

Market Potential and the Location of Japanese Investment in the European Union *

Keith Head[†] Thierry Mayer[‡]

November 19, 2003

Abstract

We investigate the hypothesis that firms prefer to locate “where the markets are.” We use a theoretical model of location choice under imperfect competition to formalize this concept. The model yields a profit function incorporating a term closely connected to the market potential index introduced by Harris in 1954. The location decision is a function of demand in all locations weighted by accessibility to consumers. We also show how the spatial distribution of competitors should be factored into the location choice. We then implement the model empirically, comparing our theoretically-derived measure of market potential with Harris’ term and with ad hoc agglomeration variables. Our sample consists of firm-level location choices by Japanese firms between 1984 and 1995 and we use both the information on the choice of country and the choice of region inside each country in our analysis. Our results show that market potential does matter for location choice but that traditional agglomeration variables retain an important role in the location decision.

JEL classification: F12, F15

Keywords: Location choice, borders, European Union, market potential.

*This paper is produced as part of a CEPR Research Network on The Economic Geography of Europe: Measurement, Testing and Policy Simulations, funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00069). We would like to thank participants at the CEPR Workshop *Economic Geography of Europe: Measurement, Testing and Policy Simulations*, and at the 7th *Empirical Investigations in International Trade* Annual Conference for helpful comments. We also thank participants at seminars in Catholic University of Leuven (KUL), University of Paris XII and XIII, and University of Geneva. Richard Baldwin, Bruce Blonigen, Pierre-Philippe Combes, Matthieu Crozet, Gianmarco Ottaviano, Jean Marc Siroën, and two referees provided helpful comments on earlier versions for which we are particularly grateful.

[†]Corresponding author: Faculty of Commerce, University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T1Z2, Canada. Tel: (604)822-8492, Fax: (604)822-8477, Email:keith.head@ubc.ca

[‡]Université de Paris I Panthéon Sorbonne (TEAM), CEPII, CERAS-ENPC, and CEPR. Email:tmayer@univ-paris1.fr

“We want to build our plants where the markets are.”

In December 1997, Hiroshi Okuda, chairman of Toyota, used the quote above to justify Toyota’s decision to build a factory in northern France. At that time, analysts in the press largely attributed that decision to the low market share of the Japanese car manufacturer in France (1%) and in Europe in general (3%).¹ The Toyota example suggests that even in the zero-tariff internal market of Europe, firms still seek locations with superior market access. The managers of Toyota apparently thought that their existing production sites in Great Britain were not “close” enough to the French market. The *Wall Street Journal* reported that “Toyota ... hopes to capture 3% [of the French market] after opening its factory here in 2001.”

This paper connects two previously disparate strands of the economic geography literature. The first strand demonstrates a statistical tendency of firms to make the same location decision as previous firms with similar attributes (such as industry and national origin).² While such “agglomeration effects” appear regularly in empirical work, they are consistent with a variety of explanations. The second strand comprises a large number of theoretical papers that focus on a particular mechanism of agglomeration: Namely, that producers concentrate where demand is highest and serve smaller markets via exporting.³

We link the two strands by showing how to derive the firm’s location choice probabilities as a function of production costs and a demand variable closely linked to the measure of “market potential” introduced by Harris (1954). We then take the model to the data, investigating whether location choices of Japanese affiliates in Europe are driven by market-access motivations *à la* Krugman or some other form of agglomeration effect. We find the demand-pull mechanism has some explanatory power but it does not appear to explain away the entire

¹Douglas Lavin in the *Interactive Edition*, December 10, 1997 and *AP-Dow Jones News Service*, December 9, 1997.

²Devereux and Griffith (1998) and Head et al. (1999) are recent examples from this literature.

³Krugman (1980 and 1991) wrote the seminal papers in this literature; the monograph of Fujita et al. (1999) thoroughly analyzes the basic model and its extensions.

empirical agglomeration effect.

The impact of market potential on the location of producers and the wages they pay has been the subject of several recent papers. Davis and Weinstein (1999, 2003) find that production increases on a more than one-for-one basis with demand (the “home market effect”) in many industries. Like their 1999 paper, we focus on how demand affects the intra-national location of production. As with their 2003 paper, we construct a market potential measure that aggregates demand from multiple locations while discounting for distance using a parameter obtained from a first-step estimation using bilateral trade flows. In addition to distance effects, our measure incorporates the impact of borders on trade as well as a theoretically derived adjustment for competition.

Hanson (2001) estimates the relationship between county-level wages in the United States and market potential. Structural estimation of this equation reveals that wages in a county are increasing in demand emanating from all American counties with weights declining exponentially with bilateral distance. Redding and Venables (forthcoming) follow a similar line of reasoning for international data using a bilateral trade equation of a theoretical model to obtain estimates of bilateral trade costs and of each country’s market and supply accessibility. They find that international inequality is closely linked to differences in market access. Crozet (2000) uses the same theoretical framework and shows that migration flows also respond to market potential.

The literature on firm location choice has not previously estimated models directly derived from the Krugman model. Prior work has, of course, considered demand but typically only local demand.⁴ Knowing the size of demand in each of the districts a firm might choose is not sufficient since firms can export to nearby locations. Some studies such as Friedman et al. (1992), Henderson et al. (1995), and Head et al. (1999) consider non-local demand but not

⁴See for instance, Coughlin et al. (1991).

using measures derived from theory.⁵ In particular, theory suggests that non-local demand must be discounted based on bilateral trade impediments. Furthermore, a given amount of market access contributes less to profits when a firm's competitors have access to the same markets. We follow Krugman (1992) in adjusting the market potential measure to take into account the location of competitors.

The paper proceeds as follows. In section 1 we derive a linear-in-logs equation that relates the profitability of a location to a prospective foreign investor to a measure of that location's access to demand. We then show how to estimate the distance and border effects that impede market access using bilateral trade data. In section 2 we report the results from the trade equation and show how we use them to calculate market potential. We then discuss our sample of Japanese investors and the set of possible location choices. Our location-choice results are detailed in section 3 and we conclude and propose directions for future work in section 4.

1 The Model

Let E_r denote expenditure in a representative industry (we omit industry subscripts for notational simplicity) in region r . Consumers (who may be firms or individuals) allocate their expenditures across differentiated varieties in the representative industry. They have constant elasticity of substitution sub-utility functions for each industry. Maximizing this sub-utility function subject to expenditure, E_r , and the delivered prices from all R possible product origins, we obtain the demand curve for the representative variety in the representative industry as

$$q_{ij} = \frac{p_{ij}^{-\sigma}}{\sum_{r=1}^R n_r p_{rj}^{1-\sigma}} E_j, \quad (1)$$

⁵Friedman et al. (1992) use a distance weighted sum of per-capita incomes, Henderson et al. (1995) use distance to nearest major business district, and Head et al. (1999) sum the personal incomes of adjacent states.

where p_{ij} is the delivered price faced by consumers in region j (destination) for products from region i (origin). It is the product of the mill price p_j and the trade cost factor, τ_{ij} . Trade costs include all transaction costs associated with moving goods across space and national borders.

1.1 The Profit Equation for Foreign Affiliates

Each firm sets its mill prices to maximize profits. Following Dixit and Stiglitz (1977), firms treat the elasticity of substitution, σ , as if it were the price elasticity of demand (this may also be interpreted as the assumption that each firm has infinitesimal market share). The resulting mill prices are simple mark-ups over marginal costs, denoted c_r : $p_r = c_r(\sigma/(\sigma - 1))$. Substituting into (1), we obtain the quantity that a firm producing in region i would deliver⁶ to each destination region j :

$$q_{ij} = \frac{(\sigma - 1)}{\sigma} \frac{(c_i \tau_{ij})^{-\sigma}}{G_j} E_j, \quad (2)$$

where $G_j \equiv \sum_r n_r (c_r \tau_{rj})^{1-\sigma}$. The gross profit earned in each destination region j for a firm producing in region i is

$$\pi_{ij} = (p_i - c_i) \tau_{ij} q_{ij} = \frac{(c_i \tau_{ij})^{1-\sigma}}{\sigma G_j} E_j \quad (3)$$

This gross profit is an increasing function of the expenditure of country j on the considered industry. The fraction multiplying E_i depends on the costs of the representative firm relative to its competitors from all R regions. In the numerator, we see that profits are decreasing in local (region i) production costs. Lower trade costs to reach region j , i.e. a low τ_{ij} , also raise profits. Since the effect of trade costs is always moderated by the elasticity of substitution, we introduce the notation of $\phi_{ij} = \tau_{ij}^{1-\sigma}$ to measure exporters' from i 's access to market j .⁷ The

⁶The "iceberg" convention implies that to deliver q units, the firm must ship τq units.

⁷Baldwin et al. (2003) refer to ϕ as "freeness of trade" and mainly consider cases of symmetric trade costs: ($\phi_{ij} = \phi_{ji}$). The term market access is better suited to our case where we find asymmetries in trade costs.

denominator contains the corresponding characteristics of competing suppliers. Note that the denominator term is multiplied by σ , capturing the idea that competition is fiercer and profits are therefore lower when varieties are less differentiated from each other.

Summing the gross profits earned in each market and subtracting the fixed costs F_r necessary to establish a plant in region r , we obtain the aggregate net profit, Π_r to be earned in each potential location r :

$$\Pi_r = \frac{c_r^{1-\sigma}}{\sigma} \sum_{j=1}^R \phi_{rj} \frac{E_j}{G_j} - F_r = \frac{c_r^{1-\sigma}}{\sigma} M_r - F_r, \quad (4)$$

where

$$M_r \equiv \sum_j \frac{\phi_{rj} E_j}{G_j}.$$

We will refer to M_r as the “Krugman market potential” since an expression of it first appears in Krugman (1992). The profit equation suggests that firms face a trade-off between low production costs and high market potential.⁸

When a firm chooses its location, the only relevant information is the ordering of the profits. Invariant fixed costs do not affect the profit ordering of regions and can therefore be omitted. For tractability, therefore, we assume that fixed costs do not differ across locations ($F_r = F, \forall r$). As monotonic transformations also leave the ordering unchanged, we will make four of them to create a simple and intuitive expression for profitability. Namely, we add F , multiply by σ , raise to the power of $1/(\sigma - 1)$, and take natural logs. Denoting the result as

⁸The wage equation analyzed by Fujita et al. (1999, p. 53) can be derived from the profit equation by assuming that free entry sets equation (4) equal to zero. In their specification, production requires one unit of labor per unit of output and F units of labor as overhead. Denoting the cost of a labor unit as w_r , this implies marginal costs of $c_r = w_r$ and fixed costs of $F_r = Fw_r$. Solving for w_r , we have

$$w_r = \left[\frac{1}{\sigma F} M_r \right]^{1/\sigma}.$$

That is, for firms to be indifferent between locations, wages must be a power function of market potential. Hanson (2001) and Redding and Venables (forthcoming) estimate variations of this relationship.

U_r we obtain

$$U_r \equiv \frac{\ln \sigma + \ln(\Pi_r + F)}{\sigma - 1} = -\ln c_r + (\sigma - 1)^{-1} \ln M_r. \quad (5)$$

Equation (5) expresses the profitability for a firm of locating in region r as a very simple function that is decreasing in production costs and increasing in the Krugman market potential term.

Considering the cost term first, let us suppose that the variable cost function is assumed to be Cobb-Douglas with constant returns, using labor at cost w_r and “other inputs” (such as land and intermediates) at cost v_r . Labor’s share is α and A_r represents total factor productivity. Thus, log marginal costs are given by

$$\ln c_r = \alpha \ln w_r + (1 - \alpha) \ln v_r - \ln A_r. \quad (6)$$

Substituting (6) into (5) and rearranging, we have

$$U_r = -\alpha \ln w_r + (\sigma - 1)^{-1} \ln M_r - (1 - \alpha) \ln v_r + \ln A_r. \quad (7)$$

We observe wages, w_r , and will calculate M_r using a method described in subsection 1.2. We do not observe v_r and A_r and they will be captured with several proxies (specified in subsection 2.3) and a random term observed by firms but not by the econometrician (detailed in subsection 2.2).

The Krugman market potential aggregates the expenditures of all regions while adjusting for region r ’s access ϕ_{rj} , and for competition from firms located in other regions, G_j . Analyzing the function, we can determine the assumptions that were implicit in the original Harris (1954) formulation of market potential. Specifically, if we set $G_j = 1$ and $\phi_{rj} = 1/d_{rj}$, M_r reduces to $\sum_j E_j/d_{rj}$, i.e. the inverse-distance weighted sum of incomes. In the Krugman market

potential function, M_r , the denominator, G_j takes into the account the competition that firms from region r face from rival firms exporting from other regions to serve the demand in each export market j . This competition adjustment increases with the number of rivals and decreases with their trade and production costs.

The competition adjustment can help explain why otherwise identical firms would not all select the same location. As more firms choose one region, the market there becomes more “crowded,” lowering M_r , until another region is more profitable. This is not the only mechanism causing dispersion, however. Firms are not identical and will therefore differ in their views of the prospective profitability of each region. This heterogeneity is analogous to “matching” in labor markets. In the context of our model, this heterogeneity can be thought of as firm-specific variation in regional productivity ($\ln A_r$).

The Krugman market potential has the advantage of being derived rigorously from theory. However, unlike the Harris form, its calculation requires estimates of the unknown parameters ϕ_{ij} and G_i . Our strategy will be to use information from international trade flows to estimate these parameters.

1.2 The Trade Equation

We do not observe trade flows between regions and must instead rely upon trade between nations to estimate the parameters that determine trade costs. Results from Wolf’s (2000) study of intra-national trade in the United States are reassuring in this respect since he finds distance effects resembling those found using international trade data. We reinterpret equation (2) as giving the quantity exported by a representative firm in *country I* to *country J* (we reserve the lower case i and j to denote origin and destination *regions*). The aggregate value of country I ’s exports to country J , denoted X_{IJ} , is given by the quantity exported by a

representative variety firm from I multiplied by the price and the number of varieties from J :

$$X_{IJ} = p_{IJ}q_{IJ}n_I = n_I \frac{c_I^{1-\sigma} \phi_{IJ} E_J}{G_J}$$

Taking natural logs and grouping variables based on subscripts, we obtain

$$\ln X_{IJ} = \ln(n_I/c_I^{\sigma-1}) + \ln(E_J/G_J) + \ln \phi_{IJ}. \quad (8)$$

Following Redding and Venables (forthcoming), we will estimate the first two terms using exporter and importer fixed effects, denoted EX_I and IM_J .⁹ Bilateral market access, (ϕ_{IJ}) will be estimated as a function of distance (d_{IJ}), borders ($B_{IJ} = 1$ for $I \neq J$), the sharing a common language ($L_{IJ} = 1$ if I and J share a language and $I \neq J$ and 0 otherwise) and an error term, ϵ_{IJ} . Parameters capturing the effect of distance, borders and language on trade volumes are denoted δ , β_J and λ respectively: $\phi_{IJ} = d_{IJ}^{-\delta} \exp(-(\beta_J - \lambda L_{IJ})B_{IJ} + \epsilon_{IJ})$. We interpret border effects as comprising home bias in consumer preferences and government procurement, differential technical standards, exchange rate uncertainty, and imperfect information about potential trade partners.¹⁰ The specification allows border effects to differ across importing countries, which is largely supported by the empirical evidence (see Chen, forthcoming). The estimated equation will therefore be

$$\ln X_{IJ} = EX_I + IM_J - \delta \ln d_{IJ} - \beta_J B_{IJ} + \lambda L_{IJ} \cdot B_{IJ} + \epsilon_{IJ}. \quad (9)$$

⁹An earlier version of this paper, Head and Mayer (2002, available at cepr.org), used the number of Japanese firms to measure n and calculated G_j at the regional level. It constructed the $c_i^{1-\sigma}$ term using wage data and estimates of σ obtained from a first-step trade equation. Our current approach, suggested by a referee, imposes fewer and more plausible assumptions but arrives at the same empirical conclusions.

¹⁰See Anderson and Van Wincoop (2003b) for a survey of the now large literature measuring and explaining border effects.

The estimated parameters on trade costs and importers' fixed effects are then used to construct the market potential variable that will be included in the location choice analysis of Japanese firms in Europe. Recall that the market potential of region i is $M_i \equiv \sum_j \phi_{ij}(E_j/G_j)$, where ϕ_{ij} is the accessibility of market j to goods shipped from region i . The formulas for calculating inter- and intra- regional access are

$$\begin{aligned}\hat{\phi}_{ij} &= \exp(-\hat{\beta}_J + \hat{\lambda}_{L_{IJ}})d_{ij}^{-\hat{\delta}} && \text{when } i \in I, j \in J \text{ and } I \neq J \\ \hat{\phi}_{ij} &= d_{ij}^{-\hat{\delta}} && \text{when } i \text{ and } j \text{ belong to the same country,} \\ \hat{\phi}_{ii} &= d_{ii}^{-\hat{\delta}} = (2/3\sqrt{\text{area}_i/\pi})^{-\hat{\delta}} && \text{for intra-regional trade.}\end{aligned}$$

The last equation models the average distance between a producer and a consumer based on a stylized geography where all producers are centrally located and the consumers uniformly distributed across a disk-shaped region.

The second component of market potential calculation is regional-level competition-weighted expenditure. Taking the exponential of importer's fixed effects' coefficients, we obtain $E_J/G_J = \exp(\text{IM}_J)$, that is the competition-weighted expenditure of country J . We calculate E_j/G_j for each region j of country J by allocating E_J/G_J to the different regions in proportion to their share of national GDP (y_j/y_J), i.e. $E_j/G_j = (y_j/y_J) \exp(\text{IM}_J)$.

We also compare the Krugman market potential with the simpler version proposed by Harris (1954). This variable sums distance-weighted industry-level expenditures, i.e. $\sum_j E_j/d_{rj}$. Again, we allocate national expenditure to regions according to GDP shares of regions ($E_j = (y_j/y_J)E_J$). National expenditure is calculated using apparent consumption, i.e., $E_J = \text{production} - \text{exports} + \text{imports}$ in the considered industry. It comprises final and intermediate demand from all sectors.

2 Econometric model and data

We estimate a model of location choice of 452 Japanese-owned affiliates that were established in 57 regions belonging to 9 European countries (Belgium, France, Germany, Ireland, Italy, the Netherlands, Spain, Portugal and the United Kingdom) during the period 1984–1995. As can be seen in equation (5), we hypothesize that market potential is a key determinant of this decision. We construct market potential for 18 industries using the parameters estimated from the trade equation (9).

2.1 Estimation of the Trade Equation

We estimate (9) using Eurostat data on bilateral trade matched with production at the NACE70 2-digit level. Production data is needed because internal trade flows, X_{II} , are constructed by subtracting total exports ($\sum_{J \neq I} X_{IJ}$) from national production. Our sample runs from 1980 to 1995 and includes the fifteen 1995 members of the European Union, plus Switzerland and Norway. Although the set of EU countries in the location choice analysis only includes 9 members, it is important to incorporate the demand emanating from the rest of the European Economic Area in the calculation of the market potential. For instance, Switzerland and Austrian consumers are not trivial components of Southern German and Northern Italian regions' market potential that we do consider in the location choice of Japanese firms. Incorporating the demand from those countries involves obtaining first estimates of their fixed effect as an importer and bilateral trade costs parameters, and we therefore incorporate them in the trade equation.

Both internal (d_{II}) and external (d_{IJ}) distances are weighted averages of point-to-point distances between sub-national regions. Head and Mayer (2000) provide greater detail on its construction and the distance matrix for EU12 countries (coordinates and population shares

of regions for Switzerland, Austria, Norway, Sweden and Finland have been collected using a combination of Eurostat and national sources). The common language variable, L_{IJ} takes a value of one for the U.K. and Ireland, Germany and Austria, Germany and Switzerland, France and Switzerland, Belgium and France, and Belgium and the Netherlands combinations. The regressions are run for each of the 16 years and 18 industries. Each component of the market potential gathered through the trade equation (the fixed effects of importing countries and bilateral trade costs estimates) are thus industry, year and country-specific. More details on the data and implementation of those regressions can be found in the data appendix.

Table 1 summarizes the estimated border, distance, and language effects estimated for each 2-digit industry. This table gives the average coefficients over two subperiods of our sample, 1980–1987 and 1988–1995, chosen because of the start of the Single Market Programme in 1987, which was expected to yield a fall in border effects. The border effects presented average over the four large sources of demand inside our sample, Germany, France, the U.K. and Italy.

We find distance effects that average -1.4 across the two periods. This number aligns closely with the results of Redding and Venables (forthcoming) while being slightly superior to usual estimates of gravity equations¹¹ and suggests Harris’s assumption of an inverse distance rule is a rough but reasonable approximation. Where Harris’ specification appears inappropriate for Europe is in its omission of the impact of national borders.¹²

Border effects for the four core EU countries average 1.84 in the first period and 1.63 in the later period. Expressing their magnitude in the McCallum (1995) manner, within-border trade after 1987 remains more than five ($\exp(1.63) = 5.1$) times as large as cross-border trade after controlling for the effect of relative distance and characteristics of the trading partners

¹¹Anderson and van Wincoop (2003a) also use this fixed effects specification in the sensitivity analysis reported in their Table 6, p.187. Interestingly, they mention a substantial rise in the (absolute value of the) distance decay parameter which becomes -1.25 for US-Canada trade flows.

¹²His pioneering study considered the market potential of counties *within* the United States.

Table 1: Estimates of Border and Distance Effects. Averages for the large EU Nations (France, Germany, Italy, and the U.K.)

| Industry Name | NACE Code | 1980–1987 | | | 1988–1995 | | |
|-------------------------------|-----------|-----------------------------|------------------------------|------------------------------|-----------------------------|------------------------------|------------------------------|
| | | Border ($\hat{\beta}$) | Dist. ($-\hat{\delta}$) | Lang. ($\hat{\lambda}$) | Border ($\hat{\beta}$) | Dist. ($-\hat{\delta}$) | Lang. ($\hat{\lambda}$) |
| Metal-Primary | 22 | 1.3 | -1.5 | .04 | .80 | -1.55 | .17 |
| Non-metallic Mineral Products | 24 | 2.29 | -1.67 | .24 | 2.11 | -1.68 | .22 |
| Chemicals & Fibres | 25/26 | 1.81 | -1.19 | .06 | 1.62 | -1.22 | .06 |
| Metal-Fabricated | 31 | 2.72 | -1.54 | .39 | 2.61 | -1.47 | .48 |
| Machinery | 32 | 1.56 | -1.08 | .26 | 1.42 | -1.06 | .32 |
| Office Machines | 33 | .75 | -.83 | .15 | .85 | -.79 | .15 |
| Electronics | 34 | 2.09 | -.99 | .32 | 1.76 | -1.01 | .33 |
| Motor Vehicles and Parts | 35 | .95 | -1.68 | .44 | .89 | -1.52 | .37 |
| Cycles | 363 | .96 | -2.12 | .27 | .66 | -1.87 | .53 |
| Precision Instruments | 37 | 1.51 | -1.00 | .34 | 1.05 | -.98 | .15 |
| Food, Drink, & Tobacco | 41/42 | 2.94 | -1.29 | .59 | 2.77 | -1.38 | .48 |
| Textiles | 43 | 2.72 | -1.28 | .46 | 2.45 | -1.26 | .59 |
| Leather | 44 | 1.76 | -1.43 | .62 | 1.23 | -1.49 | .62 |
| Clothing and Footwear | 45 | 1.95 | -1.51 | .92 | 1.87 | -1.48 | .91 |
| Wood and Wooden Furniture | 46 | 2.56 | -1.91 | .74 | 2.4 | -1.96 | .68 |
| Paper, Printing, & Publishing | 47 | 2.66 | -1.55 | .6 | 2.54 | -1.46 | .56 |
| Rubber and Plastics | 48 | 1.91 | -1.35 | .20 | 1.85 | -1.36 | .23 |
| Toys & Sports | 494 | .72 | -1.23 | .52 | .52 | -1.23 | .74 |

(in particular their economic size). While sizeable, these effects are considerably smaller than the value of 20 first reported by McCallum for the Canada-US border, but more comparable to the re-scaled estimates obtained by Anderson and van Wincoop (2003a) when using a specification more closely linked to theory. In contrast with distance effects, there appears to be a downward trend in the impact of national borders in the EU; all but one border effects declined. We also find large common language effects. The trade-creating effect of common language is slightly increasing over time, from an average of .39 in the first period to .42 in the second one. Interestingly, we observe large variation across industries in the effect of sharing a language. Goods bought mainly as intermediate inputs tend to have smaller language effects than goods destined for final consumption. For instance, country pairs sharing a language trade 2.5 times more clothing and footwear than pairs lacking a common language.

2.2 Specification of the Location Choice Model

We estimate the parameters of the profit equation (7) using a discrete choice model. As we do not observe the potential profitability of each location, we rely upon the assumption that firms choose the country yielding the highest profit. The location choice literature makes extensive use of the conditional logit model (CLM). This model requires error terms that are independent across locations. As it seems likely that the unobserved component of profitability is correlated among regions in the same nation, we use a generalization that permits such a structure of the random term, the nested logit model (NLM).¹³

For estimation purposes it is useful to decompose equation (7) into components that are observable at the nation-state level, denoted V_s , at the region level, denoted W_r , as well as

¹³Train (2003) provides a clear description of the nested logit methodology. Mayer and Mucchielli (2002) use the same sample used here to confirm the validity of a structure nesting region choice under nation choice.

remaining random variation that the econometrician does not observe, denoted ξ_r .

$$U_r = V_s + W_r + \xi_r.$$

V_s includes national policies—corporate tax rates and social charges in this study— that affect all regions. W_r includes wages and market potential as well as proxies for other input prices and productivity (detailed in subsection 2.3). We envision the random term, ξ_r , as a shock to $\ln A_r$ that is specific to firm-region pairs (we continue to suppress firm subscripts for readability). It also contains any other influences on the attractiveness of a location that matter to firms but are not included as controls by the econometrician. McFadden (1978) shows that if the distribution of ξ_r is given by a multivariate extreme value with parameter ρ , then the conditional probability that firms choose region r conditional on choosing state s is $P_{r|s} = \exp(\rho^{-1}W_r - Z_s)$, where $Z_s \equiv \ln \sum_{i \in s} \exp(\rho^{-1}W_i)$ is termed the “inclusive value” for state s . The probability of choosing nation s is $P_s = \exp(V_s + \rho Z_s - \tilde{Z})$, where $\tilde{Z} \equiv \ln[\sum_I \exp(V_I + \rho Z_I)]$. Parameter ρ measures the degree of independence between the unobserved portion of profitability of regions in a given state. For $\rho = 0$, regions are perfect substitutes inside a nation whereas for $\rho = 1$ there is full independence and patterns of substitution are the same within and between nations. We also consider this non-nested version of the model corresponding to $\rho = 1$. In that case the probability of choosing a region r is $P_r = P_{r|s} \cdot P_s = \exp(V_s + W_r - \tilde{Z})$. Consequently, the NLM collapses to the CLM.

The parameters on the components of U_r will be estimated by substituting the probabilities (P_r) of the actual location choices made by Japanese firms in Europe into a log-likelihood function and maximizing.

2.3 Implementation of the location choice model

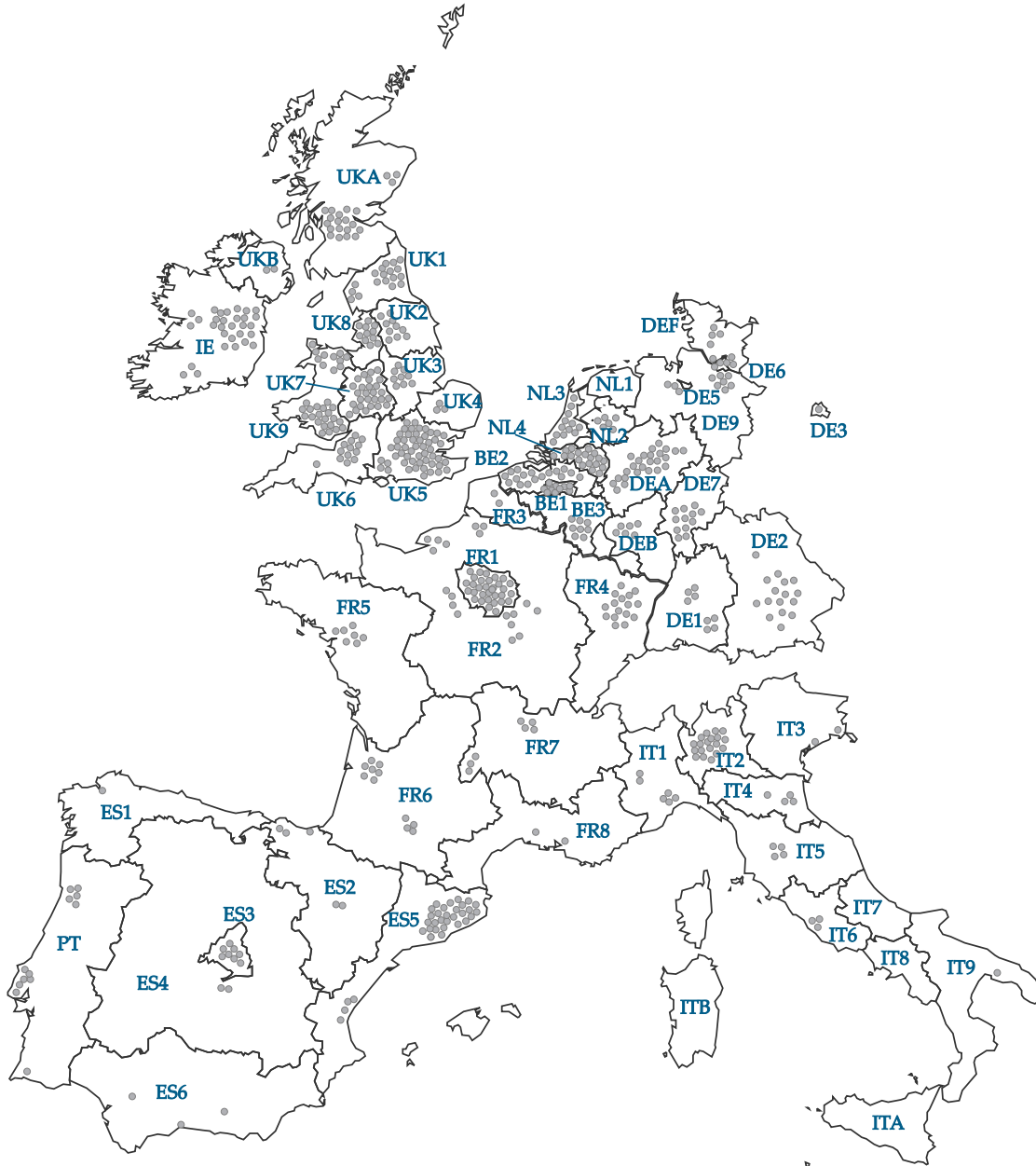
The sample of Japanese firms is extracted from the 1996 *Survey of Current Manufacturing Operations of Japanese Firms in Europe* issued by the JETRO. More than 700 manufacturing Japanese investments are listed in this survey with corresponding date when operation started, country of location, employment and other details including a detailed description of the product manufactured. In order to assign investments to sub-national regions, the *Directory of Japanese-Affiliated Companies in the EU: 1996–1997* also issued by the JETRO was used to determine the precise city where the plant was located. Almost all explanatory variables come from industrial statistics issued by Eurostat either at the national or at the regional level. The selection of Japanese investments (452 retained location decisions over the years 1984-1995) was essentially driven by the availability of regional and national data. Further details concerning the data can be found in the data appendix which includes Table 5 showing the regions in our choice set and the number of Japanese firms each one received.

Figure 1 uses this data to plot the Japanese affiliates in the NUTS1 region where they invested.¹⁴ Several important features of Japanese investment patterns are immediately apparent: The strong attractiveness of the United Kingdom as a whole, the agglomeration in the Northern part of Europe, as well as a tendency of investors to locate in the economic core of each country (Japanese investors cluster in London in the United Kingdom, Paris in France, Milan in Italy, and Barcelona in Spain).

Although the theoretical model of the location decision focuses on wages and market potential, the large literature on location choice includes a number of other explanatory variables. We include the standard controls. For wages, we use the total wage bill divided by number of employees in the 2-digit industry-region. Wages do not provide a complete description of

¹⁴NUTS is the official classification for EU regions; it has five levels of geographical detail, ranging from 15 NUTS0 nations to more than 100,000 NUTS5 areas.

Figure 1: Japanese investors in Europe at the end of 1995



labor costs since the functioning of the labor market (measured by unemployment rates) and government-imposed charges also contribute to the true cost of workers. Government subsidies and taxes affect the cost of capital (part of v in the model), we control for those through corporate tax rates and eligibility to EU regional policy funding. Last, we follow Coughlin et al. (1991) and add land area of the region, intended to control for differences in land supply and therefore land prices which also enter v .

Much evidence suggests that related firms tend to cluster in the same regions. We consider three forms of relatedness:¹⁵ (1) establishments in the same industry; (2) affiliates in the same industry originating from the same country (Japan); (3) affiliates owned by the same parent company or affiliated in the same supplier groups (known as *keiretsu* in Japan). Clusters of related firms may form regional production networks, selling intermediate inputs to each other and thereby lowering v_r . They may also share knowledge, raising A_r . It is also likely that clusters will form around the same exogenous sources of low input costs or high productivity. Of particular interest to this paper is the hypothesis that clusters form in areas with high market potential in the relevant industry. This hypothesis would receive support if, after controlling for market potential, the presence of same industry firms *lowers* the attractiveness of a region. Regardless of the underlying mechanism, we will refer to the three cluster variables as agglomeration effects. All variables used in our specification of production costs are described in Table 2.

3 Location choice results

We begin with an assessment of the performance of our market potential variable in a conventional specification used in the literature. We therefore start with the non-nested conditional

¹⁵Head et al. (1999) find evidence that all three forms matter for Japanese location choices in the United States.

Table 2: Production costs specifications

| Geographic level | Variable | Definition |
|----------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Region-specific (W_r) | Wages | Total wage bill divided by number of employees in the 2-digit industry-region |
| | Unemployment Rate | Unemployed as percent of the labor force |
| | Obj 1 Eligibility | Equals one when the per capita income of the region is less than 75% of the EU average, the critical value for qualifying for European Union structural funds |
| | Area | Land area of region |
| | Domestic Industry Count | Number of establishments in the 2-digit industry-region |
| | Japan Industry Count | Number of Japanese affiliates in the 3-digit industry-region |
| | Network Count | Number of affiliates owned by same Japanese parent or members of same vertical <i>keiretsu</i> |
| Nation-specific (V_s) | Corporate Tax Rate | Statutory Rate of taxation for profits made by foreign investors |
| | Social Charges Rate | Non-wage labor costs, such as payroll taxes and pension contributions, divided by the total labor cost (wage bill + social charges). |

logit estimation of the region choices of Japanese firms in Europe (Table 3). We then turn to a nested logit specification in which we first estimate the choice of region within a given nation and then estimate the choice of nation taking into account the attractiveness of its constituent regions (Table 4).

3.1 Region Choice: Non-nested Logit

Table 3 provides results for six different conditional logit estimations of the location choice of Japanese affiliates among EU regions. Specification (1) shows coefficients for the standard set of production costs variables, excluding the difficult-to-interpret agglomeration effects. Specifications (2) to (4) add successively more sophisticated measures of demand, culminating with the Krugman market potential. Column (5) adds nation-level fixed effects to the analysis, which is a first way to capture unobserved correlations in the characteristics of regions belonging to the same nation. Column (6) adds the counts of various types of related firms.

While conventional wisdom would predict a negative effect for wages and a positive effect for unemployment, the results in specification (1) provide just the opposite. As we shall see, neither “perverse” result is robust to the inclusion of appropriate controls. The effect of wages is small and statistically insignificant after using any of the controls for demand. The absence of a significant negative impact of regional wages on location choice in all specifications is disappointing. However, the result is not out of line with other studies (e.g. Devereux and Griffith, 1998, and Head et al., 1999). Since the standard model of wage determination (the Mincer equation) explains differences in wages with differences in human capital (education, experience, and ability) that are presumably valuable to the firm, ambiguous results should perhaps be expected.¹⁶

¹⁶In unreported regressions we attempted to control for this effect of human capital differences by using the wage of the clothing industry as a measure of unskilled wages and the chemical industry as a measure of skilled wages. The idea (for which we thank a referee) was to purge the wage measure of cross-regional human capital

Table 3: Conditional Logit Model of Region Choice

| Specification | 452 Firms Choosing Between 57 Regions | | | | | |
|---------------------------------------------------------|---------------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|-----------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ln Wages | 0.47 ^c (0.25) | -0.20 (0.26) | -0.12 (0.28) | 0.17 (0.25) | 0.50 (0.34) | 0.13 (0.36) |
| Unemployment Rate | -8.90 ^a (1.69) | -4.50 ^a (1.70) | -1.57 (1.95) | -3.22 ^c (1.78) | -4.34 ^c (2.28) | -1.35 (2.43) |
| Obj 1 Eligibility | -0.25 (0.21) | -0.12 (0.22) | 0.25 (0.22) | 0.01 (0.22) | -0.22 (0.24) | -0.24 (0.25) |
| ln Regional Area | 0.31 ^a (0.05) | 0.05 (0.06) | 0.58 ^a (0.06) | 0.59 ^a (0.06) | 0.58 ^a (0.07) | 0.21 ^b (0.08) |
| Social Charges Rate | -2.26 ^a (0.38) | -2.28 ^a (0.38) | -2.25 ^a (0.38) | -1.56 ^a (0.38) | 0.24 (1.83) | -0.01 (1.86) |
| Corporate Tax Rate | -4.82 ^a (0.59) | -4.80 ^a (0.58) | -5.03 ^a (0.60) | -4.96 ^a (0.61) | -0.40 (2.36) | -0.34 (2.34) |
| ln Regional GDP (ln y_r) | | 0.80 ^a (0.08) | | | | |
| ln Harris Market Potential (ln $\sum_j E_j/d_{rj}$) | | | 1.88 ^a (0.21) | | | |
| ln Krugman Market Potential (ln M_r) | | | | 1.11 ^a (0.13) | 1.07 ^a (0.14) | 0.34 ^b (0.16) |
| ln 1 + Domestic Industry Count | | | | | | 0.52 ^a (0.08) |
| ln 1 + Japan Industry Count | | | | | | 0.86 ^a (0.11) |
| ln 1 + Network Count | | | | | | 1.24 ^a (0.22) |
| National fixed effects | No | No | No | No | Yes | Yes |
| Likelihood Ratio Index | 0.054 | 0.079 | 0.077 | 0.073 | 0.079 | 0.126 |

Note: Standard errors in parentheses with ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

Larger regions attract significantly more investors than small ones in all specifications except (2). The elasticity is significantly below the unitary value that Coughlin et al. (1991) refer to as the “dartboard” approach to location decision. We see these results, which appear in the nested estimation as well, as indirect support for the importance of land costs in location decisions (using land area rather than land values probably makes sense since the latter is likely to reflect unobserved qualities of the location). Regions eligible for Objective One subsidies have, by definition, low per capita incomes and these two effects seem to cancel each other out, yielding insignificant effects.

The social charge rate enters with a consistently negative sign until nation-specific fixed effects are added in column (5). The negative and mainly significant effects of social charges make sense since this variable incorporates variation in labor costs that is unrelated to human capital. Unfortunately, it does not vary across regions within nations and its time-series variation is inadequate to identify a clear result. The effect of corporate taxation appears to be large, in line with recent estimates surveyed in the meta-analysis on FDI and corporate taxes by Mooij and Ederveen (2001). The coefficient in column (1) means that a one point rise in the national corporate tax rate yields a fall of about five percent in the probability that regions in that country will be chosen. The comparable average semi-elasticity in the literature reported in Table 3.1 in Mooij and Ederveen (2001) is 4.8. As with social charges, the tax effect is not robust to the inclusion of country fixed effects. The evolution of corporate tax rates over the period did not seem to influence the location choices by Japanese investors.

The main results of interest lie in columns (2) to (5), where we introduce different demand variables and compare results of Krugman market potential with alternative proxies for demand. We start, in column (2) by adding the most widely used demand measure, regional

variation. Unfortunately, the two wages were highly correlated and neither singly nor jointly yielded negative and significant effects.

GDP. As expected, the coefficient is positive and highly significant. The measure is hardly an adequate proxy for demand since few firms would go to the trouble of setting up an overseas factory to serve a single region. Column (3) substitutes the simple Harris market potential calculation to regional GDP. The coefficient is again significantly positive and its magnitude more than doubles compared to local GDP, reflecting the attractiveness of “central” regions in Europe (the ones with a combination of high local demand and proximity to other important sources of demand). The overall fit of the model is however slightly reduced. The Harris measure does not take into account border effects, variation in distance costs, or market-crowding due to local competition. The Krugman market potential, which handles all three issues, is introduced in column (4). The coefficient has the expected sign and significance, but its magnitude is lower than the coefficient on Harris’s version, and the overall fit of the model is lower (as seen in the .004 decline of the likelihood ratio index). The coefficient is however robust to the inclusion of national fixed effects in column (5).

How substantial is the impact of market potential on location choice? There are two ways to evaluate the size of the effect. First the coefficient itself is closely tied to the probability elasticity, which is given by $\hat{b}(1 - P_r)$ where \hat{b} is the coefficient and P_r is the probability of choosing region r . On average, P_r is the inverse of the number of choices. Using column (5)’s estimate of \hat{b} , we find a probability elasticity of $1.07(1 - 1/57) = 1.05$. Thus, 10% increases in market potential increase the probability of attracting investment by about 10%. While useful, this way of expressing economic significance does not take into account the actual variation in the explanatory variable.

A second approach is to imagine a hypothetical region with the mean level of market potential. Denote this region’s initial probability of being chosen as \bar{P} . Then redistribute demand so as to raise the region’s market potential by one standard deviation. Denote the new probability as \bar{P}' . Suppose that the demand redistribution leaves the overall attractiveness

of European regions unchanged, that is the inclusive value, \tilde{Z} , is fixed. Then the rise in the probability of attracting investment will be given by

$$\bar{P}'/\bar{P} = \exp(\hat{b}[\ln(\text{mean}(M_r) + \text{stdev}(M_r)) - \ln(\text{mean}(M_r))]) = [1 + \text{cv}(M_r)]^{\hat{b}},$$

where $\text{cv}(M_r) = \text{stdev}(M_r)/\text{mean}(M_r)$ is the coefficient of variation of market potential. For electronics (NACE 34), the recipient of the most Japanese manufacturing, we have a coefficient of variation of M_r of 0.58 in the year 1995. Based on the $\hat{b} = 1.07$ from specification (5), this implies a one standard deviation shock would increase the attractiveness of the average region by 63%. The impact on other industries, which mainly had greater variance in M_r are sometimes considerably larger (For motor vehicles the corresponding increase is 159%).

The agglomeration variables introduced in specification (6) may account for alternative mechanisms of spatial concentration, such as human capital externalities or technological spillovers which have been the primary emphasis of past work on location choice using this type of variables. The current paper is the first to control for market potential using the exact functional form dictated by theory. Thus, if prior findings of agglomeration effects merely reflected the absence of adequate controls for variation in demand, the agglomeration terms would *not* have significant positive effects in our specification and could even enter negatively to the extent that firms wish to avoid over-crowded markets.

Although estimated agglomeration effects are often interpreted as evidence of spatial externalities, this is not the only possible interpretation. Even with the many controls and national fixed effects employed in specification (5), it is likely that omitted variables remain a problem. Fixed effects for each region are not feasible because they cannot be estimated for regions that received no investment. Counts of firms in the same industry will partly reflect omitted variables, which form part of the unobserved attractiveness of regions and therefore

correlate with high numbers of domestic and Japanese firms. Adding agglomeration variables therefore also has the advantage of mitigating the inevitable omitted variable bias in this type of estimation.

The data reject the hypothesis of zero agglomeration effects, despite the presence of the Krugman market potential term. Market potential retains a significant positive sign but the coefficient is divided by more than three and the overall fit of the estimation is very much improved by including the counts of related firms. Market potential remains an economically significant factor, however. A one standard deviation shock now increases attractiveness by 17% in electronics and 35% in motor vehicles. We interpret the result that counts of related firms have a strong impact even after controlling for market potential—which reoccurs in the nested logits—as indicating that spatial concentration arises in large part from mechanisms *other* than the demand linkages emphasized by Krugman.

Another interesting result is that the influence of previous location choices by similar firms is increasing in the degree of “relatedness” of those competitors. All three variables have a large and positive influence, but the domestic firms in each region have a notably weaker attractive power than other Japanese affiliates, which are themselves superseded by members of the same vertical keiretsu. This last variable certainly captures input-output linkages of the Venables (1996) type. Its high coefficient is a sign that this type of vertical linkages might offer more solid empirical explanatory power than the simple version of the Krugman (1991) model primarily based on final demand linkages.

3.2 Region choices nested within nation choices

The use of the country dummies in specifications (5) and (6) of Table 3 helps to mitigate the problem associated with non-independent errors across regions belonging to the same nation. However, the country fixed effects do not resolve problems associated with cross-industry and

inter-temporal differences in the attractiveness of nations. By considering the choice of region for a given choice of nation, we condition on all aspects of the nation that do not vary across its constituent regions from the perspective of a given investor. The drawback is that we must omit the national tax and social charges variables. However, we re-introduce them when estimating the upper level of the decision tree (nation choice).

The nested region choice forces us to remove Portugal and Ireland, since they lack sub-national regions in our dataset. This reduces the set of choosers from 452 to 421. The choices sets vary in size from 4 (Belgium and Netherlands) to 11 (German, Italy, and the U.K.). On average there are 9.22 choices per Japanese investor.

Table 4 re-estimates the same specifications as Table 3 in a nested structure. The results are broadly similar. As before, controlling for demand eliminates a spurious positive effect for wages. Controlling for both market potential and agglomeration, the wage and unemployment effects conform to conventional wisdom, albeit without statistical significance. Market potential has a statistically significant effect even after controlling for agglomeration effects; however, the inclusion of the latter improve the fit considerably. Both market potential variables have higher impacts on region choice within nations than they do on non-nested region choice.¹⁷

We again find that “theory doesn’t pay,” in the sense that the Harris market potential outperforms the Krugman market potential in both magnitude and fit. In contrast, Hanson (2001) finds that when he augments the Harris function to include relationships derived from the Krugman model, the fit improves. However, Hanson’s simulations show that income shocks in the Harris-based formulation of market potential have larger effects on the wages paid in

¹⁷Due to the smaller number of alternatives in the nested logit (9.22 on average versus 57 in the non-nested estimation), direct inspection of the coefficients is adequate to make the comparison. Adjusting coefficients from specification (4) shows $1.41(1 - 1/9.22) = 1.26 > 1.11(1 - 1/57) = 1.09$. The difference is larger for specification (6).

Table 4: Nested Logit Model of Region/Nation Choice

| Specification | 421 firms choosing <i>within</i> 7 nations | | | | |
|---------------------------------------------------------|---------------------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| | (1) | (2) | (3) | (4) | (6) |
| ln Wages | 1.21 ^b (0.48) | -0.53 (0.53) | 0.12 (0.50) | -0.10 (0.50) | -0.34 (0.52) |
| Unemployment Rate | -9.77 ^a (2.39) | -3.97 (2.52) | 2.15 (2.76) | -1.03 (2.63) | 0.53 (2.76) |
| Obj 1 Eligibility | -0.89 ^a (0.33) | -0.71 ^b (0.35) | -0.81 ^b (0.35) | -0.74 ^b (0.35) | -0.54 (0.36) |
| ln Regional Area | 0.29 ^a (0.06) | 0.07 (0.07) | 0.59 ^a (0.07) | 0.69 ^a (0.08) | 0.26 ^b (0.10) |
| ln Regional GDP (ln y_r) | | 0.82 ^a (0.09) | | | |
| ln Harris Market Potential (ln $\sum_j E_j/d_{rj}$) | | | 2.30 ^a (0.26) | | |
| ln Krugman Market Potential (ln M_r) | | | | 1.41 ^a (0.17) | 0.51 ^b (0.22) |
| ln 1 + Domestic Industry Count | | | | | 0.53 ^a (0.11) |
| ln 1 + Japan Industry Count | | | | | 0.65 ^a (0.13) |
| ln 1 + Network Count | | | | | 1.46 ^a (0.26) |
| Likelihood Ratio Index | 0.054 | 0.097 | 0.095 | 0.093 | 0.146 |
| Specification | 421 firms choosing <i>between</i> 7 nations | | | | |
| | (1) | (2) | (3) | (4) | (6) |
| Social Charges Rate | -2.66 ^a (0.43) | -2.54 ^a (0.43) | -2.68 ^a (0.43) | -2.11 ^a (0.44) | -2.27 ^a (0.45) |
| Corporate Tax Rate | -3.20 ^a (0.65) | -3.69 ^a (0.63) | -2.79 ^a (0.63) | -3.32 ^a (0.62) | -2.08 ^a (0.55) |
| Inclusive Value (Z_s) | 0.59 ^a (0.07) | 0.78 ^a (0.07) | 0.53 ^a (0.06) | 0.65 ^a (0.06) | 0.69 ^a (0.06) |
| Likelihood Ratio Index | 0.115 | 0.134 | 0.117 | 0.132 | 0.139 |

Note: Standard errors in parentheses with ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

nearby counties than in the Krugman-based formulation. The broad similarity in results for the two formulations suggests that we should be cautious in interpreting the positive effect of market potential as evidence for the particular mechanisms of Krugman-style models of economic geography. Other models of economic geography could generate observationally equivalent results.

The lower frame of Table 4 reports estimates for the choice of nation. These estimations directly consider the national variables (taxes and social charges) but indirectly consider all the regional determinants of attractiveness as they enter the inclusive value, Z_s , for each nation. We do not include country dummies because there would not be sufficient remaining variation in the data.

We find consistently negative impacts of social charges. Corporate income taxes have large negative impacts. As with the case of the non-nested specification, the social charges and tax results are fragile. Including a dummy for Ireland and the U.K. makes both effects become much smaller and insignificant. This result raises the question of whether low charges/taxes or some other effect (the use of English?) made these countries so attractive to Japanese investors.

The inclusive value, Z_s , an index of the maximum expected profitability from locating in a given country considering the underlying characteristics of its regions, obtains reasonable values in all specifications, differing quite significantly from the value of one that would obviate the need for nesting and the value of zero in which investors are indifferent between regions inside a given country. In other words, the coefficient on the inclusive value supports the validity of our country-region nesting structure.

4 Conclusion

We analyze the determinants of location choices by Japanese firms in Europe. Our work is particularly concerned with the appropriate manner to take into account the spatial distribution of demand in location decisions. We rigorously link the optimal location choice of Japanese investors to a theoretical model of imperfect competition in a multi-location setting. The underlying profit equation incorporates a term that is closely related to the market potential index originally introduced and used by geographers (Harris, 1954). Our theory-derived term aggregates the spatial distribution of demand weighted by trade costs and the location of potential competitors. We estimate the border and distance effects that determine market accessibility using a bilateral trade equation implied by the same model that generates the profit equation.

We find that demand does matter for location choice: A ten percent increase in our market potential term raises the chance of a region being chosen by three to eleven percent, depending on the specification. Despite the fact that we bring theory to empirical implementation in a structural way, the “correct” measure of market potential actually underperforms the atheoretical Harris (1954) measure. Moreover, non-structural agglomeration variables retain a robust influence. These results suggest that the downstream linkages emphasized in Krugman (1991) are not the only or even the main cause of agglomeration. Future research should probably consider other reasons why firms cluster. It does not seem possible to falsify the hypothesis that observed agglomeration effects merely reflect omitted exogenous location attributes. However, a natural follow-up to this paper would be to estimate structural location choice models that implement the Venables (1996) setup with upstream and downstream linkages based on an input-output matrix.

References

- ANDERSON, J. AND E. VAN WINCOOP, 2003a, "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review* 93(1):170–192.
- ANDERSON, J. AND E. VAN WINCOOP, 2003b, "Trade Costs", Boston College manuscript.
- BALDWIN, R., R. FORSLID, P. MARTIN, G. OTTAVIANO AND F. ROBERT-NICOUD, 2003, *Public Policies and Economic Geography* (Princeton University Press, Princeton).
- CHEN, N., forthcoming, "Intra-national versus international trade in the European Union: why do national borders matter?", *Journal of International Economics*.
- COUGHLIN, C.C., J.V. TERZA AND V. ARROMDEE, 1991, "State Characteristics and the Location of Foreign Direct Investment in the United States", *Review of Economics and Statistics* 68 : 675–683.
- CROZET, M., 2000, "Do migrants believe in market potential?" Cahier de la MSE Série Blanche Université de Paris I # 2000-30.
- DEVEREUX, M. AND R. GRIFFITH, 1998, "Taxes and the Location of Production : Evidence from a Panel of US Multinationals", *Journal of Public Economics* 68(3) : 335–367.
- DEVEREUX, M., R. GRIFFITH, AND A. KLEMM, 2002, "Corporate Income Tax Reforms and International Tax Competition", *Economic Policy*, 35 : 449–488.
- DAVIS, D. AND D. WEINSTEIN, 1999, "Economic Geography and Regional Production Structure: An Empirical Investigation", *European Economic Review* 43(2):379–407.
- DAVIS, D. AND D. WEINSTEIN, 2003, "Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment", *Journal of International Economics*, 59(1):1–23.
- DIXIT, A. AND J. STIGLITZ, 1977, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*. 67(3) : 297–308
- FRIEDMAN, J., D. GERLOWSKI AND J. SILBERMAN, 1992, "What Attracts Foreign Multinational Corporations", *Journal of Regional Science* 32(4): 403-418.
- FUJITA , M., P. KRUGMAN AND A. VENABLES, 1999, *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge : MIT Press.
- HANSON, G., 2001, "Market Potential, Increasing Returns, and Geographic Concentration" University of California, San Diego manuscript.
- 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 T. MAYER, 2000, "Non-Europe : The Magnitude and Causes of Market Fragmentation in Europe" *Weltwirtschaftliches Archiv* 136(2): 285–314.
- HEAD, K. AND T. MAYER (2002), "Market Potential and the Location of Japanese Investment in the European Union," Centre for Economic Policy Research Discussion Paper # 3455.

- HEAD, K., J. RIES AND D. SWENSON , 1999, “Attracting Foreign Manufacturing: Investment Promotion and Agglomeration”, *Regional Science and Urban Economics* 29(2) : 197–218.
- HENDERSON, J.V., A. KUNCORO AND M. TURNER, 1995, “Industrial Development in Cities”, *Journal of Political Economy* 103(5) : 1067–1090.
- KRUGMAN, P., 1980, “Scale Economies, Product Differentiation, and the Pattern of Trade”, *American Economic Review* 70:950–959.
- KRUGMAN, P.R., 1991, “Increasing Returns and Economic Geography”, *Journal of Political Economy* 99(3) : 483–499.
- KRUGMAN, P.R. , 1992, “A Dynamic Spatial Model”, National Bureau of Economic Research Working Paper #4219.
- MAYER T. AND J-L. MUCCHIELLI, 2002, “Hierarchical Location Choice and Multinational Firms’ Strategy” in Dunning J. and J-L. Mucchielli (eds.), *Multinational Firms, The Global and Local Dilemma*, pp. 133–158. London: Routledge.
- MCCALLUM, J., 1995, “National Borders Matter: Canada-US Regional Trade Patterns”, *American Economic Review* 85:615–23.
- McFADDEN, D., 1978, “Modelling the Choice of Residential Location”, in A. Karlquist et al. (eds), *Spatial Interaction Theory and Residential Location.*, pp. 75-96. Amsterdam: North-Holland.
- MOOIJ R.A. AND S. EDERVEEN, 2001, “Taxation and foreign direct investment A synthesis of empirical research,” CPB Discussion Paper No. 003.
- REDDING, S. AND A. VENABLES, forthcoming, “Economic geography and International Inequality”, *Journal of International Economics*.
- TRAIN, K., 2003, *Discrete Choice Methods with Simulation*, pp. 81–90, Cambridge: Cambridge University Press.
- VENABLES, A.J., 1996, “Equilibrium Locations of Vertically Linked Industries”, *International Economic Review* 37(2) : 341–359.
- WOLF, HOLGER C., 2000, “Intranational Home Bias in Trade”, *Review of Economics and Statistics* 82(4):555–563.

Data Appendix

Trade equation estimation

Most data used in estimating equation (9) come from Eurostat databases, and were in part already used in Head and Mayer (2000) where more details can be found. The COMEXT database provides bilateral trade flows. The VISA database provides the production data used to calculate internal trade flows of a country, subtracting its value to the value of total exports of the country. Production values are adjusted in order to account for the fact that some countries reported data only for firms larger than twenty employees in the years we are considering, while trade flows are exhaustive. Trade and production figures are then both converted into NACE 2-digit industries in order to match with the level of detail of the

subsequent location choice estimation. While this is straightforward for production data and trade flows data post 1987 (provided by Eurostat in NACE 3-digit), a large concordance work is needed to convert trade flows for the previous period from NIMEXE to NACE 3-digit. This has been done using a correspondance available from the site <http://www.eiit.org/>.

Distances calculations are crucial in this paper for both the trade equation and profit equation estimations. We calculate the distance of one nation to another—or itself—as weighted averages of sub-national distances. Considering two countries I and J (the origin and destination countries of a given flow), respectively consisting of regions indexed $i \in I$ and $j \in J$, the following formula provides both external and internal distances.

$$d_{IJ} = \sum_{i \in I} \left(\sum_{j \in J} \omega_j d_{ij} \right) \omega_i$$

We define d_{ij} as the distance between the centers of regions i and j and ω_i as the weight of region i , calculated as the share of population in 1990 for both origin and destination weights. The distance of a region to itself is obtained using a simple geographical approximation. Each region is approximated as a disk in which all production concentrates in the center and consumers are uniformly distributed throughout the rest of the area. The average distance between a producer and a consumer is then given by

$$d_{ii} = \int_0^R r \cdot 2r/R^2 dr,$$

where R denotes the radius of the disk, and $2r/R^2$ is the density of consumers at any given distance r to the center. We obtain R as the square root of the area, A , divided by π . Integrating, we obtain $d_{ii} = (2/3)R = (2/3)\sqrt{A/\pi} = .376\sqrt{A}$.

The estimation procedure consists of 288 OLS regressions (18 industries \times 16 years), providing the estimates to be used for the construction of market potentials. Each regression yields the importers' fixed effect, the estimated impacts of bilateral distance and common language, and a set of importer-specific border effects. Those border effects are identified using internal trade flows and therefore cannot be estimated for observations where production figures are missing. This is in particular the case for non-EU members countries, for which Eurostat does not report data. We apply the estimated German border effect to Northern European countries (Sweden, Finland, Norway, Switzerland and Austria) and the estimated French border effect for Southern European countries: Greece is assigned the French border effect for the whole period, Spain and Portugal have missing data before their membership in 1986. For this year, we calculate a ratio of French to Spanish/Portuguese border effects and apply this ratio to the French border effect to get Spanish and Portuguese values for the preceding years. After those adjustments, there are a few remaining "holes" in the data, mainly resulting from the well known confidentiality issues in Belgium and Netherlands for production data. Those missing figures are filled in taking an average of the industry and country average border effects for the corresponding year.

Location choice estimation

Regions and years used

The regional level choice sets incorporate 57 regions in Europe using NUTS 1 level of detail for Germany (11 regions), France (8 regions), Italy (11 regions), the UK (11 regions), Spain (6 regions), the Netherlands (4 regions), and Belgium (4 regions). Ireland and Portugal are

considered as single region countries. Out of those 57 regions, 50 were chosen at least once by Japanese investors.

Industry-level regional data availability limits the sample to the years 1984–1995. Although there are some Japanese investments in the late seventies and early eighties, the vast majority of the investments took place in the late eighties.

Affiliates

The location choices of Japanese affiliates are mainly extracted from JETRO’s *Survey of Current Manufacturing Operations of Japanese Firms in Europe*, 1996. This source provides in particular the country chosen and the date at which operations started for all manufacturing affiliates which had established operations in Europe by the end of 1995. Dropping investments in a set of countries (Luxembourg, Denmark, Greece, Austria, Finland, Sweden, Norway, Switzerland, and Iceland) and years (before 1984 and after 1995) for which explanatory variables were not available, we obtain 452 location choices to be explained.

We then identified for each firm the city in which the production unit was located. This information appears in a larger document also issued by JETRO: *The Directory of Japanese-Affiliated Companies in the EU: 1996-1997*.

A crucial matter for our study is the quality of this information: It had to be checked that, in the directory, the affiliate’s location reported was not the headquarters but the actual production units. Fortunately, the directory almost always specifies both the location of the headquarters and the location of the plant. However, the information was double checked using three alternative sources: The database used in Yamawaki et al. (1996), mostly using data from Toyo Keisai and kindly made available by Hideki Yamawaki was of great help. Table 5.4 in Strange (1992) also confirmed the locations of Japanese subsidiaries in the UK. Finally, a document from the DATAR helped to check locations for France.

The Japanese affiliates’ location choices is of course our dependent variable, but is also used to construct the Japanese counts agglomeration variable. This variable consists, for each choice, of a cumulative count of same 3-digit industry affiliates that chose each region from the first year where Japanese FDI started in Europe until the year preceding the choice under consideration. We also use this data to calculate the network counts. For each prospective investor it counts the number of *affiliated* firms that already chose the region in question. We defined “affiliated” so that it includes investments with the same Japanese parent company (for example two different Sony factories) or investments from parent companies that are members of the same industrial group (for example a Toyota assembly factory and a Nippondenso automotive electronics factory). We defined industrial groups using Dodwell’s (1988) lists of vertical *keiretsu*.

Market potential calculation

As detailed in the text, the calculation of regional Krugman market potential involves 1) a regional allocation of competition-weighted expenditure estimated at the nation/industry-level in the trade equations detailed above. 2) measures of bilateral distances between regions across the European Union.

We allocate the national competition-weighted expenditure of each industry between its regions according to the share of national GDP. Regional GDP shares come from Eurostat’s REGIO database. Bilateral regional geodesic distances are calculated using the coordinates of the main city in each region, which have been collected manually. Distance inside a region only requires data on the land area of the region, also obtained from the REGIO database.

Note finally that market potential calculation involves six countries that are not present in the location choice set of Japanese affiliates. Austria, Denmark, Finland, Norway, Sweden and Switzerland are included in the analysis in order to account for the fact that some regions in the choice set have their market potential enhanced by demand emanating from those six countries more than others.

The calculation of Harris market potential also involves bilateral distances between all regions, and national apparent consumption in each industry, allocated to regions using the same regional shares of GDP as in the Krugman market potential. National apparent consumption is obtained using the same Eurostat data as the one used in the trade equations and detailed above.

Industry-level regional data

The main source of industry-level regional data comes from Eurostat publication *Structure and activity of industry annual inquiry, principal results, regional data*. It consists of 2-digit NACE data, essentially available for NUTS 1 regions. This database contains number of establishments, employment and the wage bill for each industry-region combination. For single-region countries, national level data is used.

An electronic version exists with regional data for the years 1989 to 1992 but in fact 1992 has many missing values. We additionally used the printed version for years 1984 and 1987. Observations for years 1984 to 1986 are matched with the 1984 data. Observations for years 1987 and 1988 are matched with the 1987 data. 1989, 1990 and 1991 observations are matched with same-year data. Observations from 1992 to 1995 are matched with 1991 data. NACE 26 (Man-made fibres industry) was excluded of the sample because too few data were available. When the data was missing for a particular NUTS 1 region, the following procedure was adopted: The missing values are often due to missing values in NUTS 2 small sub-regions (for instance Corsica in Méditerranée or Val d'Aoste in Nord Ovest have many missing employment and wages values because there are only one or two firms. In this case we just sum up the remaining NUTS 2 regions to get what appears to be a very precise approximation of the true data. In other cases too many data from sub-regions was missing for a particular year, we then replaced the figure with its value for the nearest year available. As a general pattern, the main problems in data availability concerned Netherlands and even more Belgium.

National level data

Variables at the national level are the corporate tax rate and the social charges rate. Social charges rates use Eurostat data on non-wage labor costs (such as payroll taxes and pension contributions) at the 2-digit industry-country level. The source is the national version of the database *Structure and activity of industry*, also used to obtain regional data. The variable consists of the share of those charges in the total labor cost of the industry. The corporate tax data is the national statutory tax rate taken from the dataset put together by Devereux et al. (2002) and made available at <http://www1.ifs.org.uk/corptaxindex.shtml>.

Table 5: The location of Japanese affiliates in European regions in 1996

| Region | Code | Jpn. Firms | Region | Code | Jpn. Firms |
|--------------------------|-------------|-----------------------|--------------------|-------------|-----------------------|
| Belgium | | | Netherlands | | |
| Brabant | BE0 | 7 | Noord-Nederland | NL1 | 0 |
| Bruxelles-Brussels | BE1 | 4 | Oost-Nederland | NL2 | 6 |
| Vlaams Gewest | BE2 | 17 | West-Nederland | NL3 | 12 |
| Region Wallonne | BE3 | 8 | Zuid-Nederland | NL4 | 19 |
| Germany | | | Italy | | |
| Baden-Wuerttemberg | DE1 | 8 | Nord Ovest | IT1 | 6 |
| Bayern | DE2 | 16 | Lombardia | IT2 | 19 |
| Berlin | DE3 | 1 | Nord Est | IT3 | 2 |
| Bremen | DE5 | 0 | Emilia-Romagna | IT4 | 4 |
| Hamburg | DE6 | 6 | Centro (i) | IT5 | 4 |
| Hessen | DE7 | 12 | Lazio | IT6 | 3 |
| Niedersachsen | DE9 | 10 | Abruzzi-Molise | IT7 | 0 |
| Nordrhein-Westfalen | DEA | 25 | Campania | IT8 | 0 |
| Rheinland-Pfalz | DEB | 6 | Sud | IT9 | 1 |
| Saarland | DEC | 0 | Sicilia | ITA | 0 |
| Schleswig-Holstein | DEF | 4 | Sardegna | ITB | 0 |
| United Kingdom | | | France | | |
| North | UK1 | 18 | Ile de France | FR1 | 32 |
| Yorkshire and Humberside | UK2 | 10 | Bassin Parisien | FR2 | 18 |
| East Midlands | UK3 | 9 | Nord-Pas-de-Calais | FR3 | 2 |
| East Anglia | UK4 | 3 | Est | FR4 | 16 |
| South East (uk) | UK5 | 51 | Ouest | FR5 | 6 |
| South West (uk) | UK6 | 13 | Sud-Ouest | FR6 | 11 |
| West Midlands | UK7 | 27 | Centre-Est | FR7 | 7 |
| North West (uk) | UK8 | 11 | Mediterranee | FR8 | 2 |
| Wales | UK9 | 31 | Ireland | | |
| Scotland | UKA | 21 | Ireland | IE | 30 |
| Northern ireland | UKB | 2 | Portugal | | |
| Spain | | | Portugal | PT | 12 |
| Noroeste | ES1 | 1 | | | |
| Noreste | ES2 | 5 | | | |
| Madrid | ES3 | 8 | | | |
| Centro (e) | ES4 | 2 | | | |
| Este | ES5 | 34 | | | |
| Sur | ES6 | 3 | | | |