RIETI Discussion Paper Series 22-E-105

Does Industry Agglomeration Attract Productive Firms? The role of product markets in adverse selection

BELDERBOS, René
KU Leuven / UNU-MERIT / Maastricht University

FUKAO, Kyoji
RIETI

IKEUCHI, Kenta
RIETI

KIM, Young Gak
Senshu University

KWON, Hyeog Ug
RIETI

Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 22-E-105

November 2022

Does Industry Agglomeration Attract Productive Firms? The Role of Product Markets in Adverse Selection[1]

René Belderbos (KU Leuven, UNU-MERIT, and Maastricht University)

Kyoji Fukao (Hitotsubashi University, RIETI, and Institute of Developing Economies, JETRO)

Kenta Ikeuchi (RIETI)

Young Gak Kim (Senshu University)

Hyeog Ug Kwon (Nihon University and RIETI)

Abstract

Do high or low productivity firms self-select into locations characterized by high industry agglomeration? On the one hand, productive firms may benefit more from the availability of specialized (labour) inputs and they are also more likely to survive heightened competition. On the other hand, productive firms face greater risks of knowledge dissipation to collocated rival firms, as they may contribute more than they receive in terms of knowledge spillovers. We examine location decisions for new plant establishments by firms in Japan with established productivity records (multi-plant firms) at the fine-grained level of towns, wards, and cities where knowledge spillovers are most likely to occur. We find that the adverse selection effects of industry agglomeration–the process of agglomerated areas attracting weaker rather than stronger firms–dominate if knowledge spillovers are most harmful to productive entrants when the focal firm and local incumbent establishments target the same (domestic) product market. We conclude that negative sorting processes do occur, but that these can only be uncovered in a more fine-grained analysis that takes into account ex ante measures of firm heterogeneity and the nature of product markets.

Keywords: productivity, location, agglomeration, spillovers, adverse selection

JEL classification: D24, O32

[1]This study is conducted as a part of the Project "East Asian Industrial Productivity" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the micro data of the questionnaire information based on "Census of Manufacture" which is conducted by the Ministry of Economy, Trade and Industry (METI), and the Census of Manufacture converter, which is provided by RIETI. The authors are grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI. An earlier report on this research project was published as NISTEP Discussion Paper # 115 (Belderbos et al., 2015). The authors also thank participants in seminars at REITI and Hitotsubashi University for comments on previous drafts. René Belderbos gratefully acknowledges financial support from HIAS, NISTEP, RIETI, and the Centre for Economic Institutions and the Joint Usage and Research Center at the Institute of Economic Research, Hitotsubashi University.

Introduction

Recent empirical evidence has suggested that the locational agglomeration of firms is associated with productivity benefits (e.g. Rosenthal and Strange, 2020; Melo et al., 2009; Combes et al., 2012; Lavoratori and Castellani, 2021; Andersson et al., 2019). Two competing explanations have been put forward for this correlation. The predominant explanation is the notion of Marshallian agglomeration externalities, which contends that firms can enjoy positive externalities stemming from geographic industry clustering. These can occur on the input side, as increased demand for inputs stimulates the provision of specialized (labour) inputs and specialized business services. Externalities may also occur on the demand side, as co-location of firms lowers search costs for customers and thus heightens local industry demand, or through locally bounded spillovers of technological and organizational knowledge. These possible externalities motivate firms to choose locations where similar establishments are clustered, an intuition that has been supported by formal economic models (Krugman, 1991; David and Rosenbloom, 1990) and empirical work (e.g. Belderbos, Olffen, and Zou, 2011; Head et al., 1994; Alcácer and Delgado, 2016; Alcácer and Chung, 2007; Frenken et al., 2015; Belderbos et al., 2021).

A second explanation that has been put forward is a selection effect associated with the increased competition within clusters. Collocation of firms in local markets leads to tougher competition, forcing the exit of weaker firms with lower productivity (Melitz and Ottaviano, 2008; Syverson, 2004). In addition, the more productive firms may benefit more from agglomeration, for instance because hiring the more specialized, productive workers provides relatively large productivity benefits to firms that operate more efficiently (Combes et al., 2012), or because efficient firms benefit more from the presence of specialized suppliers (Baldwin and Okubo, 2006). This leads to a positive sorting effect, with the more productive firms overrepresented in agglomerated areas.

The notion that there are greater advantages and chances for survival in higher density locations for productive firms would imply that these firms self-select into high-density locations. However, this conjecture has received little substantive support. Faberman and Freedman (2013) find no evidence of positive sorting effects for U.S. establishments in metropolitan areas. Combes et al. (2012) similarly fail to find evidence of positive selection effects. Gaubert (2018) does calibrate a positive sorting model on French data at the broader city level, but builds on the assumption that efficient firms benefit relatively more from local agglomeration externalities (Gaubert, 2018, p. 3118). Studies on (foreign) market entry have, by and large, concluded that larger and more R&D-intensive firms are less, rather than more, responsive to locational agglomeration than smaller, less R&D-intensive firms (Shaver and Flyer, 2000; Belderbos and Carree, 2002; Alcácer and Chung, 2007). The explanation for this pattern relates to the role of knowledge spillovers in local agglomerations. Productive firms with the most innovative technologies and organizational and process skills have more to contribute to local knowledge spillovers than less productive firms. This gives collocated less productive competing firms the possibility to learn and increase market share if they are able to mimic product designs and organizational approaches, or acquire knowledge through employee mobility. Evidence on the effects of plant openings on local productivity has confirmed that such local productivity effects can be substantial (Greenstone et al., 2010). An asymmetry in knowledge spillovers due to productivity differences suggests a process of negative sorting (adverse selection) in which firms with relatively weaker (rather than stronger) productivity are more likely to locate within clusters.

The (co-)existence of two contrasting sorting effects may explain the inability of prior studies to identify net sorting effects. In this paper, we aim to reconcile these two contrasting views and empirical findings on firm heterogeneity and agglomeration. We examine how the relationship between agglomeration and location choice differs in accordance with entrants' *ex ante* productivity. We posit that the sorting effects of agglomeration depend on whether entrant and incumbent firms compete on the same product market. While productive firms should be attracted to locations providing agglomeration benefits, the presence of incumbents competing in the same product market is likely to discourage productivity leaders (relative to laggards) to collocate due to the asymmetry in knowledge spillovers that can disadvantage firms vis-à-vis incumbent rivals.

Empirically, we identify market competition by distinguishing between exporting and non-exporting firms within four-digit industries of the Census of Manufacture – which is considered a relevant delineation of product markets (Bloom et al., 2013).[1] While entrants selling on the domestic

[1] Bloom et al. (2013) suggest that R&D spillovers have profoundly different effects if they occur between market

market will be competing in a narrowly defined industry with local incumbents, exporting firms are more likely to be active in a variety of (geographic) end markets and hence are less likely to be direct market competitors. We examine plant location decisions (2002-2008) of Japanese multi-plant firms at the fine-grained location level of over 1,000 towns, wards, and cities. By focusing on multi-plant firms, we can identify (adverse) selection effects in detail by relating location decisions to firms' *ex ante* productivity, as measured in their existing plants. The fine grained geographic level is the most likely locational context in which non-market mediated knowledge spillovers occur (Rosenthal and Strange, 2020).

We estimate conditional and mixed (random coefficients) logit models of location choice and control for Marshallian agglomeration mechanisms (Glaeser and Kerr, 2009; Alcácer and Chung, 2013), intra-firm collocation advantages (e.g. Alcácer and Delgado, 2016) and local land rental costs (e.g. Puga, 2010). We find that the more productive firms respond significantly less positively to industry agglomeration, suggesting that, overall, adverse selection effects outweigh positive sorting effects. When we differentiate entries between firms selling on export markets and firms selling on domestic markets, we observe that adverse selection effects only occur if entrants have a domestic market focus, as incumbents and entrants are more likely to compete directly for market share. Adverse selection is most pronounced for non-exporting entrants' response to incumbent industry establishments with a domestic market orientation. These findings provide strong support for the notion of adverse selection due to the risk of knowledge dissipation: if existing establishments and high productivity entrants compete on the same product market, knowledge dissipation concerns are salient as increases in the competitiveness of incumbent rivals directly affect the market share and profitability of the entrant. Industry agglomeration reduces, rather than increases, the likelihood of entry. In contrast, if entrants and incumbents are less likely to compete on the same end markets – if entrants target export markets – positive agglomeration effects dominate.

Our work bears some resemblance to studies examining the relationship between agglomeration and the formation of new firms (e.g. Rosenthal and Strange, 2003; Glaeser and Kerr, 2009). Compared

rivals.

to this line of research, we abstract from *de novo* entrants that face uncertainty concerning their productivity and focus on self-selection processes of known productivity leaders or laggards in the industry. We conclude that sorting processes do occur, but that they can only be uncovered in a more fine-grained analysis that takes into account measures of firm heterogeneity and the nature of product markets. Overall, our results provide substantial support for adverse selection due to asymmetries in knowledge spillovers.

Data and Empirical Model

We draw on the *Census of Manufacture* in Japan to identify new plant establishments. We have access to data for new entries, productivity and exports of manufacturing plants for the period 2002-2008. Only from 2001 onwards shipment data for Japanese plants in the census distinguish between exports and domestic shipments – an important distinction in our research. In order to identify whether adverse selection occurs, we need reliable data on heterogeneity in firm productivity and have to apply several sample screens. Since we cannot use focal plant productivity after entry, as this is likely to be endogenous to agglomeration, we use information on existing, pre-entry, plant productivity, and hence focus on multi-plant firms.

We are interested in entry decisions for locations that are new to the investing firm in the industry. In existing locations, new plants establishments will not differ much from existing plant expansions, with external agglomeration potential or knowledge spillover considerations less likely to be relevant. Hence, we focus on the location choice for firms' new plants in locations in which no existing plant of the firm in the focal industry is operating. We do maintain new plant establishments in the analysis if the focal firm has existing operations in a region, if such operations are in different industries or concern headquarter operations

During the period 2002-2008 we observe 2381 new manufacturing plant establishments in new to the firm regions by 2177 firms with existing plants in the industry (Table 1). We prefer to apply an accurate measure of the productivity premium by requiring that we observe existing productivity levels at least for 2 plants. In doing so we avoid potential 'contamination' of the productivity level of a single plant by idiosyncratic locational characteristics. If a multi-plant firm is able to achieve higher productivity levels across locations, this will reflect capabilities, technologies, and knowledge that can be transferred across plants, and that are potentially put at risk of spillovers when the firm establishes a new plant in the vicinity of rival firms. The criterion of operating two plants in the industry reduces the number of firms and new plant establishments to 749 and 883 respectively.

INSERT TABLE 1

Table 2 provide details on the locational choice set and the distribution of entries across industries (aggregated to 45 industries for exposition). While the (theoretical) maximum choice set consists of the 2467 different towns, wards, cities, and villages in Japan, the actual choice sets are smaller because we conservatively only include locations in the choice set of a four-digit industry if there is evidence that they are 'at risk' of receiving investments. Specifically, we include locations in the choice set if during the period 1997-2008 we observe at least one establishment in in the industry. Omitting region-industry combinations without any establishments or entries keeps the models convergent and computationally feasible, whilst inclusion of locations that do not have a realistic probability of receiving investments runs the risk of violating the IIA assumption characterizing conditional logit models (see below). On average, the choice set for the entries in an industry consist of 403 regions. The choice set ranges from 5 regions for highly concentrated industries (flour manufacturing industry within the flour and grain milling industry) to 1054 regions for geographically distributed industries (food manufacturing industry not classified elsewhere within the miscellaneous foods industry). The 883 new entries are spread over 525 individual regions.

INSERT TABLE 2

TFP and exports

Plant-level TFP is measured using the index number method, based on data available from the Japan Industrial Productivity Database (Fukao et al, 2006; Belderbos et al., 2013; RIETI, 2018). One of the main advantages of the index number method is that it allows for heterogeneity in the production technology of individual firms, while other methods controlling for the endogeneity of inputs (e.g. Olley and Pakes, 1996; Levinsohn and Petrin, 2003) assume an identical production technology among firms within an industry (Van Biesebroeck, 2007; Aw et al., 2001). The productivity index captures TFP of a plant relative to a representative firm in the industry in a base year. We investigate what position plants occupy in the distribution of productivity levels across plants in the same 4-digit industry in Japan during the year prior to entry. We calculate, on a yearly basis, the *TFP premium* as the log of the difference between the output weighted average TFP in the firm's existing plants and the mean level of TPF in the industry (using plant output as weights). Leading firms (those with TFP above the mean) have positive values for the TFP premium, while lagging firms (those with TFP below the mean) have negative values for the TFP premium.

Table 1 shows that existing multi-location plant firms are more productive than single-location plant firms, with the difference between (average) TFP of firms with plants in one or more existing locations 0.063 points versus 0.037 points above the industry mean on average. Our focus on the former firms is most likely to raise the bar for a finding of adverse selection related to productivity heterogeneity - since we focus effectively on firms in the top of the distribution and hence diminish variation. In robustness analyses, we will also test models with the single location plant firms included. As to be expected (Bernard and Jensen, 1999; Delgado et al., 2002), exporting firms exhibit higher productivity levels than firms with a domestic market orientation (0.140 versus 0.049).

Agglomeration Variables

In order to disentangle agglomeration mechanisms from the agglomeration effects related to spillovers to rivals, we adopt the specification of Alcácer and Chung (2013) and Gleaser and Kerr (2009). We separate an industry 'volume' effect from the characteristics of the broader set of establishments across industries providing agglomeration advantages through supplier linkages, buyer linkages, labour pooling, or knowledge spillover from related industries. Industry agglomeration is measured as output in the industry at the four-digit level in the region, relative to industry-year average (e.g. Alcácer and Chung, 2013).

We follow Gleaser and Kerr (2009) and Alcácer and Chung (2013) in constructing regional specialization variables to measure agglomeration externalities. Supplier (buyer) fit measures the locations' relative specialization in related supplier (buyer) industries. To establish specialization, we use yearly input-output tables provided by the JIP database, such that weights vary by year. Formally, we measure buyer fit for industry *i* in location *l*:

$$\text{Buyerfit}_{ilt} = \left[\sum_k output_{i\to k} \times \frac{Y_{klt}}{Y_{kt}}\right] \times \left[\frac{Y_{lt}}{Y_t}\right]^{-1}$$

The subscripts i, k denote industries, l location, t time, Y_{klt} output of industry k in location l at time t, and $output_{i\to k}$ is the share of industry i's output that i supplied to industry k, $Y_{kt} = \sum_l Y_{klt}$, $Y_{lt} = \sum_k Y_{klt}$, and $Y_t = \sum_k \sum_l Y_{klt}$. The measure multiplies the output share of industry *k* with the output share of that industry in the focal region. The smaller the deviations of the two across industries, the stronger the 'fit' between local industry structure and the buyer profile of industry *i*. Agglomeration externalities are likely to occur if the region is responsible for a relatively large share of manufacturing output specifically in those industries that are important buyers of the focal industry i. As suggested by Gleaser and Kerr (2009), the expressions are multiplied by the inverse sum of the location's output in total national output, to reduce correlation with industry size.

Supplier fit is a variation on this. The fit variables compares the share of the region in the production of industry k with the share of industry k in the inputs for the focal industry *i:*

$$\text{Supplier}_{ilt} = -\sum_k \left(\left| input_{i\leftarrow k} - \frac{Y_{klt}}{Y_{lt}} \right|\right)$$

Labour fit is measured for industry *i* in location *l* as:

$$\text{Laborfit}_{ilt} = -\sum_o \left| L_{io} - \left(\sum_k \frac{E_{klt}}{E_{lt}} L_{ko}\right) \right|$$

where o indexes occupations, L_{io} and L_{ko} are the percentage of industry i's and industry k's employment in occupation o, respectively, which are obtained from the Population Census 2005, E_{klt} is the employment of industry k in location l at time t and E_{lt} is the total employment (across industries) for location l at time t, which are obtained from the Census for Manufacture. Hence, labor fit increases if the region is specialized in those industries that share similar occupations as the focal industry.

For knowledge fit we make use of cross-industry patent citations to identify industries that are most likely to provide relevant knowledge to the focal industry.

$$\text{Knowledgefit}_{ilt} = \left[\sum_{k} citaion_{i\leftarrow k} \times \frac{P_{klt}}{P_{kt}}\right] \times \left[\frac{P_{lt}}{P_t}\right]^{-1}$$

Where $citaion_{i\leftarrow k}$ is the share of backward citations that patents in industry i make to industry k, P_{klt} is the stock of patents generated by firms in industry k in location l at time t, and $P_{kt} = \sum_l P_{klt}$, $P_{lt} = \sum_k P_{klt}$, $P_t = \sum_k \sum_l P_{klt}$. Patent citation and patent stock data are obtained from the IIP patent database of the Japan Patent Office. Knowledge fit is constructed analogously as buyer and supplier fit, and takes higher values if the region is specialized in patenting in those domains that are more often cited by firms in the focal industry.

Other Variables

In order to measure congestion effects, we include a measure of land prices. We obtain information on land prices from Toyo Keizai (2013).[2] The analysis also controls for 'internal agglomeration' or collocation effects due to previous establishments of the firm in the location. We include a dummy variable for the presence of other plants of the firm in a different industry and the presence of the firm's headquarter operations, in addition to the distance of the region from headquarters (in cases in which headquarters is located in a different region), and the distance to the nearest other plant of the firm.[3]

Empirical Model

[2] We also experimented with a regional wage premium variable, using estimated regional wage premiums from a Mincer-type wage function by Kawaguchi and Kambayashi (2009). The wage premium data are available only for one year, however, and are strongly correlated with the land price variable. Joint inclusion left the wage variable insignificant.

[3] Regional subsidies and incentives may affect location decisions, although their effectiveness has been questioned (e.g. Head, Ries, and Swenson, 1999; Gaubert, 2018). Unfortunately no systematic information on regional incentive policies exist for Japan.

The location choice literature (e.g. Alcácer and Chung, 2007, 2013; Head et al., 1994; 2004) has primarily used the conditional logit model (McFadden, 1974) to analyze the location determinants of investments. The conditional logit model can be derived from a profit maximization framework under suitable assumptions concerning the distribution of the error term. A drawback of this model is the restrictive assumption of the independence of irrelevant alternatives (IIA). The IIA property states that for any two alternatives, the ratio of probabilities is independent of the characteristics of any other alternative in the choice set. This characteristic also implies the absence of correlations between error terms across alternatives. At the detailed regional level of analysis, the likelihood of spatial correlation is high, as regional boundaries do not necessarily demarcate the border of agglomeration externalities.[4] One solution to this is to estimate mixed logit models that relax the IIA assumption by allowing coefficients to vary. We do so in supplementary analysis. Another approach that we will follow is to examine distance-weighted variables measured across (neighbouring) regions to the models. In such models, industry agglomeration becomes the sum of all industry establishments in the focal region and all other (neighbouring) regions weighted by the geographic distance between regions, with weights taken as $3/2r$ (where r represents distance).[5] We extend the reach of the agglomeration variables to 5 and 10 kilometers from the focal region's core.

We test whether adverse selection or positive sorting occurs by including the interaction between the four-digit industry agglomeration measure and the TFP premium variable. We also interact the TFP premium variable with the agglomeration mechanisms (specialization) variables to allow for heterogeneity in the agglomeration benefits due to these externalities. We subsequently examine location decisions separately for entrants with and without an export orientation, and for different kinds of agglomeration (exporting incumbents versus non-exporting incumbents).

[4] Jofre-Monseny et al. (2011) and Rosenthal and Strange (2020) observed that knowledge spillover considerations only play a role in entry decisions as at a fine-grained, highly localized level. Similar observations are made by Andersson et al. (2019) and Lavoratori and Castellani (2021).

[5] This follows Head and Mayer (2004). The weight assumes that demand is equally distributed in a circle of radius r.

Descriptive Statistics

Table 3 shows the descriptives and correlations of the variables. Continuous variables, except for the agglomeration fit variables (which can take negative values), are expressed in natural logarithm. Table 4 provides some prima facie evidence of potential adverse selection. Panel I shows the values of industry output agglomeration of the regions of the focal entries, distinguishing between non-exporting plant agglomeration and exporting plant agglomeration. The agglomeration levels are compared between leading TFP firms (above the median among the entrants in the industry) and lagging TFP firms (below the median). There is a clear pattern, with agglomeration levels higher for laggards (1.693) than for leaders (0.881); this difference is significant ($p < 0.05$). Similar patterns are observed for non-exporter agglomeration (1.588 versus 0.758, $p < 0.05$) while for exporter establishment agglomeration, in contrast, there are no such differences observed (0.657 versus 0.669, $p = 0.96$). For non-exporting firm entries (panel II), comparable patterns are observed, with the differences more strongly significant. For exporting firm entries, agglomeration levels of any kind are lower for TFP leaders but the differences insignificant. The difference for exporter agglomeration, where competitive considerations are most likely to play a role, just does not reach conventional significance levels ($p = 0.051$).

INSERT TABLES 3 and 4

Empirical Results

The results of the conditional logit models of new plant location are presented in Table 5. Model 1 includes the agglomeration and control variables, model 2 adds the interaction of agglomeration and TFP of the investing firm, while model 3 adds the interactions with TFP and the agglomeration mechanisms are included. Models 4 and 5 distinguish between exporting firms and non-exporting firms.

INSERT TABLE 5

In model 1, entry probabilities are positively affected by buyer fit, supplier fit, labor fit, industry agglomeration, and prior activities in the region by the firm (headquarters and other plants), while land prices and distance to other establishments of the firm exert negative influences. The coefficients of the variables in logarithm can be interpreted as average elasticities (Head et al., 1994). If overall industry agglomeration increases by 1 per cent, the probability that the region is chosen for investment increases by 0.037 percent. The elasticity of the land price is larger, at -0.25 percent.

The TFP premium in interaction with industry agglomeration represents the sorting effect. The negative and marginally significant coefficient (β = -0.056) suggests that, overall, there is weak adverse selection: productivity leaders are less attracted to industry agglomeration than productivity laggards are. The interaction coefficient becomes more strongly negative and statistically significant in model 3 (β = -0.068), in which the agglomeration mechanisms are included in interaction with the TFP premium. Productive firms are relatively more attracted to regions with a better labour fit, but less attracted to regions with supplier fit. The latter finding does not suggest evidence for the theory of Baldwin and Okubo (2006).

If we distinguish exporters and non-exporters in model 4, we observed that the negative interaction effect only exists for non-exporters (β = -0.110), while for exporters, in contrast, the TFP premium increases the positive response to industry agglomeration. In model 5 we further examine the evidence of negative sorting more specifically as a function of non-exporting firms in response to the agglomeration of competing non-exporting plants. While both exporter and non-exporter agglomeration attract non-exporting firms' plants, the analysis confirms that it is the non-exporter agglomeration that generates adverse selection of less productive firms. This provides further evidence that negative sorting occurs when firms focus on the same product markets. For exporting firm entries a marginally significant effect of exporter agglomeration is identified, while there is no evidence of adverse selection, which we posit is due to the large potential variety in firms' export markets, which reduces direct market rivalry.

The range of elasticities of entry location choice with respect to industry agglomeration as a function of the TFP premium is shown in Figure 1. Figure 1a illustrates these elasticities for all entrants (results of model 3). For non-exporting investing firms with a negative TFP premium the agglomeration elasticity is declining in the premium but always positive and significant. For firms with a high TFP premium above 0.6, the adverse selection effects of industry agglomeration even lead to a net negative – although insignificant - entry elasticity. Stronger adverse selection patterns are observed in Figure 1b for non-exporting firms' location choice elasticity with respect to non-exporting industry agglomeration, based on the results of model 5. In this case the decline in the entry location choice elasticity is sharper, and a negative elasticity is observed at a TFP premium of 0.35, which represents the top 9 percentile of entrants. A statistically significant negative influence of agglomeration is identified at a TFP premium of 0.9 (the top percentile of entrants).

INSERT FIGURE 1

Supplementary analysis

We conducted a range of supplementary analysis for the specification of model 5 to examine the robustness of our findings. First, we estimated random coefficient mixed logit models that allow for general patterns of investor heterogeneity. Results are consistent with those reported in Table 6. The estimated interaction effect between non-exporter agglomeration and the productivity premium is substantially larger than in Table 5 (β = -0.326) but has a higher standard error and is marginally significant. There is a significant positive effect of non-exporter agglomeration on entry location choice by exporters. Second, we broaden the geographic scope of the agglomeration to allow for more varied spatial decay effects (e.g. Puga, 2010; Cainelli and Ganau, 2018; Verstraten et al., 2018) by adding adjacent regions within a 5 or 10 kilometre radius of the centre of the focal region and weighing the agglomeration variables with the inverse of distance. Results indicate that adverse selection patterns weaken in distance, with the coefficient of the interaction effect taking a smaller but consistently negative value (β = -0.094 and β = -0.075, respectively). This is in line with earlier findings on the proximate effects of knowledge spillovers (Jofre-Monseny et al. 2011; Rosenthal and Strange, 2020; Andersson et al., 2019; Lavoratori and Castellani, 2021).

INSERT TABLE 6

Third, we estimate models with regional fixed effects. While these control further for idiosyncratic locational factors that could encourage or discourages new plant investments in a region, the inclusion of fixed effects implies that these regions that did not attract any focal entry during the period of investigation have to be omitted from the analysis, as the absence of entries is fully explained by the fixed effect. This leads to a sample selection and attrition effect: it reduces the mean choice set to 182, down from 403, and it more than halves the number of observations. Nevertheless, results are consistent, with an only slightly smaller negative interaction coefficient for non-exporters' entry choice as a function of non-exporter agglomeration (β = -0.088).

Fourth, instead of distinguishing exporting firms from non-exporting firms, we distinguish exporting entries (plants) from non-exporting entries. Although the exporting decisions for new plants will be taken simultaneously with the entry location decision and are partly endogenous, we do expect and find consistent patterns, with the adverse selection prevalent only for non-exporting new plants and estimated to be of a similar magnitude (β = -0.093) as in the main model 5 in Table 5. There is a significantly positive effect of exporter agglomerating on exporting plant location choice, again suggesting that the diffuse markets here allow avoiding direct competition. Fifth, we broadened the set of plant entries to include new location decisions of firms with only a single existing plant in the industry. This increased the number of entries to 2381 and the number of firms to 2177. We expect that the TFP premium is less well measured in this case, as the characteristics of the existing location may introduce an important non-firm specific heterogeneity. We find that the interaction coefficient between the TFP premium and non-exporter agglomeration for non-exporting firms is consistently negative, but insignificant, demonstrating the importance of accurate measurement of productivity premiums.

Finally, we examined the conjecture that it is not productivity but firm size that drives sorting effects. Large firms may be less attracted to agglomerated areas because they face greater congestion charges, and perhaps they can attract their own suppliers or buyers if they invest in less densely agglomerated regions. We found no evidence for this conjecture. The mean agglomeration level of the chosen regions by firms with an above median scale is not significantly different from regions chosen by firms with a scale below the industry median. Adding an interaction term in the models of Table 5 between industry agglomeration and the log difference between investing firm plant size and the industry median plant size produced no significant effects for this term, while the estimates for the interaction between the TFP premium and industry agglomeration were left unchanged.

Conclusions

The literature has produced ambiguous findings concerning the salience and direction of the sorting process on entry in response to industry agglomeration. We posit that this is related to the variety of sorting influences at play. On the one hand, productive firms may be more likely to survive heightened competition in agglomerated areas and may benefit more from locally available specialized inputs. On the other hand, productive firms face greater risks of knowledge dissipation to collocated rival firms and contribute more than they receive in terms of knowledge spillovers from plants in close proximity. Our study sought to provide more insights into these relationships by relating entry location choices to a clear indicator of *ex ante* productivity, examining entries by multi-plant firms for which existing productivity levels can be accurately observed, and by distinguishing the influence of industry agglomeration of direct market rivals -firms selling in the same domestic market- from the influence of industry agglomeration of firms that are not directly competing in the same end market (exporters).

Our results for the location choice of manufacturing entries in Japan at the detailed geographic level of towns, wards, and cities provide strong support for the notion of adverse selection of manufacturing entries related to the risk of knowledge dissipation. If existing establishments and high productivity entrants share the same (domestic) market, knowledge dissipation concerns are salient, as increases in competitiveness of incumbent rivals are likely to directly affect the market share and profitability of the entrant. Industry agglomeration reduces, rather than increases, the likelihood of entry for the most productive firms. If entrants and incumbents do not share the same end markets – i.e. entrants target export markets and incumbents domestic market or vice versa – positive agglomeration and sorting effects dominate. We conclude that sorting processes on entry do occur, but that these have to be uncovered in a more detailed analysis that takes into account ex ante measures of firm heterogeneity and the nature of product markets.

Our finding suggest some interesting avenues for future exploration. If the most productive firms invest in less agglomerated areas, can they keep their productivity at a high level or does the avoidance of rival firms have a cost in terms of noticeably smaller agglomeration benefits? Can TFP leaders maintain their distance to rival firms after investments in a new location, or do we observe subsequent entries by rivals in the vicinity of the new plant of the leader?

Future research may also address some of the limitations of our research, e.g. by bringing in the influence of urbanization and diversity at the broader city level (Lavoratori and Castellani, 2021), by including wage costs in addition to land costs (Verstraten et al., 2018), and by examining spatial clusters of establishments rather than regions with administrative boundaries (Puga, 2010). Our paper contributes to an expanding strand of research that examines agglomeration effects at the fine grained spatial level to uncover heterogeneous influences of agglomeration (Monseny et al., 2011; Andersson et al., 2019; Lavoratori and Castellani, 2021; Cainelli and Ganau, 2018; Verstraten et al., 2018). We suggest that in addition to focusing on fine grained spatial levels, future research should provide ample consideration to the nature of product market competition to better understand the complex relationship between establishment density and productivity through sorting effects, competition, and agglomeration externalities.

References

Alcácer, J., and Chung, W. 2013. Location Strategies for Agglomeration Economies. *Strategic Management Journal*, 35(12), 1749-1761.

Alcácer J. and Chung W. 2007. Location Strategies and Knowledge Spillovers, *Management Science*, 53(5), 760-776.

Alcácer, J., & Delgado, M. 2016. Spatial Organization of Firms and Location Choices Through the Value Chain. *Management Science,* 62 (11): 3213-3234.

Andersson, M., Larsson, J. P., and Wernberg, J. 2019. The economic microgeography of diversity and specialization externalities—Firm-level evidence from Swedish cities. *Research Policy* 48(6), 1385–1398.

Aw, B. Y., Chen, X., and Roberts, M. J. 2001. Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing. *Journal of Development Economics*, 66, 51-86.

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

Belderbos, R. and Carree, M. 2002. The Location of Japanese Investments in China: Agglomeration Effects, Keiretsu, and Firm Heterogeneity, *Journal of the Japanese and International Economies* 16, 194-211.

Belderbos, R., Fukao, K., Ito, K., Letterie, W. 2013. Global Fixed Capital Investment by Multinational Firms, *Economica*, 80(318), 274-299.

Belderbos, R., Ikeuchi, K., Fukao, K., Kim, Y. G., Kwon, H. U. 2015. Agglomeration and Adverse Selection: Evidence from multi-plant firms, *NISTEP Discussion Paper* No. 115.

Belderbos, R., Kazimierczak, M. and Goedhuys, M. 2021. Trademarks, patents, and the appropriation strategies of incumbents: The scope of new firm formation in European regions, *Regional studies*, 56 (2), 210-226, 2021.

Belderbos, R., Olffen, W. van, and Zou, J. 2011. Generic and Specific Social Learning Mechanisms in Foreign Entry Location Choice. *Strategic Management Journal*, 32(12), 1309-1330.

Bernard, A., and Jensen, J.B., 1999. Exceptional export performance: cause, effect, or both? *Journal of International Economics*, 47: 1-25.

Biesebroeck, J. van. 2007. Robustness of Productivity Estimates. *Journal of Industrial Economics*, 55(3), 529-569.

Bloom, N., Schankerman, M., and Van Reenen, J. 2013. Identifying Technology Spillovers and Product Market. *Econometrica*, 81(4), 1347-1393.

Cainelli, G., and Ganau, R., 2018. Distance-based agglomeration externalities and neighbouring firms' characteristics. *Regional Studies*, 52 (7), 922–933.

Combes, P-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. 2012. The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection, *Econometrica*, 80(6), 2543-2594.

David, P.A. and Rosenbloom, J.L. 1990. Marshallian Factor Market Externalities and the Dynamics of Industrial Location, *Journal of Urban Economics*, 28(3), 349-370.

Delgado, M., Farinas J. and Ruano S. 2002. Firm Productivity and Export Markets: a Nonparametric Approach. *Journal of International Economics* 57: 397-422.

Faberman, J. and Freedman, M., 2013. *The Urban Density Premium across Establishments,* WP 2013-01, Federal Reserve Bank of Chicago.

Frenken, K., Cefis, E., and Stam, E. 2015. Industrial dynamics and clusters: a survey. *Regional studies*, 49(1), 10-27.

Fukao, K. and Kwon, H. U., 2006. Why Did Japan's TFP Growth Slowdown in the Lost Decade? An Empirical Analysis Based on Firm-level Data of Manufacturing Firms. *Japanese Economic Review*, 57(2), 195-228.

Gaubert, C. 2018. Firm Sorting and Agglomeration. *American Economic Review*, 108(11):3117-53.

Glaeser, E., and Kerr, W., 2009. Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain? *Journal of Economics and Management Strategy*, 18(3), 623-663.

Greenstone, M., Hornbeck, R., and Moretti, E. 2010. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings, *Journal of Political Economy*, 118(3), 536-598.

Head, K. and Mayer, T., 2004. Market Potential and the Location of Japanese Investment in the European Union, *Review of Economics and Statistics*, 86(4), 959-972.

Head, C. K., Ries, J. C., & Swenson, D. L. (1999). Attracting foreign manufacturing: Investment promotion and agglomeration. *Regional Science and Urban Economics*, *29*(2), 197-218.

Head, K., Ries, J., and Swenson, D., 1994. Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States, *Journal of International Economics*, 38(3/4), 223-247.

Jofre-Monseny, J., Marin-Lopez, R., and Viladecans-Marsal, E. 2011. The Mechanisms of Agglomeration: Evidence from the Effect of Inter-industry Relations on the Location of New Firms, *Journal of Urban Economics,* 70 (2-3), 61-74.

Kawaguchi, D. and Kambayashi, R., 2009. 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 Manufacture* and the *Basic Survey on Wage Structure*], in Yukinobu Kitamura (ed.), *Ohyo Mikuro Keiryo Keizai-gaku* [*Applied Micro-Econometrics*], Tokyo, Japan: Nihon Hyoron-sha.

Krugman, P., 1991. Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483-499.

Lavoratori, K., & Castellani, D. 2021. Too close for comfort? Microgeography of agglomeration economies in the United Kingdom. *Journal of Regional Science*, 61(5), 1002-1028.

Levinsohn, J., and Petrin, A. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2), 317-341.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics* Ed P. Zarembka (Academic Press, New York). 105-142.

Melitz, M. and Ottaviano, G. I. P. 2008. Market Size, Trade and Productivity, *Review of Economic Studies*, 75(1), 295-316.

Melo, P. C., Graham, D. J., and Noland, R. B. 2009. A Meta-Analysis of Estimates of Urban Agglomeration Economies, *Regional Science and Urban Economics*, 39(3), 332-342.

Olley, S., and Pakes, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64, 1263-1297.

Puga, D., 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1), 203–219.

RIETI, 2018. Japan Industrial Productivity Database 2018. Research Institute of Economics, Trade and Industry, Tokyo. https://www.rieti.go.jp/en/database/JIP2018/index.html#01

Rosenthal. S. and Strange, W.C. 2020. How close is close? The spatial reach of agglomeration economies. *Journal of Economic Perspectives*, 34(3), pp: 27–49.

Rosenthal, S. and Strange, W. C. 2003. Geography, Industrial Organization, and Agglomeration, *Review of Economics and Statistics*, 85, 377-93.

Shaver, J. M. and Flyer, F. 2000. Agglomeration Economies, Firm Heterogeneity, and Foreign Direct Investment in the United States, *Strategic Management Journal*, 21(12), 1175-1193.

Syverson, C. 2004. Market Structure and Productivity: A Concrete Example, *Journal of Political Economy*, 112(6), 1181-1222.

Toyo Keizai Inc. 2013. *Chiiki-keizai-deta CD-ROM* (*Regional Economy Data CD-ROM*). Toyo Keizai Inc., Tokyo.

Verstraten, P., Verweij, G., and Zwaneveld, P. J. 2018. Complexities in the spatial scope of agglomeration economies. *Journal of Regional Science*, 59(1), 1–27.

Table 1. Multiplant firms' new plant establishments and productivity, 2000-2008

	# firms	# new plants	TFP					
			Mean	Min.	p25	p50	p75	Max.
All multiplant firms with entries into new-to-the-firm regions	2,177	2,381	0.037	-1.121	-0.109	0.009	0.160	1.534
of which firms firms with two existing plants	749	883	0.063	-1.105	-0.075	0.038	0.170	1.022
of which non-exporting firms	639	747	0.049	-1.105	-0.088	0.024	0.164	1.009
of which exporting firms	110	136	0.140	-0.585	0.016	0.094	0.223	1.022

Note: TFP is the ln difference between focal the plant and the industry mean.

Table 2. New plant entries and number of locations in the choice set per industry

2-digit industry	# entries	# firms	# obs.	Choice set: # locations Mean	S.D.	Median	Min.	Max.
Livestock products	39	25	12,087	312	94	298	198	476
Seafood products	9	9	3,184	354	189	479	110	526
Flour and grain mill products	5	5	731	146	79	180	5	186
Miscellaneous foods and related products	101	83	70,278	693	335	926	27	1054
Prepared animal foods and organic fertilizers	10	8	1,231	123	3	122	119	127
Beverages	12	12	2,174	181	134	176	27	316
Textile products	19	15	5,159	281	172	275	68	674
Lumber and wood products	8	8	2,117	265	117	323	21	351
Furniture and fixtures	2	2	1,292	646	8	646	640	652
Pulp, paper, and coated and glazed paper	21	11	775	37	12	31	22	67
Paper products	33	30	10,440	316	83	343	35	374
Printing	43	40	30,400	704	291	829	154	912
Rubber products	5	5	1,472	294	156	354	17	379
Chemical fertilizers	2	2	91	46	2	46	44	47
Basic inorganic chemicals	9	8	966	107	2	107	104	110
Organic chemicals	5	5	187	37	40	12	7	97
Miscellaneous chemical products	18	17	1,655	92	59	117	20	186
Pharmaceutical products	19	14	1,787	94	58	129	10	142
Petroleum products	5	3	82	16	14	11	8	41
Coal products	9	6	890	99	75	55	46	209
Glass and its products	19	19	2,201	116	39	130	16	138
Cement and its products	43	35	17,240	401	130	404	135	627
Miscellaneous ceramic, stone and clay produ	9	9	696	77	102	43	28	346
Miscellaneous iron and steel	40	35	8,001	200	128	170	9	380
Smelting and refining of non-ferrous metals	4	4	200	50	18	57	23	63
Non-ferrous metal products	7	7	926	132	31	141	70	158
Fabricated constructional metal products	26	23	21,117	812	88	802	709	994
Miscellaneous fabricated metal products	45	37	9,058	201	111	120	40	481
General industry machinery	12	12	2,965	247	91	292	80	312
Special industry machinery	22	21	6,306	287	92	262	157	412
Miscellaneous machinery	14	14	6,722	480	254	605	15	681
Office and service industry machines	5	5	1,023	205	60	164	156	272
Electrical distribution and industrial	28	25	11,260	404	107	437	176	580
Household electric appliances	4	1	1,468	367	.	367	367	367
Electronic data processing machines	2	2	749	375	19	375	361	388
Communication equipment	3	3	505	168	45	194	116	195
Electronic (measuring) equipment	3	3	558	186	3	188	182	188
Semiconductor devices and integrated circui	17	13	1,249	73	11	78	43	80
Electronic parts	27	24	11,473	417	246	372	16	703
Miscellaneous electrical machinery equipmer	6	4	1,291	215	95	269	42	274
Motor vehicles	1	1	69	69	.	69	69	69
Motor vehicle parts and accessories	104	81	82,614	793	49	834	733	844
Other transportation equipment	4	3	377	94	30	98	61	120
Precision machinery & equipment	8	8	848	106	39	114	58	162
Plastic products	53	49	21,081	398	197	366	6	627
Miscellaneous manufacturing industries	3	3	557	186	128	159	73	325
All (total, mean)	883	749	357,552	403	303	344	5	1054

Notes: Choice sets are the locations for which a least one (new) plant has been observed during 1997 to 2008 for a 4-digit industry

Table 3. Descriptive statistics and correlations

		mean	SD	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
[1]	Entry	0.002	0.050													
[2]	TFP premium (ln)	0.057	0.228	0.001												
[3]	Industry agglomeration	0.127	4.516	0.013	0.000											
[4]	non-exporting industry agglomerati	0.126	4.499	0.012	0.001	0.974										
[5]	exporting industry agglomeration	0.009	3.531	0.009	-0.001	0.267	0.151									
[6]	Buyer industry fit	24.410	66.617	0.005	-0.002	0.141	0.129	0.171								
[7]	Supplier industry fit	-27.504	35.365	0.016	-0.015	-0.030	-0.034	0.034	-0.166							
[8]	Knowledge fit	1.787	8.136	0.006	-0.011	0.035	0.033	0.026	0.038	0.053						
[9]	Labor fit	-1.214	0.236	0.009	-0.025	0.213	0.199	0.181	0.261	-0.259	0.029					
[10]	HQ of the firm	0.000	0.021	0.070	0.004	0.011	0.011	0.011	0.007	0.006	0.001	0.001				
[11]	Firm plant other industry	0.001	0.032	0.032	0.012	0.016	0.016	0.009	0.008	0.003	0.001	0.011	0.093			
[12]	Min. distance from firm's other plan	5.178	1.141	-0.040	-0.035	-0.090	-0.088	-0.051	0.007	-0.024	0.032	-0.002	-0.026	-0.032		
[13]	Distance from firm's HQ	5.541	1.065	-0.046	-0.034	-0.103	-0.101	-0.056	0.001	0.051	0.045	-0.059	-0.166	-0.040	0.688	
[14]	Land price	-0.465	0.957	0.007	0.030	0.199	0.199	0.081	-0.022	-0.029	-0.046	0.002	0.035	0.016	-0.248	-0.320

Notes: all variables except dummy variables and fit measures are in natural logarithm

Table 4. Differences in agglomeration levels of entry locations: TFP leaders versus TFP laggards

I. All firms' entries	All	laggards (less than median)	leaders (above median)	p value difference
agglomeration	1.287	1.693	0.881	0.028
non exporter agglomeration	1.172	1.588	0.758	0.024
exporter agglomeration	0.663	0.657	0.669	0.967
# entries	883	441	442	

II. Non-exporter entries	All	laggards (less than median)	leaders (above median)	p value difference
agglomeration	1.132	1.628	0.579	0.008
non exporter agglomeration	1.050	1.512	0.535	0.013
exporter agglomeration	0.565	0.444	0.699	0.373
# entries	747	394	353	

III. Exporter entries	All	laggards (less than median)	leaders (above median)	p value difference
agglomeration	2.135	2.241	2.079	0.876
non exporter agglomeration	1.846	2.224	1.646	0.585
exporter agglomeration	1.205	2.443	0.551	0.051
# entries	136	47	89	

Table 5. Conditional logit estimates of the location choice for new plant entries: exporting versus non-exporting firm entries and industry agglomeration

	model 1	model 2	model 3	model 4		model 5	
				non-exporters	exporters	non-exporters	exporters
Industry agglomeration	0.0372***	0.0410***	0.0421***	0.0400***	0.0365		
	[0.00797]	[0.00828]	[0.00831]	[0.00897]	[0.0229]		
TFP premium * Industry agglomeration		-0.0564*	-0.0679**	-0.110***	0.179*		
		[0.0297]	[0.0307]	[0.0330]	[0.0931]		
non-exporting industry agglomeration						0.0352***	0.0278
						[0.00885]	[0.0219]
TFP premium * non-exp. industry agglom.						-0.102***	0.1060
						[0.0321]	[0.0840]
exporting Industry agglomeration						0.0235**	0.0324*
						[0.0109]	[0.0193]
TFP premium * exporting industry agglom.						-0.0207	-0.0419
						[0.0386]	[0.0796]
Buyer industry fit	0.0856**	0.0842**	0.0894**	0.0828**	0.245**	0.0808**	0.253**
	[0.0356]	[0.0356]	[0.0362]	[0.0402]	[0.111]	[0.0400]	[0.113]
Supplier industry fit	0.329*	0.332*	0.448**	0.435*	0.0172	0.403*	-0.0684
	[0.199]	[0.199]	[0.206]	[0.229]	[0.496]	[0.230]	[0.503]
Knowledge fit	0.000762	0.000794	0.0004	0.00132	-0.03610	0.0013	-0.0387
	[0.00236]	[0.00238]	[0.00271]	[0.00249]	[0.0292]	[0.00250]	[0.0299]
Labor fit	0.542**	0.545**	0.397*	0.294	1.196*	0.2570	1.046*
	[0.225]	[0.225]	[0.238]	[0.262]	[0.616]	[0.265]	[0.636]
TFP premium * Buyer industry fit			-0.0428	0.0018	-0.668	-0.0010	-0.6140
			[0.209]	[0.227]	[0.699]	[0.227]	[0.681]
TFP premium * Supplier industry fit			-1.877**	-2.529**	1.455	-2.548**	1.5000
			[0.936]	[1.056]	[2.166]	[1.060]	[2.160]
TFP premium * Knowledge fit			-0.0090	-0.00694	-0.0428	-0.0073	-0.0296
			[0.00962]	[0.00972]	[0.166]	[0.00971]	[0.165]
TFP premium * Labor fit			1.777**	2.253**	-1.237	2.271**	-0.5350
			[0.865]	[0.935]	[2.522]	[0.943]	[2.523]
HQ of the firm	1.387***	1.415***	1.405***	1.733***	-1.643	1.725***	-1.6160
	[0.353]	[0.354]	[0.354]	[0.374]	[1.442]	[0.374]	[1.458]
Firm plant other industry	1.717***	1.717***	1.684***	1.684***	2.009***	1.674***	2.040***
	[0.299]	[0.299]	[0.299]	[0.352]	[0.557]	[0.351]	[0.557]
Min. distance from firm's other plant	-0.585***	-0.586***	-0.587***	-0.620***	-0.369***	-0.620***	-0.373***
	[0.0350]	[0.0350]	[0.0350]	[0.0374]	[0.102]	[0.0374]	[0.102]
Distance from firm's HQ	-0.233***	-0.232***	-0.232***	-0.211***	-0.384***	-0.210***	-0.380***
	[0.0311]	[0.0311]	[0.0310]	[0.0331]	[0.0956]	[0.0331]	[0.0955]
Land price	-0.251***	-0.251***	-0.254***	-0.250***	-0.323***	-0.254***	-0.328***
	[0.0410]	[0.0410]	[0.0411]	[0.0443]	[0.114]	[0.0445]	[0.114]
Wald Chi-square	881.3***	884.8***	892.1***	921.4***		918.9***	
logLikelihood	-4487.1	-4485.3	-4481.7	-4467		-4468.3	
Locational choice-set (average)	402	402	402	402		402	
Number of entries	883	883	883	883		883	
Observations	357552	357552	357552	357552		357552	

Notes: ***p<0.01, **p<0.05, *p<0.10. Standard errors in brackets.

Table 6. Supplementary Analysis

	mixed logit		5 km geo scope		10 km geo scope		fixed effects		exporting plants		all multi-plant entries	
	exporting firms	non-exp. firms	exporting firms	non-exp. firms	exporting firms	non-exp. firms	exporting firms	non-exp. firms	exporting plants	non-exp. plants	exporting firms	non-exp. firms
non-exporting industry agglomeration	0.290***	0.222***	0.029	0.0303***	0.019	0.0237**	0.011	0.006	0.027	0.0360***	0.017	0.0138***
	[0.0807]	[0.0282]	[0.0220]	[0.00892]	[0.0234]	[0.00965]	[0.0224]	[0.00922]	[0.0256]	[0.00867]	[0.0141]	[0.00514]
TFP premium * non-exporting ind. aggl.	0.530	-0.326*	0.089	-0.0946***	0.152	-0.0748**	0.062	-0.0880***	0.076	-0.0938***	0.052	-0.016
	[0.673]	[0.183]	[0.0831]	[0.0315]	[0.0947]	[0.0335]	[0.0814]	[0.0317]	[0.0933]	[0.0319]	[0.0510]	[0.0186]
exporting industry agglomeration	0.018	-0.019	0.0338*	0.0194*	0.011	-0.003	-0.002	-0.003	0.0446**	0.0216**	0.0252*	0.0327***
	[0.0221]	[0.0262]	[0.0187]	[0.0105]	[0.0180]	[0.00957]	[0.0207]	[0.0120]	[0.0220]	[0.0105]	[0.0137]	[0.00616]
TFP premium * exporting ind. aggl	-0.076	-0.055	-0.013	0.007	0.028	-0.005	-0.043	-0.040	-0.023	-0.019	-0.014	0.007
	[0.224]	[0.137]	[0.0736]	[0.0360]	[0.0660]	[0.0339]	[0.0823]	[0.0411]	[0.0831]	[0.0384]	[0.0475]	[0.0223]
Buyer industry fit	0.057	0.040	0.264**	0.104***	0.193	0.0983**	0.232*	0.164***	0.076	0.052	0.052	0.0728**
	[0.177]	[0.0508]	[0.122]	[0.0394]	[0.127]	[0.0425]	[0.135]	[0.0632]	[0.137]	[0.0424]	[0.0846]	[0.0312]
Supplier industry fit	-0.424	-0.003	0.265	0.453*	1.115*	0.763***	0.988*	1.319***	0.081	0.419*	0.094	0.069
	[0.585]	[0.264]	[0.536]	[0.234]	[0.634]	[0.275]	[0.538]	[0.273]	[0.595]	[0.224]	[0.359]	[0.140]
Knowledge fit	-0.048	-0.002	-0.039	0.002	-0.052	0.002	0.007	0.003	0.028	0.001	-0.008	0.000
	[0.0391]	[0.00627]	[0.0315]	[0.00242]	[0.0372]	[0.00255]	[0.0399]	[0.00323]	[0.0397]	[0.00247]	[0.0156]	[0.00194]
Labor fit	0.541	-0.298	0.030	-0.009	0.068	-0.009	0.890	0.397	1.280*	0.326	1.071**	0.326**
	[0.782]	[0.305]	[0.0704]	[0.0206]	[0.0688]	[0.0216]	[0.725]	[0.340]	[0.733]	[0.261]	[0.446]	[0.162]
TFP premium * Buyer industry fit	-0.793	0.004	-0.414	0.089	0.212	0.047	-0.412	-0.189	0.978	-0.267	-0.051	-0.010
	[1.277]	[0.271]	[0.680]	[0.216]	[0.788]	[0.242]	[0.773]	[0.333]	[0.953]	[0.243]	[0.423]	[0.147]
TFP premium * Supplier industry fit	0.822	-2.063	0.571	-1.693	-0.609	-2.426*	0.444	-2.375**	2.310	-2.539**	-0.373	0.216
	[3.035]	[1.258]	[2.275]	[1.095]	[2.684]	[1.258]	[2.205]	[1.092]	[2.505]	[1.032]	[1.505]	[0.568]
TFP premium * Knowledge fit	-0.009	-0.010	-0.069	-0.005	-0.042	-0.009	-0.019	-0.002	-0.417	-0.007	0.044	0.003
	[0.241]	[0.0121]	[0.192]	[0.00957]	[0.217]	[0.0144]	[0.156]	[0.0166]	[0.286]	[0.00967]	[0.0420]	[0.00565]
TFP premium * Labor fit	-1.272	1.608	-0.032	0.065	-0.259	0.066	0.577	2.994**	-4.721	2.957***	-1.338	0.242
	[3.426]	[1.220]	[0.286]	[0.0664]	[0.400]	[0.0697]	[2.721]	[1.247]	[2.956]	[0.918]	[1.717]	[0.616]
HQ of the firm	-2.746	-7.093	-1.398	1.760***	-1.215	1.792***	-2.904	1.147***	-0.463	1.613***	1.917***	2.578***
	[3.024]	[6.586]	[1.453]	[0.373]	[1.421]	[0.373]	[1.961]	[0.435]	[1.342]	[0.366]	[0.600]	[0.184]
Firm plant other industry	-9.279	2.485***	2.097***	1.754***	2.089***	1.772***	1.607***	1.482***	3.048***	1.452***	1.888***	0.787***
	[9.359]	[0.359]	[0.558]	[0.353]	[0.559]	[0.354]	[0.614]	[0.383]	[0.603]	[0.339]	[0.385]	[0.255]
Min. distance from firm's other plant	-0.294**	-0.533***	-0.375***	-0.621***	-0.361***	-0.623***	-0.412***	-0.627***	-0.277**	-0.629***	-0.754***	-0.911***
	[0.147]	[0.0558]	[0.102]	[0.0373]	[0.102]	[0.0374]	[0.109]	[0.0425]	[0.121]	[0.0367]	[0.0645]	[0.0231]
Distance from firm's HQ	-0.501***	-0.289***	-0.376***	-0.210***	-0.362***	-0.208***	-0.557***	-0.354***	-0.367***	-0.216***	-0.188***	-0.271***
	[0.131]	[0.0495]	[0.0953]	[0.0330]	[0.0949]	[0.0329]	[0.109]	[0.0419]	[0.109]	[0.0325]	[0.0545]	[0.0202]
Land price	-0.394***	-0.291***	-0.337***	-0.261***	-0.311***	-0.243***	-0.101	-0.060	-0.397***	-0.249***	-0.348***	-0.386***
	[0.136]	[0.0507]	[0.115]	[0.0453]	[0.118]	[0.0469]	[0.584]	[0.569]	[0.140]	[0.0433]	[0.0771]	[0.0282]
Wald Chi-square	-8656.2***		892.6***		882.6***		1407.9***		933.2***		5124.0***	
logLikelihood	-4328.1		-4481.4		-4486.4		-3602.6		-4461.1		-10701.5	
Locational choice-set (average)	402.6		402.6		402.6		181.8		402.6		394.3	
Number of entries	883		883		883		883		883		2381	
Observations	357552		357552		357552		161312		357552		941720	

Notes: ***p<0.01, **p<0.05, *p<0.10. Standard errors in brackets.

Figure 1. The elasticity of the probability of entry with respect to industry agglomeration for different level of the TFP premium

Figure 1a. All entrants and industry agglomeration

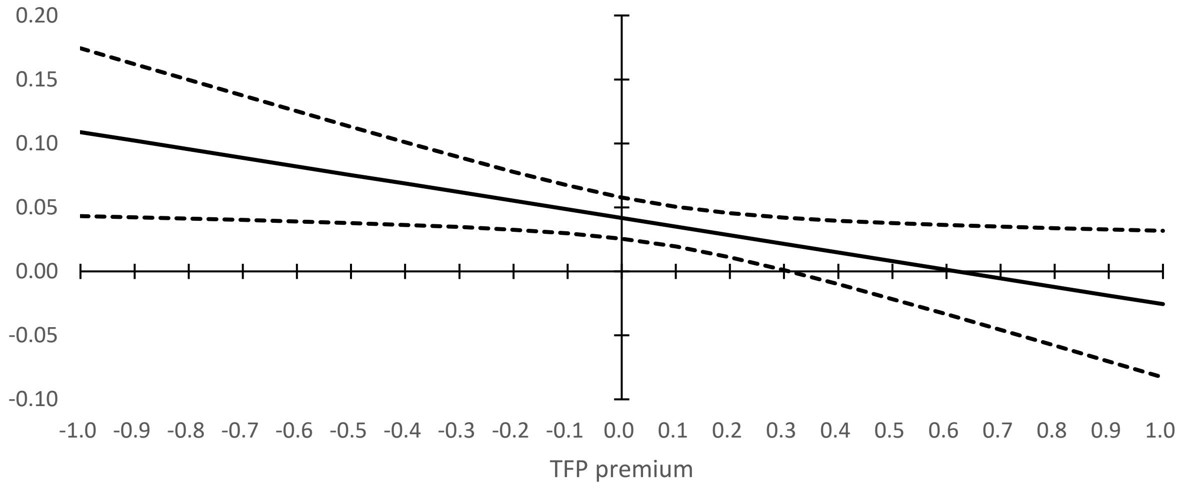
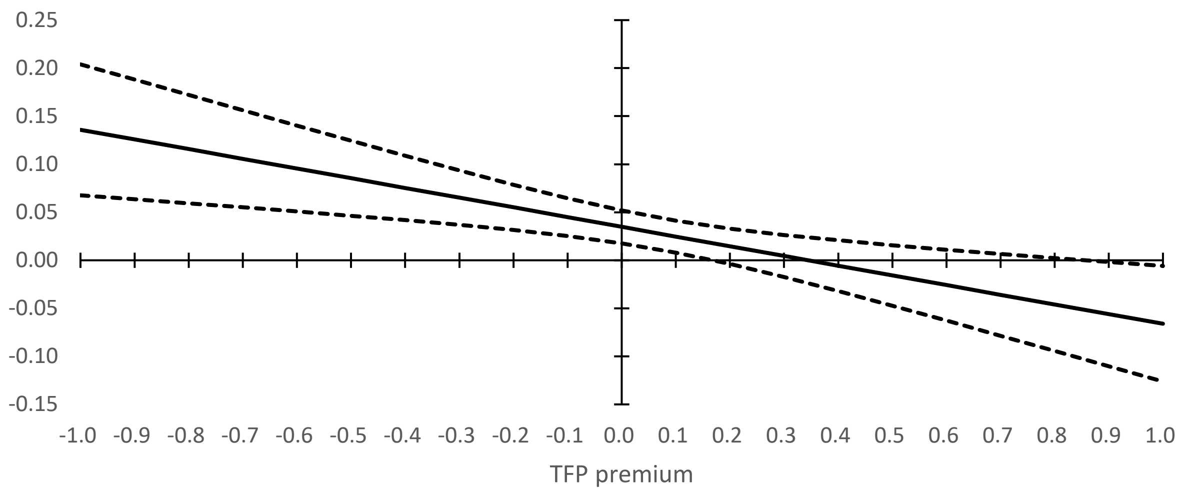


Figure 1b. Non-exporters and non-exporting industry agglomeration



Notes: the dashed lines indicate the 95% coefficient interval, the solid line the point estimate of the elasticity.