The Role of Market Potential in Location Choice: Evidence from Japanese Investment

Kazunobu HAYAKAWA
Faculty of Economics, Keio University, Japan
Research Institute of Economy, Trade and Industry, Japan

Toshiyuki MATSUURA* §
The Institute of Economic Research, Hitotsubashi University, Japan
Research Institute of Economy, Trade and Industry, Japan

Abstract: The purpose of this paper is to examine the validity of market potential as a proxy for demand linkages, from the viewpoint of "*who are consumers of goods*". Our claim is that an appropriate proxy for demand linkages crucially depends on the underlying theoretical models. Particularly in location choice of intermediate goods plants, we demonstrate that the widely-used Harris market potential is not an appropriate proxy for demand linkages. Thus, using consistent estimators obtained by restricting our sample only to finished goods plants, we strictly quantify the effect of the market potential on their location choice.

***Keywords*:** Market potential; Location choice; Japan
***JEL Classification*:** F23; H32; R34

---

1. Introduction

In the present global era in which firms choose the location of their plants beyond national borders, existence of agglomeration becomes one of the most important location elements. In addition to low business risks, proximity to large

---

* Corresponding author. Toshiyuki Matsuura, address: The Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8603 Japan. Phone: 81-42-580-8369; Fax: 81-42-580-8333. E-mail: matsuura@ier.hit-u.ac.jp
§ This research was conducted as a part of the project of Research Institute of Economy, Trade and Industry (RIETI). We thank the Ministry of Economy, Trade, and Industry of the Japanese government for releasing the micro data used in this study. The authors would like to thank Kyoji Fukao and other seminar participants at RIETI for helpful discussions. The opinions expressed in this paper are the sole responsibility of the authors and don't reflect the views of RIETI.

market of final consumption and formation of agglomeration are essential for developed countries in which labor costs are expensive. In Japan, for example, Ministry of Economy, Trade and Industry has facilitated the formation of agglomeration as "Industrial Cluster Project" since 2001. Also in developing countries which experience a gradual rise in wages, its formation is crucially important to deter the drain of multinational firms to developing countries with lower wage rates. Indeed, there are a large number of agglomerations such as Hsinchu in Taiwan, Jurong industrial park in Singapore, Samut Prakan and the Eastern Seaboard in Thailand, and Penang and Shah Alam in Malaysia, which seem to contribute to deter their drain to some extent.

A large number of studies have empirically investigated the role of agglomeration benefits such as demand linkages and cost linkages in firms'/plants' location choice. The recent references are the following: Head, Ries, and Swenson (1999), Castellani and Zanfei (2004), Crozet, Mayer, and Mucchielli (2004), Head and Mayer (2004), and Basile, Castellani, and Zanfei (forthcoming). In those studies, GDP or market potential introduced by Harris (1954), i.e., sum of distance-weighted GDP, has been used as a proxy for demand linkages[1], while output, value-added, or the number of firms in each industry as a proxy for cost linkages. The proxy for cost linkages is also sometimes constructed as sum of distance-weighted output (see, for example, Bekes, 2006). As a result, those studies consistently find the positive effects of demand and cost linkages on location choice.

Against those studies, this paper examines the validity of market potential as a proxy for demand linkages, from the viewpoint of "*who are consumers of goods*". We claim that an appropriate proxy for demand linkages crucially depends on the underlying theoretical models. Supposing the location choice of finished goods plants, their outputs are purchased by all people living in a region/country as household consumption or fixed capital investment. On the other hand, in the case of intermediate goods plants, the consumers of their outputs may be not only finished goods producers but also intermediate goods producers themselves. That is, a source of demand linkages, i.e., income of consumers, is qualitatively different according to

[1] Head and Mayer (2004) examine the effect of market potential directly derived from new economic geography model (Krugman's model) on location choice. That is, they employ the market potential measure taken into account the location of competitors (i.e., price index) rather than the simple sum of distance-weighted GDP. Such a measure is constructed by employing estimators of importing country dummy variables in the well-known gravity equation, as in Redding and Venables (2004). As a result, they find that "theory doesn't pay", in the sense that the Harris market potential outperforms the Krugman's market potential in both the magnitude of its coefficient and the fit of the model to be estimated.

the type of plants. Thus, while examining the effect of GDP/sum of distance-weighted GDP on location choice of finished goods plants seems to be plausible due to the principle of equivalent of three aspects[2], examining that of intermediate goods plants may be misleading.

This paper empirically investigates the effect of market potential on Japanese plants' location choice by using appropriate proxy variables for the underlying theoretical models. We employ the somewhat unique and finely-compiled dataset in the sense that it includes data on both plants in Japan and Japanese overseas plants.[3] Furthermore, we restrict our sample to finished goods plants. This restriction is because, to strictly analyze the location choice of intermediate goods plants, we need the relatively unobtainable data; total production values of finished goods and intermediate goods. We first theoretically demonstrate that applying estimation equation for finished goods plants to intermediate goods plants sample yields serious econometric and economic problems: aggregation bias, inconsistency due to the omitted variable and errors-in-variable problems, and the difficulty of interpretation for estimators. Last, we empirically show that a coefficient for market potential tends to suffer from downer bias in the sample of all types of plants compared with the sample of only finished goods plants.

The rest of this paper is organized as follows. The next section develops a model in which firms choose locations of their plants among regions. The location choice of finished goods plants and intermediate goods plants are separately examined. Section 3 presents our empirical equations for those two types of plants separately, and discusses econometric problems in applying the equations to a sample of the other type of plants. In section 4, we report empirical results, and in section 5, we conclude.

2. Location Choice Model

This section develops a model in which firms choose locations of their plants among regions. First, we take location choice of finished goods plants, and next that of intermediate goods plants is examined.

2.1. Finished Goods Plants

The representative consumer in each region is assumed to have a two tier utility

---

[2] GDP in the aspect of expenditure is demand for final products.

[3] Such a type of dataset is used also in Mayer, Mejean, and Nefussi (2007), which investigate whether investment abroad by multinational firms substitutes for investment at home.

function, which becomes the standard in international trade and new economic geography literature. The upper tier is a Cobb-Douglas function of the utility derived from consumption of finished goods. Specifically, we apply the following utility function of the consumer in region *r*:

$$U_r = \prod_{h=1}^{H} \left(C_r^h\right)^{\alpha_h}, \quad \sum_{h=1}^{H} \alpha_h = 1,$$

where C_r^i is the aggregate consumption of finished good *i* in region *r*.

We formalize expenditure allocation across finished machinery goods consisting of multiple varieties and omit the subscript representing the name of finished goods for now. The consumer has the following preference specified as constant elasticity of substitution (CES) function over the varieties:

$$C_r = \left(\sum_{i=1}^{R} \int_0^{N_r} x_{r,i}\left(j\right)^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}},$$

where *R*, *N*, and $x_{r,i}$ (*j*) are the number of countries, the number (mass) of finished varieties, and the demand of region *r* for the finished variety *j* produced in region *i*, respectively. σ is the elasticity of substitution between finished varieties and is assumed to be greater than unity. The utility maximization yields:

$$x_{r,i} = \alpha t_{r,i}^{1-\sigma} p_i^{-\sigma} P_r^{\sigma-1} Y_r, \tag{1}$$

where p_i and P_r denote the price of the variety produced in region *i* and the price index in region *r*, respectively. Y_r is total expenditure in region *r*. Transactions in finished goods between regions *r* and *s* is modeled as facing Samuelsonian iceberg costs, $t_{r,s}$.

The market structure in finished goods sector is assumed to be Chamberlinian monopolistic competition. The finished goods producer of each region combines a composite index aggregated across varieties of intermediate inputs and primary factors, e.g., labor and physical capital, using Cobb-Douglas model. The composite enters the cost function for each producer through a CES aggregator. Specifically, we have the following cost function:

$$C(x_r) = w_r^{1-\mu} G_r^{\mu} x_r + F_r, \quad G_r = \left[\int_0^{M_r} q_r\left(j\right)^{\frac{1}{1-\upsilon}} dj \right]^{1-\upsilon},$$

where w_r denotes price index for primary factors employed by each producer to produce finished output x_r. G_r is price index for intermediate goods, and F_r is fixed costs. μ is a linkage parameter between finished and intermediate goods. M_r, $q_r(j)$, and ν are the number (mass) of intermediate varieties produced in region *r*, the price of *j*-th variety produced in region *r*, and the elasticity of substitution between intermediate goods, respectively. The elasticity is again assumed to be greater than unity. Notice that intermediate goods market is assumed to be segmented; transaction costs of intermediate goods across countries are prohibitively high. It will be clear later that this assumption is only for the easy comparison of profit function between finished and intermediate goods producers. Each firm maximizes its profit with respect to quantity to derive producer prices:

$$p_r = \left(\frac{\sigma}{\sigma - 1}\right) w_r^{1-\mu} G_r^{\mu} . \tag{2}$$

Suppose that firms choose locations of their plants among *R* regions. Using (1) and (2), we can derive a profit function of the firm locating its plant in region *r* :

$$\pi_r = \alpha \sigma^{-\sigma} (\sigma - 1)^{\sigma-1} w_r^{(1-\mu)(1-\sigma)} G_r^{\mu(1-\sigma)} \left(\sum_{i=1}^{R} \frac{P_i^{\sigma-1} Y_i}{t_{i,r}^{\sigma-1}} \right) - F_r .$$

We hereafter call the second bracket of the RHS, i.e., $\sum_{i=1}^{R} \frac{P_i^{\sigma-1} Y_i}{t_{i,r}^{\sigma-1}}$, market potential, which is denoted by MP_r. Thus the profit function is rewritten by:

$$\pi_r = \alpha \sigma^{-\sigma} (\sigma - 1)^{\sigma-1} w_r^{(1-\mu)(1-\sigma)} G_r^{\mu(1-\sigma)} MP_r - F_r . \tag{3}$$

This profit function indicates that low price index for primary factors, low price index for intermediate goods (cost linkages), and good access to total expenditure (demand linkages) have a positive impact on the profit of finished goods plants.

2.2. Intermediate Goods Plants

In the case of location choice of intermediate goods plants, the profit function (3) qualitatively changes. We consider the production technology with horizontal linkages (see, for example, Krugman and Venables, 1995). That is, intermediate goods are produced not only with primary factors but also with intermediate goods themselves. As in the finished goods producer, intermediate goods producer of each region combines a composite index aggregated across varieties of intermediate inputs and primary factors using Cobb-Douglas model. The composite enters the cost function for each producer through a CES aggregator. Then the profit function is given by:

$$\pi_r^I = \nu^{-\nu}(\nu-1)^{\nu-1} w_r^{(1-\nu)(1-\lambda)} G_r^{\lambda(1-\nu)} \left(\sum_{i=1}^{R} \frac{G_i^{\nu-1}(\mu X_i + \lambda Z_i)}{\tau_{i,r}^{\nu-1}} \right) - F_r^I , \quad (4)$$

where λ is a linkage parameter among intermediate goods. τ and F^I are iceberg costs and fixed costs, respectively. X_i is equal to $N_i p_i \sum_r x_{r,i}$. Z_i is equal to $M_i q_i z_i$, where z_i denotes total production of intermediate varieties produced in region *i*. In this case, demand linkages become a complicated composition. The magnitude of intermediate goods production as well as finished goods production is positively related to the profit of intermediate goods plants.

Assuming the prohibitively high ice-berg costs, we obtain:

$$\pi_r^I = \nu^{-\nu}(\nu-1)^{\nu-1} w_r^{(1-\nu)(1-\lambda)} G_r^{(\lambda-1)(1-\nu)} (\mu X_r + \lambda Z_r) - F_r^I . \quad (5)$$

Using the relationship:

$$X_r = \left(\frac{\sigma\alpha}{\sigma-1} \right) n_r p_r^{1-\sigma} MP_r = \alpha\sigma^{2-\sigma}(\sigma-1)^{\sigma-2} n_r w_r^{(1-\mu)(1-\sigma)} G_r^{\mu(1-\sigma)} MP_r ,$$

the profit function can be rewritten as:

$$\pi_r^I = \nu^{-\nu}(\nu-1)^{\nu-1} w_r^{(1-\nu)(1-\lambda)} G_r^{(\lambda-1)(1-\nu)} \left[\alpha\mu \left(\frac{\sigma}{\sigma-1} \right)^{2-\sigma} n_r w_r^{(1-\mu)(1-\sigma)} G_r^{\mu(1-\sigma)} MP_r + \lambda Z_r \right] - F_r^I . \quad (6)$$

The operating profit of intermediate goods plants is now related non-log-linearly to market potential.

3. Empirical Issues

3.1. Empirical Specification

Our purpose is to evaluate the effect of demand linkages on location choice of plants. To this end, we must use the appropriate proxy variable for the expenditure of goods that the plants produce. In the location choice of finished goods plants, Gross Domestic Expenditure becomes the good proxy for that since consumers of finished goods are all people living in the region. On the other hand, in the location choice of intermediate goods plants, the total production values of finished goods and those of intermediate goods are good proxies though non-linear estimation techniques are necessary in order to estimate the profit function. In addition, the use of a direct measure, i.e., the expenditure of intermediate goods in the region, may be better than those variables. In either case, to obtain those data is a difficult task, and the input-output table seems to be the only data source. Moreover, our sample covers many countries, so those data are unavailable. As a result, we focus on the location choice of only finished goods plants.

The location model can be estimated by using conditional logit. In the profit function (3), F_r is assumed to be identical across regions for tractability, as in Head and Mayer (2004). As monotonic transformations leave ordering of the profit unchanged, the firm chooses the region in which the following log-function is maximized:

$$\ln \Pi_r = V_r + \varepsilon_r = (1-\mu)(1-\sigma)\ln w_r + \mu(1-\sigma)\ln G_r + \ln MP_r + \varepsilon_r, \qquad (7)$$

where ε_r denotes unobservable regional characteristics. McFadden (1974) demonstrates, when ε_r is independent and follows identical type I extreme value distribution across regions, the probability that the firm locates its affiliate k in region r is given by:

$$P_r^k = \frac{\exp(V_r^k)}{\sum_i \exp(V_i^k)},$$

where V_r^k denotes the affiliate k's V_r. The coefficients are estimated by maximum likelihood procedures.

It is worth pointing out econometric problems in estimating equation (7) for the location choice of intermediate goods plants. First, taking a look at the functions (5) and (6), we can find that the demand component, i.e., $\mu X_r + \lambda Z_r$, is not log-linearly related to MP_r. Therefore, an errors-in-variable problem, which results in inconsistency of estimators, emerges in estimating equation (7) for location choice of intermediate goods plants. Second, the powers of w_r and G_r in the function (5) are different from those in the function (3). In particular, the power of G_r becomes positive, implying that the magnitude of its coefficient may suffer from a serious aggregation bias in applying the equation (7) to all plants. Third, relating to the second point, G_r in the function (5) plays a role to capture a part of demand linkages, not cost linkages. Above all, small G_r implies not good access to input markets but the existence of many competitors for the intermediate goods plant, resulting in its low operating profit.

Without horizontal linkages among intermediate goods (i.e., $\lambda = 0$), furthermore, clearer contrasts emerge. Then, the function (6) can be rewritten as:

$$\pi_r^I = \alpha\mu\nu^{-\nu}(\nu-1)^{\nu-1}\left(\frac{\sigma}{\sigma-1}\right)^{2-\sigma} n_r w_r^{(1-\nu)+(1-\mu)(1-\sigma)} G_r^{(\nu-1)+\mu(1-\sigma)} MP_r - F_r^I .$$

Although this function is similar to the function (3), this contains the number of finished varieties. Omitting the number in estimation equation yields a well-known omitted variable problem, resulting in making estimators inconsistent. In addition, the powers of w_r and G_r in this function are again different from those in the function (3).

In sum, the estimation of equation (7) for the location choice of intermediate goods plants yields the econometric problems of omitted variable and errors-in-variable, making estimators inconsistent. In addition, the magnitude of some coefficients suffers from the aggregation bias and the difficulty of interpretation. Moreover, to strictly analyze the location choice of intermediate goods plants, we need the relatively unobtainable data; total production values of finished goods and intermediate goods. To avoid these problems, we restrict our attention only to the location choice of finished goods plants.

3.2. Location Choice of Japanese Plants

Our data sources of location choice are the longitudinal data sets of “Census of Manufactures” and “Basic Survey of Overseas Business Activities”. In the Census of Manufactures, data on establishments locating in Japan (e.g., location, the number of employees, tangible assets, and value of shipments) are available. Those data in Japanese overseas affiliates are available in the Basic Survey of Overseas Business Activities.[4] The information on parent firms of establishments/affiliates, e.g., the number of employees, can be obtained from the Basic Survey of Japanese Business Structure and Activities. These censuses and surveys are conducted and published by Ministry of Economy, Trade and Industry (METI). As a result, our dataset is somewhat unique one in the sense that it includes data on both plants in Japan and Japanese overseas plants. Such a dataset is useful in investigating global location choice among possible all regions. As for more detail data compilation procedures, see Appendix 1.

By using the dataset, we analyze the effects of regional characteristics on the location choice of Japanese finished goods plants in 30 regions during the period 1989-2003. 8 regions of them are Japanese regions, and the rest of them are foreign

[4] In this survey, subsidiaries and sub-subsidiaries are collectively referred to as overseas affiliates. The overseas affiliates include foreign affiliates in which a Japanese corporation(s) has invested capital of 10% or more, and in which a subsidiary, funded more than 50% by a Japanese corporation(s), has invested capital of more than 50%, and in which a Japanese corporation(s) and a subsidiary funded more than 50% by a Japanese corporation(s) have invested capital of more than 50%.

countries. The foreign countries are 8 East Asian countries and 10 developed countries. Mainland China is further divided into 4 regions. The regional definition is listed in the Appendix 2. We focus on the Japanese finished goods plants in machinery industry, which consists of several machinery sectors such as office machines, household machines, and electrical machines. In particular, based on the more detailed sectoral classification, we can pick up finished goods subsectors such as household electric appliances and motor vehicles. We can also identify some intermediate goods subsectors, which are electronic parts and motor vehicles parts and accessories. The rest of the subsectors include both finished goods and intermediate goods.[5] As for the sectoral classification, see Appendix 3.

It is worth noting two points in our location data. First, we linked the Census of Manufactures and the Basic Survey of Overseas Business Activities with the Basic Survey of Japanese Business Structure and Activities, of which targets are firms engaged in business with both a minimum capital of 30 million yen and 50 or more employees. Thus the establishments/affiliates with small parent firms are excluded from our sample. Second, in this paper, affiliates and establishments are collectively referred to as "plants", though we use affiliate-level data in overseas location due to the data availability. In order to address such inconsistency to some extent, we introduce an overseas dummy variable, which takes unity if the chosen region is overseas and zero otherwise, into our equation to be estimated. In addition, this dummy variable would play a role to partly control differences in fixed costs between domestic location and overseas location.

New entry of Japanese finished goods plants is reported in Table 1. In the table, that of intermediate goods plants is also reported. Taking a look at the entry of finished goods plants, we can find their concentration in some regions: Tohoku, North Kanto, and South Kanto in Japan, East China including Shanghai, and the U.S. Their new entry remarkably decreases from the former half to the latter half of our sample period. Only South Kanto in Japan has experienced its increase. On the other hand, the new entry of intermediate goods plants increases in most of the regions compared with finished goods plants. In East Asia, the Philippines, Thailand, Indonesia, Korea, and South China have experienced a certain increase of intermediate goods plants' entry. Consequently, most of the Japanese plants have located in domestic regions

[5] Our definition of finished goods subsectors is the ones in which a share of manufactures' intermediate demand in total domestic demand is less than 30%. On the other hand, that of intermediate goods subsectors is the ones in which it is greater than 70%. Such a share is calculated by using the Input-Output Tables (Ministry of Internal Affaire and Communications of Japan).

though the number of their entry in Mainland China has been relatively large.

=== Table 1 ===

3.3. Independent Variables

Our independent variables are price index for primary factors, intermediate goods, and market potential. For price index for primary factors, we use the average wages in each region, which are estimated by aggregating the Census of Manufactures and the Basic Survey of Overseas Business Activities. In addition, to control the other fundamental economic/social conditions to some extent, we introduce country risk index, which is drawn from "Institutional Investor". This index is the aggregate of bankers' evaluation on risk of default, and the larger index indicates that the risk of default in the country is smaller.

As for market potential, we use the Harris market potential index, i.e., sum of distance weighted-real GDP, rather than the Krugman-type variable used in Head and Mayer (2004). The construction of the latter variable requires inter-regional transaction values, which are unavailable in our sample regions. Specifically, the Harris market potential index is the following:

$$MP_r^{Hariss} = \sum_{i=1}^{R} \frac{GDP_i}{dist_{i,r}},$$

where $dist_{i,r}$ denotes a great distance between regions i and r. Following border effect literature (see, for example, Head and Mayer, 2000), we use (2/3) times a radius of surface area in the region for the intra-regional distance. Our formulation of market potential implies that δ (σ-1) is assumed to be unity in log-trade costs function, $\ln t_{i,r} = \delta \ln dist_{i,r}$. This assumption is plausible since δ (σ-1) is estimated to be 1.08 in machinery industry in Head and Mayer (2004). Data on distance are drawn from CEPII website.[6] The data on GDP and GDP deflator in each country can be obtained from "World Development Indicator" (World Bank). Those in Taiwan, Japanese regions, and Chinese regions are from "Statistical Yearbook of the Republic of China" (Taipei: Directorate-Genral of Budget, Accounting and Statistics, Executive Yuan, Republic of China), "Annual Report on Prefectural Accounts" (Cabinet Office of Japanese Government), and "China Statistical Yearbook" (National Bureau of Statistics of China), respectively.

As usual, the data on price index for intermediate goods are unavailable. In this paper, we use the variable reflecting the magnitude of agglomeration on intermediate

[6] http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

goods producers as a proxy for that. Since from the theoretical point of view the price index for intermediate goods is low in the regions with such large agglomeration, this proxy is plausible enough. Although such a variable is often constructed by using the number of employees or value-added, those data are unavailable in finished goods and intermediate goods separately in our sample regions. Thus, in this paper, we construct the agglomeration variable by using the aggregated data of finished goods and intermediate goods. Active input-output relationship between intermediate goods producers and finished goods producers and that among intermediate goods producers would develop in large agglomeration in a sector. Therefore, our proxy variable seems to be plausible at least to control variation in price index for intermediate goods across regions.

We use intra-sectoral agglomeration index (*ISA*) as the agglomeration variable, which is often used in empirical analysis of economic geography (see, for example, Hanson, 1998; Tomiura, 2003). *ISA* in sector *h* in region *r* is given by:

$$ISA_r^h = \frac{VA_r^h \Big/ \sum_k VA_r^k}{\sum_i VA_i^h \Big/ \sum_i \sum_k VA_i^k},$$

where VA_r^h denotes value-added in sector *h* in region *r*. The data on value-added by sector in each country can be drawn from "International Yearbook of Industrial Statistics" (United Nations Industrial Development Organization). Those in Taiwan, Japanese regions, and Chinese regions are from "Statistical Yearbook of the Republic of China", "Census of Manufactures", and "China Industrial Economic Statistical Yearbook", respectively. The value-added is deflated by Japanese deflator by sector, which is compiled by using the Japan Industry Productivity (JIP) data base[7].

Last, as argued above, we introduce an overseas dummy variable. To control qualitative differences in foreign direct investments (FDIs) between East Asian countries and developed countries (i.e., vertical FDIs and horizontal FDIs), we decompose the overseas dummy variable into East Asia dummy and developed countries dummy, and introduce them into our equation.

4. Empirical Results

4. 1. Baseline Results

We begin with the report of location choice results in a conventional sample used

[7] As for the details of Japan Industry Productivity (JIP) data base, which is downloadable from http://www.rieti.go.jp/jp/database/d05.html, see Fukao et al. (2006).

in the literature, i.e., all plants. Basic statistics are reported in Table 2. The result of conditional logit estimation for all plants sample is provided in the column “ALL” in Table 3.

=== Tables 2 & 3 ===

We can find that all coefficients are significantly estimated with expected signs. Contrary to the results in Head and Mayer (2004), the low price index for primary factors such as labors is one of the important elements for location choice of all plants. The Japanese plants’ entry to developing countries particular to China is considerably motivated by their low wages. The regions with large intra-sectoral agglomeration succeed in attracting Japanese plants. Although the primary purpose of *ISA* is to control variation in price index for intermediate goods, this variable captures various kinds of industrial agglomeration benefit, such as knowledge spillover. Thus, this result also indicates that such agglomeration benefit has been a key element of location choice. As in most of the previous studies, the proximity to final consumers is positively related to location choice, though this variable does not exactly reflect the proximity to demand in location choice of intermediate goods plants.

As argued before, however, the results of these coefficients in the sample of all plants may suffer from the aggregation bias and further being inconsistent. To avoid such bias and inconsistency, we next estimate for the sample of finished goods plants. The results are reported in the column “Finished” - “Sector”. All coefficients have again significantly expected signs and should be consistent estimators. Our interest here lies in the result in market potential. The magnitude of its coefficient is larger than that in the sample of all plants. As demonstrated in the previous section, access to final consumers does not affect location choice of some intermediate goods plants. Consequently, incorporating intermediate goods plants in sample produces not only inconsistency but also a downer bias in the coefficient for market potential. Last, disappointingly, the overall fit of the model is almost unchanged with that in all the plants’ samples. Pseudo R-square declines by 0.065. As in Head and Mayer (2004), *theory doesn’t pay*, in the sense that the model with the widely-used sample fits as well as that with the theory-based sample.

So far, we have not taken both theoretically and empirically firms’ heterogeneity into account. But the well-known Helpman-Melitz-Yeaple model (hereafter, HMY model) indicates that only firms with higher productivity can afford to pay expenses for entry to overseas (Helpman, et al., 2007). This argument suggests to us that we control firms' productivity in order to address sample selection bias to some extent. We introduce a large firm slope dummy both with East Asia and with developed countries into our equation. The large firm dummy takes unity if the plants belong to their firms with more than 1,000 employees and zero otherwise. Since firms with higher productivity produce more outputs and thus employ more employees in the HMY model, the use of the large firm dummy variable is plausible.

The results of conditional logit estimation with the slope dummy variable are reported in the fourth column in Table 3. Three points are to be worth noted in the results. First, the results in the coefficients are both qualitatively and quantitatively unchanged. The coefficient for market potential is again estimated to be larger than that in the sample of all plants. Second, both kinds of slope dummy are significantly estimated to be positive. This result supports the HMY argument and implies that compared with low productivity-firms, firms with high productivity tend to locate their plants in overseas. Third, the overall fit of the model greatly improves. Pseudo R-square doubles. In sum, we must take care of firms' heterogeneity in the case of location choice among domestic and overseas regions.

Last, we quantify a contribution of each location element on location choice. To this end, we first imagine a hypothetical region with the mean level of independent variables at t and denote this region's probability of being chosen as P_t. Second, we calculate region r's probability of being chosen when the independent variable i of hypothetical region approaches to that of the region r at t and denote it as P_{irt}. Last, we define (P_{irt} - P_t) as the contribution of independent variable i to the region r's probability of being chosen. We use mean values of sample periods for each independent variable. Using the result of the third column in Table 3, we can depict the contribution of each location element in Figure 1.

=== Figure 1 ===

There are three noteworthy points. First, low wages are the most crucial element in East Asian developing countries and Mexico, and market potential in most of the Japanese regions and European countries. These results seem to reflect that choosing the former regions is vertical direct investment and that choosing the latter regions (European countries) is horizontal direct investment. Second, intra-sectoral agglomeration is also a key element in some developed countries. Particularly in the regions with small market potential, e.g., Tohoku in Japan and Singapore, the significance of the intra-sectoral agglomeration is outstanding. This implies that even in the regions with immature markets of final consumption, the formation of agglomeration succeeds in attracting plants' location. Last, the lowness of country risk encourages plants' entry to developed countries.

4. 2. Robustness Checks

We perform two kinds of robustness checks. The first is to identify finished goods plants based on another criterion rather than based on the subsectors to which plants belong. For sales to final consumers, firms need to incur high costs for advertisement. In other words, firms with high expenses of advertisement seem to be engaged in the production of finished goods. In particular, we define the firms in which a share of advertisement costs in total costs is greater than its median as finished goods firms. As explained in section 3.2, we choose the finished goods plants based on the detailed industry-level input-output structure. However, at the plant or firm level, not all firms or plants produces finished goods even if they are not categorized as "finished goods sector". As a result, our sample would successfully include finished goods plants which are not categorized into either finished goods subsectors or intermediate goods subsectors.

The conditional logit results by such a criterion are reported from the fifth to sixth column in Table 3. The results are qualitatively unchanged with those by the previous criterion. All the coefficients remain to have significant expected signs. The coefficient for market potential is again estimated to be much larger than that in the sample of all plants. But disappointingly, Pseudo R-square is down by half. In sum, as in the previous results, theory does pay in the sense of magnitude but does not pay in the sense of model's fit.

The second robustness check is to estimate using nested logit. As is well known, the conditional logit model is based on the assumption of independence from irrelevant alternatives (IIA). However, our alternatives are less likely to be equally substitutable. In particular, domestic regions should be grouped into one subset and overseas regions in another. In this sense, the nested logit model is a viable alternative to the conditional logit model because it relaxes the restrictive IIA property by closely grouping substitutable alternatives into one subset.

In the upper level of the decision tree (domestic vs. overseas), three independent variables are introduced. The first one is, as before, a proxy for firms' productivity, i.e., large dummy variable. The second is a share of R&D in sales, which may be a proxy for two elements. On the one hand, R&D has an aspect of firm specific assets, yielding economies of scale at firm-level (Navaretti and Venables, 2004, pp.25). Thus, firms with its large share tend to locate their plants abroad in order to broaden a market. Such a location pattern is sometimes called horizontal foreign direct investments. For vertical foreign direct investors, on the other hand, firms with its large share may tend to fragment labor-intensive production processes into the countries with low wages and to specialize their domestic activities in R&D or high-tech processes. In either case, firms with a large R&D share seem to tend to locate their plants abroad. The third independent variable is trade-status dummy, which takes unity if firms are engaged in international trade activities (exports or imports) and zero otherwise. Compared with non-trading firms, trading firms already have much knowledge on foreign markets and thus face low costs to enter foreign markets. Therefore, plants belonging to trading firms may tend to be located abroad.

The nested logit results are reported in Table 4. First, the results of likelihood ratio test support the validity of our nesting structure, implying that national border still matters for Japanese investors. Second, the coefficients in the upper level of the decision tree are well estimated. All the coefficients are significantly negative, implying that firms with high productivity, a high R&D share, or trade experience are likely to locate their plants abroad. Third, the results in the lower level of the decision tree are unchanged with those in the previous table. All the coefficients have significantly expected signs, and in particular, the order in the magnitude of coefficients for market potential has been preserved: ALL < Finished. Fourth, Pseudo R-square in finished goods plants sample is again as high as or lower than that in all plants sample.

=== Table 4 ===

5. Concluding Remarks

In this paper, we examine the validity of market potential as a proxy for demand linkages, from the viewpoint of "*who are consumers of goods*". Our claim is that an appropriate proxy for demand linkages crucially depends on the underlying theoretical models. Particularly in location choice of intermediate goods plants, we demonstrate that the widely-used Harris market potential is not an appropriate proxy for demand linkages. As a result, we find that the coefficient for the market potential suffers from a downer bias in applying not only to finished goods plants but also to intermediate goods plants though then the fit of the model becomes better.

Our paper suggests various avenues for future research. One important direction is to examine effects of market potential in location choice of intermediate goods plants by using an appropriate proxy variable. Comparing the effects of market potential between location choice of finished goods plants and that of intermediate goods plants, we may get some fruitful insights. Then, however, we need to overcome data limitation; data of total production values on intermediate goods and finished goods are sometimes unavailable. Another direction is to ask whether or not the effects of demand linkages in location choice are different between vertical FDIs and horizontal FDIs. We did not explicitly distinguish those types of FDIs, which is beyond our scope in this paper. Although few theoretical models in the vertical FDIs shed light on the role of agglomeration benefit, cost linkages should be one of the important location advantages, in which differences between host country and home country are the key motivation of vertical FDIs. Therefore, cost linkages may be a more important regional element for attracting vertical FDIs than demand linkages. On the other hand, demand linkages would be a more important driving force of horizontal FDIs.

Appendix 1. Data Compilation Procedures

Our dataset is compiled by linking three micro data sources: the Census of Manufactures (COM), the Basic Survey of Overseas Business Activities (BSOBA), and the Basic Survey of Japanese Business Structure and Activities (BSJBSA) provided by Ministry of Economy, Trade and Industry (METI). The COM reports information on establishments locating in Japan, while the BSOBA on Japanese overseas affiliates. The information on firms locating in Japan is reported in the BSJBSA. In this appendix, we report our procedures of the link of these three data.

At first, we link plant data from the COM and firm data from the BSJBSA. Although both surveys are conducted by the METI, each survey has original firm identification (ID) code respectively and there is no matching table between the code in the COM and the code in the BSJBSA. Therefore, we match firms between the COM and the BSJBSA, referring to the firms' name, their telephone number, and their other information such as address. In addition, though the firm ID number for the

COM is available from 1994 to 2003, the firm ID number is drastically revised between 1996 and 1997. Thus, we need to make matching table by ourselves by referring to the firm ID number of continuing plants. Consequently, the result of the link between the COM and the BSJBSA seems to be good enough. The ratio of the number of matched plants data to the number of total manufacturing establishments reported in the BSJBSA is more than 95%.[8]

Next, the BSOBA is linked with the BSJBSA. First of all, since the METI has revised both parent firm code and affiliate code every year, we make matching table for parent firm code and affiliate code and complete panel dataset.[9] Second, based on the firms' information, we match firms between the BSJBSA and the BSOBA. While the BSOBA covers almost all industries except for Finance and Insurance, the coverage of the BSJBSA is restricted to mining, manufacturing, wholesale and retail, and some service industry. Therefore, not all foreign affiliates in the BSOBA are linked with BSJBSA.

Last, it is worth noting how to identify entry years of domestic establishments. While the BSOBA has a survey item of date of establishment/capital participation, the COM does not have the corresponding item. Therefore, we define the entry year of each establishment as the year in which the establishment first answers survey form of the COM.

[8] Note that since the BSJBSA covers the firms with more than 50 employee and 30 million capital amount, the establishments which belong to small enterprises, cannot be linked with firm-level data. The ratio of the number of matched plants to total number of plants in the COM is about 10%.

[9] For details of the BSOBA panel dataset, see also Kiyota, Matsuura, Urata, and Wei (forthcoming).

Appendix 2. Our Sample Regions

Country(ies)	Region	Name
Japan	Hokkaido	Hokkaido
	Tohoku	Aomori, Iwate, Miyagi, Akita, Yamagata, Fukushima
	North Kanto	Ibaraki, Tochigi, Gunma, Niigata, Yamanashi, Nagano
	South Kanto	Chiba, Saitama, Tokyo, Kanagawa
	Tokai-Hokuriku	Toyama, Ishikawa, Fukui, Shizuoka, Aichi, Gifu, Mie
	Kinki	Shiga, Kyoto, Oosaka, Hyougo, Nara, Wakayama
	Chuugoku-Shikoku	Tottori, Shimane, Okayama, Hiroshima, Yamaguchi, Tokusima, Kagawa, Ehime, Kochi
	Kyuusyu-Okinawa	Fukuoka, Saga, Nagasaki, Kumamoto, Ooita, Miyazaki, Kagoshima, Okinawa
East Asia	Malaysia	Malaysia
	Philippines	Philippines
	Thailand	Thailand
	Indonesia	Indonesia
	Singapore	Singapore
	Taiwan	Taiwan
	Korea	Korea
	Hong Kong	Hong Kong
	North China	Beijing, Tianjin, Heibei, Liaoning
	East China	Shanghai, Jiangsu, Zhejiang, Shandong
	South China	Guangdong, Guangxi,Hainan, Fujian
	West China	Hubei, Neimenggu, Jilin, Heilongjiang, Anhui
		Jiangxi, Henan, Shanxi, Hunan
		Chongqing, Sichuan, Guizhou, Yunnan, Tibet
		Shanxi, Gansu, Qinghai, Ningxia, Xinjiang
Developed countries	U.S.	U.S.
	Canada	Canada
	Mexico	Mexico
	U,K.	U,K.
	France	France
	Germany	Germany
	Italy	Italy
	Netherland	Netherland
	Spain	Spain
	Australia	Australia

Appendix 3. Our Sample Sectors and Subsectors

Sectors (ISIC Rev. 3 Code)	Sub-sectors	Type
Office, service industry and household machines (300)	Office, service industry and household machines	Finished
Electrical machinery, equipment and supplies (311, 312, 313, 314, 315, 319)	Industrial electrical apparatus	
	Household electric appliances	Finished
	Electronic equipment	
	Miscellaneous electrical machinery equipment and supplies	
Information and commnunication electronics equipment and of electronic parts and devices (321, 322, 323)	Communication equipment and related products	
	Electronic data processing machines digital and analog computer, equipment and accessories	Finished
	Electronic parts and devices	Parts
Transportation equipment (341, 342, 343, 351, 352, 353, 359)	Motor vehicles, motor vehicle bodies and trailers	Finished
	Motor vehicle parts and accessories	Parts
	Miscellaneous transportation equipment	
Precision instruments and machinery (331, 332, 333)	Optical instruments and lenses	Finished
	Watches, clocks, clockwork-operated devices and parts	
	Miscellaneous precision instruments and machinery	

References

Basile, R., Castellani, D., and Zanfei, A., forthcoming, "Location Choices of Multinational Firms in Europe: The Role of EU Cohesion Policy", forthcoming in *Journal of International Economics*.

Bekes, G., 2005, "Location of Manufacturing FDI in Hungary: How Important are Intercompany Relationships?", MNB Working Paper, 12, WP 2005/7.

Castellani, D. and Zanfei, A., 2004, "Choosing International Linkage Strategies in the Electronics Industry: the Role of Multinational Experience", *Journal of Economic Behavior and Organization*, **53**: 447-475.

Crozet, M., Mayer, T., and Mucchielli, J-L., 2004, "How Do Firms Agglomerate? A Study of FDI in France", *Regional Science and Urban Economics*, **34**: 27-54.

Disdier, A.C. and Mayer, T., 2004, "How Different is Eastern Europe?: Structure and Determinants of Location Choices by French Firms in Eastern and Western Europe", *Journal of Comparative Economics*, **32**: 280-296.

Fukao, K., Hamagata, S., Inui, T., Ito, K., Kwon, H., Makino, T., Miyagawa, T., Nakanishi, Y., and Tokui, J., 2006, "Estimation Procedures and TFP Analysis of the JIP Database 2006", RIETI, Discussion Paper, No. 07-E-003.

Hanson, G., 1998, "Regional Adjustment to Trade Liberalization", *Regional Science and Urban Economics*, **28**: 419-444.

Harris, C., 1954, "The Market as a Factor in the Localization of Industry in the United States", *Annals of the Association of American Geographers*, **64**: 315-348.

Head, K. and Mayer, T., 2000, "Non-Europe: The Magnitude and Causes of Market Fragmentation in Europe", *Weltwirschaftliches Archiv*, **136**: 285-314.

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, K., Ries, J., and Swenson, D., 1999, “Attracting Foreign Manufacturing: Investment Promotion and Agglomeration”, *Regional Science and Urban Economics*, **29**: 197-218.

Helpman, E., Melitz, M., and Yeaple, S., 2004. “Export versus FDI with Heterogeneous Firms”, *American Economic Review*, **94**(1): 300–316.

Krugman, P. R. and Venables, A. J., 1995, “Globalization and the Inequality of Nations”, *Quarterly Journal of Economics*, **110**: 857-880.

Kiyota, K., Matsuura, T., Urata, S., and Wei, Y., forthcoming, “Reconsidering the Backward Vertical Linkage of Foreign Affiliates: Evidence from Japanese Multinationals”, forthcoming in *World Development*.

Mayer, T., Nejean, I., and Nefussi, B., 2007, “The Location of Domestic and Foreign Production Affiliates by French MNCs”, CEPII, Working Paper, No. 2007-07.

Navaretti, B. and Venables, A. J., 2004, *Multinational Firms in the World Economy*, Princeton University Press.

Redding, S. and Venables, A. J., 2004, “Economic Geography and International Inequality”, *Journal of International Economics*, **62**(1): 53-82.

Tomiura, E., 2003, “Changing Economic Geography and Vertical Linkages in Japan”, *Journal of the Japanese and International Economies*, **17**(4): 561-581.

Table 1. New Entry of Japanese Plants

Goods	ALL			Finished			Intermediate		
Period	ALL	89-95	96-03	ALL	89-95	96-03	ALL	89-95	96-03
Japan									
Hokkaido	34	23	11	6	4	2	14	8	6
Tohoku	325	203	122	94	62	32	140	78	62
North Kanto	425	229	196	143	83	60	177	82	95
South Kanto	390	159	231	147	63	84	121	41	80
Tokai	513	292	221	88	53	35	259	142	117
Kinki	246	129	117	69	41	28	61	30	31
Chugoku	191	126	65	37	27	10	96	65	31
Kyuusyu	187	107	80	32	18	14	80	44	36
ASEAN									
Malaysia	173	141	32	49	38	11	82	67	15
Philippines	113	69	44	33	25	8	57	28	29
Thailand	242	128	114	49	29	20	149	66	83
Indonesia	173	100	73	41	28	13	83	38	45
NIES									
Singapore	101	77	24	43	33	10	32	23	9
Taiwan	91	52	39	30	16	14	35	20	15
Korea	66	36	30	6	3	3	38	16	22
Hong Kong	111	79	32	45	30	15	37	27	10
China									
North China	149	101	48	45	31	14	65	41	24
East China	308	165	143	98	56	42	141	72	69
South China	152	89	63	56	40	16	65	26	39
West China	54	39	15	23	19	4	25	15	10
Developed Countries									
USA	494	296	198	139	86	53	218	110	108
Canada	23	12	11	3	3	0	14	3	11
Mexico	51	33	18	18	11	7	22	14	8
UK	117	83	34	35	23	12	43	27	16
France	49	31	18	24	17	7	12	3	9
Germany	57	43	14	19	13	6	13	10	3
Italy	16	12	4	5	4	1	6	3	3
Netherland	31	16	15	14	5	9	5	3	2
Spain	21	18	3	8	8	0	7	4	3
Australia	28	21	7	17	12	5	3	1	2

Table 2. Basic Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
ALL					
Market Potential	147,930	14.00	0.73	12.33	15.53
ISA	147,930	-0.16	0.76	-6.58	1.83
Wage	147,930	2.69	1.96	0.06	6.51
Country Risk	147,930	0.75	0.17	0.22	0.95
Finished					
Market Potential	42,480	14.00	0.73	12.33	15.53
ISA	42,480	-0.23	0.95	-6.58	1.83
Wage	42,480	2.69	1.95	0.06	6.51
Country Risk	42,480	0.75	0.18	0.22	0.95

Table 3. Conditional Logit Results

Goods Criterion	ALL	Finished Sector	Finished Sector	Finished Adv. > Med.	Finished Adv. > Med.
MP	0.309***	0.414***	0.415***	0.382***	0.382***
	[0.034]	[0.065]	[0.065]	[0.048]	[0.048]
ISA	0.463***	0.338***	0.333***	0.331***	0.335***
	[0.025]	[0.041]	[0.041]	[0.033]	[0.033]
Wage	-0.372***	-0.439***	-0.442***	-0.422***	-0.421***
	[0.029]	[0.053]	[0.053]	[0.038]	[0.038]
Crisk	0.706***	0.613***	0.672***	0.864***	0.874***
	[0.115]	[0.217]	[0.222]	[0.159]	[0.160]
Asia	-0.902***	-0.882***	-1.922***	-0.409***	-1.706***
	[0.065]	[0.120]	[0.142]	[0.087]	[0.116]
Asia*Large			2.158***		2.044***
			[0.137]		[0.110]
Developed	-1.002***	-0.757***	-2.553***	-0.322***	-1.906***
	[0.043]	[0.079]	[0.178]	[0.058]	[0.127]
Developed*Large			3.074***		2.379***
			[0.204]		[0.143]
Observation	147930	42480	42480	73950	73950
Log likelihood	-15957	-4613	-4387	-8186	-7918
LR chi2	1629	406	858	396	932
Pseudo R2	0.0486	0.0421	0.0891	0.0236	0.0556

Notes: Standard errors are in parentheses. ***, **, and * show 1%, 5%, and 10% significant, respectively.

Table 4. Nested Logit Results

Goods Criterion	ALL	Finished Sector	Finished Adv. > Med.
Location choice among all regions			
MP	0.322***	0.399***	0.382***
	[0.035]	[0.067]	[0.051]
ISA	0.496***	0.368***	0.362***
	[0.025]	[0.040]	[0.033]
Wage	-0.433***	-0.396***	-0.407***
	[0.023]	[0.042]	[0.030]
Crisk	1.049***	0.759***	1.138***
	[0.121]	[0.217]	[0.161]
Location choice between Japan and Overseas			
Japan * Large	-1.905***	-2.128***	-2.026***
	[0.074]	[0.145]	[0.108]
Japan * R&D Share	-3.715***	-2.283	-2.513*
	[1.132]	[1.863]	[1.422]
Japan * Trade Status	-1.001***	-1.041***	-0.492***
	[0.075]	[0.143]	[0.109]
Observation	147,930	42,480	73,950
Log likelihood	-15179	-4366	-7888
LR chi2	3185	900	991
LR test (Prob > chi2)	0.0000	0.0000	0.0000
Pseudo R2	0.0949	0.0934	0.0591

Notes: Standard errors are in parentheses. ***, **, and * show 1%, 5%, and 10% significant, respectively.

Figure 1. Decomposition Analysis