Institute for International Integration Studies

IIIS Discussion Paper

No.220 / April 2007

Location Choices of the Pharmaceutical Industry in Europe after 1992

Frances Ruane

Economic and Social Research Institute and IIIS, Trinity College Dublin

Xiaoheng Zhang

IIIS and Department of Economics, Trinity College Dublin

IIIS Discussion Paper No. 220

Location Choices of the Pharmaceutical Industry in Europe after 1992

Frances Ruane

Xiaoheng Zhang

Location Choices of the Pharmaceutical Industry in Europe after 1992*

Frances P. Ruane†

Xiaoheng Zhang‡

DRAFT-NOT FOR QUOTATION WITHOUT PERMISSION OF AUTHORS

Abstract: Differences in regulations, technical standards and national medical cultures across EU member states created a highly segmented pharmaceutical market in Europe prior to the implementation of the Single Market Programme. The subsequent reduction in non-tariff barriers to trade would be expected to have an impact on where pharmaceutical multinationals locate production within the EU. Using discrete choice models, we study separately the determinants of multinational location choices in terms of expanded production at existing facilities and location of start-up firms. Our results support the findings of models which predict reduced rather than increased agglomeration in the face of trade-cost reductions.

JEL Classification: F15, F23, R12

11th April 2007

* We acknowledge financial support for this research through a grant to the Institute for International Integration Studies (IIIS) at Trinity College Dublin (TCD) from DEPFA BANK PLC. We are grateful to participants at seminars at the IIIS and to helpful comments and suggestions received from Ron Davis (University of Oregon, IIIS), Franco Mariuzzo (TCD), Carol Newman (TCD), Raymond Mataloni (BEA), and Matt Moran (IBEC). All the usual disclaimers apply.

† Economic and Social Research Institute, and IIIS, Trinity College Dublin.

‡ IIIS and the Department of Economics, Trinity College Dublin.

1. Introduction

The activities of multinational enterprises (MNEs) have been widely studied by economists over the past several decades. One much-researched issue, among others, is the location choice of MNEs' production. In particular, the issue arises as to why some countries attract relatively more MNEs' investment as a location for production. This question will be explored in this paper in the context of European integration in the 1990's.

Implementation of the Single Market in the European Union after 1992 resulted in the abolishment of non-tariff barriers for trade between the member states of the European Union (EU) and as consequently, trade-cost levels in the EU are at an unprecedented low point. As implied in the theories of New Economic Geography (NEG), industrial location across countries is closely connected with inter-country trade-cost levels. Some NEG theories predict that as trade costs decrease, increasing returns to scale (IRS) industries will agglomerate initially but then disperse across countries. This prediction implies that earlier industrial agglomeration across countries may influence the locational trend of industries across those countries. In this paper, we study the impact of country-level agglomeration on industrial location from a firm-level perspective, or more precisely, its impact on MNEs' location-choice decisions in a particular industry in selected EU member states in the 1990's. Since MNEs' location-choice decisions are also driven by other country-level characteristics, e.g., market size, corporate tax rates and labour market conditions etc., we also explore the impact of these characteristics on MNEs' location choice.

We focus on the pharmaceutical industry in this paper because it is a major industry[1] in Europe and one in which non-tariff barriers have been very significant in the past.[2] It also features substantial increasing returns to scale, as a R&D-intensive sector, accounting for about 17 per cent of total EU business R&D expenditures (2003 figures, EFPIA 2005). We look at the pharmaceutical MNEs' location choices in the period between 1995 and 2003. Firm-level data on location choices are drawn from *Amadeus*

[1] The pharmaceutical industry is the fifth largest industry in the European Union in terms of manufacturing value added (3.5 per cent, 2003 figure, EFPIA 2005). It is also important for European people as drugs and medicines play important roles in any national health service, which is one crucial indicator of national social welfare.
[2] See Cecchini et al., 1988.

business database, which currently has the most comprehensive data on European firms' accounts. We distinguish the location choice of where to expand existing production facilities from the location choice of where to establish new firms because (i) expansion is an important channel for MNEs to relocate production and (ii) the two kinds of choices may be motivated by different economic considerations. In our analysis, we apply two different discrete-choice models - the conditional logit model (CLM) and the mixed logit model (MXL).

The results of CLM show that, in terms of MNEs' location choice of expansion, past agglomeration of the pharmaceutical industry in one EU country reduces the probability of this country being chosen. But when it comes to MNEs' location choice of new firms, agglomeration does not have any effect. Moreover, the effects of other country-level characteristics on the probability of a country being chosen also differ for the two kinds of investment decisions. MXL produces slight different results as those of CLM for the location choice of expansion, but it produces very similar results from those of CLM for the location choice of new firms.

This research contributes uniquely to the location-choice literature because, to our knowledge, no studies on the location choice of expansion have been published. Since MXL is relatively new and is not yet popular in the location-choice studies, a further contribution of this paper is to examine the performance of MXL by comparing its results with those of CLM.

The paper is organized as follows. Section 2 gives an account of NEG theories and discusses their implications for industrial location in the context of European integration; the literature on the location choice of industries (firms) is also reviewed. Section 3 discusses the empirical methodologies and data used in this paper. Section 4 presents and discusses the results for the location choice of expansion and new-firm creation. In section 5, the paper is summarized and conclusions drawn.

2. Theoretical background

2.1 New Economic Geography (NEG) theories

One of the key features of all New Economic Geography (NEG) theories, which attempt to explain the geographic distribution of economic activities, is that production in certain sectors is presumed to exhibit increasing returns to scale (IRS). Firms in these IRS sectors generate pecuniary externalities that cause firms in the sector to agglomerate together in particular regions. While the assumption of IRS production is common to all models, the models differ in other assumptions they make, for example, regarding the degree of factor mobility (inter-sectoral and inter-regional) and the direction of trade costs. For example, Krugman (1991) assumes a two-region framework, where the manufacturing sector, which enjoys IRS, and agriculture sector (featuring constant returns to scale (CRS) production technology) exist in both regions. He supposes that some historical event happens initially, which leads one manufacturing worker to move from one region to another region, this movement will increase the demand of goods in receiving (or "core") region. Manufacturing firms reinforce this initial move by following that worker to the core region because of its larger market size and larger workforce. This relocation of the manufacturing sector reduces price index (due to saving on trade costs[3]) in the core region and raises its real wage rates. Thus more workers are further encouraged to move to the core region. Since this market size-production linkage is a self-reinforcing process, the final equilibrium in the model shows that the manufacturing sector and workers agglomerate in the core region, while leaving agriculture sector in the other ("periphery") region.

This circular causality is the essence of all NEG theories but other models are constructed on different agglomeration mechanisms. Venables (1996) suggests that final good producers prefer to cluster close to the producers of intermediate products in order to reduce production costs and, similarly, the producers of intermediate products prefer to locate close to the final good producers. Therefore, even without perfect mobility of labour, continuous agglomeration can occur driven by regional clustering

[3] If manufactured goods are available locally instead of having to be imported from the other region, the trade costs of importing can be saved. This outcome is referred to as "Pecuniary externality" in the NEG theories.

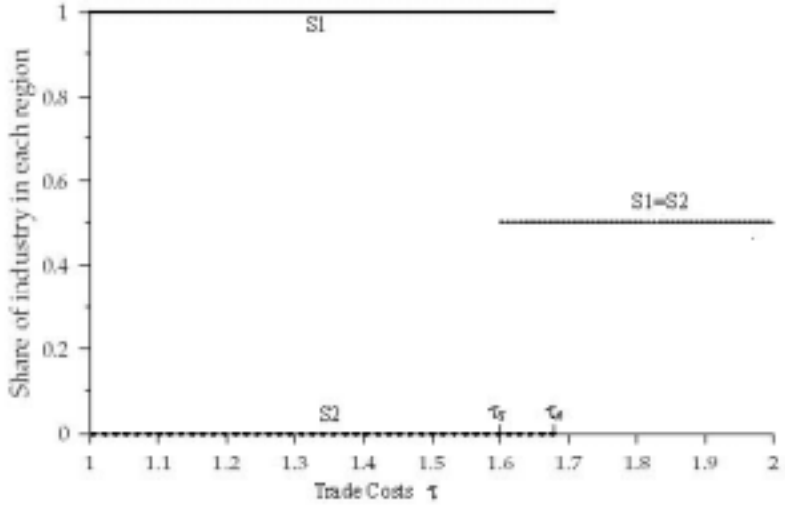
of production units, which are linked upstream or downstream. Baldwin (1999) uses a similar framework to that in Krugman's model, but focuses on endogenous capital accumulation — if some exogenous event helps to raise the profits of manufacturing firms in one region but to reduce the profits of manufacturing firms in the other region, the rate of return in the former region begins to increase and leads to more activities in this region. Therefore, the manufacturing sector in the first region will expand at expense of the agriculture sector, by absorbing agriculture workers. The reverse happens in the other region - the manufacturing sector shrinks and workers move to the agriculture sector. In this model, inter-regional labour mobility is not needed to generate the core-periphery outcome. Puga (1999) presents a model based on Krugman's Core-Periphery model, into which he incorporates input-output linkages. This model demonstrates the agglomeration processes of manufacturing sector across two regions under assumptions that either manufacturing workers are inter-regionally mobile or immobile.

Despite differences in the mechanisms employed in the NEG models, they all imply that industrial agglomeration in one or more regions is closely connected with trade costs between regions (e.g., transportation costs, tariffs, non-tariff barriers and customs efficiency, etc.), and is influenced by balance between agglomeration forces and dispersion forces (i.e., high labour cost, high land rental and severe competition in the intermediate and final goods markets.). Generally, these models generate two different patterns of relationship between industrial agglomeration and trade costs: a monotonic relationship or a non-monotonic (bell-shaped) relationship, depending on their assumption of labour mobility. The monotonic outcomes in Krugman (1991) and Baldwin (1999) are linked to the assumption of inter-regional labour mobility, which is illustrated in Figure 1. The x-axis shows the trade-cost level, ranging from $\tau = 2$ (maximum trade costs) to $\tau = 1$ (zero trade costs) and the y-axis shows the shares of industry in two regions.[4] At the right end of the figure, when trade costs between two regions are extremely high, manufacturing firms prefer to stay and serve the region where they were incorporated. Krugman assumes a special case where industry is evenly distributed across these two regions at high levels of trade costs. As trade costs

[4] τ denotes a Samuelson-type iceberg cost. If one wants to import one unit of a good from region A into region B, then one needs to ship τ ($\geq$1) units of that good , since τ-1 melts down during the trip, i.e., τ-1 is the trade cost. Therefore, τ =1 means no trade costs incur.

fall, industrial agglomeration stays unchanged until trade costs cross the critical value of τ_s. As one moves through the critical trade cost values, industry begins to agglomerate in one region and which region receives all industry depends on the exogenous shock or so called "historical event.[5] At these low trade costs manufacturing firms can locate anywhere and serve other regions without incurring any additional costs. While the excess demand for labour in the receiving region generates a wage gap between two regions, as long as labour mobility can eliminate the inter-regional wage gap, then other firms follow and eventually, complete industrial agglomeration in one region is reached.

Figure 1: Monotonic Relationship between Agglomeration and Trade Costs in Krugman (1991) and Baldwin



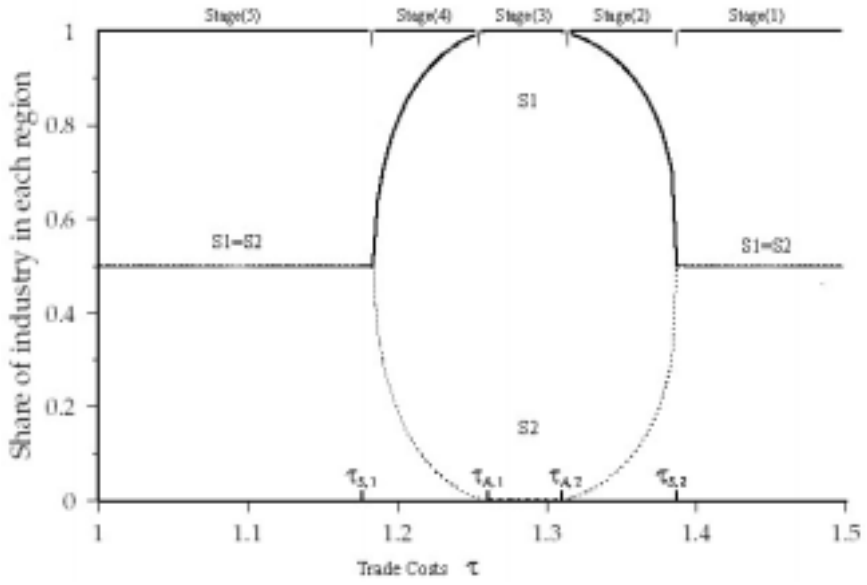
Source: Modified version of figure in Ottaviano and Puga (1997).

By contrast, the immobile labour assumption in the Venables (1996) and Puga (1999) models creates a dispersion force to counteract agglomeration advantages, so the industry-agglomeration process goes through five stages under different levels of trade costs and shows a non-monotonic (or bell-shaped) relationship. It is illustrated in Figure 2. In Stage 1, when trade costs are very high, serving the other region is not economically profitable and hence firms in the manufacturing industries distribute evenly across two regions. As trade costs reduce, the industries begin to agglomerate (Stage 2) in one region and if the costs continue to fall, then complete agglomeration is

[5] For example, suppose the authority in one region introduces a tax incentive for firms that increase their net profits, this will be the region into which the firms move.

reached between two critical values of trade costs (Stage 3). As trade costs reduce further, agglomeration begins to decline (Stage 4). In this stage, because of inter-regional labour immobility, the wage gap between the agglomerated and un-agglomerated regions causes more losses for firms than the agglomeration benefits they receive (from either the large market or the proximity of upstream or downstream industries). Consequently, it drives previously agglomerated industries back to the less agglomerated periphery region. When trade costs reduce enough (Stage 5), industries distribute evenly in two regions again.

Figure 2: Bell-shaped Relationship between Trade Costs and Agglomeration Process



Source: Modified version of original figure in Puga (1999).

2.2 Implications of NEG theories for the agglomeration of the pharmaceutical industry in the EU

Pharmaceutical agglomeration in the EU: theory and real trend

In the European Union, the major elements of trade costs - tariffs and quotas - were removed between the initial six members at the end of 1960's and then following the entry of each successive wave of new entrants. However, there were no major changes in non-tariff barriers (NTBs), which continued to grow until the launch of the Single European Market Programme, which announced the abolishment of NTBs in the whole

European Union with effect from 1993. In this context of decreasing trade costs, NEG models predict that the agglomeration level of industries in the European Union can either be strengthened (see Figure 1 and Stage 2 in Figure 2) or be weakened (see Stage 4 in Figure 2). Because the pharmaceutical industry features high increasing returns to scale (see related studies in Brülhart and Torstensson (1996), Brülhart (1998) and Midelfart-Knarvik et al. (2000) etc.), we argue that this possible outcomes regarding agglomeration apply in this industry.[6] Clearly, the question of whether the agglomeration of the European pharmaceutical industry was strengthened or weakened in the past decade depends on the level of trade costs prior to and after 1993. Unfortunately, because there is no benchmark by which to judge how high the level of trade costs was, one cannot tell the precise stage of agglomeration in pharmaceuticals in Figure 2.

However, we can measure the actual agglomeration trend of the whole European pharmaceutical industry for the past decade to shed some light on this question. The geographic concentration (agglomeration)[7] of production of the pharmaceutical industry in 14 European Union countries from 1993 to 2002[8] can be measured using the Hirschman-Herfindahl Index (HHI)[9]. Pharmaceutical production is measured by gross output and by the number of employees, using data that come from the OECD STructural ANalysis (STAN) database. Figure 3 shows changes in the HHIs for pharmaceutical production over the 10-year period, with the HHI of deflated GDP for comparison.[10] The figure clearly shows a decreasing trend of agglomeration of pharmaceutical production, especially when measured in terms of gross output. During

[6] Brülhart (1998) ranks the pharmaceutical industry as a high economies-of-scale industry following the classification of Pratten (1988). Brülhart and Torstensson (1996) and Midelfart-Knarvik et al. (2000) rank the chemical industry as high IRS industry.

[7] Concentration and agglomeration are interchangeable in this paper and both refer to the absolute level of industry production in certain geographic unit (country in this paper). However, in other research on industry location, concentration typically refers to the concentration of output amongst firms within a sector.

[8] Due to data availability, we can only calculate cross-country HHIs for 14 European countries from 1993 to 2002. These countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, and UK.

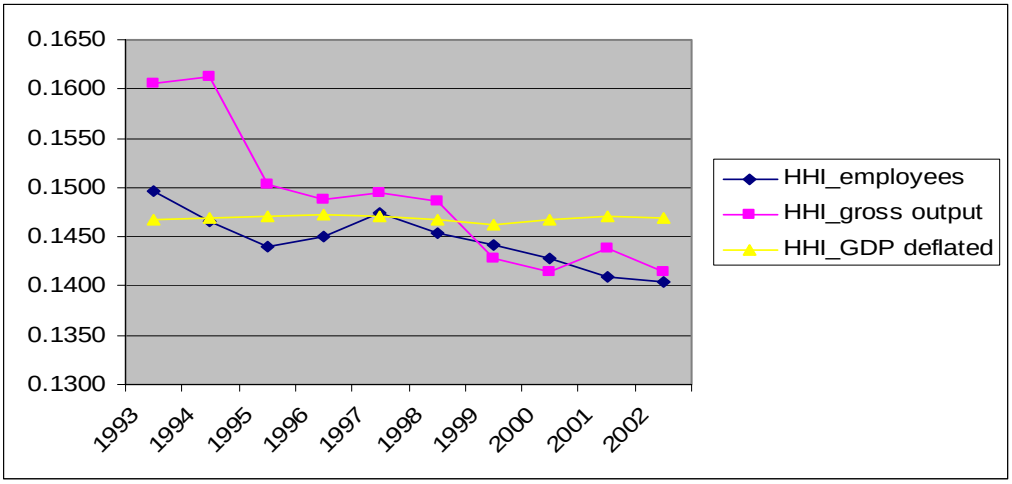
[9] The Hirschman-Herfindahl Index is defined as $HHI = \sum_{i=1}^{n} S_i^2$, where S_i is the production share of country *i* in the data set that under investigation, and *n* is the number of countries.

[10] We have some data that allow us to calculate the HHIs of pharmaceutical production for a smaller group of European countries for the years up to 1985. These countries are Austria, Denmark, Finland, France, Ireland, Italy, the Netherlands, Spain, Sweden and UK. The HHIs show a relatively stable trend prior to 1990, but a decreasing trend after 1990.

the same period, the geographic concentration of GDP is virtually unchanged, which excludes the possibility that the decreasing trend in agglomeration of pharmaceutical production simply mirrors the decreasing trend in economic agglomeration.

Given the significant lowering of trade costs, the decreasing agglomeration trend of the pharmaceutical industry suggests that this industry has been moving down along the left part of the arc (Stage 4) in Figure 2. It is consistent with Puga's model but not with Krugman and Baldwin's models, thus implying that either the wage gap between Member States or congestion costs in the agglomerated regions are driving this industry to the less agglomerated regions.

Figure 3: HHI Measures of Geographic Concentration of Pharmaceutical Production in Selected European Countries



Source: (OECD STAN Data, EU14, excl. Luxembourg)

In terms of pharmaceutical MNEs' location-choice decisions, this downward trend implies that the previous agglomeration of pharmaceutical production at country level may reduce the probability of a particular country being chosen by pharmaceutical MNEs to expand existing production facilities or to start up new subsidiaries. However, previous agglomeration is not the only driver of MNEs' location-choice decisions but operates alongside the economic and policy influences at play. Consequently in our discrete-choice models, we isolate the agglomeration effect by controlling for these other effects. We focus particularly on market (product and labour) effects and tax-policy (corporate tax rates) effects.

In terms of market effects, it is likely that any large country will have a correspondingly large market for manufactured products, and in this case, for pharmaceuticals. Given this, a manufacturing firm featuring IRS is more able to rationalize its production in a large country than in a small country, assuming some transportation costs. We can expect, therefore, that a large market will provide a relatively more attractive location for firms to expand production (if it has existing plants) or to establish new production facilities. This would suggest a trend towards increased agglomeration. However, as implied in the NEG models, the competition on local product and labour markets in the large country may also be more severe than that in the smaller country, which works to reduce tendencies for agglomeration as they operate as a dispersion force. Therefore a large country may not necessarily be more attractive to MNEs.

In terms of corporate tax rates, traditional priors would be that locations with lower corporate tax rates would be more attractive to investment. However this is more complex in models with IRS, as its existence can influence tax policy. Using their core-periphery model, Baldwin and Krugman (2004) show how corporate-tax competition between core and periphery countries is influenced by the existence of agglomeration advantages, such as input-output linkages, proximity of large workforce, etc. Instead of "race-to the-bottom" tax competition between two countries, the government of the core country (where all manufacturing industries locate) can charge an equilibrium corporate tax that is higher than that of the periphery country and still retain these industries. This prediction, which relies on the assumption of IRS, is possible as long as the tax burden on individual firms does not outweigh the agglomeration advantages they are enjoying in the core country.[11]

2.3 Empirical literature on industry location and firm location

Research, based on sectoral-level data, on industry location in Europe in the period prior to the Single Market has found mixed evidence of industrial agglomeration. Using data on 11 European countries and 18 industries (including the chemical industry) between 1980 and 1990, Brülhart and Torstensson (1996) found that indices of IRS are positively correlated with the locational Gini coefficients, which suggests

[11] This is yet another reason why the pharmaceutical industry is an interesting case to analyse.

that industries with higher levels of IRS are more concentrated in selected European countries. Noting the significant non-tariff barriers between the EU countries, they suggest that IRS industries will become more concentrated after 1990 if non-tariff barriers cease to hinder free trade. Amiti (1998, 1999) found similar results - during the period between 1968 and 1990, industries (including the pharmaceutical industry) characterized by high-scale economies and high proportions of intermediate goods in production showed an increase in geographical concentration across 12 EU countries. In contrast, using the locational Gini coefficient and controlling for the distance between countries, Midelfart-Knarvik et al. (2000) found diverse trends of concentration across industries, with a very slow process of dispersion in geographic distribution for manufacturing sectors overall from the 1970's to the 1990's.[12] Aiginger and Davies (2004) and Aiginger and Pfaffermayr (2004) examine the geographic concentration of industries in the EU for the period up to 1998. They found, for the post 1992 period, industrial concentration declined across 14 EU countries. They suggest this evidence is consistent with a non-monotonic relationship, i.e., the left part of arc in the bell-shaped curve in our Figure 2, where decreasing transport costs lead to dispersion. [13]

Research using either panel-data models or discrete-choice models has also explored agglomeration forces by focussing on the determinants of firms' location choices conditional on the receiving region/country's characteristics while controlling for firm-level characteristics. Some of the early research papers have studied firms' location choices in the US[14], while more recently there have been several studies of location choice within individual European countries.[15] Of particular relevance to this paper are studies on the location choice of MNEs across several European countries, at regional[16] and country level: Head and Mayer (2004) for Japanese firms in nine EU

[12] Specifically, in the case of the medium and high IRS industries (including the chemical industry), they find a diminishing trend in the geographic concentration in central European countries.
[13] Combes and Overman (2004) provide a comprehensive survey of studies on this topic.
[14] See Bartik (1985); Friedman et al. (1992); Woodward (1992); Head et al. (1995, 1999); and Shaver, (1998).
[15] See Barrios et al. (2002) for Ireland; Crozet et al. (2003) for France; Hogenbirk and Narula (2004) for the Netherlands, and Békés (2005) for Hungary.
[16] Regional level studies are at NUTS 1 level. (NUTS stands for Nomenclature of Territorial Units for Statistics, the official regional classification system developed by EUROSTAT to facilitate regional study in the European Union. Current version of the NUTS classifies 25 member states at the NUTS 0 (country level), and then further decomposes these countries into 89 regions at the NUTS1 level, 255 regions at the NUTS2 level and 1221 regions at the NUTS3 level.

countries at the NUTS 1 level; Basile et al. (2003) on EU and US MNEs' investments in 55 NUTS 1 regions for eight EU countries; and Disdier and Mayer (2004) for French MNEs in Western Europe and Eastern Europe.

Agglomeration forces are seen as playing an important role in many studies. For example, Bartik (1985) and Head et al. (1995) find that the existence of regional manufacturing activities encourages new manufacturing investments. Besides that, foreign MNEs are more likely to set up their new plants in the locations where has existing business from the same country or even the same business group.[17] However, the results are not uniform for regions within countries[18] and across countries within the EU.[19]

The influence of local goods markets on the probability of one region or country being chosen by MNE investors is confirmed in most research papers. The effect of the EU market is also verified indirectly by the studies on Ireland and Hungary, which find that MNEs prefer to locate in proximity to the major ports and airports in Ireland (Barrios et al., 2002), and at locations in Hungary that are close to the Western borders (Békés, 2005).

The evidence on the importance of labour-market conditions to location choice is mixed. For example, wage rates are found to be negatively associated with FDI inflows or new establishments (Bartik, 1985; Friedman et al. 1992, etc), but the effect may also be positive or insignificant (Devereux and Griffith, 1998; Head et al., 1999; Barrios et al., 2002). Since NEG models clearly indicate that higher wage rates hinder the movement of economic activities, several arguments are proposed to explain this discrepancy between theories and empirical evidence. One argument suggests that wage rates work as a signal on the quality of labour, and consequently foreign MNEs

[17] For example, Japanese firms tend to locate in the regions with other firms from the same industrial group, or Keiretsu.
[18] For example, Hogenbirk and Narula (2004) find that when comparing the Randstad region (the agglomerated region containing large cities, major ports and airports) with the rest of the Netherlands, the presence of local business seems work as a deterrent for new foreign establishments.
[19] Disdier and Mayer (2004), using data on French MNEs, find that industrial agglomeration is less important as a factor influencing their location-choice decisions in Central and Eastern Europe than in Western Europe.

may be happy to accept high wages to secure skilled labour.[20] Another argument is that the positive wage effect on FDI merely reflects an "industry bias" or "skill bias".[21]

The influence of policy factors is also explored by researchers, with the corporate-tax rate typically used as the policy instrument. Although the evidence generally indicates a significant negative relationship between tax rates and the attraction of foreign investments, some researchers find that taxes have a relatively modest impact (e.g., Friedman et al., 1992 and Crozet et al., 2003).[22] Differences in the research findings on the impact of corporate tax rates on FDI are generally due to differences in the precise tax measures used,[23] differences in tax burdens across industries,[24] and differences in the way in which the tax effects conditional on other variables are taken into account.[25]

Other factors affecting MNEs' location decisions are R&D capacity (positive effect in Neven and Siotis, 1996 and Blonigen, 1997); infrastructure (positive in Woodward, 1992; Holl, 2004 and Basile et al., 2003) and the availability of government support such as EU structural funds and state aids/promotion (positive effect in Friedman et al., 1992 for US; Shaver, 1998 for US, Barrios et al., 2002 for Ireland and Basile et al., 2003 for the EU).

[20] Guimaraes et al. (2002) argue that wage effects may be significant between countries but not necessarily important across regions within one country if investors have already decided on the country in which to invest.
[21] Békés (2005) suggests that MNEs are most likely to be in the high-tech sectors where the wage is high or, they are bringing superior technologies to host countries and consequently are hiring more skilled or managerial workers than domestic firms.
[22] The relationship is even found to be positive in some specifications in Shaver, 1998.
[23] The corporate-tax rates can be measured in different ways. The most frequently used tax rate is the statutory tax rate (see Head and Mayer (2004), Bénassy-Quéré et al (2005)). This rate ignores actual tax practices (e.g., fiscal incentives) in different countries and MNEs' responses to them. The average tax rate based on actual firm-level tax data is also increasingly used in panel studies. However, as Devereux and Griffith (2002) point out, the average tax rate is a backward-looking rate whereas investment should be based on the effective tax rate, which is estimated using the relevant parameters at the time the real investment is made. Using this tax rate, Devereux and Griffith (1998), in a study of a panel of US firms locating in European market. find that their effective tax rate has a negative impact.
[24] Many European governments target high-tech industries so that they can avail of more tax incentives, or equivalently, have lower tax burdens than the rest of industry. Unfortunately there are no papers in the existing literature that control for this effect.
[25] For example, the impact of taxation on FDI will depend on market potential, the agglomeration of business activities and labour-market conditions, and these are taken account of in very different ways in the various studies, reflecting their particular focus.

Firm-level heterogeneity is also found to be important in MNEs' location-choice decisions, and is reflected in the determinants of location choices for MNE parents from different countries (or continents). For example, Basile et al. (2003) find that, when investing in EU countries, US MNEs treat regions within a country as equivalent substitutes, whereas EU MNE location decisions respond to regional differences.[26]

3. Empirical methods and Data

3.1 The Conditional logit and the Mixed logit models

Since the purpose of this paper is to study the impact of a set of country-level characteristics on alternative countries' probability of being chosen by MNEs to expand existing production facilities or to set up new subsidiary, we first look at the *Conditional logit model* (CLM). McFadden(1974) models discrete choice in terms of an individual *i (i=1,...., I)* making a choice among *J* alternatives to maximize his/her perceived utility (U) conditional on the characteristics (X_{ij}) of each alternative. Therefore, the perceived utility generated by individual *i* from choosing alternative *j* can be expressed as

(1) $$U_{ij} = X_{ij}\beta + \varepsilon_{ij},$$

where X_{ij} is a vector of alternative *j*'s characteristics, which are observable to the individual as well as to researchers; β is a vector of coefficients measuring the influence of these characteristics on the individual's utility and ε_{ij} is the unobservable random element of the individual's utility.[27]

McFadden proposed that, if (and only if) the random element follows a type I extreme value distribution, independently and identically across *J* alternatives and *I* individuals, the probability of alternative *k* being chosen over other alternatives can be expressed as

(2) $$\Pr(y = k \mid 1,..., J) = \frac{e^{X_{ik}\beta + \varepsilon_{ik}}}{\sum_{j=1}^{J} e^{X_{ij}\beta + \varepsilon_{ij}}}.$$

[26] Similarly, Hogenbirk and Narula (2004) find US and Japanese MNEs prefer the Randstad region in the Netherlands, while EU MNEs prefer the regions which border other EU countries.

[27] ε_{ij} captures the unique taste of individual *i* to the alternative's characteristics and the contribution of any unobservable alternative characteristics to the individual's utility.

Independent and identical distribution of the random element gives CLM that property that the ratio of probabilities of any two alternatives being chosen is independent on any other alternatives. This *Independence from Irrelevant Alternatives* (IIA) property implies that all alternatives should be perfectly substitutable to the individual, after controlling for all observable characteristics. However, in the real world, the IIA assumption is often violated because a subgroup among the alternatives may share some common features that are not shared by the rest of the group. For example, in the context of the EU, regions within one country are generally more similar to each other than they are to regions in another country.[28]

A recent development in the discrete-choice model family is the *Mixed logit model* (MXL) in Train (2003). This model assumes that the unobserved random element in the utility function (eq.1) follows a distribution that can take any form.[29] MXL can be derived from a utility-maximization frame in a "random coefficients" specification,

(3) $U_{ij} = X_{ij}\beta_i + \varepsilon_{ij}$,

where β_i is a vector of coefficients for the alternative-specific characteristics X_{ij} associated with individual *i*, and ε_{ij} is a unobserved error term following IID extreme value distribution. β_i is assumed to distribute randomly across all individuals and has a density function of $f(\beta)$. The probability of alternative *k* being chosen by individual *i* over *J* alternatives is an integral of the ratio of utility derived from alternative *k* to the sum of utilities derived from all alternatives over the density function of β_i:

(4) $$\Pr_i(y = k \mid 1,...,J) = \int \frac{e^{X_{ik}\beta_i}}{\sum_{j=1}^{J} e^{X_{ij}\beta_i}} f(\beta)\, d\beta .$$

β_i reflects the idiosyncratic tastes of individuals to each alternative and is usually assumed to have a normal distribution form (see Revelt and Train (1998) and Ben-Akiva and Bolduc (1996)). Therefore, the density function can be expressed as $f(\beta \mid \mu, \sigma)$, where μ is the mean of the normal distribution and σ is the standard deviation of that distribution. These two parameters have to be estimated to evaluate

[28] This means, for example, that a region of France is not perfectly substitutable for a region of UK, when all observable country features are taken into account.
[29] In effect, CLM is a specific case of MXL where restraints are imposed on the unobserved random element.

the effect of an alternative-specific characteristic on individuals' utilities (its magnitude as well as its variation across individuals).[30]

In this paper we will estimate the mean β and the variance σ^2 for the explanatory variables that are set as random effects to identify the heterogeneity across the pharmaceutical MNEs when they make their decisions on where to expand production capability or to locate a new subsidiary. This estimation is done by using the STATA add-in programme "GLLAMM" (see Rabe-Hesketh et al., 2004).[31]

3.2 Research design and data description

Our study is focused on the location choices of MNEs taking place between 1993 and 2003 in the pharmaceutical industry[32] in EU15 countries (excluding Luxemburg).[33] We consider MNE subsidiaries whose parents can come from any country in the world.

Two types of location choices are considered. The first type is the choice of location amongst existing high-performance subsidiaries at which production is to expand,[34] where high-performance subsidiaries are defined as those with above-median rates of output growth.[35] The second type is the choice of location for new start-up subsidiaries. We distinguish between these two types of location decisions because, while they are

[30] The mixed logit model can also be derived in an "error components" specification, which assumes that utility can be decomposed into a fixed part for all individuals and a random-error part that varies across all individuals. More details refer to Train (2003). The "random coefficients" specification and the "error components" specification reflect different perspectives a researcher may have when considering a utility-maximization discrete choice problem. When regarding the coefficients of observable alternative-specific characteristics as randomly distributed across individuals, the researcher is focusing on the taste heterogeneity among any individuals. If the researcher is considering the unobservable characteristics of all alternatives, or the error components, he or she is more interested in the different attractiveness of each alternative to the individual, i.e., substitution effects.

[31] GLLAMM stands for Generalized Linear Latent And Mixed Models and it can estimate a class of multi-level latent variable models including MXL.

[32] The pharmaceutical industry is defined according to NACE Rev.1.1 industry code at 3-digit level. The 3-digit NACE code for the pharmaceutical industry is 244 and two 4-digit codes are assigned to its sub-industries: 2441 (manufacture of basic pharmaceutical products) or 2442 (manufacture of pharmaceutical preparation).

[33] They are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, and United Kingdom.

[34] The terms subsidiary (ies) or firm (s) are used interchangeably throughout this paper since they are equivalent in the sense that all firms studied in this paper are subsidiaries.

[35] Most of the existing subsidiaries expanded their turnover at different growth rates during the last decade. It is to be expected that subsidiaries would expand at higher growth rates in the more attractive country locations, and thus to identify those countries one needs to choose high-performance subsidiaries. We use above median-growth rates of all existing pharmaceutical subsidiaries as the criterion to determine high-performance subsidiaries.

both vehicles of production relocation, they reflect different considerations for MNEs. In particular, expanding production in an existing subsidiary generally involves no sunk costs while starting up a new subsidiary requires a pharmaceutical MNE to be willing to incur significant sunk costs. Thus we study both of them separately, recognising that the two kinds of location choices may be different when faced with the same country-level characteristics, e.g., agglomeration of the pharmaceutical industry.

MNEs' subsidiary firms are identified in, and their data are extracted from, a commercial dataset "*Amadeus*", which contains accounts data on firms located in Europe.[36] A subsidiary firm may have more than one MNE shareholder, and its MNE shareholders may be inter-linked.[37] Usually *Amadeus* marks one of the MNE shareholders as the ultimate owner of the subsidiary firm. In the event that it does not, we classify, from among all MNE shareholders of a subsidiary, the MNE shareholder that has the largest share (directly, or indirectly through other subsidiaries) as its ultimate owner. By so doing, each subsidiary is linked to an ultimate MNE shareholder as its MNE parent, and all ultimate MNE parents defined by this way are independent of each other. Therefore, these clean "parent - subsidiary links" allow us to study ownership effect on MNEs' location choices (through subsidiaries) later.[38]

In order to study MNEs' location choices in terms of expansions in high-performance subsidiaries and new start-up subsidiaries, we construct two samples from the cleaned firm-level data extracted from *Amadeus* file. The year 1993 is chosen to distinguish between the existing firms and new firms, so no pharmaceutical firm exists in both

[36] Compiled by the Bureau Van Dijk, *Amadeus* collects both public and private firm accounts for 38 European countries and it is able to provide researchers with comprehensive information on increasing numbers of firms from 1992. This information covers the balance sheet, the profit and loss account, various financial ratios, the ownership data, the industry classification code, address details and the year of incorporation. Therefore, it allows us to trace a firm's birth and evolution over time.

[37] In the build-in ownership database in *Amadeus*, each firm is linked to its shareholders and the value of each shareholder's share in that firm is available.

[38] By defining an ultimate owner as having the largest share in a firm, we avoid the complication of joint ventures. According to the ownership database in *Amadeus*, only one joint-venture case where two parents have exactly 50 per cent shares each in a subsidiary is found, which is Bracco Spa, an Italian company owned equally by E.MERCK (Germany) and Brafin Finanziaria Spa (Italy). We somewhat arbitrarily treat Bracco Spa as the subsidiary of E.MERCK because E.MERCK is a leading European pharmaceutical multinational.

samples.[39] The first sample is referred to as the "high-performance sample"; it contains the 224 high-performance subsidiaries that were established before 1993 and operated between 1995 and 2003 in EU-15.[40] The second sample, the "new-firm sample", contains the 129 new subsidiaries that were established between 1993 (inclusive) and 2003 and continued to operate at least until 2003 in EU-15.[41]

In Appendix 1, Tables A1 and A2 summarize the descriptive statistics and location distribution (by parent nationality) of the high-performance firms, while Tables A3 and A4 summarize the descriptive statistics and location distribution (by parent nationality) of the new firms.[42] Comparing mean and median values for the high-performance firms, we can see that the distribution of employees, turnover and fixed assets is skewed towards larger firms; the distribution of age is skewed to older firms; and the distribution of growth rates is skewed towards fast-growing firms. For the new firms, the same features can be observed, except in the case of age which is skewed towards younger firms.

Turning to the geographic distribution of firms, we see that France, Italy, Spain and UK account for the majority of high-performance firms during the period between 1995 and 2003, while France, Spain and UK account for the majority of new firms established. We also note that, relative to land and population size, both Belgium and Ireland account for a significant number of new firms. The relatively low representation of Germany amongst high-performance firms may in part be due to the fact that German firms are underrepresented in *Amadeus*.[43]

[39] Moreover, since in *Amadeus,* firms' accounts from 1995 are better (in terms of completeness) than those in 1993, we calculate the growth rates of turnover of each existing pharmaceutical firm for the period between 1995 and 2003. Those firms with a growth rate above the median growth rate estimated from these data are defined as high-performance firms.
[40] These firms were located in just 11 of the EU-15 countries: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Portugal, Spain, Sweden, and UK.
[41] These new firms were established in just 11 of the EU-15 countries: Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Spain, Sweden, and UK.
[42] The MNE parent's nationality is decided by headquarter's location.
[43] According to the help file in *Amadeus*, "Failure to comply with disclosure requirements is not normally a punishable offence; … Consequently, only about 10-20% of 450,000 companies actually do so. There is therefore no guarantee that information will be readily available on smaller and medium-sized companies in particular."

3.3 Explanatory variables

Based on the implications of the NEG theories and related empirical research, a set of explanatory variables containing the country-level characteristics and firm-level heterogeneity was constructed to test their effects on MNEs' location choices.

3.3.1 Country-level variables

Agglomeration variables: Agglomeration of the pharmaceutical industry at country-level is the primary focus of this paper as we want to test the implications of different NEG theories regarding agglomeration. Following other studies, two variables are constructed. First we use the number of employees in the pharmaceutical industry [*PHAR* in the model] in each country to measure the agglomeration of the pharmaceutical industry. Since this measure is likely to be biased due to productivity differences across countries, we use gross output of pharmaceuticals [*PHAR2*][44] as a robustness check. Since the NEG models predict different agglomeration trends as trade costs decrease, we have no apriori sign for these variables. Second, because input-output linkages in various NEG models imply that the pharmaceutical industry may co-locate with the chemical industry (with a positive sign expected) we include the two corresponding chemical-industry agglomeration variables, namely, employment level [*CHEM*] and gross output [*CHEM2*].

Market variables: We choose the national consumption of drugs and medicines (million USD), [*CDRUG*], as a variable to proxy national market size, which is expected to have a positive sign. We also include a market-concentration variable to proxy competition [*COMP*] and to test whether competition in the pharmaceutical market at country level influences MNE's location decisions. Our measure is the share of top 25 firms on the national pharmaceutical market in terms of sales.[45] Baldwin (1995) and Barbosa (2003) suggest that in a highly concentrated market, firm size is relatively larger than that in a less concentrated market. The presence of lager firms is likely to act as a deterrent to new entrants into the market and may also act as a deterrent to the expansion of exiting producers. Therefore, we suggest that this market-concentration variable may negatively affect both kinds of location choices.

[44] It may also be biased due to price difference across countries.
[45] See Gambardella et al. (2000).

Corporate tax rate: We use the effective average tax rate [*EATR*] generated by Devereux and Griffith (2003) to measure corporate tax rates. EATR is superior to the statutory tax rate as it takes account of various financial factors that a hypothetical investment project will face in a discrete location-choice context.[46] We expect a negative effect on output expansion and new set up.[47]

Geographic variables: In the NEG theories, trade costs are conceived as a mixture of various factors that hinder the free movement of goods between countries, e.g., tariffs and quotas, non-tariff barriers, customs inefficiency, transportation costs, etc. However, it is not possible to find a simple variable that can take account of all these costs.[48] Instead, we focus on a proxy for transportation costs, namely, the Euclidean distance between countries or regions. In this paper we use distances [*DIST*] of each country's capital city from Brussels to measure trade costs from each country to access the whole EU market. Brussels is chosen as it is seen by many leading pharmaceutical MNEs as the key distribution centre in Europe.[49] We expect that the larger is the distance from one country to Brussels, the higher the costs of accessing European market are for a manufacturer from this country and thus the lower the probability of this country being chosen.[50]

Following Bartik (1985), Woodward (1992) and Hogenbirk and Narula (2004), we also include land mass [*AREA*] in the specification for the new-firm sample in order to control for the dart-board concept, i.e., the larger is the land mass of a country, the more new investments it can take if all other country-level characteristics are the same.

Labour-market conditions: To take account of national differences in labour markets, labour compensation per employee [*LCOST*] in the pharmaceutical industry is

[46] These financial factors include: the statutory tax rate, the fixed asset-deprecation rate, the interest rate and the inflation rate.
[47] Devereux and Griffith (1998) tested a panel of US MNEs' location choices across three European countries and found that the effective tax rate had a negative impact on the country's probability of being chosen.
[48] We focus on distance only on the grounds that other trade costs have been reduced by the Single Market.
[49] An industry specialist identified this institutional feature of the market.
[50] A limitation of this variable is that its accuracy depends on the assumption that pharmaceutical production within a country is near its capital city. A close look at the firm-level data used here reveals that this is true for most of the countries under investigation but less true for Austria, Germany, Italy, and Spain.

derived from the OECD STructural ANalysis (STAN) database.[51] To control for labour quality we include the percentage of workers having completed tertiary education in the total manufacturing workforce [*EDU3*]. Results in the related empirical studies, e.g., Bartik (1985), Friedman et al. (1992) etc., show the negative effect of labour costs and positive effect of labour quality. Therefore, assuming that EDU3 captures the skills component fully, we expect a negative sign of LCOST and a positive sign of EDU3.

Institutional efficiency: To capture differences in institutional efficiency across countries, we use the World Bank's Aggregate Governance Indicator [*GOV*].[52] The precise variable used is an average of scores for six sub-indicators across seven years, where higher scores mean better governance. Although none of the literature on location choice cited above uses institutional efficiency variables, political economists have found evidence that institutional efficiency (or inefficiency) is associated with FDI inflow and investment patterns for many countries (see Aizenman and Spiegel (2002) for the study on a large cross-section of countries; Smarzynska and Wei (2000) for CEEC).

Familiarity: A familiarity variable [*FAM*] is generated for each firm in the sample. It equals 1 for a country if a high-performance firm or a new firm in that country has an MNE parent located in that country or has an MNE parent (out of that country) with one or more subsidiaries already in that country.[53] Potentially, this indicator variable might capture an MNE's knowledge of a particular country's business environment, its intra-country upstream-downstream linkages, or its marketing network in that country. Familiarity is confirmed as having a positive impact on MNEs' overseas investments in Rangan (2000), which finds that MNEs invest more in 'familiar' countries. [54]

[51] Labour compensation is defined as "wages as well as the costs of supplements such as employer's compulsory pension or medical payments."(OECD, STAN Indicators 2003). This variable is seen as a better measure of real labour costs than the industry-level wage rate.
[52] This indicator contains six sub-indicators of different institutional aspects of a country for seven years from 1996. They are voice and accountability, political stability, government effectiveness, regulatory quality, rule of law and control for corruption.
[53] By this definition, FAM can be regarded as a hybrid variable at both country-level and firm-level.
[54] Familiarity of an MNE with a country may be related to the number of subsidiaries it has in that country. This relationship will be studied in a future paper.

3.3.2 Firm-level heterogeneity variables

Because of limited sample size for both the high-performance sample and the new-firm sample, we use interaction terms between the four firm-heterogeneity variables and the country-level variables to isolate the responses of different firms to various country-level characteristics instead of running regressions for the various sub-samples. The firm-heterogeneity variables include three dummy variables and a continuous variable.

Firm ownership: Two dummy variables, [*EU*] and [*US*], are created to capture the nationality of the MNE parent as being an EU MNE parent and a US MNE parent respectively.[55] In addition, a dummy variable [*TOP*] is created for those firms belonging to the top 50 global pharmaceutical MNEs, to capture the dominance of these particular MNEs in the development of the European market.[56]

Firm-level production size: An MNE's decision to increase production may take the form of an expansion in production in one or more subsidiaries or the creation of one or more new subsidiaries in countries having certain characteristics. This choice may be influence by the scale of increase proposed - for example, a large expansion might be expected to take place in countries with low corporate tax rate, or large new subsidiaries might be established in countries that are close to Brussels. If the scale of the production increase does matter in MNEs' location-choice decisions, by interacting the production size with a country-level characteristic, we can identify the effect of this characteristic on location choice associated with different levels of production size. We construct a production-size variable [*SIZE*] using firm-level data from *Amadeus*. For a high-performance firm, this variable equals the difference in its turnover between 2003 and 1995. For a new firm, it equals its turnover in the third year following the year in which that new firm was incorporated.[57]

All country-level variables and firm-heterogeneity variables are listed along with their sources and expected signs in Appendix 2.

[55] MNE parent's nationality is decided by headquarter's location.
[56] Some 17 out of the 50 top pharmaceutical MNEs in the world are US MNEs, while another 17 MNEs are European MNEs.
[57] The reason of using the third year turnover instead of the first year turnover to proxy the size of a new firm is that a firm needs time to reach its designed production capacity.

3.4 Specification of equations

All explanatory variables enter the CLM and MXL equations in logarithm form except those variables that are in percentage form, such as EATR, COMP and EDU3. The use of logarithm-form variables allows us to interpret the coefficients as elasticities,[58] while the coefficients of percentage-form variables can also be roughly interpreted as elasticities as well. For the high-performance sample, we use the average values of country-level variables for as many previous years as are available over the period 1994 and 2003.[59] For the new-firm sample, if applicable, all country-level variables are lagged one year prior to the year when a firm was established, in order to minimize the simultaneity problem and to account for the fact that decisions to locate are influenced by what happens in the period prior to the investment.

For computational reasons, PHAR/PHAR2, CHEM/CHEM2, CDRUG and EATR are set as random effects in the MXL specifications, while other explanatory variables are set as fixed effects.[60]

4. Location choice results and discussions

4.1 Model specification

We begin by estimating CLM equations for various combinations of explanatory variables to identify the "best specification" in terms of goodness of fit. To capture agglomeration, we consider the two sets of variables: the number of the employees (lnPHAR and lnCHEM) and the gross output (lnPHAR2 and lnCHEM2) for the pharmaceutical and the chemical industries respectively. Because changes in pharmaceutical production due to expansion outweigh those arising from new-firm

[58] Another aim of taking logarithm of the explanatory variables is to solve a technical problem in the regression. GLLAMM cannot handle the random effect variables that have large variation in values, e.g., across country variation in market size and pharmaceutical production are large enough to cause the problem. Although this is just the requirement for MXL, for the purpose of comparison, we use the same data in CLM.

[59] Expansion took place over the period from 1995 to 2003 and it was continuously influenced by the country-level variables and their changes in every year; therefore variables in any single year cannot capture their aggregated effects on expansion.

[60] GLLAMM utilizes numerical simulation technology to maximize the log-likelihood for MXL, which implies a heavy computation load. Generally the computation load is proportional to the sample size and increases exponentially to the number of the random effects involved in the estimation. Technical issues on the estimation and simulation methods are discussed in Train (2003) and Rabe-Hesketh et al. (2004). Further information about GLLAMM is on www.gllamm.org.

creation[61], and because the sample size of the high-performance sample is larger than that of the new-firm sample, we examine the location choice for output expansion in existing plants first, and then consider the location choice for new firms. The results of the specification tests are reported in Table 1 using the high-performance sample.[62]

Specification 1 includes four explanatory variables together with the agglomeration of the pharmaceutical industry, which is proxied by the number of employees (lnPHAR). The coefficients of lnPHAR, lnCDRUG and EATR are all statistically significant at the 1 per cent significance level, while lnDIST is not statistically significant. In Specification 2, labour costs (lnLCOST), market competition (COMP), institutional efficiency (lnGOV) and familiarity (FAM) are added to the model. The coefficients of lnPHAR, lnCDRUG and EATR do not alter much but the coefficient of lnDIST becomes positive and statistically significant. The coefficient of lnLCOST is statistically significant but the positive sign differs from what we would expect were it just picking up the effect of the cost of labour on production. The coefficient of COMP is negative as expected and statistically significant at 1 per cent level, while lnGOV shows a significant negative effect in contrast to the expected positive effect. Finally, the familiarity variable FAM shows no significant effect. In Specification 3, education level (EDU3) is introduced to control for labour quality. It shows a positive but insignificant effect and its inclusion is associated with a slight reduction in the positive lnLCOST coefficient.

We judge which specification's fitness is better based on two statistics: Pseudo R^2 and Bayesian information criterion (BIC)[63], as well as on economic rationale. The larger Pseudo R^2 or smaller BIC indicate better fitness. These two statistics are reported in Table 1 below the coefficients estimated for each specification. Specification 2 has smaller Pseudo R^2 as well as smaller BIC when compared with Specification 3, so each of the two statistics supports different specifications. However, Raftery (1995) claims "…BIC tends to choose the one with fewer parameters." and that this

[61] Firm-level data in *Amadeus* show that the change in pharmaceutical production in existing firms was roughly six times that of new firms during the period from 1993 to 2003, where new firm is defined as a firm being established after 1993 and its production value in 2003 is used for comparison.
[62] We also undertook the same specification analysis with the new-firm sample. The results are similar to those here for the high-performance firms and are not reported here.
[63] A cross-check with related research using CLM shows that a Pseudo R^2 ranging from 0.15 to 0.2 (in our paper) indicates an "average" fitness. Raftery (1995) suggests that the Bayesian information criterion is a better statistic than P-values to judge model fitness.

judgement does not depend on the theoretical considerations in favour of a particular specification. Since controlling for labour quality has strong empirical support in other studies, we use Specification 3 in preference to Specification 2.

In Specification 4 we replace the number of the pharmaceutical employees (lnPHAR) by the gross output of pharmaceuticals (lnPHAR2). We see that lnPHAR2 has a much smaller coefficient than that of lnPHAR in Specification 3. In Specification 3(2) and 4(2), the impact of the agglomeration of the chemical industry is tested by including the number of employees (lnCHEM) and the gross output (lnCHEM2) separately. Looking at the various specifications, we choose Specification 3 to estimate the model with the agglomeration of the pharmaceutical industry and Specification 3(2) for the model with the agglomeration of the chemical industry.[64]

4.2 Interpretation of coefficients in discrete-choice models

In interpreting the coefficients in discrete-choice models we must account for the probability not being a linear function of the independent variables (see Equation 2 in Section 3.1). The coefficients of variables that are in logarithm form can be interpreted as elasticities.[65] To generalize the probability elasticity for all alternative countries, we calculate the average probability elasticity following the method proposed in Head et al. (1995) and Head and Mayer (2003), who define the average probability elasticity of variable x_k for all alternative countries is b_k times one minus the average probability of a country being chosen, *Pr*, i.e.,

$$\frac{\partial \ln p}{\partial \ln x_k} = b_k(1 - Pr)$$

where b_k is the coefficient of the variable x_k, and Pr is the average probability of a country being chosen, which equals to $1/L$ where L is the number of alternative countries in the estimation. In this study, the numbers of countries turn out,

[64] Since lnPHAR and lnCHEM are highly correlated (correlation coefficient=0.9771), we do not include them together. When we combined the two variables in the model at the same time, the coefficient of lnPHAR became statistically insignificant while the coefficient of lnCHEM decreased and was only statistically significant at 10 per cent level.

[65] In other words, if the coefficient is positive (negative), a 1 per cent increase in the independent variable x_k would raise (reduce) the probability of a particular location being chosen by b_k per cent as b_k is the coefficient of independent variable x_k.

coincidently, to be eleven for both the high-performance and the new-firm samples. Therefore, we multiply the coefficients of logarithm-form variables by a parameter of 0.91 to obtain the average probability elasticity.

Since the effective average tax rate (EATR), labour quality (EDU3) and market share (COMP) are measured in percentages, they are bounded between 0 and 100 and their coefficients b_k can be roughly interpreted as 1-point increase of a variable will increase (if b_k is positive) or decrease (if b_k is negative) the probability of a country being chosen by b_k per cent.

4.3 Results for the high-performance sample

The conditional logit model

Table 2 reports the results of both CLM and MXL for the high-performance sample, using as alternative agglomeration variables the number of employees in the pharmaceutical industry or in the chemical industry. In Column 2 the number of employees in the pharmaceutical industry (lnPHAR) shows a significantly negative effect, supporting the prediction of Puga's model on the industrial agglomeration in the EU, from a firm's perspective. In effect, trade costs are at such a low level that geographic distribution of the pharmaceutical industry is dispersing, with output expansion of MNEs being more likely taken place where the agglomeration rate is lower. This result is consistent with the actual dispersing trend of the pharmaceutical industry shown in Figure 3 in Section 2.2, and contrasts with most of studies on the MNEs' location choice, which find that agglomeration has a positive impact on MNEs' location choices.

The impact of the output market effects, measured by lnCDRUG and COMP, are significantly positive and negative respectively.[66] The elasticity of the effective average tax rate (EATR) is negative and significant, at approximately -0.1. Our only geographic variable, lnDIST shows no significant effect,[67] while the variable for labour costs (lnLCOST) is found having a significant positive impact on location

[66] This latter result suggests that a highly-concentrated market is associated with larger firms and these incumbents are more capable of deterring other existing firms to expand production in this market.
[67] This is consistent with pharmaceuticals being high-value products by their nature, so the location-choice decision is not heavily influenced by transportation costs.

choice, even after controlling for labour quality (EDU3).[68] This strong positive effect, while opposite to what one would generally expect, is consistent with signalling role of labour costs as a indicator of quality and knowledge-intensive nature of the pharmaceutical industry, i.e., MNEs are willing pay high wages on the understanding that the higher pay secures better quality workers.

The institutional efficiency (lnGOV) has a strong and unexpected negative effect, suggesting that something hindering pharmaceutical MNEs' development is being captured by this variable.[69] The familiarity variable (FAM) shows no effect on the location choices, which suggests that MNEs in the pharmaceutical sector are showing no sign of geographic consolidation. This result helps to explain the dispersion trend in the European pharmaceutical industry and is also in line with the effect of agglomeration (lnPHAR and lnCHEM) on location choice.

In Column 3, the replacement of lnCHEM with lnPHAR indicates that both agglomeration variables affect location choice in a similar way. This result is obviously contrary to the implications of input-output linkages in Venables (1996) or Puga (1999)'s models. One possible explanation could be that the agglomeration of the chemical industry captures, as the institutional efficiency variable perhaps does, higher costs (e.g., congestion) in the previously agglomerated countries. If the Single European Market means that the European chemical industry has begun to move out of the previously agglomerated countries during the last decade, then this might be expected to impact on any component of this industry (including pharmaceuticals). This issue merits further investigation. Coefficients of all other country-level variables have similar magnitude and signs as those in Column 2.

To account for firm heterogeneity, we introduce firm ownership and size effects into the model by interacting the ownership dummies or size variables with each of nine explanatory variables, in order to identify the heterogeneous responses of firms to various country-level characteristics. Because the interaction terms are highly

[68] However, labour quality does not have a significant effect, which may mean that the labour-quality variable, as measured by the share of tertiary education level in the workforce, is not a good proxy for labour quality in the pharmaceutical sector.

[69] For example, better governance in a country may be associated with a higher level of development, which in turn may be associated with higher congestion and regulatory costs, which may drive production to a lower cost country.

correlated with each other,[70] each interaction term enters the model separately. Only the statistically significant coefficients of the interaction terms are reported in Tables 3 and 4 separately, along with the major effects of corresponding country-level variables.

Looking at the major and marginal effects in Table 3 in the context of EU MNEs, we find that, in their location choice, they respond relatively more negatively to industrial agglomeration, are less positively influenced by local market size, and are slightly less affected by higher labour costs. By contrast with the result in Table 2, the major effect of the familiarity variable (FAM) in Table 3 is positive and significant, indicating that production expansion follows existing output. However, we find that the marginal effect of the interaction term is negative and highly significant, suggesting that EU MNEs are not expanding production in countries where they already have subsidiaries, in contrast with non-EU MNEs.[71] This result, distinguishing between non-EU and EU MNEs, throws important light on the different processes taking place in the EU following the creation of the Single Market.

The results in Column 4 suggest that the top 50 global pharmaceutical MNEs are relatively less concerned with labour quality, while Column 5 shows how MNEs' location choices of expansion are related to size, with significant but very small marginal effects which reinforce the major effect in the case of labour costs, but offset them in the case of the distance from Brussels, market competition and institutional efficiency.

We repeated the same analysis for firm-level heterogeneity using the agglomeration in the chemical industry, as measured by employment levels. The results are reported in Table 4, which shows similar heterogeneous effects for country-level variables.

The mixed logit model

Employing MXL permits us to control for (as well to estimate) firm-level heterogeneity thoroughly by allowing the effects of country-level characteristics to

[70] Usually the coefficient of correlation for each pair of interaction terms is higher than 0.8, which is seen by econometricians as a signal of collinearity.

[71] In effect, non-EU investors are attracted to those countries where they have an existing business presence, a result that is consistent with the positive marginal effect of the familiarity variable to US MNEs (Column 3).

vary randomly across all firms in the sample. Table 2 reports the MXL results for agglomeration in the pharmaceutical industry (Column 4) and agglomeration in the chemical industry (Column 5). For the four variables that are set as random effects (lnPHAR, lnCHEM, lnCDRUG and EATR), MXL estimated stronger effects than those estimated by CLM.[72] Log-likelihoods generated in the two MXLs are slightly larger than those generated by the corresponding CLMs, which suggest that the MXL model exhibits slightly better fitness than the CLM .

4.3 Results for the new-firm sample

The conditional logit model

The empirical analysis for the new-firm sample uses the same specifications as for the high-performance sample. In addition, a land mass variable [*AREA*] is added to the model to capture the "dart-board" effect – a large country can receive proportionally more new investments than a small country can, if the two countries are identical in all other respects. Results from the CLM analysis are reported in Columns 2 and 3 in Table 5.

Fewer of the explanatory variables are found to have significant effects in the new-firm sample than in the high-performance sample.[73] In Column 2, the agglomeration variable is statistically insignificant, i.e., does not support the predictions of Puga's model. The coefficients of market sizes (lnCDRUG), the tax rates (EATR), labour costs (lnLCOST), labour quality (EDU3), market competition (COMP), and institutional efficiency (lnGOV) are also statistically insignificant, in contrast with the results for the high-performance sample. However, distance from Brussels (lnDIST) is found to have the expected negative coefficient and it is statistically significant. The variable for land mass (lnAREA) shows a positive effect, although it is only significant at the 10 per cent level. The familiarity variable (FAM) has a strong and statistically significant effect, suggesting that perhaps risk aversion plays a strong role in the decision-making process for new subsidiaries. The results when we use lnCHEM in place of lnPHAR generates similar results (Column 3).

[72] Cross-firm variation of four random effect variables are estimated by MXL as variances, which are reported in Table 2 following the coefficients with their standard errors.
[73] That fewer variables having significant coefficients for the new-firm compared with the high-performance sample is not the result of including land mass as one explanatory variable. Dropping the land mass, we still get the similar results.

When we took account of firm-level heterogeneity using interaction terms in CLM, we found that none of the interaction terms showed a statistically significant marginal effect, while the major effects of nine explanatory variables do not differ much from those in CLM estimated without interaction terms.[74]

The mixed logit model

Results of MXL are reported in Columns 4 and 5 in Table 5. The four random effect variables (lnPHAR in Column 4, lnCHEM in Column 5, lnCDRUG and EATR in Columns 4 and 5) do not have statistically significant coefficients. These results are very similar to the results of CLM (Columns 2 and 3). This outcome is unsurprising given that no heterogeneous responses from different MNEs were detected using interaction terms in CLM; besides that, the estimated variances (following the coefficients) for all four random effect variables are not statistically significant from zero. Hence, it seems plausible that there is no cross-firm heterogeneity in the new-firm sample, which explains the reason for the similar performances of the two models.

4.4 Discussion

The estimated effects of all explanatory variables for both the high-performance sample and the new-firm sample using CLM and MXL are summarized in Table S1. All statistically significant effects are expressed as the average probability elasticities.

Comparison of two samples: CLM results

Empirically we find that the agglomeration variables [lnPHAR, lnCHEM] have comparatively strong effects on the location choices of the expansion in production among existing enterprises, but have no effects on the location choices of new start-up firms. Since we are controlling for the effects of local market size, the effective average tax rate, land mass and other country-level explanatory variables, it seems reasonable to conclude that the results properly capture the pure effects of agglomeration on two production-relocation channels, i.e., expansions or new start-up firms. We can also suggest that the expansion of high-performance MNEs probably

[74] Because of their similarity we do not report these results but they are available from the authors on request.

contributes to geographic dispersion of pharmaceutical production at Europe level (Figure 3 in Section 2.2), while new-firm creation by MNEs has not such effect.

In terms of the other variables, the results for two samples show that, when making decisions on where to expand production, pharmaceutical MNEs respond more to country-level characteristics than they do when they make decisions on where to start up new production facilities. This result can be understood in terms of segmentation of European pharmaceutical markets prior to 1993. As pointed out in Cecchini Report (Cecchini et al., 1988), "…, the sector (Pharmaceuticals) is highly regulated, with two areas of regulation (market registration and price controls) …Admission of new products to national markets is subject to registration procedures to … All EC countries have measures to control public expenditure on pharmaceuticals." In effect, European pharmaceutical markets were segmented in the early years in that pharmaceutical MNEs had to set up production facilities in each country in order to access local markets, no matter what the business environment was like.[75] Since the implementation of the Single Market Programme, MNEs can be more footloose than before in determining where to locate additional production in response to more/less favourable local business climates.[76] In contrast, new-firm creation after 1993 is less subject to market segmentation, so the MNEs can locate new production facilities anywhere within the EU. The empirical evidence shows that the location choices of new firms do not respond to many country specific variables with the exception of distance from Brussels (lnDIST) and land mass (lnAREA).

The familiarity variable (FAM) is important to the location choice of new firms, but does not matter for the location choices of expansion. However, using ownership dummy variables, we can isolate the effects of familiarity on either EU MNEs or US MNEs. We find that EU MNEs tend to expand production in less familiar countries, while US MNEs chose to expand production in familiar countries. This observation is consistent with foreign investors being less familiar with European countries than

[75] For a further discussion of the regulation of the European pharmaceutical industry, see Burstall (1990), Burstall et al. (1999), Gambardella et al. (2000), Danzon and Chao (2000) and Abraham and Smith (2003).

[76] In effect, they can begin to reduce production in countries with higher industrial agglomeration (negative effects of lnPHAR and lnCHEM, associated with congestion costs), higher institutional efficiency (negative effect of lnGOV, associated with congestion costs), higher tax rates (negative effect of EATR), and raise production in countries with larger markets, which is often associated with higher labour costs (positive effect of lnCDRUG and lnLCOST).

European investors, and consequently their new investments are more concentrated in countries where production is already located.

Firm-level heterogeneity and performance of CLM and MXL

Firm-level heterogeneity is introduced in the CLM by interacting firm characteristics with explanatory variables. In the high-performance sample, we found evidence of cross-firm variation in MNE's responses to various country-level characteristics, but no such evidence was found in the new-firm sample. This discrepancy suggests that firm heterogeneity is more pronounced in the high-performance sample than in the new-firm sample.

In the high-performance sample, the MXL estimates slightly stronger effects for the four random effect variables than those estimated by the CLM, while in the new-firm sample, both models estimate that random effort variables do not have effects on the location choice. How can one explain this difference in performance? CLM treats any country-level variable as having a fixed effect across MNEs, estimating only a fixed coefficient. By contrast, MXL treats any country-level variable as random effect across MNEs and estimates a mean (to depict its average effect over MNEs) and a variance (to depict the magnitude of variation across MNEs). Therefore, if any country-level variable has only fixed effect on the MNE's location choices, two models will produce the same results; if the effect of any country-level variable is randomly distributed across MNEs, MXL can fully account for the random effect but CLM only can produce biased fixed coefficient, that is to say, if significant heterogeneity exists in a sample (i.e. the high-performance sample, as captured by introducing ownership dummies in CLM, see Column 2 in Table S1), the two models should produce different results; [77] otherwise (i.e. the new-firm sample) they produce the same (or similar) results. Our conclusion on the performance comparison between the two models is that, given the possibility that one may not be able to find proper dummy variables to control for heterogeneity in samples when applying CLM, it is useful to apply MXL as a robustness check.

[77] For the high-performance sample, the estimated variances for the four random effect variables are all not significantly from zero, implying that cross-firm variation of the four random effect variables cannot be confirmed. This is contrast with the evidence from CLM that cross-firm variation exists. We accept the evidence from CLM because using interaction terms to identify cross-firm variation is a simple and reliable way.

5. Summary and conclusions

New Economic Geography theories predict two possible types of relationship between industrial agglomeration and the level of trade costs in a two-region framework. The monotonic type (Krugman, 1991 and Baldwin 1999) suggests that if trade costs are high, industries agglomerate entirely in one region, whereas when trade costs are low, industries distribute evenly across the two regions. In contrast, a non-monotonic type (Puga, 1999 and Venables 1996) shows that prohibitive or negligible trade costs lead to symmetric distribution of industries between two regions, but at intermediate levels of trade costs, complete agglomeration of industries in one region can be an equilibrium, as can partial agglomeration in any one region. The non-monotonic type implies that there are stages when, as trade costs decrease, agglomerated industries in one region may move between regions in such a way that the overall distribution of industries become more or less dispersed.

Since 1993, when the Single Market Programme effectively reduced the level of trade costs among EU Member States, a dispersed trend of production location can be observed in the European pharmaceutical industry. This outcome is consistent with the Puga-Venables non-monotonic relationship, and suggests that we should test the hypothesis that, during the past decade, the agglomeration of pharmaceutical production at country level in the EU may have negatively impacted on the location choices of pharmaceutical MNEs that were expanding production or starting up new subsidiaries.

We test this hypothesis, as well as examining the effects of a set of country-level characteristics and firm-level heterogeneity variables (other than agglomeration) on pharmaceutical MNEs' location choices, using the conditional logit and mixed logit models. Estimations are made based on two samples of MNE pharmaceutical subsidiaries in Europe. The first sample consists of subsidiary firms that experienced high levels of output expansion in the period between 1995 and 2003, and the second sample consists of only new subsidiary firms that were established after 1993.

The results show that, for MNEs wishing to expand production in their existing firms, agglomeration of the pharmaceutical industry, market competition, the corporate tax rate and institutional efficiency have negative impacts on a country's attractiveness, while market size for medicines and labour costs show positive impacts. Location decisions for establishing new firms are found to respond to fewer country-level characteristics, with only country land mass and distance from Brussels having significantly positive and negative effects respectively. This outturn is consistent with the notion that in choosing locations for new subsidiaries, MNEs are thinking on an EU-wide basis.

The results on firm-level heterogeneity show that if an MNE has one or more subsidiaries in a country, it is more likely to invest in that country again, when starting up a new firm. In the case of expansions, EU pharmaceutical MNEs are found more likely to be attracted by low labour costs, while US MNEs are more likely expand production in familiar countries, while the top 50 global pharmaceutical MNEs are less likely to expand in countries with high labour quality, often associated with high labour costs. These differences are consistent with the different levels of information available to both sets of investors.

Finally we can summarise our major findings as follows:

(i) Evidence is found to support the Puga and Venables models of a non-monotonic relationship between industrial agglomeration and trade costs;

(ii) The expansion in production at existing plants in Europe may contribute to Europe-level geographic dispersion of pharmaceutical production;

(iii) MNE's responses to country-level characteristics are found to be different in the location choices of expansion at existing plants compared with establishing new firms. We suggest this discrepancy can be explained by the segmentation of the European pharmaceutical markets prior to 1993;

(iv) The performance of the conditional clogit and mixed logit models are compared using the same data. Differences in their performance for the high-performance sample can be explained by firm-level heterogeneity. We suggest that if one cannot find effective methods to control for heterogeneity in the conditional logit model, the mixed logit model can work as a robustness check.

References

Abraham, John and Smith, Helen L. *Regulation of the Pharmaceutical Industry.* Basingstoke: Palgrave Macmillan, 2003.

Aizenman, Joshua and Spiegel Mark M. "Institutional Efficiency, Monitoring Costs, and the Investment Share of FDI." NBER Working Papers no. 9324, 2002, National Bureau of Economic Research.

Aiginger, Karl and Davies, Stephen W. "Industrial Specialisation and Geographic Concentration: Two Sides of the Same Coin? Not for the European Union." Journal of Applied Economics, Nov 2004, VII(2), pp. 231-248.

Aiginger, Karl and Pfaffermayr, Michael "The Single Market and Geographic Concentration in Europe." Review of International Economics, 2004, 12, pp.1-11.

Amiti, Mary "New Trade Theories and Industrial Location in the EU - A Survey of Evidence." *Oxford Review of Economic Policy*, 1998, 14(2), pp. 45-53.

Amiti, Mary "Specialisation Patterns in Europe." *Weltwirtschaftliches Archiv*, 1999, 135, pp. 573-593.

Baldwin, Richard E. *The Dynamics of Industrial Competition: A North American Perspective.* Cambridge, New York and Melbourne: Cambridge University Press, 1995.

Baldwin, Richard E. "Agglomeration and Endogenous Capital." *European Economic Review*, 1999, 43(2), pp. 253-280.

Baldwin, Richard E. and Krugman, Paul "Agglomeration, Integration and Tax Harmonisation." *European Economic Review*, February 2004, 48(1), pp. 1-23.

Barbosa, Natália "What Drives New Firms into an Industry? An Integrative Model of Entry." Working Paper Series no. 23, 2003, Nucleo de Investigacao em Microeconomia Aplicada, Universidade do Minho.

Barrios, Salvador; Görg, Holger and Strobl, Eric "MNEs' Location Choice, Agglomeration Economies and Public Incentives." Research Papers 33, 2002, Nottingham University.

Bartik, T. J. "Business Location Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States." *Journal of Business and Economic Statistics,* 1985, 3(1), pp. 14-22.

Basile, Roberto; Castellani, Davide and Zanfei, Antonello "Location Choices of MNE Firms in Europe: The Role of National Boundaries and EU Policy.", Economics, Mathematics & Statistics Working Paper no. 78, December 2003, University of Urbino

Békés, Gábor "Location of Manufacturing FDI in Hungary: How Important are Inter-company Relationships?" MNB Working Papers 2005/7, Magyar Nemzeti Bank (The Central Bank of Hungary).

Ben-Akiva, Moshe and Bolduc, Denis "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure." Department of Civil Engineering, Working Paper, 1996, MIT

Bénassy-Quéré, Agnès; Gobalraja, Nicolas and Trannoy, Alain "Tax Competition and Public Input." CEPII Working Paper no. 2005-08, June 2005.

Blonigen, Bruce A. "Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment." *American Economic Review*, 1997, pp. 447-465.

Brülhart, Marius "Trading Places: Industrial Specialization in the European Union." *Journal of Common Market Studies*, September 1998, 36(3), pp. 319-346.

Brülhart, Marius and Torstensson, Johan "Regional Integration, Scale Economies and Industry Location in the European Union." Centre for Economic Policy Research Paper no. 1435, July 1996.

Burstall, Michael L. *1992 and the Regulation of the Pharmaceutical Industry.* IEA Health Series, no.9, London : The IEA Health and Welfare Unit, 1990.

Burstall, Michael L.; Reuben, Bryan G. and Reuben, Anthony J. "Pricing and Reimbursement Regulation in Europe: An Update on the Industry Perspective." *Drug Information Journal*, 1999, 33, pp. 669–688.

Cecchini, Paolo; Catinat, Michel and Jacquemin, Alexis *The European Challenge: 1992: The Benefits of A Single Market.* Aldershot: Gower Publishing Company Limited, 1988.

Combes, Pierre-Philippe and Overman, Henry G. "The Spatial Distribution of Economic Activities in the European Union," in Vernon Henderson and Jacques - FranHois Thisse, eds., *Handbook of Urban and Regional Economics*. 4, Amsterdam: Elsevier-North Holland, 2004.

Crozet, Matthieu; Mayer, Thierry and Mucchielli, **Jean-Louis** "How Do Firms Agglomerate? A study of FDI in France." Technical report, 2003, TEAM, University of Paris.

Danzon, Patricia M. and Chao, Li-Wei *Prices, Competition and Regulation in Pharmaceuticals: A Cross-national Comparison.* London: Office of Health Economics, 2000.

Devereux, Michael P. and Griffith, Rachel "Taxes and the Location of Production: Evidence from a Panel of US MNEs." *Journal of Public Economics,* 1998, 68, pp. 335–367.

Devereux, Michael P. and Griffith, Rachel "The Impact of Corporate Taxation on the Location of Capital: A Review." *Swedish Economic Policy Review*, 2002, 9, pp. 79-102.

Devereux, Michael P. and Griffith, Rachel "Evaluating Tax Policy for Location Decisions." *International Tax and Public Finance*, March 2003, 10(2), pp. 107-26.

Disdier, Anne-Celia and Mayer, Thierry "How Different is Eastern Europe? Structure and Determinants of Location Choices by French Firms in Eastern and Western Europe." *Journal of Comparative Economics*, June 2004, 32(2), pp. 280-296.

Duranton, Gilles and Overman, Henry G. "Testing for Localisation Using Micro-Geographic Data." *Review of Economic Studies*, 2005, 72(4), pp. 1077-1106.

Figueiredo, Octávio; Guimaraes, Paulo and Woodward, Douglas "Home-field Advantage: Location Decisions of Portuguese Entrepreneurs." *Journal of Urban Economics*, September 2002, 52(2), pp. 341-361.

Friedman, Joseph; Gerlowski, Dan. A. and Silberman, Jon. "What Attracts Foreign MNE Corporations? Evidence from Branch Plant Location in the United States." *Journal of Regional Science,* 1992, 32(4), pp. 403-418.

Gambardella, Alfonso; Orsenigo, Luigi and Pammolli, Fabio "Global Competitiveness in Pharmaceuticals – A European Perspective." Report Prepared for the Directorate General Enterprise of the European Commission, November 2000.

Guimaraes, Paulo; Figueiredo, Octávio and Woodward, Douglas "Modeling Industrial Location Decisions in U.S. Counties." ERSA Conference Papers, 2002, ersa02p060, European Regional Science Association.

Head, Keith C. and Mayer, Thierry "Market Potential and the Location of Japanese Firms in the European Union." *Review of Economics and Statistics,* 2004, 86(4), pp. 959–972.

Head, Keith C.; Ries, John C. and Swenson, Deborah L. "Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States." *Journal of International Economics,* 1995, 38(3-4), pp. 223-247.

Head, Keith C.; Ries, John C. and Swenson, Deborah L. "Attracting Foreign Manufacturing: Investment Promotion and Agglomeration." *Regional Science and Urban Economics*, March 1999, 29(2), pp. 197-218.

Hogenbirk, Annelies and Narula, Rajneesh "Location and Agglomeration of FDI in the Netherlands: Implications for Policy." MERIT (Maastricht Economic and Social Research and Training Centre on Innovation and Technology)-Infonomics Research Memorandum series, 2004-006.

Holl, Adelheid "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain." *Regional Science and Urban Economics*, 2004, 34(3), pp. 341-363.

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

McFadden, Daniel L. "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka, ed., *Frontiers in Econometrics*, pp. 105-142, Academic Press: New York, 1974.

Midelfart-Knarvik, K.H.; Overman, H.G.; Redding, S.J. and Venables, A.J. "The Location of European Industry." *European Economy*, 2000, Economic Papers 142.

Neven, Damien and Siotis, George "Technology Sourcing and FDI in the EC: An Empirical Evaluation." *International Journal of Industrial Organization*, 1996, pp. 543-560.

Ottaviano, Gianmarco I. P. and Puga, Diego "Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'." Centre for Economic Performance Discussion Paper no. 356, August 1997, The London School of Economics and Political Science

Puga, Diego "The Rise and Fall of Regional Inequalities." *European Economic Review*, February 1999, 43(2), pp. 303-334.

Rabe-Hesketh, Sophia; Skrondal, Anders and Pickles, Andrew *GLLAMM Manual.* Division of Biostatistics Working Paper Series. Working Paper 160, 2004, U.C. Berkeley

Raftery, Adrian E. "Bayesian Model Selection in Social Research." *Sociological Methodology*, 1995, 25, pp. 111–195.

Rangan, Subramanian "Search and Deliberation in International Exchange: Microfoundations to Some Macro Patterns." *Journal of International Business Studies*, 2000, 31(2), pp. 205-222.

Revelt, David and Train, Kenneth "Mixed Logit with Repeated Choices." *Review of Economics and Statistics,* 1998, 80, pp. 647–657.

SCRIP 100 "An analysis of the Pharmaceutical Industry's Performance and Prospects." Winter 2005/2006, KPMG.

Shaver, J. Myles "Do Foreign-Owned and US-Owned Establishments Exhibit the Same Location Pattern in the US Manufacturing Industries?" *Journal of International Business Studies,* 1998, 29(3), pp. 469-492.

STAN Indicators 1980-2001, OECD, 2003

The Pharmaceutical Industry in Figures, The European Federation of Pharmaceutical Industries and Associations (EFPIA), 2005 update.

Train, Kenneth *Discrete Choice Methods with Simulation.* Cambridge, MA: Cambridge University Press, 2003.

Venables, Anthony J. "Equilibrium Location of Vertically Linked Industries." *International Economic Review,* 1996, 37, pp. 341-359.

Smarzynska, Beata K. and Wei, Shang-Jin "Corruption and Composition of Foreign Direct Investment: Firm-Level Evidence." NBER Working Paper no. 7969, October 2000, National Bureau of Economic Research

Woodward, Douglas P. "Locational Determinants of Japanese Manufacturing Start-ups in the United States." *Southern Economic Journal,* 1992, 58(3), pp. 690-708.

Yago, Glenn; Hall, Thomas and Harrington, Michael "Think Locally Act Globally: Domestic Capital Market Restructuring and Sustainable Global Economic Growth." Capital Studies Group, Milken Institute, March 8, 2000.

Table 1. Choice of Model Specifications: High-Performance Sample and CLM

Explanatory Variables	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 3(2)	Spec. 4(2)
lnPHAR	-1.395*** *(0.299)*	--2.947*** *(0.574)*	-2.583*** *(0.779)*			
lnCHEM					-2.293*** *(0.567)*	
lnPHAR2				-0.66** *(0.33)*		
lnCHEM2						-0.699** *(0.284)*
lnCDRUG	2.031*** *(0.259)*	2.337*** *(0.349)*	2.167*** *(0.429)*	1.271*** *(0.289)*	1.912*** *(0.308)*	1.255*** *(0.252)*
EATR	-0.121*** *(0.022)*	-0.099*** *(0.033)*	-0.099*** *(0.032)*	-0.109*** *(0.029)*	-0.078** *(0.032)*	-0.112*** *(0.03)*
lnDIST	-0.002 *(0.087)*	0.364* *(0.211)*	0.333 *(0.209)*	0.064 *(0.176)*	0.105 *(0.175)*	0.092 *(0.183)*
lnLCOST		3.637*** *(1.047)*	3.233*** *(1.173)*	1.763* *(0.941)*	3.206*** *(1.057)*	1.511* *(0.852)*
EDU3			0.013 *(0.021)*	0.058*** *(0.017)*	0.048*** *(0.015)*	0.057*** *(0.016)*
COMP		-0.08*** *(0.023)*	-0.073*** *(0.025)*	-0.015 *(0.021)*	-0.135*** *(0.035)*	-0.043* *(0.022)*
lnGOV		-1.618*** *(0.515)*	-1.739*** *(0.559)*	-2.09*** *(0.618)*	-2.333*** *(0.56)*	-2.108*** *(0.604)*
FAM		-0.025 *(0.19)*	-0.03 *(0.19)*	0.001 *(0.19)*	-0.051 *(0.191)*	-0.006 *(0.19)*
# of Obs.	224	224	224	224	224	224
Pseudo R^2	0.1766	0.1941	0.1945	0.1868	0.1995	0.1891
Bayesian Information Criterion	-306.057	-303.131	-298.148	-289.936	-303.566	-292.439

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; standard error in parentheses. Larger Pseudo R^2 and smaller Bayesian Information Criterion indicate better fitness.

Table 2. High-Performance Sample: CLM and MXL

Explanatory Variables	Column 2 CLM	Column 3 CLM	Column 4 MXL	Column 5 MXL
lnPHAR	-2.583*** *(0.779)*		-3.12*** *(0.949)*	
lnCHEM		-2.293*** *(0.567)*		-2.531*** *(0.676)*
lnCDRUG	2.167*** *(0.429)*	1.912*** *(0.308)*	2.632*** *(0.613)*	2.078*** *(0.372)*
EATR	-0.099*** *(0.032)*	-0.078** *(0.032)*	-0.155*** *(0.059)*	-0.077* *(0.045)*
lnDIST	0.333 *(0.204)*	0.105 *(0.175)*	0.294 *(0.208)*	0.18 *(0.204)*
lnLCOST	3.233*** *(1.173)*	3.206*** *(1.057)*	3.287*** *(1.189)*	3.706*** *(1.296)*
EDU3	0.013 *(0.021)*	0.048*** *(0.015)*	0.006 *(0.022)*	0.051*** *(0.015)*
COMP	-0.073*** *(0.025)*	-0.135*** *(0.035)*	-0.09*** *(0.03)*	-0.146*** *(0.037)*
lnGOV	-1.739*** *(0.559)*	-2.333*** *(0.56)*	-1.775*** *(0.576)*	-2.496*** *(0.63)*
FAM	-0.03 *(0.19)*	-0.051 *(0.191)*	-0.037 *(0.196)*	-0.091 *(0.197)*
Variance				
lnPHAR			1.138 *(1.601)*	
lnCHEM				0.084 *(0.311)*
lnCDRUG			1.462 *(1.604)*	0.001 *(0.032)*
EATR			0.0126 *(0.015)*	0.008 *(0.01)*
# of Obs.	224	224	224	224
Log-likelihood	-432.678	-429.969	-431.285	-429.292

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; standard error in parentheses; lnPHAR, lnCHEM, lnCDRUG and EATR are set as random effects; their variances are reported following the coefficients.

Table 3. High-Performance Sample: CLM with Interaction Terms[⊕]

Explanatory variables	Column 2 EU dummy	Column 3 US dummy	Column 4 Top dummy	Column 5 Size
lnPHAR	-2.294*** *(0.792)*			
Interaction w/ lnPHAR	-0.532** *(0.248)*			
lnCDRUG	2.432*** *(0.465)*			
Interaction w/ lnCDRUG	-0.324* *(0.196)*			
EATR				
Interaction w/ EATR				
lnDIST				0.423** *(0.212)*
Interaction w/ lnDIST				-2.67e-07** *(1.08e-07)*
lnLCOST	4.063*** *(1.288)*			3.03** *(1.198)*
Interaction w/ lnLCOST	-1.01* *(0.546)*			1.89e-06*** *(6.60e-07)*
EDU3			0.035 *(0.024)*	
Interaction w/ EDU3			-.042** *(0.021)*	
COMP				-0.087*** *(0.027)*
Interaction w/ COMP				3.73e-08* *(2.10e-08)*
lnGOV				-2.098*** *(0.582)*
Interaction w/ lnGOV				1.16e-06** *(5.27e-07)*
FAM	1.267*** *(0.386)*	-0.183 *(0.207)*		
Interaction w/ FAM	-1.81*** *(0.438)*	0.955* *(0.507)*		

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; standard error in parentheses; EU dummy equals one for a firm having a European MNE parent; US dummy equals one for a firm having US MNE parent; Top dummy equals one if a firm's parent is one of the top 50 global pharmaceutical MNEs; Size is the difference of turnover between 2003 and 1995 for a high-performance firm.

[⊕] Only significant marginal effects are reported; agglomeration is measured using lnPHAR.

Table 4. High-Performance Sample: CLM with Interaction Terms[⊕]

Explanatory variables	Column 2 EU dummy	Column 3 US dummy	Column 4 Top dummy	Column 5 Size
lnCHEM	-1.978*** *(0.58)*			
Interaction w/ lnCHEM	-0.579** *(0.247)*			
lnCDRUG	2.168*** *(0.349)*			
Interaction w/ lnCDRUG	-0.328* *(0.196)*			
EATR				
Interaction w/ EATR				
lnDIST				0.2 *(0.181)*
Interaction w/ lnDIST				-2.63e-07** *(1.07e-07)*
lnLCOST	4.112*** *(1.194)*			3.041*** *(1.07)*
Interaction w/ lnLCOST	-1.077* *(0.563)*			1.98e-06*** *(6.77e-07)*
EDU3			0.069*** *(0.019)*	
Interaction w/ EDU3			-0.041** *(0.02)*	
COMP				-0.149*** *(0.035)*
Interaction w/ COMP				3.73e-08* *(2.11e-08)*
lnGOV				-2.73*** *(0.587)*
Interaction w/ lnGOV				1.24e-06** *(5.43e-07)*
FAM	1.265*** *(0.388)*	-0.207 *(0.207)*		
Interaction w/ FAM	-1.834*** *(0.44)*	0.981* *(0.512)*		

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; standard error in parentheses; EU dummy equals one for a firm having a European MNE parent; US dummy equals one for a firm having US MNE parent; Top dummy equals one if a firm's parent is one of the top 50 global pharmaceutical MNEs; Size is the difference of turnover between 2003 and 1995 for an high-performance firm.

[⊕] Only significant marginal effects are reported; agglomeration is measured using lnCHEM.

Table 5. New-Firm Sample: CLM and MXL

Explanatory Variables	Column 2 CLM	Column 3 CLM	Column 4 MXL	Column 5 MXL
lnPHAR	0.609 *(0.487)*		0.618 *(0.827)*	
lnCHEM		0.828 *(0.713)*		-0.126 *(1.124)*
lnCDRUG	-0.748 *(0.719)*	-1.134 *(0.942)*	-0.742 *(1.07)*	0.307 *(1.613)*
EATR	-0.012 *(0.028)*	-0.012 *(0.028)*	-0.015 *(0.03)*	-0.01 *(0.033)*
lnDIST	-1.038** *(0.438)*	-1.024** *(0.442)*	-0.879* *(0.529)*	-0.551 *(0.632)*
lnAREA	1.032* *(0.566)*	1.217** *(0.603)*	0.964 *(0.653)*	0.589 *(0.886)*
lnLCOST	-0.355 *(1.153)*	0.289 *(1.203)*	0.587 *(1.546)*	0.285 *(1.263)*
EDU3	0.033 *(0.04)*	0.023 *(0.04)*	0.051 *(0.047)*	0.073 *(0.06)*
COMP	-0.016 *(0.038)*	-0.008 *(0.041)*	-0.019 *(0.042)*	-0.0005 *(0.044)*
lnGOV	-0.985 *(0.821)*	-1.55 *(1.086)*	-1.545 *(1.327)*	-0.415 *(1.555)*
FAM	1.143*** *(0.245)*	1.155*** *(0.245)*	1.146*** *(0.257)*	1.088*** *(0.258)*
Variance				
lnPHAR			2.075 *(3.02)*	
lnCHEM				2.524 *(4.361)*
lnCDRUG			1.199 *(1.745)*	1.128 *(1.875)*
EATR			0.001 *(0.003)*	0.01 *(0.015)*
# of Obs.	129	129	129	129
Log-likelihood	-247.609	-247.707	-246.890	-246.762

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; Standard Error in parentheses; lnPHAR, lnCHEM, lnCDRUG and EATR are set as random effects; their variances are reported following the coefficients.

Table S1. Comparison of Results between the High-Performance Sample and the New-Firm Sample.

	The High-Performance Sample		The New-Firm Sample	
	Column 2 CLM	Column 3 MXL	Column 4 CLM	Column 5 MXL
lnPHAR	-2.4***	-2.8***	no effect	no effect
lnCHEM	-2.0***	-2.4***	no effect	no effect
lnCDRUG	2.0***	2.5***	no effect	no effect
EATR	-0.1***	-0.2**	no effect	no effect
lnDIST	no effect	no effect	-0.9**	-0.8*
lnAREA	na	na	0.9*	no effect
lnLCOST	2.9***	3.0***	no effect	no effect
EDU3	no effect	no effect	no effect	no effect
COMP	-0.1***	-0.1***	no effect	no effect
lnGOV	-1.6***	-1.6***	no effect	no effect
FAM	no effect	no effect	1.1***	1.1***
EU parents	Do not prefer familiar location, high labour costs and large market	na	no variation	na
US parents	Prefer familiar location	na	no variation	na
Top MNE	Do not prefer high labour quality	na	no variation	na
Size	negligible variation	na	no variation	na

Note: * significant at 10 per cent level, ** significant at 5 per cent level, *** significant at 1 per cent level; Numbers in the table are the average probability elasticity that can be interpret as one per cent increase in the variable leads to certain per cent increase (if sign is +) or certain per cent decrease (if sign is -) in the probability of a country being chosen, except for EATR, EDU3 and COMP, it should be read as one percentage point increase of the variable leads to certain per cent increase (if sign is +) or certain per cent decrease (if sign is -) in the probability of a country being chosen; no effect: the corresponding coefficient is not statistically significant from zero; na: not applicable; no variation: the coefficient of interaction term with the country-level characteristic is not statistically significant from zero; negligible variation: the coefficient of the interaction term is far smaller than that of the corresponding major effect.

Appendix 1. Descriptive Statistics and the Locational Distribution of Two Samples

A1: Summary Statistics for the High-Performance Sample

Existing Pharmaceutical Firms Experiencing Above-median Growth in Terms of Turnover between 1995 and 2003, Value in 2003, Number=224

Variable	Mean	Median	Std.Dev.	Min	Max
No. of Employees	736.2	281	1,275.7	11	10,076
Turnover (thousand USD)	465,977.4	98,443	948,096.3	2,523	6,669,416
Fixed Assets (thousand USD)	295,857.9	25,834	1,581,229.4	8	18,724,261
Age (to 1993)	30.8	26	21.8	3	122
Growth Rate of Turnover (Ratio of Turnover in 2003 to Turnover in 1995)	23.1	3.2	180.7	2.1	2601

A2: Geographic Distribution of the High-Performance Sample (by Ownership)

Nationality \ Location	Belgium	Denmark	France	Germany	Greece	Ireland	Italy	Portugal	Spain	Sweden	Great Britain	Sum (Share)
EU MNE Parent	6	4	46	5	4	2	23	7	40	2	12	151 (67)
US MNE Parent	3	0	10	2	1	0	9	1	8	0	13	47 (21)
Other Non-EU MNE Parent	1	0	8	1	0	1	9	0	3	0	3	26 (12)
Sum (Share)	10 (4.5)	4 (1.8)	64 (28.6)	8 (3.6)	5 (2.2)	3 (1.3)	41 (18.3)	8 (3.6)	51 (22.8)	2 (0.9)	28 (12.5)	224 (100)

A3: Summary Statistics for the New-Firm Sample

New Pharmaceutical Firms Established between 1995 and 2003, Value in 2003, Number=129

Variable	Mean	Median	Std.Dev.	Min	Max
No. of Employees	304	152	449	14	2,960
Turnover (thousand USD)	147,373	44,933	407,725	1	3,304,014
Fixed Assets (thousand USD)	44,046	12,566	107,456	0	841,936
Age (to 2003)	5.6	6	3.0	1	11

A4: Geographic Distribution of the New-Firm Sample (by Ownership)

Nationality \ Location	Austria	Belgium	Denmark	Finland	France	Germany	Ireland	Italy	Spain	Sweden	Great Britain	Sum (Share)
EU MNE Parent	1	6	2	3	29	5	4	4	11	4	7	76 (59)
US MNE Parent	0	4	0	0	9	2	3	2	0	1	4	25 (19)
Other Non-EU MNE Parent	0	2	1	1	9	2	2	0	3	0	8	28 (22)
Sum (Share)	1 (0.8)	12 (9.3)	3 (2.3)	4 (3.1)	47 (36.4)	9 (7.0)	9 (7.0)	6 (4.7)	14 (10.9)	5 (3.9)	19 (14.7)	129 (100)

Appendix 2. List of Country-Level Variables and Firm-Heterogeneity Variables

Variable	Description	Expected sign	Source
PHAR	Number of employees in the pharmaceutical industry (100 persons)	?	OECD STAN industry data
PHAR2	Gross output in the pharmaceutical industry (million euros, deflated to base year 1994)	?	OECD STAN industry data
CHEM	Number of employees in the chemical industry (100 persons)	+	OECD STAN industry data
CHEM2	Gross output in the chemical industry (million euros, deflated to base year 1994)	+	OECD STAN industry data
CDRUG	National consumption of drugs and medicines (millions USD, deflated to base year 1994)	+	OECD Health Data
EATR	National effective average tax rate (per cent)	-	The Institute for Fiscal Studies
DIST	Geographic distance from capital city to Brussels (km)	-	CEPII's dyadic
AREA	Land mass (sq. km) excluding water area	+	CIA World Factbook 2006
LCOST	National labour compensation per worker in the pharmaceutical industry (euros)	-	OECD STAN industry data
EDU3	National share of workers with a tertiary level education in manufacturing workforce (per cent)	+	Eurostat
COMP	Market share of top 25 corporate in national pharmaceutical market (not including the Netherlands)	-	IMS International
GOV	The Governance indicator, the higher the score is, the better governance is.	+	World Bank
FAM	Dummy variable =1 if for a firm, there are at least one other firm from the same MNE existed in a country.	+	Ownership Database in *Amadeus*
EU	Dummy variable=1 if EU MNE		*Amadeus*
US	Dummy variable= if US MNE		*Amadeus*
TOP	Dummy variable=1 if the top 50 global pharmaceutical MNEs (2004 rank by sales)		SCRIP 100 (2005/2006), compiled by KPMG
SIZE	Turnover change between 1995 and 2003 for the firms in the high-performance sample. Turnover in the third year after birth for the firms in the new-firm sample.		*Amadeus*

Institute for International Integration Studies
The Sutherland Centre, Trinity College Dublin, Dublin 2, Ireland