CESifo Working Papers
www.cesifo.org/wp

The Spatial Organization of Multinational Firms

Fabrice Defever

CESIFO WORKING PAPER NO. 3304
CATEGORY 8: TRADE POLICY
DECEMBER 2010

CESifo *Working Paper No. 3304*

The Spatial Organization of Multinational Firms

Abstract

Using six years of firm-level data covering 224 regions of the enlarged European Union, we evaluate the importance to a firm of locating its activities (production, headquarters, R&D, logistics and sales) close together. We find that, after controlling for regional characteristics, being closely located to a previous investment positively affects firm location choice. However, the impact of distance is dependent on the type of investment (production or service). While within-firm co-location is important for both service and production activities, only production plants are likely to be located close to prior production investments. In this latter case, national borders have a surprisingly positive effect, increasing the probability of choosing a nearby location, but on the other side of the border.

JEL-Code: F23, L22, R30.

Keywords: functional fragmentation, vertical linkages, location choice.

Fabrice Defever
University of Nottingham
GEP and Centre for Economic Performance at the London School of Economics
B13, Sir Clive Granger Building
University Park
UK – Nottingham NG7 2RD
United Kingdom
fabrice.defever@nottingham.ac.uk

December 2010
I am very grateful to Pol Antràs, Andrew Clark, Keith Head, Laura Hering, Philippe Martin, Thierry Mayer and Jean-Marc Siroen for helpful comments as well as participants at the Second CEPR Conference of GIST “Globalization, Investment and Services Trade” (University of Ljubljana, June 2010), the GEP conference on “Trade Costs” (University of Nottingham, June 2009), the “Agglomeration and Growth in Knowledge-Based Societies” conference (Kiel, 2007) and GDR Urban and Spatial Economics (Lille, 2007) workshops. I am grateful to Helen Durrant for carefully proofreading this work. I also thank Nicolas Lambert and Celine Rozenblat for helping me to find the maps, and Thierry Mayer for helping me to calculate the distances between regions. An analysis of the quality of the database and its management was carried out at the Invest in France Agency (AFII) under the supervision of Fabrice Hatem and Edouard Mathieu. I thank them for their help with this task. I am also grateful to the consulting group Ernst & Young, and more specifically to Barry Bright and Mark Hughes. The usual disclaimer applies.

1 Introduction

Falling trade and communication costs have been one major component of the ongoing process of globalization (Fujita and Thisse, 2006). As a result, it is now increasingly possible to spatially separate production stages by task (Grossman and Rossi-Hansberg, 2008) or function (Duranton and Puga, 2005). We thus observe the international *slicing up of the value chain* (Krugman, 1995), whereby multinational firms break down their value chain into various stages spread across different countries or regions due to factor price and endowment differences.

Nevertheless, firms still tend to seek geographical proximity when setting up new activities. As noted in the *World Investment Report* (UNCTAD, 2004, page 152): "*Face-to-face interaction is still required at many points in the value chain of developing, marketing, delivering and maintaining a variety of services. [S]ome processes are hard to manage cross-nationally.*"

Proximity between activities allows communication and transport costs to be reduced. However, the nature of these costs depends on the type of activity under consideration (production or services). Whereas splitting up the production of a good across locations implies transport (or trade) costs from one location to another, a stand-alone service activity will cost more in terms of coordination costs than will integrated services.

In this paper, we analyze the spatial organization of multinational firms and the incentives for firms to locate activities close to each other. Our contribution is to consider the spatial dependencies between existing elements of the firm's value chain and its location choice for a new activity. We show that multinational firms spatially organize service and production activities differently. While the locations of both new service and new production units are affected by the location of the firm's existing investments, the effect of existing production units on new production investments continues to be felt even when located relatively far away. To show this, we consider the location choice of production and service activities, starting at conception (headquarters and R&D centers) and ending with delivery (logistics and sales offices). The location determinants of these different stages of the value chain of multinational firms in the enlarged European Union are evaluated at a very detailed geographical level. With the increased importance of regional policies in the EU, this geographical scale provides an interesting and relevant tool to analyze the location of multinational firms' investments.

Previous econometric work has already considered the importance of locating new activities near to existing investments. Using data on Japanese foreign investments in the United States and European regions respectively, Head et al. (1995) and Head and Mayer (2004) find that the regional co-location between affiliates of the same industrial *Keiretsu* is an important determinant of firms' location choices. While this work informs us about the local impact of firms' networks, the analysis of the spatial organization of multinational firms has been limited to the study of the co-location of investments. Only Smith and Florida (1994) have considered the influence of distance to Japanese assembly plants on the location choice of Japanese Auto-related parts suppliers. In addition, existing work has focused on production-plant location only. Recent contributions have extended these analysis to the international location of service activities by multinational firms. Defever (2006) shows that firms tend to locate different activities within the same country, specially R&D and production. Using a more detailed geographical level, Basile et al. (2009) analyze the location of production and service investments in European regions. To take into account spatial autocorrelation, they use spatial econometrics. The drawback of this methodology is the need to aggregate data on the number of investments, which prevents the analysis of individual firm effects.[1]

[1] Work on the location choice of service activities has also appeared in Urban Economics. Henderson and Ono (2008) and Aarland et al. (2007) consider the location choice of Headquarters, and the choice between co-locating with production plants or locating in

To analyze the spatial organization of both production and service activities **at the individual firm level**, we start by setting out a simple model of the relationship between activity location and firm performance. For each type of activity, multinational firms choose over a set of locations, considering characteristics such as factor prices, access to intermediate service inputs and agglomeration effects. In addition, multinational firms are likely to spatially organize their production process and take into account their existing investment locations. We here evaluate how transport/communication costs may prevent firms from setting up activities in remote locations and lead them to locate their functions within a spatially-limited area.

Using a recently-collected data set from the consulting group Ernst & Young of almost 11000 location choices at the individual firm level over the 1997-2002 period, we are able to identify the exact location of each investment in a set of 224 NUTS-2 regions. The data set provides information not only at the firm level but also on the type of activities located (production, headquarters, R&D, logistics and sales activities). Using conditional and mixed logit models, we examine the regional determinants of location for both production and service activities. We are further able to distinguish between regional co-location, i.e. the location of two firm investments in the same NUTS-2 area, and location in neighboring regions to an existing investment (but not in the same region). The co-location of firm activities in the same region turns out to be very important for all functions, except for sales and marketing activities which are spread across locations. The location in neighboring regions to an existing investment plays no role in the location of service activities, and existing service investments do not affect firms' location choices. Neighboring investments only matter for the location of production, and only if the existing investment is also a production

a remote large city to gain access to a variety of service inputs. Strauss-Kahn and Vives (2009) study the relocation of headquarters to another city, and find that distance to the historical location plays an important role in new location choice.

unit. Surprisingly, an additional positive effect comes from locating a new production plant close to an existing production investment, but in another country. It is possible that firms locate their production plants in different countries to benefit from their respective comparative advantages, but that they nevertheless choose locations that are relatively close together in order to minimize transport costs.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 outlines a simple model and explains the construction of the dependent and explanatory variables. The estimation results are presented in Section 4, and Section 5 concludes.

2 Definition and descriptive statistics

This section provides an overview of the different activities of the firm's value chain, and presents some empirical evidence of location geography.

2.1 The data

We exploit a database developed by Ernst & Young, called the EIM (European Investment Monitor), which identifies project-based foreign inward investment announcements. The main sources of information are newspapers, financial information providers (such as Reuters), and national investment agencies (such as the Invest in France Agency).[2] When a new project is discovered, they track it in order to determine the exact location at the city level. The dataset is mainly purchased by economic development agencies wishing to identify trends, significant movements in jobs and industries at the local level. Regarding the representativeness of the dataset - there are no

[2]Projects included in the database have to comply with several criteria to be considered as international investments. The database excludes acquisitions, license agreements, and joint ventures (except in the case where these operations lead to an extension or a new establishment). It also excludes retail, hotel and leisure facilities, fixed infrastructures, extraction facilities, and portfolio investments.

minimum size criteria stated for selecting investment announcements, however the number of investments where less than 10 jobs are created turns out to be very low. In addition, the newspaper announcements are likely to mainly focus on large projects of large multinational firms. As a result, our empirical evidence probably identifies the location strategies of major multinational companies, while smaller multinationals could follow different strategies.

The database covers multinational firms' location choices over the 1997-2002 period in 23 countries of the enlarged European Union, including the members that entered the EU in May 2004, but excluding Malta and Cyprus. The investment-project data provide information at individual-firm level on multinational-firms' investments in Europe. In this paper we only consider investments from non-European multinational-firms and not investments from European firms. A major reason of this restriction is that the dataset does not include information on European multinational-firms investments in their own home country. Considering only the investments from non-European multinational-firms enables us to obtain a coherent and homogeneous subset of data.[3] In addition the location of investments from non-European multinational-firms - mostly American and Japanese in the dataset - in Europe is unlikely to be determined by the distance between the home country of investors and the region of location of investments. So considering investments from non-European multinational-firms only, enables us to study the location choice independently from home-country characteristics. This is relatively common in the literature. For example, Head et al. (1995) consider Japanese investment in the US, while Head and Mayer (2004) consider Japanese investment in Europe.

The data set includes the name of the firm, the name and origin of the

[3] As we do not observe investments in the home country, we would have to introduce an asymmetry between European and non-European investments.

parent firm and the sector of the firm's main activity. Further, we have information on the function of each investment. We consider only firms whose main activity is classified as Manufacturing and consider five different functions: *Headquarters (HQ)*, which corresponds to administration, management and accounting activities; *Research & Development centers (R&D)*, which encompasses both fundamental scientific research and applied development; *production plants*, covering anything related to the physical production of goods; *logistics* refers to all activities linked to the transport of goods, including warehousing; and *sales & marketing offices*, which includes both wholesale trade and business representative offices. Finally, and most importantly from our point of view, the dataset provides the exact location of each investment and the corresponding NUTS-2 unit. The EIM data set aggregates some of the NUTS-2 regions up to a more aggregated classification (NUTS-1).[4] All specific locations (islands and overseas locations) are excluded from the sample.[5] Our final sample includes 224 regions.[6]

2.2 Descriptive statistics

At the regional level, Table 1 shows the ranking of the top 10 locations in terms of the number of new projects over the 1997-2002 period for each function. There are significant differences between locations for production plants and service activities. Five of the top 10 locations for production plants are in Central and Eastern Europe, while the ranking for HQ and R&D centers includes only Western European locations. For example, Lon-

[4]The thirteen Greek NUTS-2 regions are aggregated up to three NUTS-1 regions. In the UK, Inner and Outer London are aggregated. This is also the case for the Provincia Autonoma Bolzano-Bozen and the Provincia Autonoma Trento in Italy, and Stredn Morava and Moravskoslezsko in the Czech Republic. In Germany, (i) Brandenburg-Nordost and Brandenburg-Sdwest and (ii) Chemnitz, Dresden and Leipzig are aggregated.

[5]These are: the Balearic Islands, the Canary Islands, Ciudad Autnoma de Ceuta, and Ciudad Autnoma de Melilla (Spain); Corsica, and the four French overseas regions (France); Guernsey (UK); Azores and Madeira (Portugal); and Aland Island (Finland).

[6]The complete database is composed of 13109 projects (extension of existing site and new creation), including all of the countries and functions available.

don and the Parisian region (Ile de France) are in the top 10 location for the four service functions but do not appear in the production ranking.

Maps of the regional distribution of the five functions are presented in Figure 1. To correct for different regional size, we calculate the number of investments for each function (between 1997 and 2002) divided by regional population (in 2000). The concentration levels are not the same for the five functions. Whereas HQ and R&D centers are highly concentrated in only a few locations, production plants are more widely dispersed.

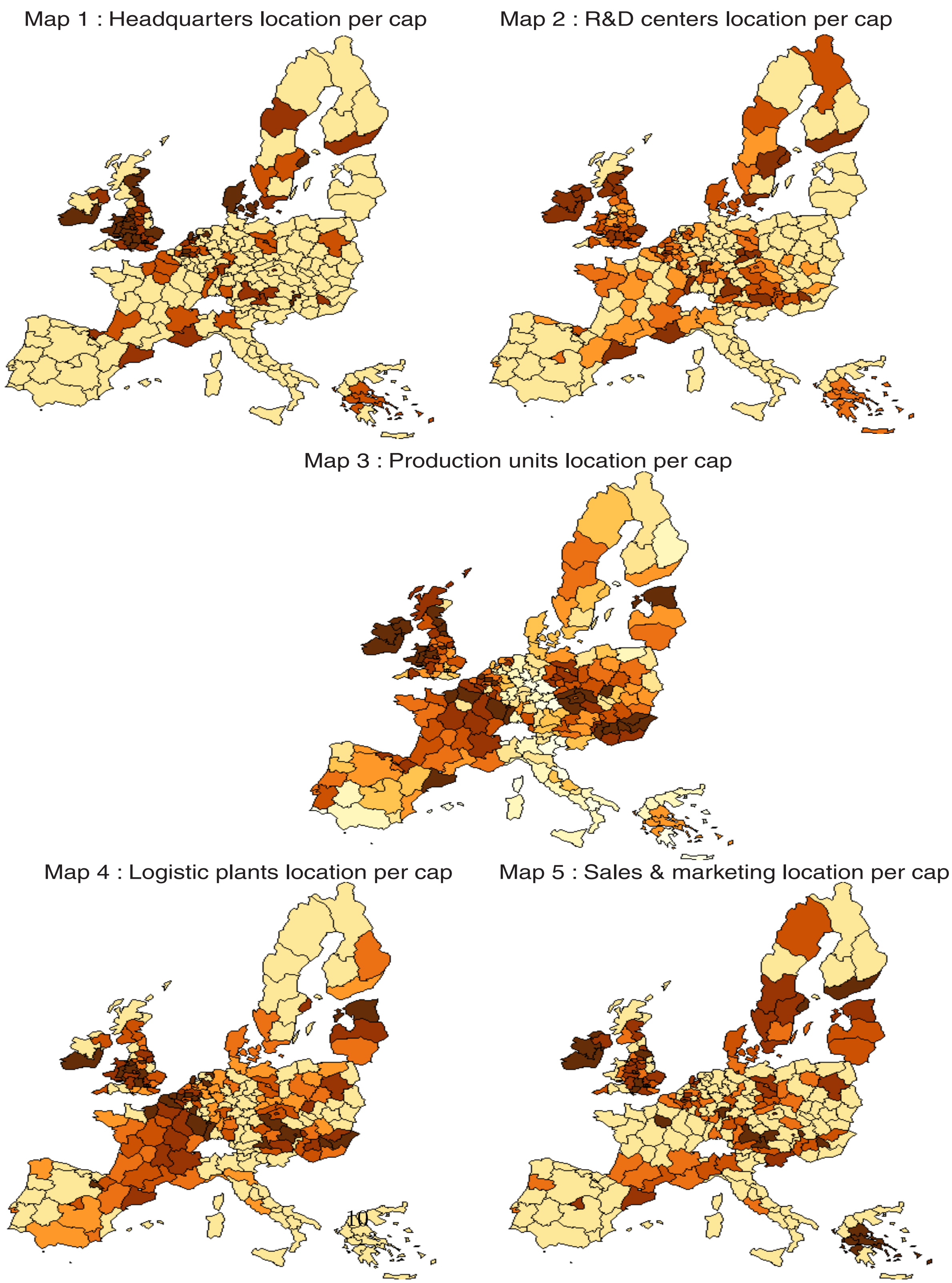
At the firm level, of the 1452 non-European parent companies in the manufacturing sectors that created new overseas establishments in the enlarged EU in 1997-2002, 1254 created new establishments for only one function. Of the other 198 firms, 125 carried out investments in two types of activities, 48 in three, 10 in four, and 15 invested in all five of the functions under consideration. Some firms invest abroad a great deal. For example, over the 1997-2002 period, Ford Motor Co and Daimler-Chrysler Corp were responsible for 38 and 34 new establishment announcements in the enlarged EU respectively. Table 2 shows the ranking of the ten largest parent companies (in terms of number of new foreign establishments). These ten firms, covering 10.5% of all new investments in the manufacturing sector, established a substantial number of new service activities to support their European production. Together, they represent 17.5% of new investments in R&D, but only 9.7% of investment in production plants.

Table 1: Top 10 overseas locations by function

Region	NUTS 2	Country	Nb of projects
		Headquarters	
London	*uki**	UK	30
Denmark	dk00	Denmark	10
Brussels	be10	Belgium	10
Ile de France	fr10	France	10
Southern and Eastern Ireland	ie02	Ireland	7
Gloucestershire, Wiltshire and North Somerset	ukk1	UK	7
Berlin	de30	Germany	7
Berkshire, Bucks and Oxfordshire	ukj1	UK	7
Vienna	at13	Austria	6
Zuid-Holland	nl33	Netherland	6
		R&D centers	
Southern and Eastern Ireland	ie02	Ireland	21
London	*uki**	UK	20
Stockholm	se01	Sweden	17
Catalonia	es51	Spain	16
Provence-Alpes-Cote d'Azur	fr82	France	15
East Anglia	ukh1	UK	14
Oberbayern	de21	Germany	11
Berkshire, Bucks and Oxfordshire	ukj1	UK	11
Ile de France	fr10	France	10
South Western Scotland	ukm3	UK	10
		Production plant	
Southern and Eastern Ireland	ie02	Ireland	50
Catalonia	es51	Spain	36
Kozep-Dunantul	hu21	Hungary	36
Kozep-Magyarorszag	hu10	Hungary	34
Border, Midlands and Western	ie01	Ireland	32
Gloucestershire, Wiltshire and North Somerset	ukk1	UK	28
Severozapad	cz04	Czech rep	27
West Wales and The Valleys	ukl1	UK	23
Severovychod	cz05	Czech rep	23
Eszak-Magyarorszag	hu31	Hungary	22
		Logistics	
Antwerpen	be21	Belgium	15
Southern and Eastern Ireland	ie02	Ireland	11
Ile de France	fr10	France	10
Rhone-Alpes	fr71	France	10
Limburg	be22	Belgium	8
London	*uki**	UK	7
Noord-Holland	nl32	Netherland	7
Berkshire, Bucks and Oxfordshire	ukj1	UK	6
Nord - Pas-de-Calais	fr30	France	5
Catalonia	es51	Spain	4
		Sales & marketing	
London	*uki**	UK	86
Ile de France	fr10	France	40
Stockholm	se01	Sweden	21
Darmstadt	de71	Germany	19
Oberbayern	de21	Germany	18
Berkshire, Bucks and Oxfordshire	ukj1	UK	17
Vienna	at13	Austria	12
Brussels	be10	Belgium	12
Mazowieckie	pl12	Poland	12
Southern and Eastern Ireland	ie02	Ireland	11

New foreign investments of non-European firms in the 224 regions of 23 countries of the Enlarged European Union (EU15 and CEE8) for the five functions over the 1997-2002 period in manufacturing sectors. Notes: * indicates the use of NUTS-1 classifications instead of NUTS-2.

Figure 1: Number of investments by function as a share of regional population in the Enlarged European Union.



Notes: These maps show nine quantiles, with darker colours representing a higher investment/population ratio. New overseas creations in the manufacturing sector in the 23 countries of the Enlarged European Union (EU15 and CEE8) over the 1997-2002 period.

Table 2: Top 10 parent companies (in terms of number of new creations)

Parent company	HQ	R&D	Production	Logistics	Sales	Total
Ford Motor Co	2	5	23	1	7	38
Daimler-Chrysler Corp	2	3	13	9	7	34
Fujitsu Corp	3	19	1	1	5	29
Toyota Motor Corp	2	1	14	9	3	29
General Motors Corp	3	3	18	2	2	28
IBM Corp	1	16	2	1	8	28
General Electric Co	3	3	15	1	4	26
Motorola Inc	0	17	7	0	1	25
Daewoo Corp	0	0	16	3	1	20
Samsung Corp	1	1	14	2	1	19
Total of the Top 10	17	68	123	29	39	276
Total No. of projects	230	389	1264	217	521	2621
Ratio "Top 10" to "All"	7.4%	17.5%	9.7%	13.4%	7.5%	10.5%

New creations of non-European firms in manufacturing sectors in the 23 countries of the Enlarged European Union (EU15 and CEE8) for the five functions over the 1997-2002 period.

3 Empirical Implementation

3.1 The Model

To help structure the discussion, we follow Duranton and Puga (2005) and Henderson and Ono (2008), and outline a simple model describing the contribution of an outlet to a firm's performance. Let $R = (1, ..., r, ...n)$ be the set of possible locations. A firm's activity $k = \{p, s\}$ can be either a production plant p or a service unit s. Both type of activities are carried out using low-skilled labor L and high-skilled labor H, in addition to a certain amount of local intermediate services M. The production of activity k in region r is given by a Cobb-Douglas function, in which the β's represent the relative importance of each factor.[7]

$$Q_{ir}^{k} = A(E_r, D_{ir}^{k'}, \mu_i) L_r^{\beta_l} H_r^{\beta_h} M_r^{\beta_m} \qquad \text{with } \; k' \in \{p, s\} \tag{1}$$

[7]To simplify notations, we drop the k subscript in the left-hand side of the equation.

$$\text{with} \qquad H_r = \left(\sum^{h_r} \mathcal{H}_r{}^{\sigma} \right)^{1/\sigma}$$

$$\text{and} \qquad M_r = \left(\sum^{m_r} \mathcal{M}_r{}^{\alpha} \right)^{1/\alpha}$$

The terms $\mathcal{H}_r$ and $\mathcal{M}_r$ denote the amount of high-skilled and intermediate-service inputs entering into the production stage k in a Dixit-Stiglitz-Ethier fashion. The terms h_r and m_r are respectively the number of varieties of skill and service inputs available at location r, and σ and α are the elasticities of substitution between varieties. In addition, we assume that production is subject to a Hicks-neutral shift factor $A(.)$, which is itself a function of a number of other variables.

We first have evidence that related multinational firms tend to cluster in the same regions due to the presence of externalities, such as information spillovers. Duranton and Puga (2005) consider two possible forms of agglomeration: the agglomeration of firms belonging to the same sector and the agglomeration of outlets belonging to the same function (or activity, e.g. headquarters' activity). We define E_r as the the number of outlets with the same function k of other multinational firms belonging to the same sector as firm i at location r.

Second, we have the geographical relationship between the activity's location and the existing activities of firm i. Denote by $D_{ir}^{k'}$ the distance matrix summarizing the geographical relationship between firm i's activities k located at r and the location of the firm's existing activities $k' \in \{p, s\}$ located in all other regions $r' = (1...j...n)$. Distance between the firm's activities naturally increases the transport (trade) costs of the inputs that are physically shipped; it also affects how efficiently each activity can be supported, managed and monitored by service activities.

Third and last, we have unobserved firm characteristics, μ_i.

As the next step is to define the contribution of an outlet to firm profits, we now look at the firm's sub-problem of choosing the optimal level of its outlet's activity, given by the inputs L, $\mathcal{H}$ and $\mathcal{M}$ the outlet uses, taking the locations of the outlet and the firm's previously-located outlets as given. Assuming that the outlet employs labor and purchases intermediate service inputs from local markets, we denote low-skilled wages in location r by ω_r. The high-skilled H and service inputs M are produced using high-skilled workers, with wage ω_r^h, under monopolistic competition, leading to standard results for their respective prices: $p_r^h = \omega_r^h/\sigma$ and $p_r^m = \omega_r^h/\alpha$.[8] For simplicity, we assume in the rest of the paper that ω^h is constant across locations and is normalized to one.[9] We can then write the contribution of an outlet located at r to the overall profit of firm i as $\pi_{ri} = Q_{ri} - (\omega_r L_i + p_r^h \mathcal{H}_i + p_r^m \mathcal{M}_i)$. Maximizing π_{ri} with respect to L_i, $\mathcal{H}_i$ and $\mathcal{M}_i$, we obtain:

$$\pi_{ri}^k = \xi A(E_r, D_{ri}^{k'}, \mu_i)(\omega_r)^{-\beta_l}(h_r)^{\beta_h}(m_r)^{\beta_m} \quad (2)$$

where we have set $\beta_h + \beta_l + \beta_m = 1$ and $\sigma = \alpha = 1/2$ for simplicity. The term ξ is a collection of parameters.

3.2 Empirical strategy

Multiplying equation (2) by ξ and taking natural logs leaves the profit rankining between locations unchanged and allows us to obtain a simple expression for profitability.

$$\ln(\pi_{ir}^k) = -\beta_l \ln(\omega_r) + \beta_h \ln(h_r) + \beta_m \ln(m_r) + ln(E_r) + \ln(D_{ir}^{k'}) \quad (3)$$

[8]Under monopolistic competition, production costs include both a fixed and a variable component, both paid in labor units, and price is a simple mark-up over marginal cost. This latter is equal to one divided by the constant elasticity of substitution between the symmetric varieties.

[9]We can motivate this assumption by the mobility of skilled workers. Practically speaking, it is difficult to measure high-skilled wages. For this reason, we enter H in production as a set of varieties produced under monopolistic competition. This allows us to focus on the quantity of varieties instead of the wage. In the empirical section, the set of varieties will be proxied by the ratio of the skilled workers to the total population.

Equation (3) expresses the profitability of activity k located in region r. The firm looks across feasible locations and chooses the location which maximizes the outlet's contribution to its profits. To construct our dependent variable, we consider only real creations (also known as *greenfield*)[10] carried out by non-European (mostly American and Japanese) multinational firms[11] in manufacturing over a set of 224 European regions. This leaves us with 2621 investments. Each location decision is a discrete choice made among several alternatives.

Profit π_{ir} is decreasing in local wages and increasing with the availability of high-skilled workers and intermediate inputs in the local market. However, the relative importance of both types of labor and intermediate service inputs largely depends on the type of outlet. The coefficients reflecting high-skilled labor (β_h) and intermediate inputs (β_m) are likely higher for service than for production activities, which latter uses more low-skilled labor.[12] We expect the location of service activities to be influenced by local human capital and other intermediate service inputs, so the corresponding coefficients for the four service functions should be higher than those for production. Since labor costs are the most important factor for production plants, we expect this variable to have a negative and significant effect on location choice. In our empirical implementation, local labor cost is measured by *Unit Wage Costs*, which is total wages and salaries in the manufacturing sectors per worker divided by productivity (value added per head) at the

[10]The expansion of existing sites represents one-third of the total number of projects. These expansions are not directly linked to location choice. Consequently, we use only real creations for the construction of the dependent variable.

[11]To obtain a coherent and homogeneous subset, we exclude European multinational firms. We would like to study location choice independently of home-country characteristics. Considering European investments within Europe requires the consideration of other location determinants, such as the distance between the home country and the location. In addition, as we do not observe investments in the home country, we would have introduced an asymmetry between European and non-European investments.

[12]This is in line with the stylized facts in Maurin and Thesmar (2004), that both upstream and downstream service activities are skill-intensive.

NUTS-2 level. The *Education* variable corresponds to the percentage of 25 to 64 year-olds with tertiary education.[13] We use the *Density* of population of each region as a proxy for access to intermediate service inputs. Being located in large cities can be advantageous for service activities since it facilitates face-to-face relationships (see Holmes and Stevens, 2004). All of these variables are provided by Eurostat.

Profit π_{ir} also increases in the number of other local outlets E_r, due to positive scale *e*xternalities. The agglomeration variable, *function-sector count*, is defined as the logarithm of the stock[14] of foreign establishments of all firms in the same sector and function as the new investment.[15]

Finally, π_{ir} varies with firm-region variables, $D_{ir}^{k'}$, which characterize each county's geographical relationship to the other outlets of firm i. Vertical linkages between the different stages of the value chain are likely to encourage multinational firms to locate their activities close to each other, in order to reduce transport and communication costs. We capture this spatial dependence in a very simple way by considering the influence of nearby existing firm investments. D_{ir} is set equal to 1 if an activity was previously set up by the parent company in a region whose centroid is less than a certain number of miles away, d. This picks up the impact of prior investments for various value of d.[16] The variables $D_{ir}^{k'}$ are decomposed as follows:

$$D_{ir}^{k'} = C_{ir}^{k'} + N_{ir}^{k'} + N_{ir}^{k'} \times F_{ir}^{k'} \qquad \text{with} \quad k' \in \{p, s\} \tag{4}$$

where C (for *C*o-location) equals 1 if an activity k' was previously set up by the parent company in the same NUTS-2 region, and N (for *N*eighbor) captures location in a region whose centroid is less than a certain number of

[13]Groups 5-6 in the International Standard Classification of Education (ISCED).

[14]In order to use logs, even with zero investments, this figure is actually one plus the stock.

[15]The construction of the stock of previous investments is described in the Appendix.

[16]The construction of the bilateral distances between each pair of regions is described in the Appendix.

miles from the centroid of the region r under consideration, but in a different NUTS-2 region. Neighboring locations which are in a different country to the prior investment are picked up by F, for *F*oreign neighbor; this will reflect any additional effect due to the presence of a national border. To estimate the respective impacts of *p*roduction and *s*ervice activities, we construct two distinct variables, D^p_{ir} and D^s_{ir}, reflecting the type of the prior investment $k' \in \{p, s\}$.[17]

3.3 Econometric implementation

We now present an econometric model of firm-location choice. The most widely-used econometric technique for this type of problem is the conditional logit model (CLM) of McFadden (1984). The CLM focuses on the attributes of the choices in the set: here the characteristics of the NUTS-2 regions of the European Union. These attributes can be constant across all investors, such as wages or average education, or can vary across firms, such as their own prior investments in the same or neighboring regions. The conditional logit model is specified as follows. While the true profits from different locations $R = (1, ..., r, ...n)$, are not observed, we do see firms' actual choices and the characteristics of the alternative locations. Each location is associated with a profit of π_{ir} such that:

$$\pi_{ir} = \theta_r + \delta D_{ir} + \varepsilon_{ir}, \tag{5}$$

with $\theta_r = \beta X_r$, where X_r are region r-level control variables common to all firms (e.g. regional wage), D_{ir} is a vector of firm-region independent variables (e.g. the firm's prior investments in the vicinity),[18] and ε_{ir} is the

[17]Service activities consist of the four service functions described above, plus all other service functions, e.g. call centers.

[18]We could have dealt with spatial autocorrelation between investments via spatial econometrics, as in Bloningen et al. (2007) and Basile et al. (2009). Contrary to the approach here, these latter use aggregate data to create the dependent variable. As

error term.

From equation (2), π_{ir} also depends on firm characteristics μ_i; however our imposition of separability within A(.) implies that these will not influence location decisions. However, in the case where the unobserved characteristics of the choosers are correlated across alternative location choices, this heterogeneity will affect the error term and produce inconsistent estimates; it will in fact lead to violation of the IIA assumption. Taking this problem seriously, we introduce individual random effects into the estimation via a Mixed Logit Model (MLM: (Train, 2003)). In this case, the utility return to firm i from choice r is as specified in equation (5), but now with $\theta = \beta' X_{ir} + \mu_i X_{ir}$, where the X_{ir} are observed variables which include some of the unobserved firm characteristics, and μ_i is a random coefficient with zero mean which varies between firms. This is equivalent to a random-parameter model where the coefficients on X_{ir} can be thought to vary randomly with mean β' and the same distribution, given by μ_i, around this mean. We estimate the β' and μ using simulation methods, under the assumption that μ is normally distributed.[19] As the distance matrix is directly linked to firms, δ can be considered as fixed coefficients (see Defever (2006) and Basile et al. (2008) for recent applications of the mixed logit model to the location of multinational firms).

Another strategy is to capture the regional characteristics θ_r via fixed effects for locations (these pick up the attraction of location r that is common to all investors, independent of the parent company's prior investments). This also removes some forms of bias which potentially arise from the IIA assumption.

The coefficient vectors θ_r (and the β that they represent) and δ are estimated by maximum likelihood. The firm chooses to locate in r if the

noted by Fleming (2004), spatial econometrics with qualitative dependent variables is still developing, which prevents us from using it here.

[19]These estimations results from 250 simulations.

profit there is higher than that obtained in any other location. Assuming a Type I extreme-value distribution for the error term, ε_{ir}, we obtain the simple probability of choosing r:

$$P_{ir} = \frac{e^{\theta_r + \delta D_{ir}}}{\sum_{j=1}^{n} e^{\theta_j + \delta D_{ij}}}. \tag{6}$$

The coefficient on the firm-specific dummy variable D_{ir} can be interpreted as an odds-ratio. Everything else equal, the exponential of δ is the estimated probability ratio of choosing region r, which is close to an existing investment by the firm, over the probability of choosing region j, which is not close to an existing investment.

$$\frac{P_{ir}(y = 1/D_{ir} = 1)}{P_{ij}(y = 1/D_{ij} = 0)} = \frac{e^{\delta D_{ir}}}{e^{\delta D_{ij}}} = \exp[(D_{ir} - D_{ij})\delta] = \exp[\delta]$$

4 Econometric results

Section 4.1 presents the results of Conditional Logit Model (CLM) and Mixed Logit Model (MLM) regressions which estimate the role played by the different regional and firm-region variables (in equation (3) of Section 3) in multinational firms' new investment location choices. We then explore in Section 4.2 various distance bands for the firm-region variables via CLM estimation with region fixed-effects. Finally, Section 4.3 distinguishes between co-location, neighboring locations and foreign neighbors, as in equation (4).

4.1 Basic specification

The regression results in Table 3 show the effect of education, unit wage, population density and an agglomeration variable on location choice for each function using CLM. Table 4 presents the Mixed Logit Model estimation results: these are mostly similar to those from CLM.[20]

Regional characteristics The positive and significant *Education* coefficient in columns 1, 2, 4 and 5 shows that service activities are located in more skilled regions. This variable is negative and significant at the 10 percent level only for production. This is consistent with our prior expectation. *Unit wage cost* does not significantly affect location choice for services, but is strongly negative and significant with respect to (labor-intensive) production. However, the introduction of a simple East-West dummy or country fixed-effects render wages insignificant.[21] The negative wage coefficient essentially reflects the wage gap between Central and Eastern European (CEE8) and Western European countries.[22]

The coefficient on *population density*, used as a proxy for urban economies, is positive and significant for all service activities, but not for production. The Urban Economics literature, such as Duranton and Puga (2005), has highlighted the importance of a large, service-oriented area in order to benefit from local service input suppliers. The importance of density in the location of upstream stages was also highlighted by Holmes and Stevens (2004), who argue that service activities largely depend on face-to-face relationships. For downstream activities, such as Sales & Marketing, the result is, unsurprisingly, due to the advantage of being located close to demand.[23] This is also consistent with Holmes and Stevens (2004) and the results in Holmes (2005) regarding the location of Sales offices.[24]

[20]As we do not interpret the heterogeneity terms here, they are not shown in the table. The results are available upon request.

[21]The estimation results are not presented for space reasons.

[22]Wage is also negative for Sales & Marketing. This function is largely located in countries with high demand-potential, such as Ireland, Greece, and Hungary (See Table 1).

[23]There are a number of arguments in the literature underlining the importance of market size for upstream activities. In particular, the literature on the internationalization of R&D centers by multinational firms (Kuemmerle, 1997) suggests that centers can be dedicated either to creating new products or adapting existing products to the local market. For the latter, market size may be an important determinant of R&D location choice.

[24]A market-potential variable would also capture the importance of being close to mar-

As in previous work, e.g. Head et al. (1995), we find that *agglomeration* regarding the prior location of multinational firms' investments plays an important role in location choice for all functions. As we consider the sector-function count of previous investments, this highlights both the functional and sectoral dimension of agglomeration, in line with Duranton and Puga (2005).

Prior firm investments in the vicinity: The two last variables in Table 3 are the firm-region variables. The first controls for prior service investments located in the 75-mile area around the chosen location, while the second considers analogous prior production investments. The coefficients are interpreted as odds ratios. In the last line of column 3 of Table 3, the probability that a multinational firm locate its new production plant in region r increases by a factor of $\exp(1.63) \simeq 5.1$ if the firm had previously located an investment in region r or one of the surrounding regions i within a radius of 75 miles. The analogous probability factors for locating a HQ, an R&D center or a logistic plant in r are $\exp(0.94) \simeq 2.6$, $\exp(0.53) \simeq 1.7$, and $exp(0.93) \simeq 2.5$ respectively for there having been a prior production investment in the vicinity. Prior service activity also positively affects the probability that a multinational firm locate an R&D center, production plant or logistic plant in the vicinity (see the first from last line in Table 3). Sales & Marketing is the only function whose location choice is independent of prior local investments.

kets, especially for downstream activities. To calculate regional potential demand, we use the simple methodology inspired by Harris (1954), which consists of the sum of the GDPs of all other countries weighted by their distance to the chosen location. We then add the internal distance, as in Head and Mayer (2004). As was the case for density, market potential is strongly significant for all service activities, but not for production. This result contrasts with Head and Mayer (2004), who find a very significant market-potential coefficient when considering the production-plant location of Japanese firms in Europe. The fact that we also consider Central and Eastern European countries largely explains this difference.

Table 3: Conditional Logit Model.

		Dependent Variable: Location choice				
		HQ	R&D	Production	Logistics	Sales
Education (%)	h_r	0.09[a]	0.06[a]	-0.01[c]	0.04[a]	0.07[a]
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Log (unit cost)	w_r	0.35	0.02	-0.42[a]	-0.10	-0.43[a]
		(0.24)	(0.21)	(0.12)	(0.25)	(0.16)
Log (density)	m_r	0.51[a]	0.16[a]	-0.00	0.32[a]	0.44[a]
		(0.05)	(0.04)	(0.03)	(0.06)	(0.04)
Log (function-sector	E_r	1.01[a]	1.42[a]	1.35[a]	1.13[a]	1.02[a]
count +1)		(0.14)	(0.10)	(0.04)	(0.26)	(0.07)
		Firm - Region level				
Prior service activity	D^s_{ir}	0.40	0.74[a]	0.57[a]	0.90[a]	0.05
in a 75-mile vicinity		(0.30)	(0.17)	(0.14)	(0.25)	(0.21)
Prior production plant	D^p_{ir}	0.94[a]	0.53[a]	1.63[a]	0.93[a]	0.17
in a 75-mile vicinity		(0.31)	(0.19)	(0.09)	(0.23)	(0.24)
Observations		230	389	1264	217	521
Log likelihood		-1027	-1875	-6096	-1089	-2337

Notes. Standard errors in parentheses. [a], [b] and [c] represent respectively the 1%, 5% and 10% significance levels. Dependent variable: Location choice in the 224 regions of the Enlarged European Union (EU15 and CEE8) of the five functions over the 1997-2002 period. New creations of non-European firms in the manufacturing sector.

Table 4: Mixed Logit Model.

		Dependent Variable: Location choice				
		HQ	R&D	Production	Logistics	Sales
Education (%)	h_r	0.09[a]	0.06[a]	-0.01[c]	0.04[a]	0.07[a]
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Log (unit cost)	w_r	0.36	0.00	-0.42[b]	-0.11	-0.43[b]
		(0.33)	(0.24)	(0.13)	(0.35)	(0.19)
Log (density)	m_h	0.51[a]	0.16[a]	-0.00	0.32[a]	0.44[a]
		(0.05)	(0.04)	(0.03)	(0.07)	(0.04)
Log (function-sector	E_r	0.94[a]	1.42[a]	1.35[a]	0.93	1.02[a]
count +1)		(0.21)	(0.10)	(0.04)	(0.62)	(0.08)
		Firm - Region level				
Prior service activity	D^s_{ir}	0.43	0.74[a]	0.57[a]	0.89[a]	0.05
in a 75-mile vicinity		(0.27)	(0.18)	(0.13)	(0.27)	(0.21)
Prior production plant	D^p_{ir}	0.96[a]	0.53[a]	1.63[a]	0.94[a]	0.17
in a 75-mile vicinity		(0.27)	(0.19)	(0.09)	(0.22)	(0.24)
Observations		230	389	1264	217	521
Log likelihood		-978	-1810	-6013	-1058	-2319

Notes. Standard errors in parentheses. [a], [b] and [c] represent respectively the 1%, 5% and 10% significance levels. Dependent variable: Location choice in the 224 regions of the Enlarged European Union (EU15 and CEE8) of the five functions over the 1997-2002 period. New creations of non-European firms in the manufacturing sector.

Table 5: Conditional Logit Model with region fixed-effects.

		Dependent Variable: Location choice				
		HQ	R&D	Production	Logistics	Sales
		Firm - Region level				
Prior service activity	D^s_{ir}	0.33	0.61^a	0.57^a	0.75^a	0.06
in a 75-mile vicinity		(0.29)	(0.18)	(0.14)	(0.25)	(0.22)
Prior production plant	D^p_{ir}	1.02^a	0.61^a	1.80^a	0.71^a	0.29
in a 75-mile vicinity		(0.30)	(0.20)	(0.09)	(0.23)	(0.25)
Region fixed-effects		Yes	Yes	Yes	Yes	Yes
Observations		230	389	1264	217	521
Log likelihood		-873	-1588	-5790	-881	-2018

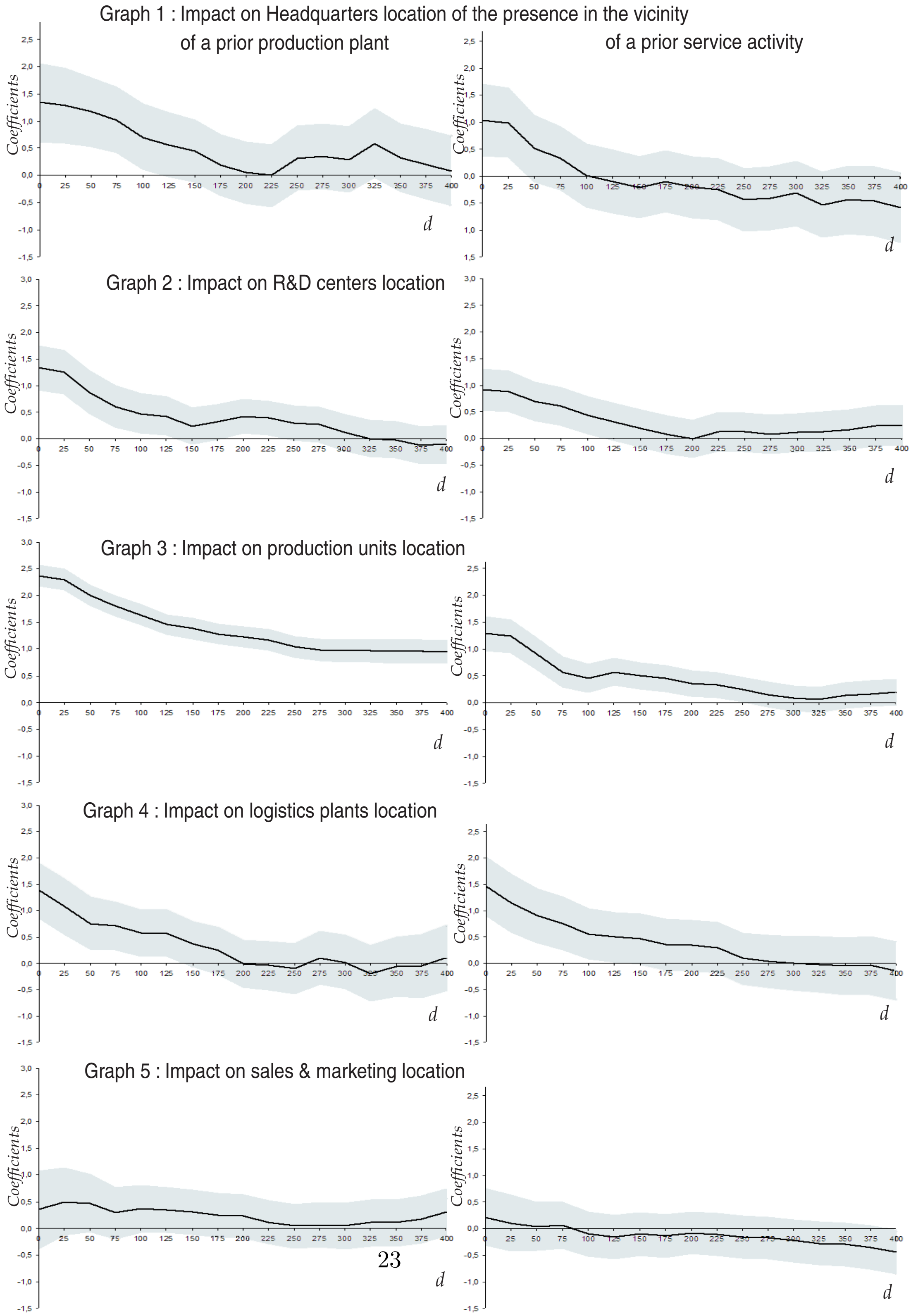
Notes. Standard errors in parentheses. ^{a}, ^{b} and ^{c} represent respectively the 1%, 5% and 10% significance levels. Dependent variable: Location choice in the 224 regions of the Enlarged European Union (EU15 and CEE8) of the five functions over the 1997-2002 period. New creations of non-European firms in the manufacturing sector.

4.2 Different distance bands

Table 5 presents a CLM with region fixed-effects, which capture the regional characteristics analyzed in the previous subsection, in addition to the two *prior investment in a 75-mile vicinity* firm-region variables. The results are robust to this specification change. Of course, the choice of the 75-mile radius is arbitrary, and it is possible that the coefficients on *prior production in the vicinity* and *prior service in the vicinity* depend on the size of the area under consideration. To check, we repeat the analysis in Table 5 with different distance bands, ranging from 0 to 400 miles, in 25-mile steps. The co-location of activities in the same region is considered to represent zero distance.

The coefficients from these regressions for different distance bands are depicted in Figure 2. The left-hand side panel of Graph 1 depicts the impact of prior production plants on the location of a new headquarters, depending on distance; the right-hand side of Graph 1 repeats the analysis for prior service activities. The ensuing graphs then consider the impact of prior pro-

Figure 2: Impact of prior local investments, depending on the radius d of the area considered.



Notes. The graphs show the coefficient estimates for each distance area with the 5% confidence intervals from conditional logit regressions with region fixed-effects. New creations of non-European firms in the manufacturing sector in the 23 countries of the Enlarged European Union (EU15 and CEE8) over the 1997-2002 period.

duction/service investments on the location of R&D, Production, logistics and sales offices location.

For a distance of zero, i.e. regional co-location, and short distances, multinational firms seem to place a great deal of importance on prior investments. Only Sales & Marketing is unaffected by previous investments, whatever the distance (see Graph 5 in Figure 2). As distance increases, both the coefficients and their significance fall. Beyond 125 miles, the area is too large and prior investments are irrelevant for the location choice of headquarters, R&D centers and logistics (see Graphs 1, 2 & 4 in Figure 2).

When considering an area with a radius superior to 125 miles, only the location of production plants exhibits spatial dependence to the geographical distribution of prior investments. From the left panel of Graph 3 in Figure 2, we can see that if region r has at least one production plant in the radius of a maximum of 200 miles, its probability of being chosen by the firm for the location of a new production plant increases by $\exp(1.23) \simeq 3.4$. Even when considering a wider area of 400 miles, the probability still rises by $\exp(0.97) \simeq 2.7$. This implies that a region located less than 200 miles from the previous investment has only a 25% higher chance of attracting the new investment than one located between 200 and 400 miles away. While the location of service activities results in a trade-off between splitting or co-locating with prior investments, the spatial organization of production is different. Even for relatively large distances, physical proximity with prior production plants still affects production-unit location choice.

4.3 Co-location, neighbors and foreign neighbors

In this section, we first distinguish between co-location and neighboring locations. We then focus on production-plant location, and identify the impact of neighboring locations in another country.

Co-location versus neighboring firm investment: Despite our ef-

Table 6: Distinction between co-location and neighborhood.

		Dependent Variable: Location choice					
		HQ	R&D	Production		Logistics	Sales
		Firm - Region level					
Service co-location	C^s_{ir}	1.03^a	0.93^a	1.23^a	1.25^a	1.46^a	0.21
$d = 0$		(0.34)	(0.19)	(0.16)	(0.16)	(0.28)	(0.27)
Production co-location	C^p_{ir}	1.32^a	1.34^a	2.37^a	2.37^a	1.38^a	0.34
$d = 0$		(0.37)	(0.21)	(0.10)	(0.10)	(0.26)	(0.37)
Neighbor service	N^s_{ir}	-0.15	0.24	0.06	0.12	0.04	0.02
$d > 0$ & $d <= 75$ miles		(0.36)	(0.23)	(0.18)	(0.20)	(0.31)	(0.25)
Neighbor production	N^p_{ir}	0.42	-0.46	0.51^a	0.37^a	0.03	0.12
$d > 0$ & $d <= 75$ miles		(0.36)	(0.34)	(0.13)	(0.14)	(0.29)	(0.29)
Foreign (Service)	F^s_{ir}				-0.20		
					(0.38)		
Foreign (Production)	F^p_{ir}				0.63^b		
					(0.26)		
Region fixed-effects		Yes	Yes	Yes	Yes	Yes	Yes
Observations		230	389	1264	1264	217	521
Log likelihood		-869	-1566	-5692	-5690	-867	-2018

Notes. Standard errors between parentheses. [a], [b] and [c] represent respectively 1%, 5% and 10% significance levels. Dependent variable: Location choice in the 224 regions of the Enlarged European Union (EU15 and CEE8) on the five functions during the period 1997-2002. New creations of non-European firms in the manufacturing sector.

forts to carry out our analysis at a detailed geographical level, our measure of vicinity is still broad. Some regions, for example in Spain, are very large, which makes difficult to evaluate precisely the impact of small distances. However, the considerable heterogeneity in our sample in terms of region size and the significant number of small regions, allow us to observe nearby locations between prior and new investments which are not necessarily on same region.

Table 6 therefore includes four different dummy variables for the location of prior investments: *service co-location* ($d = 0$), *production co-location* ($d = 0$), *neighbor service* ($0 < d <= 75$ miles), and *neighbor production* ($0 < d <= 75$ miles). The results in Table 6 show that the possibility of co-locating with a service or a production activity strongly affects the probability of choosing this specific region for all functions, except for sales & marketing. However, the presence of service or production activity in neighboring regions does not affect the location choice of service activities,[25] while column 3 shows that being located in the neighborhood of a production plant does impact the location choice of new production plants. Contrary to service activities, production plants are spatially organized by the firm, with their location being more detailed than a simple binary choice of co-locating in the same region or not.

Production plant networks and national borders: Distance to previous production investments matters when firms spatially organize their production network. National borders may also affect the likelihood of choosing a region in the neighborhood of a prior production investment but in an adjacent country.

Column 4 of Table 6 is identical to column 3 except that we add two new variables: Foreign (service) F^s and Foreign (production) F^p. These are equal to 1 if the firm's prior service/production investment in the neighborhood was in an adjacent country. Column 3 of Table 6 shows that the location choice of new production plants is not affected by the location of prior service investments in neighboring regions, and, logically, the border effect of *foreign (service)* in column 4 is also insignificant. The results are very different regarding the influence of prior production plants in the neighborhood, . An existing local production plant is more likely to attract a new production plant if it is located in an adjacent country. The coefficient on the variable F^p is positive and significant at the 2% level, which means that there is an additional positive effect from the prior production plant being located in an adjacent country[26]. This result is surprising. Firms find it

[25]We also carried out estimations considering a distance between regions of 50 miles or less, obtaining the same results. Distances of less than 50 miles are difficult to estimate due to the relatively few regions in this narrow distance band.

[26]The variable *foreign (production)* is only significant at the 10% level when considering a distance of $0 < d <= 125$ miles to calculate neighboring investments, and is insignificant

profitable to fragment their production process on both sides of the border. One possible explanation is that multinational firms locate their production plants in different countries to benefit from their respective comparative advantages, but close to each other to reduce transport and communication costs. This is similar to vertical specialization, where the production of a final good is made via multiple stages located in multiple countries. Hummels et al. (2001) identify this phenomenon as a major aspect of modern international trade.

5 Conclusion

This paper extends the existing literature on the location choices of multinational firms by studying at a detailed geographical level both upstream and downstream production and service activities. Focusing on 224 regions of the enlarged European Union over a period of 6 years, we show that production and services are very differently spatially organized.

Firms' location decisions depend strongly on the geography of prior investments, and firms tend to reinvest in the same region as before. However, nearby production plants matter only for the location of new production plants. For service activities, the physical distance to other functions, including production plants, does not seem to play any role. Finally, firms locate their production plants on either side of national borders to benefit from the respective comparative advantages, but still close to each other in order to reduce transport and communication costs.

for $0 < d <= 175$. We do not show these estimations for space reasons.

References

Aarland, K., J. Davis, V. Henderson, and Y. Ono (2007). Spatial organizations of firms: the decision to split production and administration. *Rand Journal of Economics 38*(2), 480–494.

Basile, R., L. Benfratello, and D. Castellani (2009). Location determinants of foreign firms' business functions in the enlarged Europe: Evidence from negative binomial additive models. *mimeo*.

Basile, R., D. Castellani, and A. Zanfei (2008). Location choices of multinational firms in europe: The role of EU cohesion policy. *Journal of International Economics 74*(2), 328–340.

Bloningen, B., R. Davies, G. Waddell, and H. Naughton (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review 51*(5), 1303–1325.

Defever, F. (2006). Functional fragmentation and the location of multinational firms in the enlarged Europe. *Regional Science and Urban Economics 36*(5), 658–677.

Duranton, G. and D. Puga (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics 57*(2), 343–370.

Fleming, M. (2004). *A Review of Techniques for Estimating Spatially Dependent Discrete Choice Models*. Advances in Spatial Econometrics, Springer.

Fujita, M. and J.-F. Thisse (2006). Globalization and the evolution of the supply chain: Who gains and who loses? *International Economic Review 47*(3), 811–836.

Grossman, G. M. and E. Rossi-Hansberg (2008). Trading tasks: A simple theory of offshoring. *American Economic Review 98*(5), 1978–97.

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 (2004). Market potential and the location of Japanese firms in the European Union. *Review of Economics and Statistics 86*(4), 959–972.

Head, K., J. Ries, and D. Swenson (1995). Agglomeration benefits and location choice: Evidence from Japanese manufacturing investment in the United States. *Journal of International Economics 38*, 223–257.

Henderson, J. and Y. Ono (2008). Where do manufacturing firms locate their headquarters? *Journal of Urban Economics 63*(2), 431–450.

Holmes, T. J. (2005). The location of sales offices and the attraction of cities. *Journal of Political Economy 113*(3), 551–581.

Holmes, T. J. and J. J. Stevens (2004). Geographic concentration and establishment size: Analysis in an alternative economic geography model. *Journal of Economic Geography 4*(3), 227–250.

Hummels, D., J. Ishii, and K. Yi (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics 54*, 75–96.

Krugman, P. (1995). Growing world trade: Causes and consequences. *Brookings Papers on Economic Activity* (1), 327–342.

Kuemmerle, W. (1997). Building effective R&D capabilities abroad. *Harvard Business Review 2*(75), 61–70.

Maurin, E. and D. Thesmar (2004). Changes in the functional structure of firms and the demand for skill. *Journal of Labor Economics 22*(3), 639–664.

McFadden, D. (1984). Econometric analysis of qualitative response models. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2. Amsterdam: Elsevier/North-Holland.

Smith, D. and R. Florida (1994). Agglomeration and industry location : An econometric analysis of Japanese-affiliated manufacturing establishments in automotive-related industries. *Journal of Urban Economics 36*, 23–41.

Strauss-Kahn, V. and X. Vives (2009). Why and where do headquarters move? *Regional Science and Urban Economics 39*(2), 168–186.

Train, K. (2003). *Discrete Choice Methods with Simulation.* Cambridge University Press.

UNCTAD (2004). *World Investment Report: The Shift Towards Services.* New York and Geneva: United Nations Conference on Trade and Development.

Appendix

A Distance matrix between regions

Bilateral distances between each pair of regions are calculated as great circle distances between the centroid of each region. The EIM data set provides the exact coordinate for each investment. We simply take the average longitude and latitude of investments in each region and use this as the centroid. For the few regions without any investment, we consider the coordinates of the biggest city of the region.

B History of past investments

We construct the past investment history for each parent company and for each function in all regions r. In order to study precisely the history of location of a specific site, we consider for each function only one possible investment by each parent company and for each city.[27]

We then construct the stock of investments carried out by multinational firms between 1997 and 2002. We take into account all the projects of the sample (greenfield and expansion of existing sites). More precisely, we include all the established extensions (which represent about one third of the total number of projects) carried out in the 1997-2002 period and which were not created during this period. We have to be careful not to consider the same project more than once. For example, a site extended in 2000 with no creation reported during the period 1997-2002 would be considered as anterior to 1997. However, a production plant created in 1999 and extended in 2001 has to be treated as existing since 1999. A shortcoming of our data is that we do not observe plant exit. We thus assume that any created and/or expanded activity is active over the whole period.

This allows us to consider these investments as anterior, to which we add the new establishment creations carried out during the years before the specific investment under consideration. This allows us to construct the stock of investments. *Joint ventures* are considered as a previous investment for each parent company engaged in this investment. Finally, we exclude all projects carried out by affiliates of the parent company of the firm making the investment.

[27]We count as just one investment all the projects in a specific function and in a particular city (the most detailed geographical level) for each parent company. For example, if a firm decides to locate two production plants in the same city, we only consider this investment once. The problem arises if we observe a firm carrying out a number of extensions within with the same city. We do not know whether this reflects a number of extensions of the same site, or a number of different sites.This allows us to establish an investment history at the city level, and to avoid double counting.

C Variable Definitions

Table 7: Dependent and independent variable definitions

Variable		Definition	Origin
Y		Location choices among 224 regions (*greenfield* only) of non-European firms from manufacturing sectors.	EIM
		Regional variables (NUTS-2)	
Unit Wage Cost	w_r	Total wages and salaries per worker divided by value added per worker.	Eurostat
Education	h_r	% of 20 to 65 year olds with tertiary education levels (ISCED 5-6)	Eurostat
Population density	m_r	Population divided by land area.	Eurostat
Function-sector count	E_r	Stock (plus one) of firms location in the same function and in the sector of the investing firm (both *greenfield* and expansion of existing sites).	EIM
		Firm-region variables. Distance to other functions	
Prior investment in the vicinity (e.g. $d <= 75$ miles)	D_{ri}	1 if prior investment located in the vicinity (in a radius of d miles) and 0 otherwise.	EIM
Function Co-location $d = 0$	C_{ri}	1 if prior investment located in the same region and 0 otherwise.	EIM
Neighbor firm investment (e.g. $d > 0$ & $d <= 75$ miles)	N_{ri}	1 if prior investment located in a neighboring region (in a radius of d miles) but not in the same region ($d > 0$) and 0 otherwise.	EIM
Foreign neighbor	F	1 if *neighbor* equals 1 and if the prior investment in the neighboring region was carried out in a different country and 0 otherwise.	EIM