Location of manufacturing FDI in Hungary: How important are business-to-business relationships?*

Gábor Békés
Central European University
Budapest, Nádor utca 7., Hungary

June 23, 2004

Abstract

Contributing to the new economic geography literature, this paper sets up a simple model with monopolistic competition and estimates the determinants of location choice. This research addresses three central questions: Is there an agglomeration effect strong enough to explain co-location? Does market access matter for the location of foreign investment even within a small country? Are input-output linkages the key motive for location? In this paper, I apply a discrete choice methodology to find out to what extent various factors like wages or market access influence location choice within one country. I use two detailed datasets of Hungarian firms and wages between 1992 and 2001 considering all the new-born companies in manufacturing with foreign ownership.

JEL classification: F23, R3, R55

Keywords: economic geography, industrial location, FDI, regional policy, discrete choice models

1 Introduction

Over the past twenty years, study of international location of production and trade has gone through a remarkable development. New trade theory in the eighties, and economic geography in the nineties offered new modelling techniques and economic explanations. Geographical location was placed in the centre of thinking as a new breed of models on urban, regional and international economics of agglomeration were developed.

Economic geography, simply put, is "all about where economic activity takes place - and why" (Fujita, Krugman and Venables [1999, p.14.]). Models aim at explaining cross regional agglomeration patterns amid transaction

*This is work in progress and is not to be quoted. For comments and suggestions I thank Gianmarco Ottaviano, Laszlo Halpern, Fabrice Defever and Almos Telegdy.

costs of doing interregional business. The subject has relevance both at a firm level and at macro level. At the firm level, it gives arguments for location decisions including moving into a foreign region, following peers and business partners or being a pioneer in an unknown area. At a macro level it helps understanding dissimilar regional development patterns or causes of backwardness. It also helps see what happens when regions come closer as market integration proceeds or new motorways are built. Importantly, it may give new consideration for policymakers when deciding on regional development policy.

In Central and Eastern Europe rapid changes and restructuring in manufacturing have taken place since 1990. Thus, countries like Hungary offer a laboratory experiment to study the geographic properties of a large number of new investments by firms entering a region previously closed to foreigners. As for the development, foreign direct investment was instrumental in transforming the industry of transition economies. In Hungary, for example, the stock of FDI reached 40% of GDP by 2003. The rapid appearance of foreign-owned manufacturing sites offers a great opportunity: studying the geographic properties of a large number of new firms entering a region previously closed to foreigners.

In this paper I will consider foreign direct investments in Hungary and investigate why the presence of a firm has an impact on investment by another. This research addresses three central questions: Is there an agglomeration effect strong enough to explain co-location? Within a small country does market access matter for the location of foreign investment? Are input-output linkages the key motive for location? In this study, I apply a discrete choice methodology to find out to what extent factors like wages or input-output linkages influence location choice within one country. I use two detailed datasets of Hungarian firms and wages between 1992 and 2001 and I consider all the new-born manufacturing companies with foreign ownership.

The paper is organised as follows. First I summarize the related literature analysing results of firm location in general and FDI location specifically. Second, I present a small model of location choice. Third I report key findings about location choice within a country, describe the empirical method and the data to be used. Finally, the results of the empirical investigation are presented and a few points for economic policy are made. For this paper is a work in progress, ideas for future research are noted, too.

2 Related literature

2.1 The economic geography contribution

With recent developments in new economic geography (or NEG) modelling, location theory experienced a marked revival. Over the past two centuries, from von Thünen through Marshall to Krugman, serious efforts have been invested into studying locational patterns of firms. The key trade-off firms have to bear in mind was established by von Thünen as early as 1826: being close to customers versus being close to the source of inputs. Further, the fact that the transportation cost is of paramount importance was laid down a century ago. Also, the key idea that firm location depends on the proximity of demand was introduced a long ago by Harris [1954] who devised the simplest aggregate market-potential function.

2.1.1 Basic intuition

By the neoclassical model of economics textbook, economic activity is spread out evenly through space since the flow of production factors levels out differences in development and prices alike. Wherever there is a scarcity in one good or factor, its relative price will be higher making it worthwhile to ship goods from other places in the world as long as prices are equalized. Equalisation may be reached via trade and/or capital investment and labour migration. It is easy to see that this is not the case in reality: there is a concentration of activity in cities, industrial or financial centres, and there is a marked difference between developed and underdeveloped regions even within one country. There are many reasons for the concentration of production (i.e. marked co-location of firms) and models of new economic geography aim at uncovering the essential reasons behind both agglomeration and dispersion of economic activity.

Let us give here a bit of economic intuition that lies behind these theories. Most of the models assume that firms produce with an increasing returns to scale technology, market transactions are costly and these costs determine whether firms benefit from settling close to one another thereby giving rise to agglomerations. In the lack of transaction (trade) costs, production would be determined by supply side consideration (such as efficient scale size) only. However, if transportation is costly, demand side becomes a determining factor of location choice given that being close to customers yields lower operating costs. Accordingly, a shift in transaction costs may lead to relocation of industries as both optimal level of concentration and optimal distance from customers is altered.

To better grasp the key ideas of the new economic geography, let us consider a simple framework with two regions (e.g. as in Fujita, Krugman and Venables [1999, Chapter 4]). Firms can decide whether to settle in one region, the other one or in both. Let us see what are the forces in the economy that determine concentration or dispersion of firms and their activity.

Let us start with one region having slightly more firms than the other. The more firms are present in a region the more easily can they find the required intermediate goods locally. Hence, there is a lower import share and saving in transport costs will make final prices lower, too. Greater competition among firms will also lead to higher wages that, along with lower prices help raise living standards. Better prospects will magnet migrants from the other region and the labour pool will rise. This will lower wages to some extent but the size of the market will rise thus helping firms to sell more allowing to lower prices. Also, greater market (more customers locally and the possibility to make an even better use of increasing returns to scale) will make new firms enter the region. Thus, in this case labour market development and capital flows reinforce each other: efficiency of production and stronger purchasing power of customers will offset rising wages and agglomeration forces lead to a growing concentration of activity in one region. This is what Nobel-laureate Gunnar Myrdal dubbed "cumulative causation" (Myrdal [1957])

Of course, agglomeration forces do not prevail without boundaries, there are dispersion forces in action, too. First and foremost, high wages will make certain wage-sensitive industries incapable to offset rising costs. These companies will at some point opt to locate in the other region. Although they will face much higher transaction costs when selling to the larger (and richer) region, but production costs will be much lower in the other region. Another reason to move is falling final prices as a result of greater competition. In this case, benefits of lower competition in the other region will offset disadvantages of loosing suppliers and some customers in the larger region. As we have seen, the size of transaction costs and thus the distance between markets plays a pivotal role. Note, that remoteness does not only incorporate physical distance " as the crow flies" but also the quality of transport network, language and cultural barriers, differences in corporate management styles or regulatory environment.

An excellent survey of key hypotheses emerging from models of new economic geography and their mixed empirical support can be found in Head and Mayer [2003*a*].

2.1.2 Focus on location choice

Over the past few decades there has been a renewed interest in location theory. Indeed, in the age of cheap transportation costs and complex production methods as well as business to business relationships, it emerged that "new firms have a high propensity to settle at places where economic activities are already established" as posited in Ottaviano, Tabuchi and Thisse [2003, p.7.].

New economic geography models have been employed to explain location of overall economic activity by Krugman [1991], industrial clusters in Fujita, Krugman and Venables [1999, Ch. 16.], location of various manufacturing sectors by Midelfart-Knarvik, Overman and Venables [2000]or individual sectors such as those of accordions by Robert-Nicaud [2002, Ch. 5]. Further, a multi-regional setting was soon applied to the NEG models. Krugman extended his earlier model to have multiple regions and the core-periphery model was extended in the Fujita, Krugman and Venables [1999, Ch. 16.] volume to have three, four, or many regions.

The notion that inter-company sales should be taken into account is more recent but input-output (I-O) linkages were incorporated into the new economic geography models about a decade ago. In the literature there are two distinct ways to introduce I-O linkages. Venables [1996] posits that there are downstream and upstream industries *per se* and there is a flow of goods between these two. The other stream, initiated by Krugman and Venables [1995] considers a single imperfectly competitive industry where the final good of one firm is the intermediate good of an other. Here, the purchase of intermediate goods enters the cost function via a composite price index of such products, while demand is augmented with corporate purchasing power. Input-output linkages appeared in some recent work by Ottaviano and Robert-Nicaud [2003] who compared the theoretical as well as welfare implications of I-O models in comparison with labour migration.

In the vertical linkage models centrifugal and centripetal forces lead to the emergence of two important externalities. First, when a firm enters a region and starts production, it also increases the demand for upstream activities thus expanding the home market. Second, it also increases local supply of downstream output, leading to the so called market crowding effect. These two forces work against each other, and agglomeration takes place when market expansion effect dominates the market crowding effect.

There are other drivers of industrial clustering. One such reason that makes worth locating close to one another is the potential of knowledge spillover. This is true for human as well as physical capital. The attraction to work close to other people is noted in Marshall [1920] and the importance of face to face communication is discussed in Leamer and Stolper [2000]. As for firms, proximity allows to exchange inventions while technology spillovers help increase productivity using other firms' knowledge.

Another such agglomeration force is labour pooling: firms enjoy the presence of a larger set of labour where the specific knowledge required by the firm, may just be fished out easily Amiti and Pissarides [2001]. In a transition economy, one reason for settling where the old industrial base used to operate may be the presence of such labour pool.

Market potential has been first investigated in an international context; proximity to key markets and suppliers have explicitly featured in empirical works explaining overall economic activity or per capita income. Redding and Venables [2001] argues that a country's wage level (proxied by per capita income) is dependent on its capacity to reach export markets and manage to get hold of the necessary intermediate goods cheaply. For the European Union, Head and Mayer [2003*b*].look at Japanese investments. Results show that apart from a very important market potential measure, a number of traditional explanatory variables (e.g. taxes) and agglomeration variables turn to be significant.

2.2 Location of FDI within a country

Location of industry and more specifically that of the foreign-owned manufacturing has long been in the limelight of economic research. Mainstream international economics focused on the interaction of direct investment and international trade. The interest in intra-national location choice, inspired by the marriage of international and urban/regional economics, is more recent.

Various studies have researched location choice of foreign investors, mostly in manufacturing, using some version of a discrete choice model. An inspiring piece is Crozet, Mayer and Mucchielli [2003] who study location of FDI in France. They use a simple model of oligopolistic competition and a conditional logit model to simulate corporate choice of location. They find that firms of the same nationality like to group together, locations close to home country are chosen more frequently, and some industries (like car plants) have a strong tendency to agglomerate. The authors also find that firms of the same nationality like to settle close to one another. Similarly, a study by Head and Ries [1999] looks at Japanese investments in the US and finds that firms belonging to the same *keiritsu* tend to settle close to each other.

Agglomeration has been indeed found to be an important determinant of location in developed countries. Coughlin and Segev [2000] looked at the geographic features of employment of newly established foreign owned plants in US states. They found that "size, labour force quality, agglomeration and urbanization economies, and transportation infrastructure are found to affect positively the location of new foreign-owned plants, while unit labour costs and taxes are found to deter new plants". For example the agglomeration effect was proved to be especially important in explaining the attractiveness of the Southeast region. In a similar study, Coughlin and Segev [1999] considered FDI in China finding location to be important (proximity to the Coast) but transportation infrastructure (thus costs) to have a negligible impact only. Given the concentration of FDI in regions that contain the preferential zones created in the early eighties, the "history matters" argument (i.e. in the presence of virtuous circles, a small difference between regions may give an ever increasing advantage to the slightly better one) found some support, too.

Barrios, Strobl and Gorg [2002] look at multinationals' location choice in Ireland with special interest in the role of agglomeration forces as well as state support. They find that agglomeration forces contributed substantially to location choices but proximity to major ports and airports was also helpful. More importantly they find evidence that higher public incentives in designated areas have increased the probability of multinational investment. According to the results, regional policy has been effective in attracting low-tech multinationals to the designated areas.

Anecdotal evidence confirms that agglomeration forces are active in transition economies of Central and Eastern Europe (CEE). The presence of industrial clusters is an easy to spot feature of new manufacturing base in the region, including the motor vehicle cluster in North-West Hungary, West Slovakia or South-West of the Czech Republic. Also, there is some evidence showing that large multinationals lured in their usual suppliers[1].

Results with data on developing or transition economies have just started to emerge of late. Disdier and Mayer [2003] compares French investment in Western and Eastern part of Europe. They find that location choice is positively influenced by local demand and proximity to France increases the probability of the given country being chosen. Cieslik [2003] used a Poisson model on 50 Polish regions to find that proximity of key export targets, industry and service agglomeration and road network are the key magnets for foreign investment.

As for Hungary, Fazekas [2003] looked at the concentration of FDI from a labour market point of view to study what impact capital inflow had on the regional structure of the country. The paper finds that concentration pattern of foreign-owned enterprises (FE) is just marginally higher (and unchanged through time) than that of the domestically owned (DE) ones. However, FEs are concentrated in a different pattern, being located closely to the Western border. Then, Fazekas estimates a regression of the concentration indices being dependent of education, industrial base and distance from the border. He finds significant coefficients with the expected signs. My approach is somewhat different of Fazekas [2003] in that I only concentrate on FDI and investigate the agglomeration patterns of these firms.

Barta [2003] describes regional differentiation in post-transition Hungary. She gives two good examples of agglomeration forces in work. In Central and Eastern Europe manufacturing of electronic devices by firms such as Flextronics can be found in a fairly narrow band from north Poland through the Czech Republic, West Slovakia, West and Central Hungary down to North Slovenia and Croatia. In Hungary, suppliers to the car plant of Suzuki are shown to be settled in neighbouring counties of Komarom-Esztergom megye, where the Suzuki plant is located. Further, second wave of suppliers that settled directly to service the plant are on average much closer to the factory than the suppliers during the first half of the nineties.

[1] The latest example is Hyundai motors in Slovakia where eight other Korean firms announced to follow Hyundai.

3 The model with I-O linkages

In this model we put an emphasis on business to business relations. The main relationship between any two firms is a potential of supplier-buyer link, i.e. one firm's output is the intermediate good of another. Modelling and measuring this potential will be in the centre of this analysis. The model is using the "classic" ingredients of new economic geography or Dixit-Siglitzt-Krugman (referred as DSK) world: Cobb-Douglas utilities and a market structure á la Dixit and Stiglitz [1977]. One key aspect of firm-to-firm relationship here is related to input-output linkages that were introduced by Krugman and Venables [1995] in order to model the fact that firms sell goods not only to consumers but other firms as well. This paper follows the concise display of multi-country and multi-industry model in Fujita, Krugman and Venables [1999, Chapter 15A].

3.1 Market structure

There is monopolistic competition in all sectors producing a range of differentiated goods. We focus on manufacturing and overlook agriculture here. True, we will miss a dispersion force but hope that wages and local consumer demand will be enough, owing to very limited migration.

There are $r = 1...R$ regions, $j = 1...J$ industries, with n_r^j firms producing a variety each of industry j in region r.

Each consumer enjoys manufacturing goods and the composite good consumed comes from a constant elasticity of substitution (CES) function of the available varieties. The elasticity of substitution between goods is measured by σ_j. Theoretically it measures to what extent goods are close to each other, i.e. whether consumers are easily willing to replace one with another. If it is small, products differ, in case of $\sigma_j = \infty$, the products are homogenous, and the market structure is identical to perfect competition.

Importantly, all firms use a set of goods produced by other industries that are aggregated by a CES subutility function into a composite good. The intermediate good price index, G_r^j denoting the minimum cost of purchasing a unit of this composite good, is a key variable in this setup for firms benefit from supplier proximity. If more of necessary intermediate goods are produced locally, less transportation cost will have to be paid, hence production costs will be lower, too. This creates a forward linkage. Here, the intermediate price index is weighted average (with n_l^j being the number of relevant firms[2]) of f.o.b. prices $\left(p_l^j \tau_{l\_r}^j\right)$ that already include an iceberg type transport costs, $\tau_{l\_r}^j \geq 1$: (i.e. when $\tau_{l\_r}^j = 1$ for the home region only, $l = r$).

[2]Later, the number of firms may be replaced with volume of output.

$$G_r^j = \left[\sum_{l=1}^{R} n_l^j \left(p_l^j \tau_{l\_r}^j \right)^{1-\sigma_j} \right]^{\frac{1}{1-\sigma_j}} \tag{1}$$

This way of entering the price index implies the love of variety effect. The intermediate price index for a firm in industry j of region r is G_r^j, where io_{ij} is the input-output coefficient, i.e the share of industry j in all output used by industry i. In a small country, industry buys goods and service from abroad and the import coefficient, io_{j*}, for each industry gives the share of a composite imported good (priced G_W^j). Since data come from a complete IO table, $\sum_{i=1}^{J} io_{ij} + io_{j*} = 1$, $\forall j \in J$.

$$GP_r^j = \left[\prod_{i=1}^{J} \left(G_r^i \right)^{io_{ij}} \right] (G_W^i)^{io_{i*}} \tag{2}$$

The marginal cost function of a representative firm in industry j and region r is defined as follows:

$$mc_r^j = w_r^{a_j} (GP_r^j)^{\mu_j} (\mathbf{b}_r^j)^{\delta_j} \tag{3}$$

where w is the nominal wage, G_r^j is the composite price index of intermediate goods and $\mathbf{b_r}$ is a vector of other location dependent non-wage factors of production. At the moment I simply consider b_{1r} as the presence of business assisting services such as banks and consultants. Thus, b_{1r} may now be taken as a proxy to urbanisation, too. Second, b_{2r} is taken as the availability of a composite natural resource good. For the time being it is assumed that these are consumed locally only.

Admittedly, in a short-sighted model like ours, wage formation is ignored as agents are assumed to be myopic. Assume now that fixed cost of starting a new business is the same in all regions, and the cost of capital is unchanged through space as well - this can be considered as one key difference between national and international models. Firms pay taxes and receive investment support and the net state involvement (denoted by t_r) is allowed to be regionally different. The profit is simply:

$$\Pi_r^j = (1 - t_r)(p_r^j x_r^j - mc_r^j x_r^j) \tag{4}$$

As it is the case in models following the Dixit-Stiglitz tradition, profit maximisation yields a price that equals marginal cost and a markup, Φ^j:

$$p_r^j = mc_r^j \Phi^j \tag{5}$$

In our case, the markup depends on the elasticity of substitution. Assuming that firms have the same size, and there are N_r^j firms in region r, the markup is:

$$\Phi_r^j = \frac{\sigma_j}{\sigma_j - 1 + (\sigma_j - 1)/N_r^j} \tag{6}$$

Indeed, if two products are close substitutes, the monopoly power to set prices should be small, hence the low markup. In the DSK world, it is assumed that there is a large enough number of firms, hence: $\Phi_r^j (= \Phi^j) \simeq \sigma_j/(\sigma_j - 1)$, i.e. the markup is not dependent on consumption. This assumption is crucial for it yields that mill-pricing is optimal.

Firms sell their product to consumers and firms who use other firms' output as their input. This latter gives rise to a system of input-ouput linkages - a key agglomeration force. As in Fujita, Krugman and Venables [1999], demand can be derived from the Cobb-Douglas utility. Consumption in a region l for a unit of industry j output produced in region r is:

$$q_{rl}^j = (p_r^j)^{-\sigma_j} (\tau_{l\_r}^j)^{1-\sigma_j} E_l^j (G_l^j)^{\sigma_j - 1} \tag{7}$$

Expenditure on the $j - th$ industrial goods for a given region comes from two sources: consumers (who spend a μ fraction of their income on l region, j industry goods) and other firms coming from all industries.

$$E_l^j = \mu_l^j Y_l + \sum_{i=1}^{J} i o^{ji} X_r^i \tag{8}$$

In equilibrium, the supply of an industry j in region r will be equal to demand from Hungary and the rest of the world.

$$X_r^j = \sum_{l=1}^{R} q_{rl}^j + QW_r^j \tag{9}$$

where QW represents foreign demand.

Note that the way expenditure is set up creates a backward linkage: firms want to be close to their markets that may well be spread out across space.

Unlike in Fujita, Krugman and Venables [1999], this paper does not intend to end up with a set of equations and simulate results. Instead of a general equilibrium approach we need to be "short sighted" and consider a partial equilibrium without dynamic effects of an investment. This will allow to describe a profit function that will be related to the number of firms per region. Indeed, the main goal of this model is to yield a corporate profit function that will be linked to the settlement decisions of firms in the empirical work.

So, the profit function can now be rewritten:

$$\Pi_r^j = (1-t_r) mc_r^j (\Phi^j - 1) \left[\sum_{l=1}^{R} (mc_r^j \Phi^j)^{-\sigma_j} (\tau_{l\_r}^u)^{1-\sigma_j} E_l^j (G_l^j)^{\sigma_j - 1} + (\tau_{x\_r}^u)^{-1} QW_r^j \right] \tag{10}$$

where $\Psi^j := (\Phi^j - 1)(\Phi^j)^{1-\sigma_j}$ is a monotonically decreasing function of the industry specific elasticity of substitution, σ_j. Note, that at the moment this measure is industry-dependent only.

Let us define the aggregate demand variable, AD_r^j as

$$AD_r^j := \left[\sum_{l=1}^{R} \left((\tau_{l\_r}^u)^{1-\sigma_j} \left(\mu_l^j Y_l + \sum_{j=1}^{j} io^{ji} X_l^j \right) (G_l^j)^{\sigma_j - 1} \right) + (\tau_{x\_r}^u)^{-1} QW_r^j \right] \tag{11}$$

So, the profit function is:

$$\Pi_r^j = (1 - t_r) \left(w_r^{a_j} (GP_r^j)^{\mu_j} (\mathbf{b}_r^j)^{\delta_j} \right)^{1-\sigma_j} \Psi^j \left[AD_r^j \right] \tag{12}$$

3.2 Input-output linkages: access to supply and demand

3.2.1 Market access

Market access is relevant for purchasing power of consumers and firms. There are various ways to measure market access. In his seminal paper, Harris used the simple formula in which, the market potential of the k-th area is sum of the purchasing power of all other regions weighed by a function of the distance. Measuring inner distance (i.e. a region's distance from itself) is problematic hence, it may be useful to separate own and foreign expenditures and this is what we do here. (True, there are some concerns with this approach (e.g. Redding and Venables [2001]) but it is quite helpful.)

Indeed, aggregate demand as defined in (11) may be proxied by three market access variables. Here, we start with this:

$$MAC_r^j = \zeta_1 (\mu_r^j Y_r) + \zeta_2 (\sum_{l \neq r}^{R} \frac{\mu_l^j Y_l}{\tau_{l\_r}^j}) \tag{13}$$

where, Y_r is regional income, μ is the share of income spent on the particular industry weighted by the transport cost. I dropped the time subscript here, but of course in practice MAC_{rt}^j is the relevant measure. Also, we consider that other firms use our products as well:

$$MAF_r^j = \sum_{j}^{K} co^{ij} \left(\zeta_3 (X_r^j) + \zeta_4 (\sum_{l \neq r}^{R} \frac{X_r^j}{\tau_{l\_r}^j}) \right) \tag{14}$$

Finally, the third case takes into account that export is a crucial determinant of the revenue of Hungarian firms. Accordingly market access to foreign firms and customers should be taken into account. The key determinant of location for export purposes is the distance from the border, and closeness to the few motorways that provide access to the rest of Europe. [3] Amiti and Javorcik [2003] face the same challenge for Chinese subsidiaries of multinational firms that typically produce a great deal of their output for foreign markets. In their paper, access to foreign markets is proxied by the tariff rate but European free trade in manufactured goods makes this unnecessary. Thus, in this paper, I proxy access to foreign markets by taking into account the distance to borders and the airport.

3.2.2 Proxy for the intermediate price index

Amiti and Javorcik [2003] posited that G can be proxied by calculating the supplier access, a weighted average of potential suppliers locally, at national and international level. Then they simply took intermediate good production and purchase together and created supplier access variables. However, since the input-output table is not symmetric (i.e. textile manufacturing uses a great deal of cotton, but cotton production uses little textile input), one should proxy G and the AD variable separately.

Given the market structure, the intermediate price index will be negatively correlated with the supply of these goods. Hence we can use some SA variable, such as:

$$SA_r^j = \sum_j^K co^{ji} \left(\vartheta_1(X_r^j) + \vartheta_2(\sum_{l \neq r}^R \frac{X_r^j}{1+\tau_{l\_r}^j}) \right) \tag{15}$$

3.2.3 Other business-to-business relations

Previously I noted that competition may be a determining force of the intermediate good price index, but is left out of this analysis for the time being. However, there may be other potential forces such as knowledge spill-over and labour market pooling that should somehow be included in the broader model.

Several NEG models incorporate some sort of knowledge spill-over as a chief externality explaining agglomeration (e.g. Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud [2003, Ch.7]). Clusters of sectors such as the Silicon Valley or Hollywood would be explained by proximity to firms and people who innovate. Also, sharing knowledge and not only about technology may help reduce costs of administration, for example . Crozet, Mayer and Mucchielli [2003], studying location of FDI in France find that

[3] Hungary has (unfortunately) no access to the Sea.

firms of the same nationality like to group together, locations close to home country are chosen more frequently, and some industries (like electronics) have a strong tendency to agglomerate.

One, admittedly simple measure of these forces that I propose now, is the size of own industry agglomeration once we controlled for input-output linkages.

4 The econometric model

In the empirical section of the paper, I will aim at explaining why certain areas of the country proved to be popular site for investment. Since I have a panel data set, for every year analysed, I can regress the choice variable on the labour market and market structure (with firms already settled) that exists at that time (i.e. the preceding year to the decision). Thus, we can take into account the fact that corporate landscape was changing through this time. For a transition economy, this is essential. To investigate the agglomeration effect I now apply the widely used binary choice model.

4.1 Conditional logit

First, I will estimate a conditional logit (CL) model to study the influence of input-output linkages, labour market conditions and market access on investment decisions in Hungary. A key result that allows for such a structure to be used here is the Random Utility Maximisation framework of McFadden [1974]. In this framework, firms are assumed to make decisions maximising expected profits, but given less than perfect information and errors made by analysts maximisation per se is less than perfect. McFadden assumed that profit (or utility for consumers) is a random function. [4]

The methodology widely applied in spatial probability choice modelling is the conditional logit model based on Carlton [1983]. Decision probabilities are modelled in a partial equilibrium setting with agents pursuing profit maximization behavior. Thus, they maximise a profit function like (12) subject to uncertainty. Apart from observed characteristics of firms, sector and location (entering the profit equation) unobserved locational characteristics, measurement errors or improper maximization will determine actual profits. Note, that we do not observe either derived or actual profits, but perceive locational decisions of firms.

Taking all potential effects into account, a firm i (where $i \in \{1, ..., N\}$) of sector j (where $j \in \{1, ..., J\}$) locates in region r (where $j \in \{1, ..., R\}$) will attain the profit level of dependent on industry, sector firm and cross-specific variables. Importantly, not all of these variables matter, as the choice of region is independent on individual firm or industry characteristics. Thus,

[4] For details, see Maddala [1983]

if agents maximise expected utility in this partial equilibrium setting, the number of firms in a region is related to the expected profit, as laid down in the profit function. The expected profit of firm i in industry j and region r is:

$$\pi_{ij}(r) = A + \gamma' b_r + \lambda' d_{rj} + \varepsilon_{irj} \tag{16}$$

In order to be able to use results of McFadden [1974], we need to assume that the error term, ε_{irj}, has a type I extreme value (or Gumbel) distribution. Note, that using this distribution is empirically hardly distinguishable from a Normal, Gumbel gives slightly fatter tails only. (Train [2003, p. 39.]) More important an assumption is that errors are independent of each other, i.e. the error for one alternative provides no additional information about the error for another one.

For every spatial option, the investor will compare expected profits for all other ones and chose region r provided that the following condition is fulfilled for $\forall l \neq r$:

$$prob[\pi_{ij}(r) < \pi_{ij}(l)] = prob[\varepsilon_{irj} < \varepsilon_{ilj} + A_r - A_l + \gamma' b_r + \lambda' d_{rj} - \gamma' b_l - \lambda' d_{lj}] \tag{17}$$

If this is the case, we can posit that the investor's probability of selecting location r provided she opted to invest in sector j is:

$$P_{r|j} = \frac{\exp(\gamma' b_r + \lambda' d_{rj})}{\sum_{l=1}^{R} \exp(\gamma' b_l + \lambda' d_{lj})} \tag{18}$$

The probability $P_{r(j)}$ is the logit model itself. Estimation is carried out by maximising the log-likelihood:

$$\log L = \sum_{j=1}^{J} \sum_{r=1}^{R} n_{jr} \log P_{r(j)} \tag{19}$$

where n_{rj} denotes the number of investments carried out in sector j of region r.

It is important to note that parameter values do not correspond to marginal effects the same way as one is used to in a linear regression framework. Instead, coefficients need to be transformed to yield odds-ratios that are easier to interpret.

4.2 Variables

One record in the database is one company with three sets of variables for each firm. Feature variables include the year of birth, the county of birth, a two digit NACE code for the industry sector, and the size of the output (sales). Basic variables are region and mostly industry specific and include the wage rate or the unemployment rate. Access variables are determined for industry-region pairs using various sources of information such as input-output tables and freight data. All are based on industrial production figures per county and sector and these numbers are determined by aggregating sales figures from the balance sheet data for all the relevant firms: $IP_{jr} = \sum_i x_{i(jr)}$.

The explained variables are the location choices of firms, along with the county of the location site, year of investment and code of the industrial sector. The choice variable is 1 if the investment took place in that particular county and 0 for the remaining 19 counties.

Explanatory variables are lagged one year for two reasons. The economic rationale (see "time-to-build" models) is that firms may be assumed to spend a year between investment decision and actual functioning (that is picked up by the data). The econometric support stems from a requirement to avoid endogeneity, and lagging will free the model of simultaneity bias.

To get a linear relationship, all variables are taken in logs. Access variables are denoted with ACC, while w stands for various labour cost measures. The basic estimable equation (for a given year, i.e. without time subscripts) is as follows:

$$\pi_r^j = \gamma AD_r + \alpha IP_r^j + \sum_z \beta_{(z)1}^j w_{(z)r}^j + \sum_h \delta_{(h)1}^j ACC_{(h)r}^j \tag{20}$$

Overall there are six market access variables. $SA1$ and $MA1$ are local while $SA2$ and $MA2$ are national (non-local) measures of supply and market access, respectively. These four measures are industry dependent. For their special nature, two additional supply access variables are used; $RA1$ and $BA1$ are to measure raw material and business access. Purchasing power of consumers is measured by INC: local income (relative to national average) variables are weighed by distance. The INC variable is decomposed into the number of inhabitants, $Size$ and income per capita, YPC. Industrial output (IP_r^i) of the given sector in the given region is introduced to check if there is an other agglomeration force in play other that has not been accounted for by the access variables. If significant, it may signal that competition should be introduced as a determinant of intermediate good price or that other sort of business to business externalities are present that were overlooked by the formal model. Various types of wage measures (blue-collar, white-collar and managerial) will be used, indexed as $z = 1, 2, 3$. At the moment, I disregard any local tax incentive.

Measuring the advantages of being close to export markets proved to be a hard one. At the moment a distance measure was calculated from key borders to each directions plus Ferihegy airport. Distance from individual borders were weighed by share of directions estimated from international trade statistics (KSH). [5]Note that this is a rough measure for it does not take into account diversity in industrial export destinations. Also it lacks imports.

4.3 Firm data

The dataset that I am using is based on annual balance sheet data and was compiled by the Institute of Economics of the Hungarian Academy of Sciences. It contains information on more than 10 thousand firms for a time-span of about 10 years. Although this is a representative set of data, on average 80% of firms with employment over 10 people will be included. Data include industry code, size of employment the share of foreign ownership and a county code.

As for the corporate database, for a given year, say 1999, there are 18330 firms all together with 5761 companies being manufacturing firms out of which foreign stake (defined as foreign ownership of more than 10% of the equity capital) is found for 1634 firms. There are data for the years 1988-2001. However statistical comparability of data through time is very questionable, and this is why I opted to use data for 1992-2001.

For there is no appropriate dataset of individual firm establishments, I assumed that the year of birth equals the year of first appearance in the dataset. (To be precise, new-born firms are found by choosing the first year of submitting a report to the Tax Authority.) Unfortunately this is not always the case for various reasons. For example, small firms are randomly chosen and in certain cases even medium sized firms might be left out. However, a one-year lag in a few cases should not cast doubt on the validity of the procedure. More importantly, it can safely be assumed that omissions are exogenous.

Overall, my dataset is composed of 1760 location settlements by firms with foreign ownership. Of this, 405 events seem to be a foreign acquisition of an existing company.

A final question before looking at the results: is our dataset large enough? For logit a model, Disdier and Mayer [2003] use 1879 location decisions for 19 regions (countries) over the time period of 20 years. Crozet, Mayer and Mucchielli [2003] use a sample of almost 4000 location choices over 10 years and 92 administrative locations. Basile [2001] uses a dataset of about 1400 foreign investments in 91 Italian regions over 12 years (mainly via mergers and acquisitions). For a smaller country, Portugal, Figueiredo, Guimaraes and Woodward [2002*a*] work with 759 investment decisions in 258 small regions. Using a negative binomial model, Coughlin and Segev [2000] consider 380 manufacturing plants in 48 states (in 8 regions) over 5 years. Overall, I

[5]We used a weighted average of distnace to the borders: West/Austria: Hegyeshalom; South/Croatia :Letenye; North/Slovakia: Komárom, East/Ukraine: ,Airport: Ferihegy/Budapest.

reckon that in this exercise, we used a dataset comparable in size with other empirical studies.

4.4 Finding coefficients for regressors

There are various ways to measure distance between counties. Some (e.g. Crozet, Mayer and Mucchielli [2003]) search for the central point and approximate the county area by a circle only to proxy the distance by the average of any two points within those circles. Here I chose a simpler method. Using the TSTAR database on settlements, I picked the most important city per county (i.e. with the largest number of manufacturing plants) and determined the road distance between these two points. In all but one case, the largest city was at least twice the second. In the remaining one county (Pest), there are 3 settlements with comparable size, but they are very close to each other. Transport distance is measured as the shortest route by car between two settlements. It is assumed that goods are transported by trucks only and that vehicles move at the same speed and cost no matter what the road is.

There are some coefficients that are not estimated here but used from other sources: Input-output table comes from Hungarian Statistics Office's publication on 1998 data (KSH [2001]). This is the only IO table available for the time period used. However, the assumption that input requirements per sectors have not greatly changed in a decade seems acceptable. The data indeed show that production is specialized, about half the value of output comes from purchasing goods and services from other producers. Out of domestic input, some 40% comes from buying goods, 55% from market services (including construction) and 5% from non-market services. Hungary is a small and open country with a production sector that relies greatly on imports: the average import share is 34%, but for some branches of manufacturing (e.g. electronics), it reaches as high as 80%.

Unit transport costs are estimated by assuming a very simple relationship:

$$(1/\tau^u_{l\_p})^\sigma = 1/(dist_{l\_p} * V^u) \tag{21}$$

i.e. it depends on the distance and on the cost of transporting one dollar worth of good by one kilometer. All data refer to distance by car, thus the road network that is crucial for transportation of goods is indeed taken into account of. The value of a typical package of industrial output $V^u = (\$/kg)^u$ on 1km comes from the World Bank database. True, these figures are based on more developed market data, and aggregation will mask many features. However, I assume it helps correct for the fact that it is cheaper to ship €100 worth of laptop PC than the same value of steel.

Studies with international data make use of the availability of cross-regional (i.e. international) figures for trade. This allows explicitly to estimate transportation costs. Proximity to key markets and suppliers have featured in empirical works explaining overall economic activity or per capita income. Redding and Venables [2001] argues that a country's wage level (proxied by per capita income) is dependent on its capacity to reach export markets and manage to get hold of the necessary intermediate goods cheaply. For the European Union, using bilateral trade data, Head and Mayer [2003*b*] also estimates trade costs first and only then do they run regressions with market access for Japanese firms. However, this is not an option when working with a country for which inter-regional data do not exist for commerce. Hence, one has make assumptions and check their robustness.

5 Estimation results

Estimation results are presented in tables 3 and 4. Below I emphasise the key findings.[6]

5.1 Access, demand and clusters

Theoretically all access variables should enter our equation. However, there is a strong correlation between $SA1$ and $MA1$ and $SA2$ and $MA2$ and I decided to use $SA1$ and $MA2$ for they worked the best overall.

Results of regressions with the access variables are reported in Table 3. Overall, consumer demand, local supply and national access are in most case significant and so is the distance from the key Western external border for exporters. When used separately, both relative wealthiness and size of a county increases the likelihood of investment, just as expected.

Industries like to cluster for other reason than input-output linkages as it is proven by the strong significance of the industrial output variable of the actual sector (IP), even when other access variables are controlled for. It is impossible to separate the key motives, such as labour pooling, knowledge spillover or a decrease in business costs due to information sharing. There are two consequences of this finding. First, the formal model needs to be extended to take this into account. Second, local competition needs to be measured empirically.

As for other access variables, access to business services does not seem to induce more firms to enter but this variable is strongly correlated with the per capita income variable. Access to raw materials is significantly negative in a few cases, possibly picking up the fact that traditional industries might be built on them that later turned out to be an impediment to development.

The border measure is overwhelmingly significant with the expected negative sign. However, simply adding the variable to the equation (see equa-

[6] To run regressions, I used statistical package *Stata 8.2,* while calculating my access variables was carried out by programming language *Gauss 4*.

tions (1) and (4) in Table 3.) alters the sign of market access variable. For this is just a rough measure at the moment, it is early to comment.

5.2 The labour market

One has to approach the impact of labour market on location choice with great care. The theoretical prediction of the wage coefficient is clear, wages are positively related to costs and hence negatively to locations. However, the empirical evidence is mixed with a slight leaning towards the opposite sign. For example, in Figueiredo, Guimaraes and Woodward [2002*b*] local wage has the expected sign, while in Holl [1993], the wage coefficient is insignificant but is significantly positive for the food sector while negative for paper and printing. There are various explanations. For example Figueiredo, Guimaraes and Woodward [2002*a*] argue that firms consider the wage level as a determinant to locate in cheaper country like Portugal (or even more so, Hungary) bur within the country it has no effect.

Let me describe just two more possible explanations here. First, note that if local labour markets are interlinked, wages should be close to the marginal product of local labour and have no influence on investment decisions. Wages should then be a function of skills and education and lower wages in a region should simply mirror the skill and education composition of the regional labour pool. Accordingly the local wage is a weighted average of various type of workers. A positive coefficient on wages may just imply that multinationals need white-collar workers a great deal, and are willing to pay them. I dub this the "composition bias".

Second, individual industries use different type of labour in different shares. The share of white-collar workers may vary a great deal among sectors. Moreover, "blue-collar" workers may differ greatly depending on how skilled they are. Thus, wages may well reflect an "industry bias". An insignificant or a positive coefficient may just imply that investors are bringing in superior technology and hence, require more skilled and educated (i.e. more expensive) sort of labour, reflected in higher wages.

Crozet, Mayer and Mucchielli [2003] control for the industry bias and uses industry specific local wages. They find the expected negative coefficient. Unlike in most studies in the literature, Barrios, Strobl and Gorg [2002] used the local wage level explicitly as proxy to local pool of skills and found a significantly positive coefficient. In order to control for some aspects of the above described biases, I first distilled three categories from the wage data: blue-collar, white-collar and managerial type labour.

In the sample, using simple county level wages gives an insignificant coefficient. Note, that for other specifications, the sign alters with the coefficient remaining always insignificant. Upgrading for industry specific wage helps a great deal. The coefficient is now significantly below unity: lower wage is what helps bring in new firms. If the job independent industry specific wage is replaced by a job specific wage variable, the cost of white-collar jobs turns out to be a dispersion force. Interestingly, blue-collar workers' wage is positively correlated with new investments. In my view this reflects the fact that there is a great diversity among such workers, and the wage may reflect its skill content.

Finally I consider the effect of labour pool quality by creating a new variable, the share of people (working the region) with a degree in higher education. A larger share is expected to make firm location more likely and indeed this what early results suggest.

5.3 Robust checks

In order to find out more about the robustness of results, I am checking for the role of foreign acquisitions and individual industry features. One must note, that the dataset is not large enough to support massive robust checks.

First, I considered firms that are likely to have some predecessors, i.e. may be closer to a foreign acquisition than a greenfield investment. There are limitations to this approach because of the moderate reliability of this feature of the data. Local demand seems more important while the national corporate market access is less important for foreign acquisitions. This may be explained by the fact that many acquisitions, e.g. in light industries were carried out in the early nineties to occupy the consumer market. In this case cheap labour was a plus indeed.

Second, I grouped some interesting industries into two categories: light industry (e.g. textile, clothing, etc.) and electronics/equipment (inc. electric machinery, audio-video manufacturing, etc.), and ran regression for one group at a time. As expected results vary substantially. Equipment and electronics industries prefer wealthier and more developed sites located to have an access to national markets. Being close to suppliers and low wages are important for light industries but not for the second group.

5.4 Are there larger regions?

The conditional logit modelling has some important limitations. An important restriction is

$$p_j(y_j)/p_h(y_h) = \exp((y_j - y_h)\beta) \tag{22}$$

so that "relative probabilities for any two alternatives depend only on the attributes of those two alternatives" (Wooldridge [2002, p. 501]). This is called the assumption of Independence of Irrelevant Alternatives (IIA). In our case, this posits that all locations are considered similar (having controlled for explanatory variables) by the decision making agent, yielding independent errors across individuals and choices. When IIA is assumed, an investor will look at all regions as equally potential places for investment.

Thus complex choice scenarios cannot be included. Indeed unobserved site characteristics (such as actual geography) may well give way to correlation across choices.

One solution to solve the problem caused by an unrealistic assumption of IIA is the application of a nested logit model. In this case, investors first choose among large entities and then pick a smaller region within that entity. This is sometimes a natural distinction - when working with international data, assuming that a firm chooses a country first and a region second seems quite realistic. However, in other cases setting the layers is rather artificial. One can argue that investors first decide to settle in a NUTS2 region based on a few parameters and only then compare potential NUTS3 counties - within the chosen region. Unfortunately, with so few regions, this story does not seem to have any solid background.

To check whether IAA assumption is strong enough, I first ran generalized Hausman tests (Hausman and McFadden [1984]) for all NUTS2 regions. Results show that IAA fails for five regions (both at 1% and 10%) out of seven suggesting that a more complex tree structure should be used. However the fact that it holds for two NUTS2 regions suggest that NUTS2 regions may not constitute the best upper level. Second, I defined Central, West, East areas and ran tests again. Now, IAA failed at 1% for all three areas. Having tried a few key equations to see how robust the IAA failure is to specification, I found results to be indifferent to specification. This of course suggest a nested logit specification (to be completed.)

6 Conclusions and policy recommendations

In the introduction three research questions were proposed. To sum up, results suggest that there is indeed an agglomeration effect for companies in play and input-output linkages work their way through supplier and market access providing a key reason for co-location. Also, even within a small country market, access to national markets does matter although proximity to foreign markets seems to have an overwhelmingly important effect, too.

As for development policy, assisting foreign direct investment has been in the forefront of successful modernisation of developing countries. In this paper I looked at some key determinants of location choice. Although state incentives via tax breaks were not explicitly modelled, there are a few conclusions for economic policy.

First, most of the industries do have a strong tendency to settle where other similar firms have already settled. Spending money on incentives to have them established elsewhere may be inefficient, and instead labour migration should be made easier, for example via development of temporary housing conditions.

Second, input-output linkages are important. Thus, an improving the relationship between suppliers and multinationals is key to fostering more investment. With a recent experience of loosing multinationals to Eastern Europe and China, this may be ever more important.

Finally, it is worth noting that there may well be a trade-off between equality and efficiency in a geographical sense, too. One key policy tool is the development of transport infrastructure that is expected to bring cities closer to each other, and hence, foster development. We too find that proximity to export markets is key. When designing policy, economic geography consideration are sometimes taken into account, although not always every aspect of it. For example, Puga [2001] quotes a report of European Union's Committee of the Regions that emphasises positive impacts of a better infrastructure but disregards agglomeration forces that may lead to a loss of industry in the poorer region that was originally to be developed. Martin [1999] and Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud [2003, Ch. 17] look explicitly at infrastructure policies to find that there is a trade-off between spatial efficiency and equity when policies manage to reduce transportation costs.

With European Union membership there will be a large amount of regional aid directed to poorer regions that should lower regional differences. My analysis seems to suggest the importance of closeness of similar firms as well as suppliers. Thus resources should be devoted to developing transport infrastructure - bringing firms closer to eachother, thus assisting developed regions as well. It may well be beneficial for the whole country despite its inequality fostering repercussion.

7 Future research

As for future research, there are three issues. First, there are some theoretical considerations to be taken care of. Second, other econometric methods need to be applied to see how robust previous results are as well as to better treat some problems. Third, it seems possible for about half the companies to determine exact location at the plant level and thus, look at location choice within a county.

7.1 Theoretical additions

7.1.1 Competition

Since the number of firms is assumed to be infinite, competition is to some extent ignored in this model. The markup is dependent on a industry-specific factor only.[7] I have a few ideas how to remedy this problem:

[7] To remedy this, other models, for example, Crozet et al (2003) uses a single good model where firms play a strategic location game.

(1) If there are more firms in a region, the competition rises and the markup, that corresponds to the number of firms falls, along with the profits. I could assume that there is a finite number of firms that are in Cournot competition and so $\Phi_r^i = \Phi^i = \sigma_i/(\sigma_i - 1)$ does not hold, making Φ_r^i depend on N_r^i. However, this would violate a key assumption. An other option would be to take σ depend on r, too.

(2) Competition influences the price index directly. The entry of a new firm has a market crowding effect. At the moment this effect is ignored but anyway, I will have to come back to it.

(3) Also, competition could be a cost rather than a price affecting phenomenon. For example, local competition for labor would bid up wages.

7.1.2 Strategic interaction among firms

Most certainly, firms do have strategic considerations when decide upon location. In a recent paper, Altomonte and Pennings [2003] raise the question of strategic reaction in an oligopolistic setting looking at interacting of rivals' investment in country-industry pairs when uncertainty is present. The paper is a good example of ideas that may be added to the original model.Altomonte and Pennings [2003] argue that one important motive for multinational companies in less developed markets is to gain a cost advantage. It is shown that "follow-the-leader" type investment is most likely when a few firms dominates and the probability of such strategic reaction is positively related to cost uncertainty. The availability of panel data helps estimating any sort of strategic reaction.

7.1.3 Profits

In this structure, firms work out an expected profit function and evaluate it for a set of location options and then choose a location. In the dataset there are data for actual profits. Although there are several problems with the figures, it may be interesting to see how actual profits (π_t) captured by balance sheet data are related to $E_{t-1}(\pi_t)$, $E_{t-1}(\pi_{t+1})$, etc., as described in (20).

7.2 Improving methodologies

There are methods that I plan to use to check robustness and treat problems including count data approaches and a modification of the logit model.

To check robustness of the logit, one option is using count data models, such as the Poisson, allow for estimating equations where the dependent variable represent the number or frequency of a particular event. In our case, we explain the number of investments in a particular area. Discrete choice models built on CLM yield a log-likelihood function that includes the term n_{jk} that is prime facie count data variable denoting the number of investment actions. Thus, link to count data models is easy to see. Indeed, Figueiredo, Guimaraes and Woodward [2004] shows that the conditional logit equation may stem from a Poisson model where n_{jk} is the explained variable itself assumed to follow a Poisson distribution with $E(n_{jk}) = \exp(\gamma' b_j + \lambda' d_{jk} + s_k)$, where s_k is a sector dummy. If we reject the equality of the expected value and the variance, we may turn to the negative binomial model, which was used by Coughlin and Segev [2000]. Also,a crucial assumption of the Poisson model is that events (here, firm appearance) arrive independently. Alternative models such as the negative binomial allow for dependence. Note that the negative binomial nests the Poisson, and it can be tested whether the move from Poisson to binomial is warranted. Both the Poisson and the negative binomial has also been used in location research.

When applying the CLM, we simply pool data for various years. Although simultaneity bias is excluded, pooling itself may add autocorrelation if, say, wages at (t) are dependent on entrance of a firm in $(t-1)$.To remedy this problem and make use of information stemming from a panel, application of panel discrete choice methods may be advisable. Examples here include time series count data model developed in Brandt and Williams [2000] or a negative binomial panel used in Altomonte and Pennings [2003]. In this latter piece, the panel model is used to extract information on firm reaction to decisions by other firms, thus the paper should serve as a useful reference.

In a recent paper Muchielli and Defever [2004] used a mixed logit methodology to improve upon the treatment of the IAA problem. They make use of Brownstone and Train [1999] who suggest that introducing random effects would help relax the IAA assumption. Mixed logit is a flexible model that can approximate any random utility model and offer a remedy for various problems of the logit model such random variation of preferences. Hensher and Greene [2002] discuss details and application of mixed logit as well as its relationship to CL models.

7.3 Disaggregating regions

A more adequate question may be if the relevant decision structure involves a smaller region than the county level. One reason is that any geographical grouping of firms is arbitrary, and the if counties were not the true level of decision-making an important problem would arise, called the Modifiable Area Unit Problem or MAUP.

There may be two separate issues with regionally aggregated data. First, one has to decide upon the scale of aggregation of smaller units into larger entities. Second, while as rendering geographical places to countries is simple and in most cases straightforward, drawing the lines of non-national borders of areas is more problematic. For example, let us consider two plants that are a few kilometres away from each other but are separated in different

NUTS2 regions. When they are treated as region 1 and region 2 plants, their estimated distance based on the knowledge of regional borders only, can reach a few hundred kilometres. Thus, working with counties may overly simplify the setting and mask important features. This is why I plan to have a second round of estimation with greater geographic "resolution".

I expected a substantial improvement of the model when leaving the county level for the NUTS4 level that includes 150 sub-regions or "*kistérség*". In this case determining regional characteristics is more problematic but a richer set of location options should compensate for the loss of data accuracy.

Importantly, a more detailed dataset would importantly allow to study the effect of transport network and state support in the form of industrial zones.

8 Appendix

In the Appendix I describe details of data manipulation that I carried out in order to remedy some errors. Sometimes, corrections involve just a handful of firms, but including some big ones making corrections helpful.

The key problems I found and/or learned from others working with the same or similar datasets are as follows: (1) 0 is imputed instead of actual figures for sales, (2) thousands written instead of millions, (3) one digit is left out making sales figure be 1/10 of actual data, (4) sales and export sales figures swapped, (5) various other typing errors e.g. when digits are swapped. I concentrate on 1-4, estimates of problematic figures range anywhere between 1%-10%.

Out of the total 117379 records, sales equals zero for 6691. This includes both a zero entry and a "not available" or a not imputed entry.

To remedy some, I developed three methods.[8] First I take the whole balance sheet and calculate the sales figure from using the profit and determinants of it such as total costs, result of financial transactions, etc. (As a check, I calculate total cost the main determinant of profit in a similar fashion to check if there are multiple problems in the balance sheet.). I replace the sales figure with the "accounting" sales figure whenever $sales = 0$. For about 1.5% of data I make other smaller corrections on the key total cost variable based on detailed balance sheet data. This method allows to replace zero for 1633 cases.

Second, I fill in holes if there is no (non-problematic) balance sheet data, using time series data for the actual firm. For a firm for a given t year but sales are different from zero for $(t-1)$ and $(t+1)$, the average of these two is applied. Similar method is used to bridge a two or three years gap. This helps find a proxy for 540 zeros in sales data.

[8] Corrections were carried out by simple Stata Do files. Files as well as detailes of results are available from the author on request.

These two methods eliminate about one-third of zeros but leave 4521 entries when sales=0 including cases when sales is indeed zero.

Third, I tried to detect "problematic sales figures" i.e. ones that are different from zero are hard to believe for one or another reason - including various typing errors. One such situation is data blips: when the sales figure drops for a year only to jump back for the next, possibly indicating that somehow a digit was skipped. I found 141 such stories.

Fourth, I used two profitability measures, one based on the number of employees (also corrected by time series figures) and another one returns to assets ratio, to find problems of nature (2), (3) and (4). After the proceeding corrections, I found that some 2084 cases where productivity was very low for both measures and number of employees was over 10. Mostly sales were very close to zero, only for 157 entries was sales above 10. This is the most problematic situation for here, we have no reliable sales figure at all. I tried to estimate it using average industrial productivity data but results are not conclusive and are hence, unused

Overall, I made more than 2100 modifications of the data reaching almost 2% of the total dataset. This process is no free of personal judgement and arbitrary conditions. However, I believe it helps improve the data. Further, I looked at the sensitivity of my problem-signalling parameters - running regressions for a few values, and results were unchanged.

References

Altomonte, C. and Pennings, E. [2003], Oligopolistic reaction to foreign investment in discrete choice panel data models. Universita Bocconi.

Amiti, M. and Javorcik, B. S. [2003], Trade costs and location of foreign firms in china. World Bank.

Amiti, M. and Pissarides, C. A. [2001], Trade and industrial location with heterogeneous labor, Technical report, CEP/LSE.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G. and Robert-Nicoud, F. [2003], *Public Policy and Spatial Economics*, MIT Press.

Barrios, S., Strobl, E. and Gorg, H. [2002], Multinationals' location choice, agglomeration economies and public incentives, Research Papers 33, Nottingham University.

Barta, G. [2003], Developments in the geography of hungarian manufacturing. in Munkaerõpiaci Tükör.

Basile, R. [2001], The locational determinants of foreign-owned manufacturing plants in italy, WP 14, ISAE.

Brandt, P. T. and Williams, J. T. [2000], A linear poisson autoregressive model. Indiana University.

Brownstone, D. and Train, K. [1999], 'Forecasting new poduct penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Carlton, D. W. [1983], 'The location and employment choices of new firms', *Review of Economics and Statistics* **65**(440-449).

Cieslik, A. [2003], Location determinants of multinational firms within poland. Warsaw University.

Coughlin, C. and Segev, E. [1999], Foreign direct investment in china: A spatial econometric study, Wp, Federal Reserve bank of St Louis.

Coughlin, C. and Segev, E. [2000], 'Locational determinants of new foreign-owned manufacturing plants', *Journal of Regional Economics* **40**, 323–351.

Crozet, M., Mayer, T. and Mucchielli, J.-L. [2003], How do firms agglomerate? a study of fdi in france, Technical report, TEAM, University of Paris.

Disdier, A.-C. and Mayer, T. [2003], How different is eastern europe?, Wp, CEPII.

Dixit, A. and Stiglitz, J. E. [1977], 'Monopolistic competition and optimum product diversity', *American Economic Review* **67**, 297–308.

Fazekas, K. [2003], Effects of foreign direct investment on the performance of local labor markets - the case of hungary. Budapest Working Papers on the Labour Market, No 03/03.

Figueiredo, O., Guimaraes, P. and Woodward, D. [2002*a*], 'Home-field advantage: location decisions of portuguese entrepreneurs', *Journal of Urban Economics* **52**(2), 341–361.

Figueiredo, O., Guimaraes, P. and Woodward, D. [2002*b*], Modeling industrial location decision in us counties. University of South California.

Figueiredo, O., Guimaraes, P. and Woodward, D. [2004], 'A tractable approach to the firm location decision problem', *Review of Economic and Statistics* .

Fujita, M., Krugman, P. and Venables, A. J. [1999], *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge.

Harris, C. [1954], 'The market as a factor in the localization of industry in the united states', *Annals of the Association of American Geographers* **64**(315-348.).

Hausman, J. and McFadden, D. [1984], 'A specification test for the multinomial logit model', *Econometrica* **52**, 1219–1240.

Head, K. and Mayer, T. [2003*a*], The empirics of agglomeration and trade, DP 15, CEPII.

Head, K. and Mayer, T. [2003*b*], Market potential and the location of japanese firms in the european union, Discussion paper 3455, CEPR.

Head, K. and Ries, J. [1999], Overseas investments and firms exports, WP 8528, NBER.

Hensher, D. A. and Greene, W. H. [2002], The mixed logit model: The state of practice, Working Paper ITS-WP-02-01, Institute of Transport Studies, The University of Sydney.

Holl, A. [1993], Manufacturing location and impacts of road transport infrastructure:empirical evidence from spain. Department of Town and Regional Planning, University of Sheffield.

Krugman, P. R. [1991], 'Increasing returns and economic geography', *Journal of Political Economy* **99**, 483–499.

Krugman, P. and Venables, A. J. [1995], 'Globalization and the inequality of nations', *Quarterly Journal of Economics* (110), 857–880.

KSH [2001], Ágazati kapcsolatok mérlege 1998 (input-output tables), Technical report, Központi Statisztikai Hivatal.

Leamer, E. and Stolper, M. [2000], The economic geography of the internet age. UCLA.

Maddala, G. S. [1983], *Limited Dependent and Qualitiative Variables in Econometrics*, Cambridge University Press.

Marshall, A. [1920], *Principles of Economics*, Macmillan Press.

Martin, P. [1999], 'Public policies, regional inequalities and growth', *Journal of Public Economics* **73**, 85–105.

McFadden, D. [1974], *Conditional Logit Analysis of Qualititative Choice Behaviour*, Academic Press, New York, NY.

Midelfart-Knarvik, K. H., Overman, H. G. and Venables, A. J. [2000], Comparative advantage and economic geography: estimating the determinants of industrial location in the eu. LSE.

Muchielli, J.-L. and Defever, F. [2004], Functional fragmentation of the production process:a study of multinational firms location in the enlarged european union. University of Paris I.

Myrdal, G. [1957], *Economic Theory and Under-developed Regions*, Duckworth, London.

Ottaviano, G. and Robert-Nicaud, F. [2003], The genome of neg models with input-output linkages.

Ottaviano, G., Tabuchi, T. and Thisse, J.-F. [2003], 'Agglomeration and trade revisited', *International Economic Review* .

Puga, D. [2001], European regional policies in light of recent location theories, DP 2767, CEPR.

Redding, S. and Venables, A. J. [2001], Economic geography and international inequality, DP 2568, CEPR.

Robert-Nicaud, F. [2002], New Economic Geography, Multiple Equilibria, Welfare and Political Economy, PhD thesis, LSE.

Train, K. [2003], *Discrete Choice methods with Simulation*, Cambridge University Press, Cambridge.

Venables, A. J. [1996], 'Equilibrium locations of vertically linked industries", *International Economic Review* **37**.

Wooldridge, J. M. [2002], *Econometric Analysis of Cross Section and Panel data*, MIT Press.

Table 1
Industries

NACE code	**Industry**	**# of new-born firms**	**%**
17	Textiles	151	8.48
18,19	Wearing apparel, leather, luggage, etc.	171	9.61
20,21	Wood, wood products, pulp, paper, etc.	128	7.19
22	Printing, publishing	148	8.31
23,24	Refined petroleum, chemicals and chemical products	80	4.49
25	Plastic and rubber products	139	7.81
26	Non metallic minerals	95	5.34
27	Basic metal products	40	2.25
28	Fabricated metal	229	12.87
29	Machinery and equipments n.e.c.	195	10.96
30	Office machinery and computers	26	1.46
31	Electrical machinery and apparatus, n.e.c.	83	4.66
32,33	Radio, tv, telecommunication equipment	161	9.04
34,35	Motor vehicles and other transport equipment	69	3.88
36	Furniture	65	3.65

Table 2
Descriptive Statistics

Variable	Descrpition	Mean	Std. Dev.	Min	Max
INC	Log Total county income	10.52	0.62	9.34	12.76
Size	Log Number of county inhabitants	6.08	0.49	5.38	7.44
YPC	Log Income per capita	4.44	0.25	3.96	5.32
IP	Log Industrial output per industry	7.75	1.60	1.39	14.33
SA1	Log Local supplier access	5.82	1.42	1.39	11.73
SA2	Log National supplier access	8.75	1.08	4.67	12.16
MA1	Log Local market access	6.18	1.43	2.64	11.78
MA2	Log National market access	9.43	0.87	6.16	12.02
RA1	Log Local raw material access	9.91	0.60	7.94	11.16
BA1	Log Local business services access	8.43	1.41	6.05	14.26
WAGER	Log County average wage	10.07	0.36	9.66	11.34
WAGEI	Log County average wage per industry	10.02	0.41	9.03	11.87
WAGEPH	Log County average wage per industry for blue collar workers	9.87	0.39	9.03	11.55
WAGEOF	Log County average wage per industry for white collar/office type workers	10.21	0.45	9.12	12.62
WAGEMA	Log County average wage per industry for managers	10.90	0.55	9.36	14.09
Foreign type*	Size of foreign ownership 1: *10%-25%*, 2: *25%-50%,* 3: *50%+*	2.77	0.48	1	3
Fshare	Share of foreign ownership	0.76	0.28	0.1	1
Foraq	Foreign acquisition dummy	0.14	0.35	0	1
Region	NUTS2 Region	-	-	1	7

Table 3

Estimates of location choices: Conditional logit

	(1)	(2)	(3)	(4)
YPC(t,r): Per capita income	2.720***	2.632***	2.817***	1.293*
	(0.53)	(0.59)	(0.61)	(0.214)
Size(t,r): Region size - number of	1.696***	1.703***	1.811***	1.755***
people	(0.18)	(0.19)	(0.22)	(0.19)
IP(t,r,k): Own industry's output	1.222***	1.201***	1.198***	1.210***
	(0.036)	(0.037)	(0.048)	(0.386)
SA1(t,r,k): Local supplier access	1.239***	1.235***	1.217***	0.171***
	(0.73)	(0.73)	(0.086)	(0.070)
MA2(t,r,k): Non-local market access	1.245**	1.227*	1.275**	0.67**
	(0.16)	(0.17)	(0.18)	(0.11)
RA1(t,r): local raw material access	0.852**	0.850**	.901	0.92
	(0.60)	(0.59)	(0.075)	(0.068)
BA1(t,r): local business services	0.900*	0.906*	.882*	0.94
access	(0.50)	(0.53)	(0.055)	(0.052)
WAM(t,r): Local average wage		0.898		
		(0.51)		
WIP(t,r,k): Local industry wage	0.676**			0.66***
	(0.11)			(0.115)
Wf(t,r,k): Local industry specific			1.916**	
blue-collar wage			(0.505)	
Wb(t,r,k): Local industry specific			.836***	
office wage			(0.062)	
Wv(t,r,k): Local industry specific			.617**	
manager wage			(0.126)	
Border(r): Distance from borders				0.590***
(avg)				(0.050)
No of observations	1760	1760	1405	1760
Log likelihood	-4548	-4486	-3390	-4465
χ^2 (LR test)	1239	1203	1025	1246
McFadden's pseudo R^2	0.120	0.118	0.131	0.122

Notes*: Dependent variable is 1 for the actual county and 0 for all others. All in logs. Results are odds ratios and not coefficients. For all variables, time (t), county (r) and industry (k) specificity is noted when relevant. Standard errors are in brackets. *, **, *** denote significance at 10%, 5% and 1% level, respectively.*

Table 4

Estimates of location choices: *Robustness check*

	(1) Excl. foreign acquisition	(2) Foreign acquisition	(3) Light industry	(4) Equipment, etc
YPC(t,r): Per capita income	3.051***	1.92***	1.790**	3.372***
	(0.68)	(0.59)	(0.62)	(1.41)
Size(t,r): Region size - number of people	1.652***	1.90***	1.220	1.94***
	(0.20)	(0.19)	(0.25)	(0.42)
IP(t,r,k): Own industry's output	1.297***	1.00	1.020	1.206***
	(0.048)	(0.063)	(0.08)	(0.065)
SA1(t,r,k): Local supplier access	1.134**	1.650***	1.97***	1.05
	(0.078)	(0.19)	(0.27)	(0.10)
MA2(t,r,k): Non-local market access	1.340**	1.004	0.65*	3.085**
	(0.20)	(0.12)	(0.16)	(0.99)
RA1(t,r): local raw material access	0.845**	0.871	0.95	0.82
	(0.56)	(0.12)	(0.60)	(0.12)
BA1(t,r): local business services access	0.897*	0.897	0.85*	1.03
	(0.56)	(0.10)	(0.08)	(0.12)
WAM(t,r): Local average wage				
WIP(t,r,k): Local industry wage	0.68**	0.67*	0.652***	0.869
	(0.14)	(0.23)	(0.14)	(0.35)
Wf(t,r,k): Local industry specific blue-collar wage				
Wb(t,r,k): Local industry specific office wage				
Wv(t,r,k): Local industry specific manager wage				
Border(r): Distance from borders (avg)				
No of observations	1355	405	581	455
Log likelihood	-3469	-1005	-1530	-1075
χ^2 (LR test)	882	345	460	395
McFadden's pseudo R^2	0.112	0.146	0.130	0.155

Notes*: Dependent variable is 1 for the actual county and 0 for all others. All in logs. Results are odds ratios and not coefficients. For all variables, time (t), county (r) and industry (k) specificity is noted when relevant. Standard errors are in brackets. *, **, *** denote significance at 10%, 5% and 1% level, respectively.*