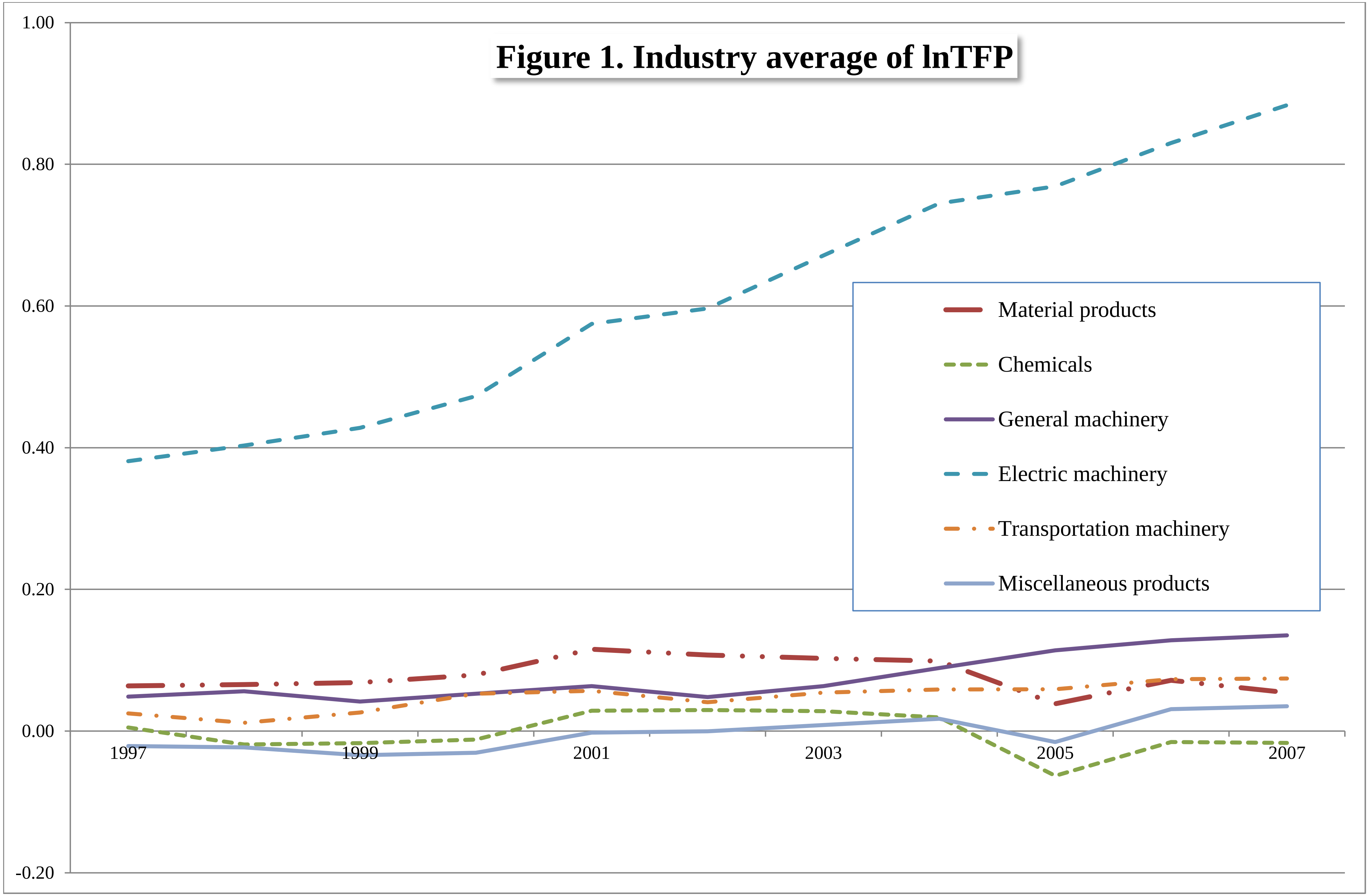

Figure 1. Industry average of lnTFP
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
1997
1999
2001
2003
2005
2007
Material products
Chemicals
General machinery
Electric machinery
Transportation machinery
Miscellaneous products

Table 2. Number of Observations

year	Total number of factories	Number of factories, of which TFP were measured	Number of factories, of which TFP data were used for measuring firm effects and location effects
1997	358,246	142,872	7,393
1998	373,713	141,379	14,615
1999	345,457	134,554	22,174
2000	341,421	130,432	22,094
2001	316,267	43,597	17,592
2002	290,848	41,657	16,818
2003	504,503	40,780	16,525
2004	270,905	40,483	17,393
2005	468,840	110,799	21,553
2006	258,543	40,837	15,168
2007	258,232	41,615	17,945
Total	3,786,975	909,005	189,270

Table 3. Number of Cities/Wards/Towns/Villages (Year=2005)

Prefecture		Total number of cities, wards, towns and villages	Total number of cities, wards, towns and villages, of which TFP were measured
1	Hokkai-do	186	108
2	Aomori	40	31
3	Iwate	35	31
4	Miyagi	36	34
5	Akita	25	21
6	Yamagata	35	32
7	Fukushima	59	49
8	Ibaragi	44	44
9	Tochigi	31	31
10	Gunma	38	33
11	Saitama	70	68
12	Chiba	56	52
13	Tokyo	62	52
14	Kanagawa	50	46
15	Niigata	30	29
16	Toyama	15	15
17	Ishikawa	19	17
18	Fukui	17	15
19	Yamanashi	28	23
20	Nagano	81	57
21	Gifu	42	40
22	Shizuoka	38	30
23	Aichi	76	72
24	Mie	29	27
25	Shiga	26	23
26	Kyoto	36	31
27	Osaka	66	62
28	Hyogo	48	47
29	Nara	39	24
30	Wakayama	30	19
31	Tottori	19	16
32	Shimane	21	16
33	Okayama	27	25
34	Hiroshima	30	28
35	Yamaguchi	20	19
36	Tokushima	24	19
37	Kagawa	17	17
38	Ehime	20	19
39	Kouchi	34	19
40	Fukuoka	76	68
41	Saga	20	18
42	Nagasaki	23	15
43	Kumamoto	47	39
44	Oita	18	16
45	Miyazaki	30	22
46	Kagoshima	45	30
47	Okinawa	41	20
	Total	1,899	1,569

Table 4. Descriptive Statistics of Location Effects and Firm Effects

（1）Location effects

Industry	Number of observation	Mean	Standard deviation	Min	Median	Max
Material products	11,588	-0.005	0.277	-2.233	0.000	1.589
Chemicals	5,045	-0.056	0.502	-3.194	-0.035	2.629
General machinery	8,476	0.106	0.504	-2.660	0.040	2.572
Electric machinery	8,827	-0.183	0.396	-2.347	-0.178	1.901
Transportation machinery	5,090	-0.001	0.297	-1.675	0.000	1.138
Miscellaneous products	12,104	-0.018	0.388	-2.529	0.003	2.104
Total	51,130	-0.025	0.404	-3.194	-0.012	2.629

（2）Firm effects

Industry	Number of observations	Mean	Standard deviation	Min	Median	Max
Material products	32,990	-0.076	0.322	-1.917	-0.065	2.208
Chemicals	8,263	0.011	0.561	-2.509	0.000	3.404
General machinery	19,771	-0.010	0.512	-2.789	0.000	3.754
Electric machinery	17,737	-0.135	0.428	-2.536	-0.084	2.032
Transportation machinery	8,752	0.001	0.285	-1.326	0.000	1.445
Miscellaneous products	32,123	-0.080	0.536	-3.257	-0.069	2.573
Total	119,636	-0.063	0.454	-3.257	-0.024	3.754

Table 5. Correlation Coefficients between Firm Variables

	Variables	1	2	3	4	5	6
1	Firm effects	1					
2	Σ shipment of the firm	0.134	1				
3	Number of employees	0.051	0.851	1			
4	Number of factories of the firm	0.022	0.465	0.512	1		
5	Weighted average of ln*TFP* of the firm	0.298	0.258	0.170	0.040	1	
6	lnTFP derived from Basic Survey Data*	0.076	0.329	0.233	-0.008	0.563	1

Note: 1. All the coefficients are statistically significant at 1% level.
2. Basic Survey is *Basic Survey of Japanese Business Structure and Activities* .

Table 6. Correlation Coefficients between Location Variables

	Variables	1	2	3	4	5	6	7
1	Location effects (by location and by subsector)	1						
2	ln(Total shipment of the same subsector in the location)	0.087	1					
3	ln(Number of factories in the same subsector in the location)	0.081	0.637	1				
4	Wage premium	0.059	0.265	0.158	1			
5	ln(Land price of the city)	0.061	0.248	0.257	0.567	1		
6	Weighted average of ln*TFP* of factories in the same subsector of the location	0.065	0.197	-0.076	0.107	0.067	1	
7	ln(Industry agglomeration)	0.088	0.855	0.579	0.365	0.504	0.184	1

Note: All the coefficients are statistically significant at 1% level.

Table 7. Results of ANOVA

	Source	Partial usm of	Degrees of	Mean square	F	Prob > F
Material products	Model	2,687.7	6860	0.392	11.0	0
	Age of the factory	2.8	1	2.795	78.8	0
	ln(number of employees)	3.8	1	3.755	105.9	0
	Location	338.9	1335	0.254	7.2	0
	Firm	1,920.0	5498	0.349	9.8	0
	Industry (JIP)	7.6	15	0.507	14.3	0
	Year	35.7	10	3.574	100.8	0
	Residual	2,079.1	58610	0.035		
	Total	4,766.8	65470	0.073		
Chemicals	Model	1,626.9	1912	0.851	17.7	0
	Age of the factory	1.8	1	1.781	37.0	0
	ln(number of employees)	0.2	1	0.208	4.3	0.0379
	Location	363.9	667	0.546	11.3	0
	Firm	940.5	1225	0.768	15.9	0
	Industry (JIP)	10.3	8	1.285	26.7	0
	Year	11.1	10	1.108	23.0	0
	Residual	696.3	14448	0.048		
	Total	2,323.2	16360	0.142		
General machinery	Model	1,340.6	4169	0.322	10.3	0
	Age of the factory	0.9	1	0.872	28.0	0
	ln(number of employees)	1.1	1	1.115	35.9	0
	Location	254.4	1014	0.251	8.1	0
	Firm	902.1	3139	0.287	9.2	0
	Industry (JIP)	0.5	4	0.114	3.7	0.0053
	Year	12.1	10	1.208	38.8	0
	Residual	888.5	28567	0.031		
	Total	2,229.1	32736	0.068		
Electric machinery	Model	2,289.8	3935	0.582	15.1	0
	Age of the factory	0.0	1	0.017	0.4	0.5104
	ln(number of employees)	0.6	1	0.564	14.7	0.0001
	Location	284.9	1070	0.266	6.9	0
	Firm	963.8	2846	0.339	8.8	0
	Industry (JIP)	1.8	7	0.257	6.7	0
	Year	343.1	10	34.306	890.8	0
	Residual	1,121.1	29112	0.039		
	Total	3,410.9	33047	0.103		
Transportation machinery	Model	490.7	1906	0.257	9.7	0
	Age of the factory	0.2	1	0.198	7.4	0.0064
	ln(number of employees)	2.8	1	2.772	104.0	0
	Location	144.2	655	0.220	8.3	0
	Firm	269.9	1237	0.218	8.2	0
	Industry (JIP)	2.2	2	1.123	42.1	0
	Year	2.9	10	0.287	10.8	0
	Residual	407.1	15268	0.027		
	Total	897.8	17174	0.052		
Miscellaneous products	Model	4,134.1	6393	0.647	18.6	0
	Age of the factory	6.1	1	6.138	176.3	0
	ln(number of employees)	4.4	1	4.393	126.1	0
	Location	415.3	1380	0.301	8.6	0
	Firm	2,605.0	4992	0.522	15.0	0
	Industry (JIP)	2.9	9	0.318	9.1	0
	Year	4.6	10	0.459	13.2	0
	Residual	1,912.1	54904	0.035		
	Total	6,046.1	61297	0.099		

Table 8. Correlation Coefficients between Location Effects and Firm Effects

Industry	Correlation Coeffcients
Material products	-0.438
Chemicals	-0.491
General machinery	-0.644
Electric machinery	-0.667
Transportation machinery	-0.620
Miscellaneous products	-0.205

Note: All the coeffcients are statistically significant at 1% level.

Figure 2. Scatter Diagrams of Location Effects and Firm Effects

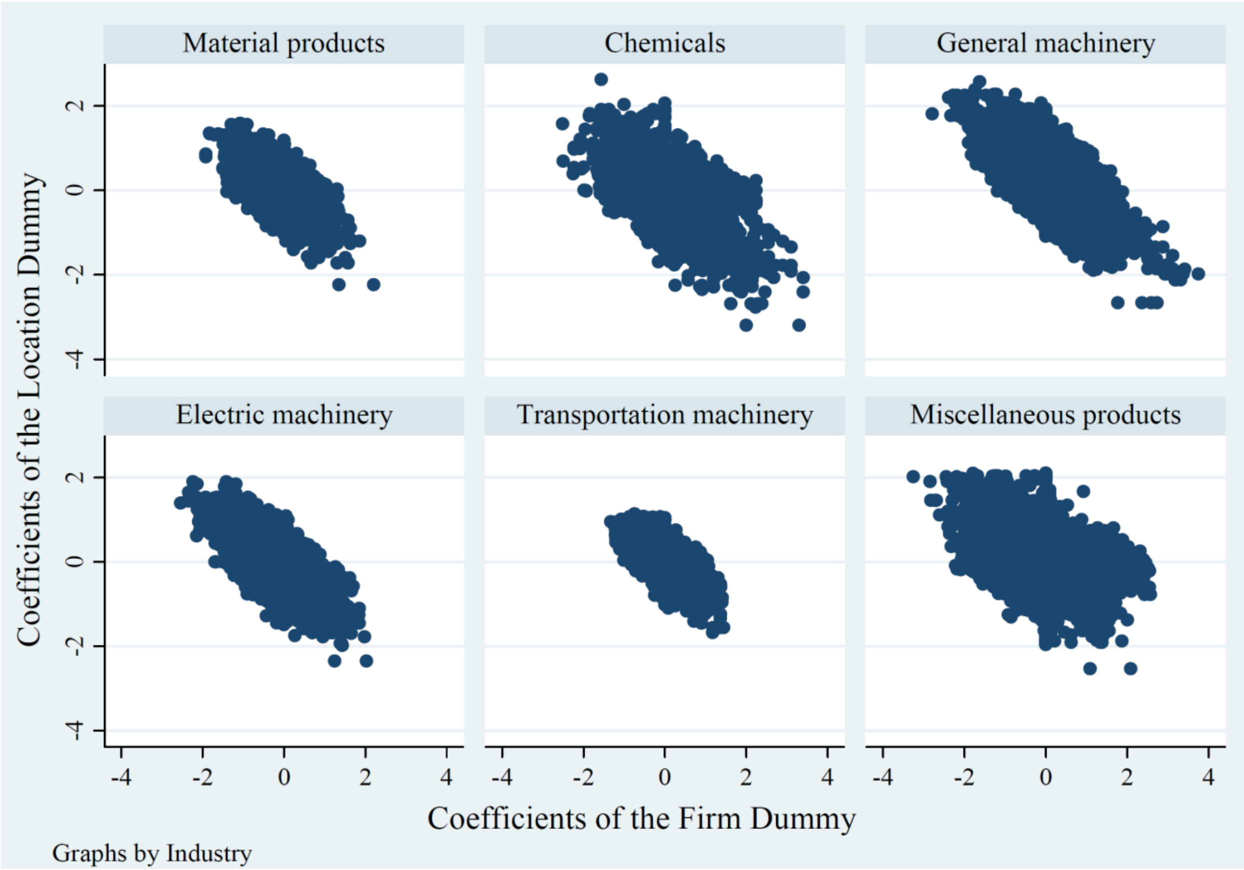

Table 9. Correlation Coefficients between Firm Effects and Location Variables

	Variables	1	2	3
Material products	1 Firm effects	1		
	2 Wage premium	0.0189***	1	
	3 Land price of the city	0.0120**	0.5709***	1
Chemicals	1 Firm effects	1		
	2 Wage premium	0.010	1	
	3 Land price of the city	0.0171**	0.4963***	1
General machinery	1 Firm effects	1		
	2 Wage premium	0.0111*	1	
	3 Land price of the city	-0.0183***	0.06145***	1
Electric machinery	1 Firm effects	1		
	2 Wage premium	0.0483***	1	
	3 Land price of the city	0.0316***	0.6373***	1
Transportation machinery	1 Firm effects	1		
	2 Wage premium	0.0156*	1	
	3 Land price of the city	0.0585***	0.5432***	1
Miscellaneous products	1 Firm effects	1		
	2 Wage premium	0.0525***	1	
	3 Land price of the city	0.1253***	0.6517***	1

Table 10. Number of Observations of Factory Startups in Location Choice Estimation

Industry	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Total
Material products	113	44	44	34	20	36	36	62	50	87	526
Chemicals	23	11	15	10	13	10	9	10	6	24	131
General machinery	44	19	21	18	19	17	10	21	15	54	238
Electric machinery	50	26	12	21	31	30	35	21	39	29	294
Transportation machinery	23	6	11	2	11	13	25	17	28	36	172
Miscellaneous product	156	38	39	43	44	42	24	47	45	139	617
Total	409	144	142	128	138	148	139	178	183	369	1,978

Notes: In our estimation, we used data of new factories, for which both the location effect data and the firm effect data are available. Therefore, the number of observations is much smaller than the number of all the startups.

Table 11.1 Correlation Coefficient between Locatoin Variables: Manufacturing Sector

Variables	1	2	3	4
1 Location effects	1			
2 Wage premium	0.075	1		
3 ln(Land price of the city)	0.089	0.565	1	
4 Indusrty Agglomeration	0.117	0.371	0.541	1

Table 11.2. Correlation Coefficient between Locatoin Variables: Machinery Sector (General Machinery, Electric Machinery, and Transportation Machinery)

1 Location effects	1			
2 Wage premium	0.061	1		
3 ln(Land price of the city)	0.058	0.582	1	
4 Indusrty Agglomeration	0.073	0.391	0.470	1

Note. All the coeffcients are statistically significant at 1% level.

Table 12.1 Results of Mixed Logit Estmation 1: Manufacturing Sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Location effect	0.146 **				0.122 *	0.128 **	-0.0318
	(0.063)				(0.063)	(0.064)	(0.070)
Wage premium (t-1)		1.14 ***			1.13 ***		
		(0.153)			(0.153)		
ln(Land price of the city) (t-1)			0.169 ***			0.167 ***	
			(0.025)			(0.025)	
Indusrty agglomeration (t-1)				0.405 ***			0.406 ***
				(0.014)			(0.014)
ln(1+Number of factory of the same firm	3.65 ***	3.62 ***	3.61 ***	3.64 ***	3.62 ***	3.61 ***	3.64 ***
in the same prefecture (t-1))	(0.068)	(0.068)	(0.069)	(0.071)	(0.068)	(0.069)	(0.071)
R-squared	0.105	0.107	0.107	0.14	0.107	0.107	0.14
Log-likelihood	-1.26E+04	-1.25E+04	-1.22E+04	-1.12E+04	-1.25E+04	-1.22E+04	-1.12E+04
Number of observations	1,887,285	1,853,868	1,791,607	1,755,951	1,853,868	1,791,607	1,755,951

Table12.2 Results of Mixed Logit Estmation 1: Machinery Sector (General Machinery, Electric Machinery, and Transportation Machinery)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Location effect	0.172 *				0.153	0.156	0.0968
	(0.098)				(0.098)	(0.100)	(0.106)
Wage premium (t-1)		1 ***			0.988 ***		
		(0.272)			(0.273)		
ln(Land price of the city) (t-1)			0.0844 *			0.0825 *	
			(0.045)			(0.045)	
Indusrty agglomeration (t-1)				0.332 ***			0.331 ***
				(0.023)			(0.023)
ln(1+Number of factory of the same firm	4.04 ***	4.03 ***	3.97 ***	4.08 ***	4.03 ***	3.97 ***	4.08 ***
in the same prefecture (t-1))	(0.121)	(0.121)	(0.121)	(0.129)	(0.121)	(0.121)	(0.129)
R-squared	0.132	0.133	0.131	0.163	0.134	0.131	0.163
Log-likelihood	-4.17E+03	-4.15E+03	-4.05E+03	-3.63E+03	-4.15E+03	-4.05E+03	-3.63E+03
Number of observations	524,215	520,626	499,568	470,761	520,626	499,568	470,761

Note 1. Standard errors are in parentheses.

2. * p<.1, ** p<.05, *** and p<.01

Table 13.1 Results of Mixed Logit Estmation 2: Manufacturing Sector

	(1)	(2)	(3)	(4)
Location effect	3.57 ***	3.44 ***		
	(0.230)	(0.235)		
Location effect*Firm effect	-1.54 ***	-1.49 ***		
	(0.101)	(0.103)		
Indusrty agglomeration (t-1)			0.244 ***	0.261 ***
			(0.062)	(0.063)
Indusrty agglomeration(t-1)*Firm effect			0.076 ***	0.066 **
			(0.028)	(0.028)
ln(1+Number of factory of the same firm		3.65 ***		3.64 ***
in the same prefecture (t-1))		(0.068)		(0.071)
R-squared	0.008	0.113	0.037	0.14
Log-likelihood	-1.39E+04	-1.25E+04	-1.26E+04	-1.12E+04
Number of observations	1,887,285	1,887,285	1,755,951	1,755,951

Table 13.2 Results of Mixed Logit Estmation 2: Machinery Sector (General Machinery, Electric Machinery, and Transportation Machinery)

	(1)	(2)	(3)	(4)
Location effect	4.4 ***	4.26 ***		
	(0.346)	(0.363)		
Location effect*Firm effect	-1.94 ***	-1.9 ***		
	(0.156)	(0.164)		
Indusrty agglomeration (t-1)			0.304 ***	0.297 ***
			(0.095)	(0.097)
Indusrty agglomeration(t-1)*Firm effect			0.019	0.017
			(0.043)	(0.044)
ln(1+Number of factory of the same firm		4.04 ***		4.08 ***
in the same prefecture (t-1))		(0.122)		(0.129)
R-squared	0.017	0.148	0.029	0.163
Log-likelihood	-4.72E+03	-4.09E+03	-4.20E+03	-3.63E+03
Number of observations	524,215	524,215	470,761	470,761

Note 1. Standard errors are in parentheses.
2. * p<.1, ** p<.05, *** and p<.01

Table 14.1 Results of Mixed Logit Estmation 3: Manufacturing Sector

	(1)	(2)	(3)	(4)	(5)	(6)
Location effect	3.41 ***	3.42 ***			3.66 ***	3.67 ***
	(0.236)	(0.238)			(0.255)	(0.259)
Location effect*Firm effect	-1.49 ***	-1.49 ***			-1.66 ***	-1.67 ***
	(0.104)	(0.104)			(0.111)	(0.113)
Indusrty agglomeration (t-1)			0.265 ***	0.326 ***	0.427 ***	0.468 ***
			(0.064)	(0.064)	(0.015)	(0.016)
Indusrty agglomeration (t-1)･Firm effect			0.072 **	0.063 **		
			(0.028)	(0.028)		
Wage premium (t-1)	1.12 ***		-0.635 ***		-0.656 ***	
	(0.153)		(0.183)		(0.183)	
ln(Land price of the city)(t-1)		0.166 ***		-0.247 ***		-0.251 ***
		(0.025)		(0.030)		(0.030)
ln(1+Number of factory of the same firm	3.62 ***	3.61 ***	3.63 ***	3.64 ***	3.62 ***	3.63 ***
in the same prefecture (t-1))	(0.069)	(0.069)	(0.071)	(0.071)	(0.071)	(0.071)
R-squared	0.115	0.115	0.141	0.144	0.148	0.152
Log-likelihood	-1.24E+04	-1.21E+04	-1.12E+04	-1.09E+04	-1.11E+04	-1.08E+04
Number of observations	1,853,868	1,791,607	1,725,699	1,675,350	1,725,699	1,675,350

Table 14.2 Results of Mixed Logit Estmation 3: Machinery Sector (General Machinery, Electric Machinery, and Transportation Machinery)

	(1)	(2)	(3)	(4)	(5)	(6)
Location effect	4.24 ***	4.28 ***			4.62 ***	4.64 ***
	(0.362)	(0.366)			(0.393)	(0.401)
Location effect*Firm effect	-1.89 ***	-1.9 ***			-2.09 ***	-2.1 ***
	(0.164)	(0.165)			(0.177)	(0.180)
Indusrty agglomeration (t-1)			0.316 ***	0.34 ***	0.349 ***	0.375 ***
			(0.097)	(0.098)	(0.025)	(0.025)
Indusrty agglomeration (t-1)･Firm effect			0.016	0.017		
			(0.044)	(0.044)		
Wage premium (t-1)	0.972 ***		-0.738 **		-0.766 **	
	(0.274)		(0.323)		(0.324)	
ln(Land price of the city)(t-1)		0.0827 *		-0.245 ***		-0.245 ***
		(0.045)		(0.053)		(0.054)
ln(1+Number of factory of the same firm	4.02 ***	3.97 ***	4.07 ***	4.01 ***	4.04 ***	3.98 ***
in the same prefecture (t-1))	(0.122)	(0.122)	(0.129)	(0.128)	(0.129)	(0.129)
R-squared	0.149	0.147	0.163	0.164	0.18	0.181
Log-likelihood	-4.08E+03	-3.98E+03	-3.62E+03	-3.52E+03	-3.55E+03	-3.44E+03
Number of observations	520,626	499,568	467,610	450,038	467,610	450,038

Note 1. Standard errors are in parentheses.

2. * p<.1, ** p<.05, *** and p<.01